

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Mineração de *Rationale* e Consenso em Fóruns
MOOC para Suporte à Engenharia da Plataforma de
Software e à Qualidade do Ensino

Tiago Lucas Pereira Clementino

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande - Campus I como parte dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Ciência da Computação

Linha de Pesquisa: Informática Aplicada à Educação, Engenharia de Software,
Engenharia de Requisitos

José Antão Beltrão Moura

(Orientador)

Campina Grande, Paraíba, Brasil

©Tiago Lucas Pereira Clementino, 19/02/2020

C626m Clementino, Tiago Lucas Pereira.
Mineração de *Rationale* e consenso em fóruns MOOC para suporte à engenharia da plataforma de software e à qualidade do ensino / Tiago Lucas Pereira Clementino. – Campina Grande, 2020.
92 f. : il. color.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2020.
"Orientação: Prof. Dr. José Antão Beltrão Moura".
Referências.

1. Engenharia de Software. 2. Informática – Estudo e Ensino. 3. Mineração de *Rationale*. 4. Fóruns MOOC – Consenso. I. Moura, José Antão Beltrão. II. Título.

CDU 004.41(043)

**MINERAÇÃO DE RATIONALE E CONSENSO EM FÓRUNS MOOC PARA SUPORTE À
ENGENHARIA DA PLATAFORMA DE SOFTWARE E À QUALIDADE DO ENSINO**

TIAGO LUCAS PEREIRA CLEMENTINO

DISSERTAÇÃO APROVADA EM 19/02/2020

JOSÉ ANTÃO BELTRÃO MOURA, Ph.D, UFCG
Orientador(a)

LÍVIA MARIA RODRIGUES SAMPAIO CAMPOS , Dra., UFCG
Examinador(a)

ADRIANO ARAÚJO SANTOS, Dr., UNIFACISA
Examinador(a)

CAMPINA GRANDE - PB

Resumo

A gestão de uma ferramenta de colaboração online como o fórum de discussão pode demandar grandes esforços à medida que o número de colaboradores cresce. No contexto educacional, podem surgir novos problemas ligados à elevada quantidade de participantes, como o suporte à aprendizagem em larga escala. Embora também surjam com isto, novas oportunidades como a utilização de conteúdo compartilhado espontaneamente por alunos como fonte de oportunidades de pesquisa ou requisitos de manutenção de software. Tais diretrizes e requisitos podem vir na forma das razões por trás de críticas, sugestões ou mesmo opiniões mais específicas a respeito da plataforma educacional. Estas opiniões e críticas, aqui definidas como *Spontaneous User Rationale*, assemelham-se em muitos aspectos ao que a engenharia de software chama de *User Rationale*.

A partir da coleta de *Spontaneous User Rationale*, definiu-se como alvo de investigação a identificação da próxima intervenção em fóruns MOOC (do inglês *Massive Open Online Courses*) por parte dos instrutores. Neste contexto, mesmo que a utilização de conteúdo compartilhado em discussões de fóruns online como fonte de informação não seja algo novo, a relação entre o nível de consenso e a qualidade da informação compartilhada em fóruns de discussão é um elemento que permanece um tanto obscuro. Assim, este trabalho apresenta como parte de seus resultados uma análise em que se concluiu que há uma correlação negativa significativa entre um alto nível de consenso e a qualidade da solução alcançada, no contexto de MOOC.

Como resultado, este documento apresenta uma ferramenta de apoio ao trabalho do instrutor em dar suporte ao aluno que busca informação em fóruns de discussão de um MOOC. Tal ferramenta tem como base a exposição de postagens e discussões agrupadas por similaridade e consenso. Seu objetivo é permitir ao instrutor atender a postagens mais urgentes prioritariamente, evitando possíveis desinformações por disseminação de conteúdo equivocado. Esta urgência é inferida a partir do consenso, com base em uma correlação identificada entre eles.

Para alcançar tais resultados, utilizaram-se modelos de análise de distância textual por Árvores KD e *Word Embeddings*, calculando o nível de consenso em discussões baseado em Lógica Difusa, além de classificar as discussões como relacionadas ao tema do curso ou não.

Para avaliar a eficácia da solução apresentada, fez-se uso de *mockups* para Validação de Face em entrevistas com profissionais de educação participantes em MOOC; verificação da solução comparando seus resultados a outras abordagens; e, validação cruzada tomando marcações de urgência em postagens como dados para validar a correlação entre urgência e consenso.

Abstract

Managing an online collaboration tool such as the discussion forum can take great effort as the number of collaborators grows. In the educational context, new problems may arise related to the large number of participants, such as support for large-scale student learning. While this also creates new opportunities such as the use of content spontaneously shared by students as a source of research guidelines or software maintenance requirements. Such guidelines and requirements may come in the form of the reasons behind more specific criticism, suggestions or even opinions regarding the educational platform. These opinions and critiques, defined here as Spontaneous User Rationale, are in many ways similar to what software engineering calls User Rationale.

Based on the collection of Spontaneous User Rationale, the aim of this study was the instructors' identification of the next intervention in MOOC forums. In this context, even though the use of shared content in online forum discussions as a source of information is not new, the relationship between the level of consensus and the quality of information shared in forum discussions remains a somewhat obscure element. Thus, this paper presents as one of its results an analysis that concluded that there is a significant negative correlation between a high level of consensus and the quality of the solution achieved in the context of Massive open Online Course (MOOC).

As its final result, this document presents a tool supporting the instructor's work in supporting the student who search for information in MOOC discussion forums. This tool is based on the exposure of posts and discussions grouped by similarity and consensus. Its goal is to enable the instructor to respond to the most urgent posts first, avoiding possible misinformation due to the dissemination of mistaken content. This urgency is inferred from the consensus, based on a correlation between them.

For achieve these goals, we used models of textual distance analysis by KD Trees and Word Embeddings, calculating the level of consensus in discussions based on Fuzzy Logic, and classify the discussions as related to the course theme or not.

To evaluate the effectiveness of the presented solution, we used mockups for Face Validation in interviews with education professionals participating in MOOCs, as well as verifying

the solution by comparing its results with other approaches and cross-validation based on Supervised Machine Learning by taking postage urgency markings as training data to validate the correlation between urgency and consensus.

Agradecimentos

Por ser aquele que colaborou mais diretamente com meu trabalho, agradeço em primeiro lugar ao meu orientador, José Antão Beltrão Moura, pela orientação durante os dois anos de mestrado na UFCG. Agradeço aos meus pais, Jario Clementino e Tereza Neumann, que sempre me deram apoio e proporcionaram os meios para que eu chegasse até aqui. Em especial, agradeço a minha esposa Thamyres Oliveria Clementino pela paciência e companheirismo. A todos os meus amigos, pelo apoio e pela compreensão em minhas ausências. Aos professores Adriano e Lívia, por terem aceito o convite para participar da banca examinadora de minha dissertação. Agradeço à Universidade Federal de Campina Grande, ao Departamento de Sistemas e Computação e à Coordenação de Pós-Graduação em Ciência da Computação por minha formação acadêmica e por tornar possível a realização da minha pesquisa. Em especial, agradeço à CAPES pelo suporte dado ao meu trabalho.

Conteúdo

1	Introdução, Fundamentação e Metodologia	1
1.1	Introdução	1
1.1.1	Relevância	2
1.1.2	Objetivos	3
1.1.3	Proposta de Aplicação	4
1.1.4	Organização da Dissertação	4
1.2	Fundamentação	5
1.2.1	MOOC	5
1.2.2	Plataformas colaborativas	6
1.2.3	<i>User Rationale</i>	8
1.2.4	Consenso	8
1.2.5	Estado-da-Arte	9
1.3	Metodologia	10
1.3.1	Passo 1: Revisão da Literatura em Ferramentas Colaborativas para MOOC	10
1.3.2	Passo 2: Exploração de <i>User Rationale</i> em Dados de Fóruns de MOOC	10
1.3.3	Passo 3: Revisão da Literatura em Consenso	11
1.3.4	Passo 4: Experimento – Consenso × Qualidade	12
1.3.5	Passo 5: Intervenção do Instrutor em Discussões de Fóruns MOOC – Revisão da Literatura	12
1.3.6	Passo 6: Proposta de Aplicação – Consenso × Urgência	12
2	<i>Rationale</i>	15
2.1	Contexto e Trabalhos Relacionados	16

2.1.1	MOOC e Análise de Fóruns	16
2.1.2	<i>Rationale</i> em Engenharia de Software	17
2.2	<i>Design</i> da Pesquisa	18
2.2.1	Questões de Pesquisa e Objetivos	18
2.2.2	Metodologia	19
2.2.3	Dados Quantitativos	19
2.2.4	Mineração de <i>Rationale</i>	20
2.3	Resultados	21
2.3.1	Fase 1	21
2.3.2	Fase 2	22
2.3.3	Fase 3: Entrevistas	22
2.4	Conclusões e Implicações	28
3	Consenso	30
3.1	Contexto e Estado da arte	31
3.2	Design da Pesquisa	33
3.2.1	Os Dados	33
3.3	Mensurando Consenso	34
3.3.1	Distância Textual	36
3.3.2	<i>Soft Consensus</i>	37
3.4	Consenso × Qualidade	43
3.5	Conclusões, Trabalhos Futuros e Limitações	46
3.5.1	Limitações	46
3.5.2	Conclusões e Trabalhos Futuros	46
4	Proposta de Aplicação baseada em Consenso para Apoio à Decisão de Instrutores sobre quando Interferir em Discussões em Fóruns MOOC	48
4.1	Objetivo	48
4.2	Solução	49
4.2.1	Validação da Solução	50
4.3	Relevância	51
4.4	O Modelo	52

4.4.1	Os Dados	54
4.4.2	Fundamentação	55
4.4.3	Busca por Vizinhos Próximos: O Algoritmo	60
4.5	Verificação e Validação	62
4.5.1	Verificação	63
4.5.2	Consenso × Urgência	64
4.5.3	Entrevistas: Validação de Face	72
4.6	Resultados e Discussão	78
4.6.1	Ameaças à Validade	78
5	Resultados, Conclusões e Trabalhos Futuros	79
5.1	<i>Rationale</i> como Fonte de Requisitos de Software em Plataformas Colaborativas	79
5.2	O Consenso em Fóruns Educacionais	80
5.3	Uma Ferramenta de Suporte ao Trabalho do Instrutor em Fóruns MOOC . .	81

Lista de Símbolos

D1 - *Base de dados históricos de postagens registradas entre 2013 e 2014*

D2 - *Base de dados de postagens extraídos diretamente da interface no final de 2018*

QP - *Questão de pesquisa: “Como ampliar o potencial de suporte da equipe de instrutores por meio do fórum de discussão em um MOOC?”*

QP' - *Questão de pesquisa modificada: “Há qualidade nas conclusões às quais discussões em fóruns MOOC possam chegar sem intervenção dos instrutores?”* Q1 - *O ‘Rationale’ subjacente às postagens nos fóruns poderia dar suporte à manutenção, atualização ou inovação em plataformas de software educacional?* Q2 - *Em fóruns de discussão educacionais, existe correlação positiva entre o nível de consenso e a qualidade da opinião consensual?*

Lista de Figuras

1.1	Fluxograma de Pesquisa	14
4.1	Diagrama de comunicação entre módulos do modelo de utilidade proposto.	50
4.2	Possíveis cenários para identificação de vizinho mais próximo	58
4.3	Distribuições Amostrais dos Dados de Verificação	65
4.4	Intervalos de Confiança: “a” apresenta a precisão de cada modelo; e, “b” a diferença entre nossa precisão e a de cada modelo concorrente	66
4.5	Matriz de Confusão da Validação da Relação Consenso \times Urgência	71
4.6	Explicação dos fatores da Validação da Relação Consenso \times Urgência a partir da ferramenta LIME	71
4.7	Diagrama de Atividades para Validação de Face. Letras são pontos de referência para elementos dos <i>mockpus</i>	73
4.8	Mockup a	74
4.9	Mockup b	75
4.10	Mockup c	75
4.11	Mockup d	76

Lista de Tabelas

2.1	Exemplos de especificações de requisitos fruto do <i>Rationale</i> extraído de D1	23
2.2	Frequência do <i>Rationale</i> e sua natureza.	24
2.3	Exemplo de <i>Release Note</i> da plataforma edX atendendo a demandas na forma de <i>Rationale</i> em D1	25
2.4	Perfis dos Entrevistados.	25
3.1	Regressão logística: apoio dos instrutores e convicção média como fatores de consenso	45
4.1	Relatório de Classificação	70
4.2	Questionário para Validação de Face.	77
4.3	Avaliação do Entrevistado pelo Pesquisador.	77

Lista de Códigos Fonte

4.1	Árvore de Busca por Vizinhos mais Próximos	60
4.2	Configuração do Modelo de classificação Gradient Boosting Recomendada Pelo TPOT	70

Capítulo 1

Introdução, Fundamentação e Metodologia

O presente documento reporta um estudo realizado entre 2018 e 2020 que investiga estratégias de suporte tanto ao aluno quanto ao instrutor em fóruns de discussão de MOOC.

1.1 Introdução

Oferecer educação e formação profissional à população em geral tem sido um desafio para governos e instituições de ensino, sobretudo no Brasil [10]. A educação tradicional considera apenas perfis específicos de alunos e seus custos geram empecilhos significativos [10]. Neste contexto, o modelo MOOC (do inglês, *Massive Open Online Course*), tem se mostrado uma alternativa viável [29].

Os MOOC são ferramentas promissoras para uma expansão do acesso à educação em um contexto global e de larga escala [29; 57], além de outras funções pedagógicas como aprimoramento profissional, apoio e complementação à educação tradicional, etc. Neste contexto, o presente trabalho partiu de uma pesquisa exploratória descrita no Capítulo 2 deste documento, em que fóruns de MOOC foram minerados em busca de diretrizes de pesquisa que apontassem para melhorias capazes de otimizar ferramentas de apoio em plataformas de MOOC. Chegou-se a conclusão de que, por se tratar de uma estratégia de larga escala, prover suporte educacional a estudantes em MOOC é um desafio ainda em aberto, sendo tratado atualmente por meio de ferramentas colaborativas como fóruns e blogs [78].

Tomando os resultados do Capítulo 2 – otimizar o suporte educacional a estudantes em MOOC como objetivo de pesquisa – como motivação, o Capítulo 3 descreve um estudo que tem como objetivo medir a relação entre o nível de consenso alcançado em discussões de fóruns de MOOC e a qualidade da solução consensual alcançada – qualidade esta definida como o apoio do instrutor à decisão consensual. Tal estudo foi feito a partir dos registros históricos de um conjunto de fóruns de MOOC e foi motivado pelas seguintes constatações: Nos dados de fóruns de MOOC analisados, 66% das postagens que demandam suporte urgente por parte dos instrutores expressam alguma opinião [6]; a dificuldade em buscar informação útil nestes fóruns é apontada como um dos principais empecilhos à navegação em fóruns, a medida que o volume de informações cresce – Capítulo 2; segundo um dos paradigmas da educação colaborativas – fóruns – a diversidade é um dos fatores de controle para sua eficácia (vide Seção 1.2.2); e, a maioria das postagens ocorrem no início do curso [14], quando os estudantes tem pouco domínio a respeito do conteúdo estudado.

Como resultado do Capítulo 3, apresentou-se evidências a respeito da relevância do suporte por parte do instrutor para o aluno em discussões de fóruns de MOOC. Isto foi concluído uma vez que o consenso mostrou-se inversamente proporcional à qualidade da opinião majoritária.

Por ultimo e como resultado de pesquisa, no Capítulo 4, é apresentado uma ferramenta de apoio ao instrutor capaz de apontar o nível de consenso em discussões de fóruns, agrupar postagens e discussões nestes fóruns por similaridade, apontar duplicatas, entre outros possíveis usos futuros. Tal ferramenta tem como principal objetivo desonerar a atividade do instrutor, priorizando – quando é necessário priorizar – a análise de postagens mais sensíveis à sua intervenção com base no nível de consenso.

1.1.1 Relevância

No Capítulo 2 deste documento explorou-se a opinião dos estudantes a respeito da plataforma de MOOC na forma de *Rationale* (vide Seção 1.2.3). Como resultado, com base na opinião dos estudantes, observou-se que navegar no fórum se torna um problema bastante relevante a medida que o volume de postagens cresce. Problema este inerente a todos os participantes do fórum – estudantes e instrutores. Problema especialmente impactante no trabalho do instrutor que não deve deixar certas postagens sem assistência. Aqui tomamos como foco de

pesquisa o suporte ao instrutor por este ter sido deixado de lado por pesquisas abordando o tema MOOC, conforme revisão da literatura realizada pelo autor [23].

Problema

Segundo [70], é indispensável promover a interação entre estudantes e instrutores para uma experiência educacional de qualidade. Diante disto e em razão da maior parte dos trabalhos identificados na literatura concentrar-se apenas no estudante, este estudo aborda especificamente o problema da mão-de-obra limitada do instrutor em um MOOC diante de um número não controlável de alunos. Com isto, identificar quais discussões demanda intervenção mais urgente é de suma importância para priorizar a atuação do instrutor. Reforçando esta afirmação, conforme [6], identificar tal urgência é apontado por instrutores como um problema relevante e em aberto. Somando-se a isto, conforme [14], a maior parte das postagens em um fórum de MOOC são realizadas no início do curso, quando os alunos ainda não dominam o conteúdo estudado e tem mais chances de compartilhar informações equivocadas. Assim, neste trabalho, evitar desinformação é tomado como objetivo por trás da abordagem dada à intervenção do instrutor.

1.1.2 Objetivos

O trabalho a seguir – a partir da investigação de relevância de demandas, sugestões e críticas de usuários de MOOC na forma de *Rationale* – apresentará um modelo de filtro e apresentação de mensagens capaz de otimizar o trabalho do instrutor em um MOOC indicando-lhe quais postagens exigem atenção com mais urgente. Urgência aqui interpretada como um esforço em evitar desinformação. Tal filtro e apresentação de mensagens é feita por similaridade e consenso. Apresentar por similaridade para permitir que o instrutor localize ideias semelhantes ao mesmo tempo; e apresentar por consenso pois, com base em resultados obtidos no Capítulo 3 deste trabalho, níveis elevados de consenso em torno de informações equivocadas podem levar a desinformação. Estes elementos foram definidos e explorados a partir da metodologia apresentada na Seção 1.3.

O modelo de filtro apresentado foi validado a partir da proposição de uma ferramenta baseada em uma interface gráfica, conforme exposto ao longo do Capítulo 4. Por meio

desta ferramenta proposta, o instrutor terá acesso a discussões e postagens agrupadas por similaridade e nível de consenso entre postagens ainda que estas postagens estejam em discussões diferentes. Para tanto, o modelo por trás desta ferramenta terá como base um algoritmo de aprendizagem não supervisionado baseado em *Word Embeddings* [61], *Soft Consensus* [37] e Árvore KD [71]. Esta abordagem por si só já destaca-se diante das soluções existentes, visto que com base no levantamento apresentado na Seção 1.2.5 deste trabalho, todas as referências da literatura à indexação de postagens em fóruns por necessidade de intervenção dos instrutores são feitas com base em algoritmos supervisionados [20; 6; 36; 79]. Para fins experimentais, a abordagem supervisionada é válida, embora seja de difícil implantação uma vez que necessitará de novos dados marcados periodicamente.

1.1.3 Proposta de Aplicação

Para concluir este estudo, criou-se um modelo de filtro de discussões baseado em consenso que foi aplicado em uma ferramenta de apoio ao instrutor. Tal ferramenta é apresentada como resultado desta pesquisa. Com o apoio de tal ferramenta o instrutor deve ser capaz de priorizar discussões valendo-se da informação a respeito do coeficiente de consenso e da similaridade entre estas. Sendo esta similaridade associada à discussão a qual o instrutor já está lendo, e agrupadas por seus coeficientes de consenso calculados em dois níveis: interno à discussão e externo / global. O nível interno define a opinião vencedora/consensual para uma dada discussão; já o nível global aponta recorrência de ideias em diferentes discussões, conclusões diferentes em diferentes discussões sobre o mesmo tema, entre outros usos avaliados e propostos.

1.1.4 Organização da Dissertação

A presente dissertação segue uma estrutura atípica. Por ter sido executada em três etapas bem distintas, este trabalho é aqui descrito em três capítulos, além deste Capítulo 1 que contém a Introdução, a Fundamentação e a Metodologia para o trabalho, e do Capítulo 5 contendo Resultados, Conclusões e Trabalhos Futuros.

No Capítulo 2 é descrito o processo que definiu os rumos e objetivos da pesquisa. Nesta capítulo foi feita uma pesquisa direta em fóruns MOOC em busca das opiniões na forma

de *Rationale* dos alunos a respeito das ferramentas de software utilizadas no curso, além de um *survey* entre ex participantes deste modelo de educação, com experiência docente. Os resultados definiram a fase seguinte, apontando o suporte ao estudante em fóruns, a análise de opiniões de alunos e a relevância do consenso como tema.

No Capítulo 3 é feita uma análise em busca de correlações entre consenso e qualidade, visando avaliar o impacto do grande volume de opiniões compartilhadas em fóruns desta natureza. Como resultado, concluiu-se com 95% de confiança que consenso e qualidade tem uma relação inversamente proporcional para elevados níveis de consenso. Isto implica que uma discussão com elevado nível de consenso leva a uma situação preocupante, onde opiniões equivocadas podem causar a desinformação.

Por ultimo, o Capítulo 4 apresenta o resultado final da pesquisa, onde uma ferramenta de software é proposta como meio de prover um melhor suporte ao aluno em fóruns MOOC. Isto se dá oferecendo um maior potencial de priorização deste suporte por parte do instrutor.

1.2 Fundamentação

1.2.1 MOOC

MOOC, do inglês, **Massive Open Online Course**, representam um conceito amplo de educação distribuído nestes quatro elementos [76]:

- **Massivo** – disponível para um número indeterminado de alunos. A expectativa por trás deste número varia bastante de uma iniciativa para outra (de poucas centenas a dezenas de milhares);
- **Aberto** – aqui várias interpretações são possíveis: Gratuito, disponível para todos, com conteúdo acessível, etc. Tais interpretações podem variar bastante caso à caso;
- **Online** – utiliza a infraestrutura digital online para funcionar. Porém, isto não quer dizer que não existam iniciativas híbridas com alguma interação presencial;
- **Curso** – este elemento final significa apenas que, geralmente, MOOC seguem a estrutura de um curso com datas de início, fim, avaliações, etc.

A partir desta definição ampla, dois paradigmas são explorados na literatura para soluções em MOOC [76]: cMOOC (*collaborative MOOC*)[76] e xMOOC (*xtreme MOOC*)[59]. O modelo xMOOC é mais focado no uso de ferramentas multimídia de suporte ao aprendizado (vídeos, quizzes, etc). Já cMOOC diz respeito à colaboração entre estudantes preenchendo lacunas do aprendizado do curso uns para com os outros (fóruns, *blogs*, redes sociais, etc). Aqui, optou-se por explorar soluções colaborativas, cMOOCs, tomando o fórum como principal ferramenta de suporte. Além disto, segundo Alario-Hoyos et al [2] o fórum é a ferramenta preferida entre todos os perfis de alunos para colaborar com conteúdo e buscar suporte em um MOOC.

1.2.2 Plataformas colaborativas

A teoria colectivista é guiada a adaptar a educação aos novos recursos tecnológicos e conexões sociais promovidas pela Web 2.0 [76]. Tais recursos e conexões de que trata este paradigma dão suporte a um processo de aprendizado de estudante para estudante [76], preenchendo lacunas no aprendizado uns dos outros.

Downes [31] propôs para a rede de aprendizado conexionista, quatro dimensões essenciais para identificar quais demandas do usuário o sistema deve suprir. Tais dimensões são:

- **Autonomia** – em que cada participante define seus próprios objetivos e ações;
- **Diversidade** – os usuários devem ser diferentes entre si e trazer diferentes pontos de vista;
- **Abertura** – a comunidade deve estar aberta à entrada e saída de novos membros e material;
- **Interatividade** – o conhecimento é gerado pela interação entre os membros da comunidade.

Dentre estas dimensões, a diversidade é especialmente abordada neste trabalho, visto que, conforme resultados obtidos no Capítulo 3, a diversidade de opiniões promove soluções de maior qualidade. Além disto, o conhecimento introduzido pela interatividade – entre possivelmente milhares de alunos – deve ser moderado pelos instrutores e é marcante a necessidade de buscar ferramentas de suporte que os auxilie nesta tarefa.

O Fórum Online

No presente trabalho partimos da seguinte definição de fórum:

Definição 1. *Fórum educacional online – aqui, simplesmente fórum – é todo e qualquer espaço de diálogo aberto estabelecido por meio da infraestrutura online e que siga uma (semi) estrutura com base em postagens sequenciadas a respeito de um tema definido.*

O Fórum como Ferramenta de Suporte

Oferecer assistência ao aluno em plataformas online tem como pré-suposto uma situação desfavorável, pois as partes não interagem fisicamente. Assim, não há o conhecimento da situação do aluno de forma ampla. Neste contexto, algumas soluções de suporte tem sido propostas [63; 66]. Contudo, sendo o fórum a principal ferramenta de suporte ao aluno em cursos online [2] é compreensível que boa parte dos estudos neste sentido girem em torno desta ferramenta.

A partir da natureza colaborativa do fórum, pode-se formular a seguinte pergunta-chave: *"Como melhorar o suporte colaborativo em fóruns de MOOC sem comprometer o requisito escala?"*

A partir dos resultados alcançados nos Capítulos 2, 3 e Seção 4.5.2 deste estudo, estabeleceu-se que o consenso em discussões online demanda maior atenção por parte dos instrutores e requer intervenção mais rápida. Aqui, conforme Seção 1.2.4, o consenso define a opinião com nível de aceitação mais elevado entre as postagens de uma discussão. A exposição do grau de consenso interno e externo à discussão, aliada ao agrupamento de discussões e postagens por similaridade podem ser instrumentos relevantes para o instrutor ao lidar com um número muito grande de alunos. Tais recursos maximizam o potencial de escala do MOOC a medida que poupam esforço da equipe de instrutores.

O Fórum como Fonte de Informação

Da Definição 1, podemos tomar o fórum como fonte de informação não apenas para os estudantes, mas também instrutores e equipe técnica em geral. Assim estes profissionais podem se valer das informações trocadas ali para aprender a respeito das necessidades dos alunos [5; 72]. Para este fim, em sua primeira etapa (vide Capítulo 2), este trabalho analisou discus-

sões em fóruns de MOOC com o objetivo de minerar comentários na forma razões por trás de críticas, recomendações e solicitações que foram compartilhadas espontaneamente pelos alunos a respeito da plataforma de software. Para analisar tais razões, fez-se uso de conceito da Engenharia de Software chamado *Rationale* [53] – descrito a seguir na Seção 1.2.3. A partir dos resultados deste processo – vide Capítulo 2.3.3 – definiu-se a forma de expor, filtrar e agrupar mensagens em um fórum como tema de pesquisa.

1.2.3 *User Rationale*

Como mencionado, *Rationale* consiste dos motivos e reflexões por trás de uma decisão de projeto em software [53]. Mas especificamente, introduzimos o conceito de *User Rationale* [50] espontâneo, definido a seguir:

Definição 2. “*User Rationale*” aqui consiste dos motivos e reflexões que levam um usuário de software colaborativo a tomar a decisão espontânea de fazer críticas ou sugestões a respeito de tal software [50].

Escolheu-se o adjetivo ‘espontânea’ para designar uma atitude não estimulada em um canal utilizado para outros fins – no caso particular deste estudo, tal canal se refere ao fórum educacional em um MOOC. Daqui em diante, sempre que utilizarmos o termo ‘*Rationale*’ na verdade estaremos nos referindo a ‘*User Rationale*’ espontâneo.

1.2.4 Consenso

Com base em dados de fóruns MOOC oferecidos pela universidade de Stanford em 2014, 66% das postagens que demandam intervenção mais urgente por parte dos instrutores expressam algum tipo de opinião [6]. Tal informação é apresentada como argumento de relevância para o estudo de alternativas de exposição do consenso como ferramenta de apoio à função dos instrutores em um fórum de MOOC. Somando-se a este, outros indicativos de relevância são apresentados como resultados preliminares deste estudo: 1) A forma como as mensagens são expostas é alvo de críticas espontâneas e persistentes por parte dos alunos em fóruns de MOOC à medida que o volume de informação aumenta; e, 2) concluiu-se – com base nos resultados do Capítulo 3 – que um nível elevado de consenso em discussões entre alunos de MOOC pode indicar conclusões de baixa qualidade.

A partir daqui, tomemos a seguinte definição de consenso:

Definição 3. “*Consenso*” é uma medida de concordância entre as respostas ou opiniões de um grupo de indivíduos sobre determinada questão ou assunto.

Observe que a partir esta definição, consenso é um número real com valores na faixa $[0; 1]$. “0” indica que não há nenhum consenso entre os indivíduos; e, “1” (ou 100%) que há consenso total entre eles.

1.2.5 Estado-da-Arte

De acordo com nossa revisão, não consta na literatura modelos não-supervisionados para detecção de urgência em discussões de fóruns de MOOC. Dentre as soluções em aprendizagem supervisionada encontradas, os trabalhos que apresentam melhores resultados abordam tal problema a partir de uma solução baseada em Redes Neurais Convolucionais e Recorrentes [36; 79]. No entanto, tais resultados são apenas experimentais, sem proposta de aplicação real. Em um ambiente real, modelos de aprendizagem supervisionada tendem a ‘envelhecer’, caso o cenário onde está inserido evolua e os padrões que ele aprendeu mudem.

Mesmo antes da era dos MOOC, o papel do instrutor como guia em discussões online já havia tido sua relevância apontada [58]. Contudo, os MOOC evidenciaram a importância em se utilizar a tecnologia para facilitar tal atividade. Almatrafi et al [6] propôs um modelo binário para classificar a necessidade de intervenção urgente por parte do instrutor em postagens de fóruns utilizando a mesma base de dados utilizada neste trabalho. Tal abordagem utiliza processamento de linguagem natural para extrair padrões linguísticos de fóruns de MOOC, então experimenta diversos modelos de Aprendizagem de Máquina – *Random Forest*, *Adaboost*, Regressão Logística, entre outros – para ensinar o modelo a prever urgência em fóruns MOOC. Chaturvedi et al [20] traz algo bem similar ao apresentar um modelo de aprendizado sequencial supervisionado para prever a intensão do instrutor e assim gerar sugestões de intervenção futuras. Tais estratégias diferem da apresentada aqui por utilizarem um modelo supervisionado de aprendizagem.

Além do problema da intervenção do instrutor, a maior parte dos trabalho tende a focar mais em questões como classificação de discussões como resolvidas e pendentes [85; 47], e completude [9]. Tais questões mostram-se pouco relevantes em fóruns educacionais como

em MOOC. Neste contexto, o fórum é totalmente direcionado ao aprendizado, não apenas à solução de um problema ou à resposta a uma pergunta [6]. Além disto, soluções voltadas a alunos ou usuários de fóruns abertos não apresentam os mesmos resultados quando aplicadas a instrutores [18].

1.3 Metodologia

Com o objetivo de explorar possíveis alternativas de melhorias em plataforma de MOOC, seguiu-se um fluxo de pesquisa em três passos: identificar uma questão de pesquisa; buscar respondê-la; e, analisar os resultados. Esta sequência foi repetida três vezes – cada uma nos Capítulos 2, 3 e 4, respectivamente. Todos os Passos realizados a cada etapa da pesquisa estão descritos textualmente a partir da Seção 1.3.1, e de forma gráfica por meio do Diagrama de Fluxo da Figura 1.1.

1.3.1 Passo 1: Revisão da Literatura em Ferramentas Colaborativas para MOOC

O primeiro passo consistiu em uma revisão da literatura em ferramentas colaborativas de suporte à aprendizagem em MOOC [23]. Tal revisão – cujos resultados estão resumidos na Seção 2.1.1, Capítulo 3 – apontou para o fórum como principal ferramenta colaborativa [2], embora tenha sido inconclusiva em formular a questão de pesquisa inicial. Para tanto, como alternativa, explorou-se a opinião do próprio aluno na forma de *User Rationale* como fonte de diretrizes de pesquisa (Passo 2).

1.3.2 Passo 2: Exploração de *User Rationale* em Dados de Fóruns de MOOC

Com base na questão de pesquisa parcial **Q1** – *O Rationale subjacente às postagens nos fóruns poderia dar suporte à manutenção, atualização ou inovação em plataformas de software educacional?* – este Passo testou fóruns de MOOC como fonte de requisitos de para atualização e evolução software. Para tanto, buscou-se a opinião dos estudantes a respeito de suas principais demandas relacionadas a plataforma de MOOC. Assim, foi realizado

uma investigação em quatro etapas baseada no conceito de *User Rationale*:

- 1: Minerar discussões de fóruns de MOOC buscando temas recorrentes em *User Rationale* a respeito da plataforma;
- 2: Buscar correlação entre novas funcionalidades, correção de bugs e melhorias em geral reportados em *Release Notes* do software de um dada plataforma de MOOC e os resultados anteriores, afim de apontar lacunas;
- 3: Comparação os resultados anteriores aos obtidos por meio de entrevistas realizadas com participantes de MOOC – descritas no Capítulo 2 deste documento – afim de buscar qualitativamente uma correlação entre o *Rationale* provocado em tais entrevistas, o *Rationale* extraído dos fóruns e lacunas identificadas na literatura.

A partir daqui, com base nos resultados deste Passo – descrito no Capítulo 2 e detalhado na Seção 2.4 – definiu-se a questão de pesquisa a ser investigada **QP**: *Como ampliar o potencial de suporte da equipe de instrutores por meio do fórum de discussão em um MOOC?*

1.3.3 Passo 3: Revisão da Literatura em Consenso

A priori, a abordagem adotada para buscar soluções que maximizem o potencial de suporte dos instrutores foi o uso consenso em discussões de fóruns aplicado a estratégias de apresentação e filtro de discussões em fóruns, sempre com foco no instrutor. Isto se deu principalmente em função de resultados obtidos no passo anterior (vide Capítulo 2) e em virtude de apontamentos extraídos da literatura (vide Seção 1.1.1). Especificamente, a partir destes resultados, pode-se reinterpretar a questão de pesquisa como **QP'**: *Há qualidade nas conclusões às quais discussões em fóruns MOOC possam chegar sem intervenção dos instrutores?*

Em tendo definido o consenso como foco de estudos, foi realizada uma ampla revisão da literatura em consenso online [22], e esta foi resumida no Capítulo 3.1 do Capítulo 4. Tal revisão proveu o ferramental necessário para investigar a correlação entre consenso e qualidade em discussões de fóruns MOOC.

1.3.4 Passo 4: Experimento – Consenso × Qualidade

Neste Passo – conforme descrito ao longo de toda o Capítulo 3 – foi realizado um novo experimento *ex-post facto* com base em dados históricos de discussões em fóruns MOOC, investigando se haveria correlação entre consenso em tais discussões e a qualidade da conclusão consensual, definindo assim a seguinte questão de pesquisa **Q2**: *Em fóruns de discussão educacionais, existe correlação entre o nível de consenso e a qualidade da opinião consensual?* A resposta para esta questão é diretamente relevante para a questão de pesquisa **QP'**.

O nível de consenso foi calculado com base em modelos baseados em quantificadores linguísticos e lógica difusa [37; 16]. Já a qualidade foi tomada como equivalente ao apoio dos instrutores à conclusão consensual. Ou seja, caso os instrutores apoiem a conclusão consensual, o consenso levou à qualidade. Caso os instrutores não apoiem a conclusão consensual, o consenso falhou ao apontar a conclusão de mais qualidade. A partir deste constructo, concluiu-se que a qualidade da conclusão é inversamente proporcional ao consenso para um nível de consenso elevado.

1.3.5 Passo 5: Intervenção do Instrutor em Discussões de Fóruns MOOC – Revisão da Literatura

Já com um direcionamento bem mais consistente, e em virtude do prazo restritivo, este Passo seguiu uma linha menos rígida. Aqui foi realizado uma revisão da literatura por *Snow Balling* a partir dos resultados obtidos no Passo 2. Buscou-se com isto identificar modelos de recomendação de intervenção por parte do instrutor. A busca inicia-se identificando os autores mais recorrentes e pesquisando entre suas publicações e citações, recomeçando em seguida a partir dos resultados anteriores. Os resultados estão descritos no Capítulo 1.2.5 desta Introdução e serviram de sustentação para os resultados finais.

1.3.6 Passo 6: Proposta de Aplicação – Consenso × Urgência

A partir dos resultados do Passo anterior desenvolvemos uma ferramenta de apresentação e agrupamento de postagens e discussões por similaridade e nível de consenso, para facilitar a gestão de urgência por parte do instrutor. Tal ferramenta, descrita no Capítulo 4 deste

documento, ainda não foi aplicada em um cenário real de ensino, embora tenha sido validada em três etapas (vide Seção 4.5).

Esta ferramenta é baseada em um algoritmo que inclui um módulo de cálculo de nível de consenso (*Soft Consensus*, vide Seção 3.3.2), vetorização de texto por *Word Embeddings* (vide Seção 4.4.2) e uma versão adaptada de Árvores KD (vide Seção 4.4.2). Aqui o objetivo é agrupar postagens e discussões similares em partições, tendo o nível de consenso como metadado, para que o instrutor possa visualmente identificar temas semelhantes em discussões distintas, revisar discussões concluídas (nível de consenso elevado) para evitar a desinformação, e evitar a leitura de postagens redundantes.

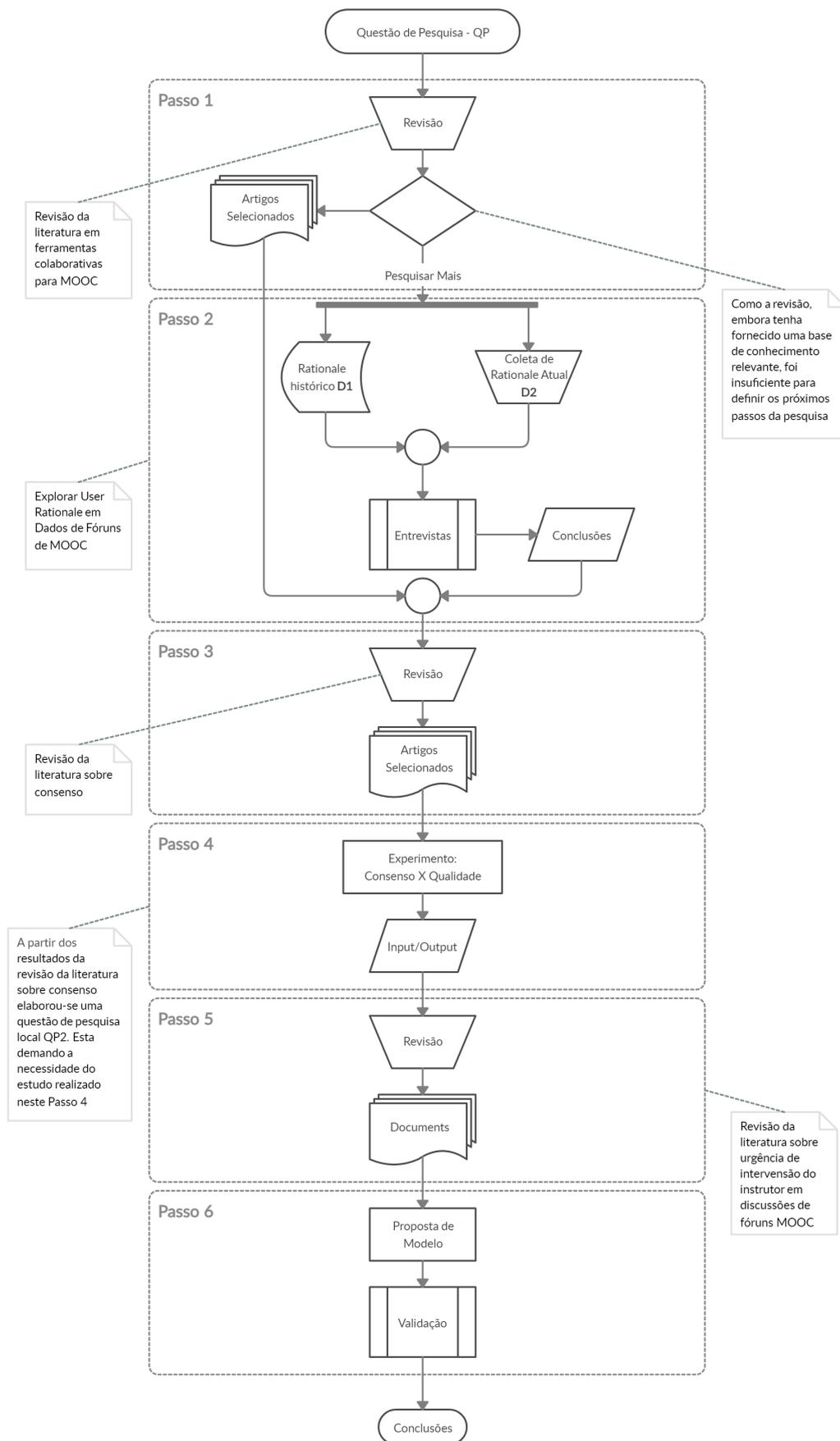


Figura 1.1: Fluxograma de Pesquisa

Capítulo 2

Rationale

Embora registros históricos de postagens em fóruns em “*Massive Open Online Course*” (MOOC) tenham sido explorados em *Learning Analytics* para aprimorar ganhos em aprendizado e *design* pedagógico [45], ferramentas de manutenção de *software* em plataformas de MOOC parecem ter sido mantidas fora do escopo de tais esforços. Para este fim, este trabalho analisa discussões de fóruns MOOC com o intuito de minerar problemas, críticas, recomendações ou solicitações que são espontaneamente compartilhadas pelos participantes de MOOC. Como já mencionado, o fruto de tal mineração pode ser tomado como análogo a “*User Rationale*”.

“*Rationale*”, de acordo com o dicionário Merriam-Webster, é a “*explanação dos princípios que norteiam uma opinião, crença ou prática*” (tradução). Assim, *User Rationale* não trata de simples demandas e opiniões do usuário, mas das razões por trás destas. Aqui, como já exposto no capítulo introdutório deste estudo, estendemos a definição proposta por [50] para acomodar reações não estimuladas de usuários a respeito de qualquer aspecto do curso online o qual possa ser explicitamente associado a questões de software e subsequentes discussões como uma forma de “*User Rationale*”. Este trabalho postula que *Rationale* de alunos de cursos online, minerados de fóruns de discussão, podem prover informação para aprimoramento do software da plataforma online a qual está inserido. Esta propriedade dos fóruns pode ser aproveitada tanto como fonte de requisitos para aprimoramento da plataforma de software, quanto como fonte de diretrizes para pesquisas aplicadas a ferramentas educacionais online, como a presente.

Para avaliar o fórum como fonte de “requisitos” de software, foram consideradas duas

bases de dados de postagens de fóruns de MOOC separadas por quatro anos de diferença: A primeira é composta por registros de fóruns em cursos oferecidos pela universidade de Stanford¹ na plataforma de MOOC edX² entre 2013 e 2014, devidamente anonimizados; já a segunda é composta por postagens observadas diretamente de 1.034 fóruns de cursos oferecidos por várias instituições de ensino pelo mundo na mesma plataforma edX em Novembro de 2018. Cursos estes deviam estar marcados como “*current*” ou “*archived*”, correspondendo a um total final de 86 cursos, além de contar com algum canal específico (discussão no fórum) para críticas e sugestões.

Em uma segunda fase, as *Release Notes*³ da plataforma edX foram examinadas em busca de alguma correlação relacionando as divergências entre os dados de *Rationale* minerados das duas bases de dados e as adições de funcionalidade ou correções de defeitos em novas versões da plataforma edX. Por último, nove usuários de MOOC com experiência em ensino foram entrevistados como parte de uma avaliação / validação mais qualitativa.

Este capítulo deste documento descreve um estudo originalmente publicado separadamente em um artigo científico no Simpósio Brasileiro de Informativa para a Educação – SBIE – em 2019 [25].

2.1 Contexto e Trabalhos Relacionados

Uma revisão da literatura não rendeu trabalhos em *User Rationale* voltados a plataformas educacionais. Embora duas áreas de pesquisa relacionadas tem reportado trabalhos com abordagens que podem ser adaptados para tal propósito.

2.1.1 MOOC e Análise de Fóruns

Análise de dados de fóruns MOOC representam um importante instrumento em busca de uma educação de qualidade e viável [5; 72]. Embora a exploração de *Rationale* espontâneo do usuário de fóruns MOOC não tenha recebido atenção na literatura, algumas estratégias de outras áreas podem ser usadas para inspirar meios de mineração em larga escala de *Ratio-*

¹<https://www.stanford.edu/>

²<https://edx.org/>

³<https://edx.readthedocs.io/projects/edx-release-notes/en/latest/index.html>

nale espontâneo em fóruns MOOC. [26] e [87] investigaram maneiras de minerar postagens relacionadas ao conteúdo estudado em fóruns MOOC. O primeiro descreve um classificador que identifica perguntas e respostas em discussões onde o instrutor participe. O segundo, também baseado em postagens relacionadas ao conteúdo, aponta que postagens fortemente relacionadas ao conteúdo compartilham padrões textuais recorrentes não relacionados ao domínio estudado. [32] propõe uma estratégia para apontar as melhores respostas para uma pergunta. Além disso, uma pesquisa em andamento em *Rationale* espontâneo de usuários em fóruns pode se beneficiar de trabalhos que visem identificação de necessidade urgência de intervenção do instrutor [6].

2.1.2 *Rationale* em Engenharia de Software

Em Engenharia de Software, *Rationale* tem recebido significativa atenção na literatura desde 2015. Algoritmos de mineração e análise de comentários em *App Stores* – e mesmo redes sociais – como fontes de *Rationale* são propostos por Williams em [86]. O trabalho de Williams difere do descrito aqui ao lidar com comentários explicitamente relacionados ao software como tema de discussão, ao invés de investigar comentários espontâneos como em fóruns educacionais, os quais requerem uma análise mais sutil e qualitativa das causas envolvidas.

Fóruns tem sido estudados em ciência de dados educacionais [5; 72], porém análises qualitativas e extração de *Rationale* de usuários permaneceram um tanto a margem. Embora muito tenha sido feito a respeito em mineração de *Rationale* em Engenharia de Software. Por exemplo, [4] apresenta um estudo empírico em *Rationale* extraído de discussões entre desenvolvedores de software livre geograficamente dispersos por meio de canais IRC (do inglês, *Internet Relay Chat*) e explora a possibilidade de mineração automática de *Rationale* em tais mensagens. Outro trabalho utilizou o *Rationale* em mensagens IRC ao separar tal conteúdo entre relacionado e não relacionado ao desenvolvimento e, assim, eliminando informação desnecessária [21].

Em [50] é apresentado uma definição completa do *Rationale* do usuário no contexto das avaliações de aplicativos em *App Stores*, além de uma metodologia completa sobre como extraí-los e classificá-los. [28] também discute abordagens sofisticadas de aprendizado de máquina para a lógica da mineração. [51] apresenta uma abordagem para investigar argu-

mentos e justificativas por trás das decisões. Além disso, [42] também considera a mineração de requisitos em avaliações de aplicativos e [46] realizou um *survey* da literatura sobre engenharia de requisitos baseada na colaboração de usuários, que menciona “*User Rationale*” como uma área de pesquisa promissora. No entanto, nenhum desses relevantes estudos considera a “*User Rationale*” em plataformas colaborativas como uma fonte de requisitos para o ferramental de software da própria plataforma, conforme realizado aqui.

O trabalho apresentado aqui está mais estreitamente relacionado a [83]; [3]; e [19]. O primeiro trabalho focou-se em explorar fóruns de discussão usando um algoritmo de aprendizado de máquina para analisar *User Rationale* em relatos de testes entre desenvolvedores para a plataforma Android⁴ por meio da plataforma de fóruns Stackoverflow⁵. O segundo realizou um estudo exploratório sobre a frequência de *Rationale* nas mensagens em chats de equipes de desenvolvimento e sobre formas de extração automática de *Rationale* e sua viabilidade. O terceiro considerou a utilidade e os desafios do uso de discussões em plataformas de Q&A para apoiar a evolução do software. Em vez de estudar o *Rationale* por trás das comunicações dos desenvolvedores de software, este estudo trata da extração de *Rationale* útil para o aprimoramento do ferramental de software em plataformas educacionais a partir de comunicações entre alunos que não necessariamente são alunos de Informática. No contexto da pesquisa científica, requisitos de software fruto de *Rationale* de usuários podem ser aplicados como direcionamento para investigação.

2.2 Design da Pesquisa

Esta seção resume a questão de pesquisa, a metodologia e o procedimento de coleta de dados usado para extrair e analisar *Rationale* das postagens de fóruns MOOC. Detalhes complementares e os resultados da Seção 2.3 estão em disponíveis em <https://bit.ly/2TMLWvp>.

2.2.1 Questões de Pesquisa e Objetivos

Este trabalho tem como objetivo reunir evidências de que as postagens em fóruns de MOOC incorporam *Rationale* dos usuários e este pode servir como fonte de requisitos para manter

⁴<https://www.android.com/>

⁵<https://stackoverflow.com/>

e/ou atualizar o software da plataforma, além de prover diretrizes para inovação nesta área. A questão de pesquisa de interesse para este capítulo do estudo pode ser formulada como:

- Q1: O *Rationale* subjacente às postagens nos fóruns poderia dar suporte à manutenção, atualização ou inovação em plataformas de software educacional?

2.2.2 Metodologia

A metodologia aplicada para responder à **Q1** envolve três fases:

1. Minerar *Rationale* dos registros históricos do fórum de uma determinada plataforma MOOC durante um período de tempo. Em seguida, comparar os temas de *Rationale* minerado dos registros históricos do fórum com aqueles extraídos da mesma plataforma atualmente;
2. Buscar evidências de que as divergências nos temas mais recorrentes de *Rationale* observado nas duas fontes de dados se correlaciona com as alterações nas *Release Notes* do software da plataforma MOOC investigada; e,
3. Entrevistar educadores experientes em participar de MOOC – como aluno ou instrutor – para validar a relevância do *Rationale* minerado. Avaliando assim o potencial de tal fonte de informação como instrumento para obter requisitos do software da plataforma que eles utilizaram.

2.2.3 Dados Quantitativos

Na fase 1, são considerados duas bases de dados de postagens de fóruns MOOC com quatro anos de diferença uma da outra: a primeira contando com 12 cursos oferecidos em inglês pela universidade de Stanford na plataforma de MOOC edX entre 2013 e 2014; e, a segundo, contando com postagens de 1.074 cursos que estavam sendo oferecidos por várias instituições por todo o mundo em novembro de 2018, também na plataforma edX. A partir daqui, os dados coletados no primeiro conjunto serão rotulados como **D1** e os do segundo, como **D2**. O *Rationale* tratado aqui foi extraída manualmente de **D1** e **D2**.

Os cursos em **D1** abrange três áreas de conhecimento: Ciências Humanas, Medicina e Educação, o que corresponde a 29.604 postagens em fóruns agrupadas em 22.804 discussões.

Essas discussões ainda podem ser caracterizadas da seguinte forma: 20.268 têm apenas uma postagem; 2.208 com 2 a 5 postagens; 179 com 6 a 10 postagens; e 48 com mais de 11 postagens⁶.

Em contrapartida, **D2** não é composto de registros históricos de fóruns do MOOC gravados automaticamente, mas sim por aqueles coletadas diretamente da interface da plataforma edX – totalizando 1.074 cursos das mais variadas áreas de conhecimento. Destes, apenas os cursos que mantêm um canal aberto para reclamações e sugestões foram utilizados. Portanto, apenas 87 cursos contribuíram para **D2**.

2.2.4 Mineração de *Rationale*

D1 e **D2** foram minerados em busca de *Rationale* separadamente, como resumido a seguir.

Minerando D1

D1 foi vasculhado manualmente usando um conjunto de palavras-chave iniciais que evoluiu por *Snow Balling*. Isto é, as primeiras pesquisas foram realizadas com base no conjunto inicial de palavras-chave e, a partir dos resultados, novas palavras-chave foram incorporadas a esse conjunto. Os conjuntos inicial e final de palavras-chave foram: **Palavras-chave Iniciais** – *platform, thread, post, discuss(ion), forum, navigat(e)(ion), user, student, MOOC*; **Palavras-chave Finais** – *helpful, platform, module, review(er), learn(ing)(er), thread, post, discuss(ion), forum, navigat(e)(ion), user, student, friendly, MOOC, Coursera, (Open)edX, interface, comment, idea, categor(y)(ies), frustrat(e)(ing)(ion), disappoint(ing), maint(aining)(einance), point, automatic(ally), feedback, system, (dis)(en)courage, (un)read, assess(ment)*.

Para o *Rationale* minado para ser “aceito” como candidato a requisito de manutenção e/ou atualização da plataforma educacional utilizada, ele tinha que dizer respeito a um *Use Case* real ou *User Story* – tais como definidos em metodologias de desenvolvimento ágil de software tais como XP ou Scrum [12; 74]. O *Rationale* aceito recebeu uma pontuação:

- 1 ponto: Se o usuário a postar demonstrou algum conhecimento técnico sobre o problema que está sendo tratado;

⁶Para mais informações a respeito destes dados, acesse: <https://datastage.stanford.edu/StanfordMoocPosts/>.

- 1 ponto: Se seus comentários correspondem àqueles feitos por outros usuários;
- 2 pontos: Se trata-se de crítica a um recurso já existente;
- 4 pontos: Se trazer uma sugestão para a manutenção e/ou atualização do software da plataforma.

Minerando D2

Os critérios que um determinado curso teve que atender para ser selecionado para **D2** foram: *ser marcado como ‘current’ ou ‘archived’, ser aberto, ter um fórum e um canal oficial no fórum para críticas e sugestões*. Tendo o curso sido selecionado para **D2**, seu *Rationale* aceito atendeu aos mesmos critérios e pontuação que aqueles oriundos de **D1**. Um total de 47 postagens foram coletadas como *Rationale*, fruto de 87 cursos.

2.3 Resultados

2.3.1 Fase 1

O resultado da fase 1 da metodologia é composto pelo *Rationale* válido minerado embutido em comentários de fóruns. Para ilustrar, a Tabela 2.1 apresenta dois casos de *Rationale* extraídos de *D1* e suas respectivas especificações de requisitos estruturadas de acordo com os procedimentos de manutenção de software, cujo formato adaptamos de [77]. As postagens na Tabela 2.1 representam *Rationale* sobre Usabilidade (requisito 1) e Apresentação de Mensagens em Discussões (requisito 2). Observe que o requisito 1 é uma sugestão, já o requisito 2 é uma crítica. Como mostrado, o *Rationale* mais completo é aquela que traz argumentos baseados em sugestões, não apenas em críticas. As postagens com críticas apresentam lacunas que devem ser preenchidas pelos engenheiros de requisitos de software.

A Tabela 2.2 categoriza o *Rationale* extraído de **D1** e **D2**. Como pode ser observado na Tabela 2.2, o padrão de *Rationale* mudou passada a lacuna de quatro anos entre **D1** e **D2**. Especificamente, a demanda por manutenção/atualização em Gamificação – uma das categorias de *Rationale* – diminuiu enquanto *Peer Assessment* passou a provocar mais *Rationale*; as demais categorias apresentam quase a mesma taxa de *Rationale* nos dois conjuntos de dados.

As demais fases da metodologia nos fornecem material para analisar essas observações com mais detalhes.

2.3.2 Fase 2

As variações (aumentos ou diminuições) em diferentes categorias de *Rationale* ao longo do tempo acontecem porque algumas demandas podem ter sido atendidas por novas funcionalidades ou correções em novas versões do software da plataforma MOOC. Para descobrir quais dessas categorias de *Rationale* tiveram suas demandas supridas por novas versões da plataforma, as *Release Notes* da plataforma edX foram examinadas. É útil para nossa questão de pesquisa **Q1** analisar se o *Rationale* no fórum realmente reflete demandas reais dos usuários, verificando se a incorporação dessas demandas cessa aquele *Rationale*.

Parte do *Rationale* observado em **D1** foi resolvida com base nas *Release Notes* e diminuiu ou desapareceu no momento em que **D2** foi observado. Esse é o caso do *Rationale* sobre Gamificação e Apresentação de Discussões. Essa última categoria, embora presente em **D2**, teve seu foco de *Rationale* modificado ao longo do tempo. Um exemplo de correlação entre uma alteração identificada nas *Release Notes* e o *Rationale* de **D1** é apresentado na Tabela 2.3. Este *Rationale* não está mais presente em **D2** e coincide com alterações feitas na plataforma no período.

2.3.3 Fase 3: Entrevistas

Para analisar a relevância real do *Rationale* minerado, foram entrevistados nove profissionais de diferentes áreas do conhecimento com experiência de ensino. Cada entrevistado deve ter tido uma experiência prática significativa em MOOC há dois anos no máximo. Os entrevistados foram recrutados pelo método *Snow Balling*, onde o primeiro participante indica o próximo e assim por diante. As entrevistas foram realizadas pessoalmente ou por videoconferência. Cada entrevistado deve ter concluído pelo menos dois MOOCs com duração de um ano ou mais. (Isso era necessário para mostrar comprometimento do participante com o curso e aumentar as chances dele ter utilizado mais ferramentas e recursos pedagógicos.) Todos os participantes vieram do Brasil, atuam em diferentes áreas de pesquisa e tem, em média, 30 anos de idade, com um desvio padrão de $\sigma = 5$ anos. A Tabela 2.4 traz mais

Tabela 2.1: Exemplos de especificações de requisitos fruto do *Rationale* extraído de **D1****Requisito 1**

Descrição	Botões separados “Save” e “Final Check” nas perguntas dos <i>quizzes</i> . Cada botão envia um comando diferente para o servidor. O botão “Save” envia um comando apenas para salvar o status atual e o botão “Final Check” envia comandos para salvar e sair simultaneamente.
Entrada	Qualquer pergunta do <i>quizz</i> total ou parcialmente preenchida
Ação	Um comando diferente para cada botão. O botão “Save” envia um comando apenas para salvar o status atual e o botão “Final Check” envia os comandos “Salvar” e “Sair”.
Pré-condição	O usuário optou por responder ao <i>quizz</i> e atendeu aos critérios anteriores de créditos e pagamento.
Pós-condição	O estado da questão do <i>quizz</i> é atualizado para “Save” e, se o botão “Final Check” foi pressionado, o processo do <i>quizz</i> é concluído e a interface é sai do <i>quizz</i> aberto e passa a exibir o <i>quizz</i> concluído.
Rationale^a	<i>"It would be helpful if each question had a separate Save and Final Check button"</i>
Tipo de Rationale	Sugestão

Requisito 2

Descrição	O usuário poderá seguir as discussões em que ele postar e qualquer outra discussão que ele marcar como “following”. E-mails diários com novas notificações de postagens devem ser enviados a ele.
Entrada	Novas postagens escritas pelo usuário ou alguma discussão marcado como “following”.
Ação	Atualize o status da respectiva discussão para “following” e um e-mail será enviado notificando cada nova postagem desta respectiva discussão.
Pré-condição	O usuário abriu uma discussão em sua interface.
Pós-condição	The user posted in the thread or marked it as “following”. Whichever situation shall change the status of thread to “following” O usuário postou no tópico ou o marcou como “following”. Qualquer destas situações devem alterar o status da discussão para “following”.
Rationale^a	<i>"If I participate in a discussion, it's nearly impossible for me to find it again later. I don't get notified if anyone responds to a post of mine, and I can't find a way to follow a discussion that interests me."</i>
Tipo de Rationale	Crítica

^a Essa linha transcreve o *Rationale* espontânea do usuário diretamente do conjunto de dados, sem nenhuma edição.

Tabela 2.2: Frequência do *Rationale* e sua natureza.

-2* <i>Categoria de Rationale</i>	D1		D2	
	Sugestões	Críticas	Sugestões	Críticas
<i>Peer assessment</i>	2	1	9	1
Classificação de Discussões ^a	2	0	0	1
Apresentação de Discussões ^b	5	0	3	2
Gamificação	9	1	2	0
<i>Feedback</i> ^c	2	4	2	3
Multimídia	0	0	0	1
Verificação de Plágio	1	0	0	0
Usabilidade	17	2	11	4
Aprendizado em Grupo ^d	1	0	0	0

a De acordo com o título e as opções que descrevem o assunto da discussão.

b Refere-se à ordem, facilidades para acompanhar a discussão, a interface de apresentação, etc.

c Manual ou automático.

d A organização dos alunos em grupos colaborativos, em vez de apenas colaboração envolvendo todos juntos.

detalhes a respeito destes perfis.

Questionário e Entrevistas

Um questionário em português baseado nas categorias de *Rationale* resultantes da Fase 1 (vide Seção 2.2.4) foi aplicado usando a plataforma Google Forms (<https://www.google.com/intl/pt-BR/forms/about/>) e está disponível em <http://bit.ly/2ZfOrrw>. Foi solicitado que fosse respondido remotamente (online) 24 horas antes da entrevista propriamente; o questionário respondido foi então usado como um guia para a entrevista. O atraso de 24 horas concedeu tempo para o entrevistado refletir sobre as categorias de *Rationale*. Cada pergunta questionava a respeito das impressões do entrevistado sobre a eficácia e as perspectivas de uma determinada categoria de *Rationale* (vide Tabela 2.2). As respostas estão em uma escala Likert de cinco níveis [55; 13]: 1 – concordo fortemente; 2 – concordo; 3 – indiferente; 4 – discordo; e 5 – discordo fortemente. Algumas perguntas usam uma variação dessa escala, sem perda de amplitude ou equivalência. O questionário é, portanto, quantitativo. As definições de cada categoria

Tabela 2.3: Exemplo de *Release Note* da plataforma edX atendendo a demandas na forma de *Rationale* em **D1**

Data da Release	Semana de 18 de Abril de 2016
Título da Release	<i>Sorting by Date for Discussion Posts</i>
Descrição da Release	<i>A problem with the method used to sort discussion posts in web browsers is now fixed. Previously, when a learner sorted discussion posts by recent activity in a web browser, the discussion posts were sorted by creation date rather than last activity date. Discussion posts are now sorted correctly. (TNL-3818).</i>
Rationale 01	“The next time Stanford talks to the people at OpenEdX, I suggest asking that they design the ability to sort threads by those which have unread posts. I, for one, follow a very large number and have to keep viewing more threads to make sure there aren’t some with posts from latecomers. Thanks!”
Rationale 02	“I didn’t think of this when filling out the exit survey, so I’ll post it here. I would like an easier (or automatic) way to follow posts and comments – you can manually star a topic you want to follow, but there is no easy way to tell if someone has replied to one of your posts within a topic, especially if you forgot to star it.”
Categoria de Rationale	‘‘Apresentação de Discussões’’.

Tabela 2.4: Perfis dos Entrevistados.

	Idade	Escolaridade	Anos decorridos \approx	Performance^a	Plataforma	Área de Atuação
1	30-39	Superior	0.3	C	Saitestore	Medicina
2	30-39	Superior	1.0	B+	AVASUS	Medicina
3	21-29	Superior	1.0	B+	Unasus	Medicina
4	30-39	Mestrado	1.5	B	Coursera	Design
5	30-39	Superior	1.0	C	Futurelearn	Direito
6	30-39	Phd	2.0	B	Coursera	Design
7	30-39	Superior	2.0	B+	edX	Artes Plásticas
8	60+	Phd	0.6	A+	Coursera	Engenharia Elétrica
9	30-39	Superior	2.0	A	edX	Cultura

a Autoavaliação de desempenho no último MOOC finalizado, o qual foi usado como parâmetro.

na Tabela 2.2 foram oferecidas apenas durante a entrevista, mas algumas perguntas no questionário buscavam deduzir o conhecimento básico do entrevistado sobre tais categorias. As questões restantes foram elaboradas para promover a discussão sobre as categorias de *Rationale* durante a entrevista. As entrevistas foram, portanto, semiestruturadas e conduzidas pelos autores.

Essas entrevistas foram realizadas esperando que *os entrevistados reproduziam espontaneamente os temas de Rationale mais frequentemente extraídos dos fóruns* e duravam entre 40 a 90 minutos cada. Entrevistas posteriores tendiam a durar mais e eram mais proveitosas, uma vez que os entrevistadores ganhavam experiência a cada nova entrevista.

A princípio, os entrevistados foram perfilados perguntando sobre seu comportamento na educação regular – na sala de aula – e sua última experiência em MOOC (vide Tabela 2.4) para combinar perfis estudantis com opiniões educacionais. Todas as entrevistas foram gravadas, com o consentimento do entrevistado, para análises posteriores. Em seguida, tais entrevistas concentraram-se em estimular o apontamento consciente, por parte dos entrevistados, de *Rationale* para especificação de ferramentas de software educacional, questionando suas experiências, casos de uso reais, reclamações e sugestões. Para fazer isso, temas de *Rationale* minerados em **D1** foram comparados àqueles minerados em **D2** durante as entrevistas.

Discussão

Identificaram-se diferentes perspectivas, espalhadas por diferentes áreas do conhecimento: “Medicina”, “Ciências Sociais” e “Ciências Exatas”. Por exemplo, ferramentas automatizadas de *Feedback* causam muita resistência nos educadores em Ciências Sociais, conforme percebido nas entrevistas. Essa observação está alinhada com o *Rationale* minerado nos dois conjuntos de dados cujo montante foi aproximadamente constante e relativos abordagens educacionais mais voltadas a subjetividade e pessoalidade. Estranhamente, apesar da popularidade de ferramentas *Multimídia*, ele estava em *Rationale* do usuário apenas uma vez e só em **D2**. Essa observação foi confirmada pelos entrevistados, que não apresentaram *Rationale* para essa categoria, mas apenas opiniões neutras em geral, o que pode indicar satisfação com o que já está disponível. Por outro lado, a categoria *Usabilidade* estimula tanto *Rationale* que não parece viável para os desenvolvedores de software levar todos em consideração. Muitos comentários parecem relativos a uma curva de aprendizado da plataforma, mas não a

problemas reais. Esses exemplos apoiam o uso de *Rationale* como fonte de requisitos e diretrizes de inovação para plataformas colaborativas, conforme confirmado pelos comentários dos entrevistados sobre as categorias de *Rationale* mineradas.

Além de Usabilidade, as únicas áreas que provocaram *Rationale* espontâneo por parte entrevistados foram Apresentação de Discussões e *Peer Assessment*. Este *Rationale*, representada principalmente por críticas, pode ser exemplificada por “..maybe a meta assessment works to fix low experience peer assessments..” para *Peer Assessment*; e “..threads with some high up-voted post would be exposed first..” para Apresentação de Discussões.

Peer Assessment foi um recurso elogiado pelos entrevistados com experiência em Exatas (como Engenharia) e Ciências Médicas, por sua viabilidade e escalabilidade. Os entrevistados das Ciências Sociais, no entanto, acham que esse método deve ser usado apenas em avaliações intermediárias devido à sua “imprecisão”. Eles recomendam cautela na aplicação, o que pode ser consequência do caráter subjetivo e de argumentação complexa em sua área. Saindo deste tema, um dos entrevistados discordou do uso de *Gamification*, provavelmente porque algumas tarefas gamificadas tenham sido utilizadas para avaliar os alunos em seu curso mais recente. Já quanto a *Aprendizagem em Grupo* e *Verificador de Plágio*, os entrevistados foram sempre vagos ou genéricos, sugerindo que possar vir a ser meros *outliers*. Embora tais entrevistas tenham nos levado a conclusões inéditas neste trabalhos, algumas reiteraram àquelas já alcançadas por meio da mineração de *Rationale*, especialmente no que se refere a recursos em *Peer Assessment*.

Ameaças à Validade

O número reduzido de entrevistados poderia influenciar o estudo. Contudo, para que um estudo qualitativo seja viável, o conjunto de *stakeholders* não deve ser muito grande. Outra preocupação diz respeito ao método de mineração manual aplicado na Fase 1 (vide Seção 2.2.4) da metodologia. Tal estratégia pode se mostrar imprecisa e ameaçar a fase principal do estudo. Esta ameaça foi aceita, tendo em mente que este estudo busca, primeiramente, evidências de que a mineração de *Rationale* tem o potencial de funcionar como fonte de requisitos para manutenção e atualização da plataforma de MOOC, para então identificar diretrizes de inovação em ferramentas de software para tais plataformas. Portanto, faz parte da metodologia de exploração, anterior ao estudo final. Assim, ainda passará por muitas outras etapas de

validação.

Embora um típico *survey* de fim de curso em MOOC possa substituir a mineração espontânea de *Rationale* do usuário, esses cursos tendem a ter altas taxas de abandono e muitos usuários que usem o recurso do fórum para compartilhar *Rationale* podem não persistir até o fim de seus cursos. O *Rationale* minerado dos fóruns oferece uma fonte contínua de requisitos de software da plataforma educacional.

2.4 Conclusões e Implicações

Este Capítulo 2 do estudo reportou a investigação do potencial da mineração da *Rationale* espontâneo do usuário em publicações em fóruns como estratégia para obter requisitos de manutenção e atualização do software MOOC. Embora apenas uma plataforma de MOOC tenha sido examinada diretamente, os resultados foram relevantes e, em alguns casos, surpreendentes – como descobrir que ferramentas multimídia geram pouco *Rationale*. Nossas observações concluíram que o *Rationale* minerado, de acordo com as opiniões dos entrevistados, é uma fonte de requisitos viável para software educacional. Embora fosse interessante já ter sido aplicado algum modelo de mineração automático de *Rationale*, isto não invalida a relevância dos resultados aqui alcançados, sobretudo no contexto de MOOC.

Especificamente para apontar problemas em Usabilidade, essa estratégia parece ser um método fértil de coleta de requisitos. Outras categorias de *Rationale* como requisitos também são promissoras, como “Apresentação e Classificação de Discussões”. Questões relacionadas à Multimídia, surpreendentemente, receberam menos atenção dos usuários em suas conjecturas. No entanto, os usuários (neste caso, os estudantes do MOOC) só podem racionalizar sobre o que sabem ou viram ocasionalmente. Consequentemente, muitas novas tecnologias e ferramentas das quais ainda não estão cientes, obviamente não lhes servem de parâmetro. Assim, embora não tenham sido alvo de *Rationale*, ferramentas Multimídia foram descartadas como tema de pesquisa apenas porque outras categorias de *Rationale* trouxeram diretrizes mais claras.

Para trabalhos futuros em *Rationale*, um método para extrair e analisar automaticamente o *Rationale* do usuário dos fóruns de MOOC forneceria uma maior confiança nos resultados. Além disso, comparar (automaticamente, desta vez) o *Rationale* minerado a séries de

commits de mais de uma plataforma de software MOOC em vez de apenas *Release Notes*, forneceria melhores conclusões sobre a utilidade de *Rationale* como fonte de requisitos para novas versões dessas plataformas.

Para este trabalho, a principal contribuição desta capítulo está em constatar que novas formas de Apresentação e Filtro de Discussões demonstraram relevância significativa junto ao usuário. Isto se apresentou ao observarmos um volume constante e inter relacionado de *Rationale* associado a este tema em todas as etapas desta pesquisa. “Apresentação e Filtro de Discussões” apresenta-se como forma de facilitar a busca por informação útil em fóruns com muitos usuários.

Capítulo 3

Consenso

Fóruns de discussão e debates online são importantes elementos de suporte ao aprendizado, onde o debate oferece ao estudante maior clareza acerca de um dado tema. Em tais fóruns, quando uma discussão é aberta, é natural criar-se uma expectativa de conclusão, seja esta fruto de um consenso imperfeito ou unânime. Diante da dificuldade de se alcançarem conclusões unânimes em certos temas de debate, o consenso – conforme Definição 3 – ganha força como instrumento de avaliação de qualidade da solução proposta.

A Definição 3 permite modelar desde uma concordância parcial até a unanimidade. Esta definição se baseia naquela de *Soft Consensus* [37] que acomoda incertezas.

Conforme levantado no Capítulo 1, 66% das postagens mais urgentes são opiniões e, com base em apontamentos apresentados nas conclusões do Capítulo 2, buscar informação útil em um fórum MOOC pode se tornar muito difícil a medida que o volume de estudantes cresce. Diante de tanta informação fruto de tantas opiniões distintas, pode se tornar difícil confiar nas conclusões apontadas em discussões neste ambiente. Neste contexto, os instrutores devem ficar atentos à qualidade do conteúdo oferecido nos fóruns. A presença de dois ou mais grupos de consenso destoantes em uma mesma discussão pode apontar uma necessidade de intervenção do instrutor. Assim, estudar o consenso pode ser uma saída viável para obter direcionamentos a cerca do quão relacionados estão o consenso e possíveis opiniões de qualidade em discussões.

Embora a pesquisa sobre consenso tenha evoluído e se adaptado aos domínios virtuais, com base em nossos levantamentos, não encontramos pesquisas que apontem se uma decisão consensual é de fato uma decisão de qualidade em ambientes de discussões online. Por con-

jectura, podemos apontar a dificuldade em se obter dados de discussões online para análise, cujas conclusões estejam classificadas em função da qualidade. Neste estudo, tal dificuldade foi contornada utilizando-se dados oriundos de fóruns de MOOC, tomando o apoio do corpo de instrutores à conclusão consensual como parâmetro de qualidade.

Este capítulo do trabalho apresenta um experimento *ex-post facto* que mede a correlação entre consenso e qualidade em discussões online, e especificamente no ambiente educacional (MOOC). Para tanto, utilizou-se a mesma base de dados **D1** de postagens de fóruns educacionais oriundos de 12 MOOC diferentes oferecidos pela Universidade de Stanford (www.stanford.edu) na plataforma edX (www.edx.org). O consenso entre estas postagens em cada discussão foi calculado com base em três diferentes métricas de distância textual. O ruído introduzido por tais métricas foi mensurado a partir da divergência entre estas. Já a qualidade das conclusões de maior consenso foi definida em função do apoio do instrutor a opinião majoritária.

Após conhecer o consenso e a qualidade das discussões por toda a base de dados, procedeu-se à análise estatística de tais dados. Para tanto, com base em Regressão Logística, encontraram-se indícios de que o consenso tem efeito negativo na qualidade da decisão. Nossos resultados confirmam os resultados obtidos na literatura [39; 41], mas agora no contexto de educação online.

Este capítulo deste documento descreve um estudo originalmente publicado separadamente em um artigo científico no Simpósio Brasileiro de Informativa para a Educação – SBIE – em 2019 [24].

3.1 Contexto e Estado da arte

Embora o consenso seja um objetivo valorizado em discussões, o simples debate entre opiniões divergentes já é um importante princípio do aprendizado [43]. A divergência leva à reflexão que promove o entendimento profundo [30; 82]. O conflito motiva a colaboração [30]. Todavia, tomar uma conclusão consensual como sendo de qualidade pode ser uma decisão precipitada. [39] questionou a necessidade real de consenso em processos de tomada de decisão com base em critérios de qualidade e satisfação entre os interlocutores. Em seu trabalho, Hirokawa utilizou um experimento controlado em ambiente educacional para

avaliar a correlação entre consenso e qualidade. Seus resultados apontaram que decisões de qualidade tendem a emergir à medida que o grupo aborda a discussão por meio de uma metodologia sistemática e “vigilante”, cenário pouco provável em fóruns online. Porém, segundo [41], de modo geral, o grupo tende a aceitar melhor decisões consensuais. Já Bass [11] testou a hipótese de que a qualidade nas decisões aumenta à medida que o consenso entre os participantes aumenta. Este chegou a conclusão que há correlação positiva neste caso, desde que haja heterogeneidade e motivação no grupo.

Sistemas de Apoio à Decisão em Grupo – GDM (do inglês, *Group Decision Making*), tem sido a principal área de aplicação de pesquisas científicas ligadas ao consenso mediante incerteza [16]. O advento da Web 2.0 as comunidades online tornaram o ambiente de discussões ainda mais flexível [37; 7]. Deste modo, alcançar o consenso com conclusões de qualidade e satisfatórias em discussões não estruturadas como fóruns online, comunidades de debate e redes sociais, é um desafio que recebe bastante atenção na literatura [65]. A análise de discurso tem mostrado resultados significativos em áreas como marketing [38; 73; 84; 75] e política [80; 67; 91; 68]. Em análise de discurso, o consenso é frequentemente um importante argumento de relevância [91; 73].

Em [67] é apresentado um modelo de predição de opiniões vencedoras em debates, além de propor um agente automático de argumentação. Embora o consenso não seja o foco em tal estudo, a opinião vencedora em um debate nada mais é do que a opinião de maior consenso parcial. Já [91] apresenta uma abordagem que agrega opiniões em núcleos de consenso com o intuito de diferenciar rumores de fatos em redes sociais. [73] explora uma série de aspectos em debates online visando identificar elementos úteis à detecção de convergência entre opiniões. Embora o consenso seja tema tão recorrente na literatura, um dos últimos trabalhos a medir o consenso em função da qualidade parece ter sido [39].

Aqui, como será descrito no próximo capítulo, o nível de consenso em discussões de fóruns MOOC é calculado e avaliado face à qualidade das postagens. Para que se possa calcular o nível de consenso automaticamente, é preciso medir a similaridade entre pares de opiniões. Para tanto, tomam-se como base as três classes de distância textual em [35]. Assim, utilizaram-se três algoritmos distintos (um para cada classe) selecionados em função de critérios de viabilidade e eficácia. O estudo da relação entre consenso e qualidade em um ambiente educacional, aliado à automação do cálculo de consenso, é a principal contribuição

desta etapa do trabalho, no sentido de se automatizar a tomada de decisão, aproveitando (ou não) o consenso em fóruns como suporte pedagógico a cursos online do tipo MOOC.

3.2 Design da Pesquisa

Esta etapa do estudo avalia a presença de correlação entre o grau de consenso alcançado pela opinião dominante (mas, não necessariamente unânime) em uma discussão e a qualidade de tal opinião avaliada com relação ao tema proposto, no ambiente educacional de fóruns de discussão em MOOC. Mais especificamente, *aqui é apresentado um experimento ex-post facto objetivando expor a correlação entre o consenso acerca da opinião dominante e o apoio do corpo de instrutores do MOOC àquela opinião*. Aqui, tal apoio é tomado como parâmetro de qualidade. Assim, sinteticamente, a questão de pesquisa a ser respondida nesta etapa deste estudo – conforme já mencionado na Seção 1.3 – é a seguinte:

- Q2: Em fóruns de discussão educacionais, existe correlação entre o nível de consenso e a qualidade da opinião consensual?

Assim definimos a hipótese nula como a ausência de correlação entre o consenso, medido como um grau de concordância entre as postagens, e a qualidade, tomando qualidade como sendo o apoio dos instrutores a opinião majoritária. Já a hipótese alternativa apontará para uma correlação, seja positiva ou negativa, entre estas duas grandezas.

3.2.1 Os Dados

Foram utilizados os dados de **D1** – vide Seção 2.2.3 – descartando discussões com menos de duas postagens.

Tratamento Manual

Os dados foram utilizados originalmente por Agrawal et al [1]. Por razões éticas, Agrawal et al filtraram todas as referências nominais aos autores de cada postagem, preservando assim seu anonimato. Em seguida, eles ponderaram manualmente as postagens dos estudantes por urgência [1, 7] – i.e., a faixa de números inteiros ≥ 1 e ≤ 7 , onde 1 representa a menor

urgência e 7, a maior; positividade [1, 7]; e, confusão [1, 7]. Com “urgência” sendo a necessidade urgente de intervenção de um instrutor; “positividade”, o quão afirmativa é a postagem; e, “confusão”, o nível de incerteza associado a afirmação na postagem. Deve-se frisar que o campo “confusão” parece ter sido mensurado em escala invertida, medindo na verdade a “convicção” em cada postagem. Em seguida, a fim de normalizar a análise, mapeamos urgência, positividade e convicção para a faixa de números decimais [0, 0; 1, 0].

As postagens foram ainda classificadas como pergunta {0, 1} – não ou sim, respectivamente; resposta {0, 1}; ou, opinião {0, 1}. Então, com os dados já previamente tratados, classificamos manualmente cada postagem como sendo de autoria ou não de um instrutor {0, 1}. Num primeiro momento, tal autoria foi identificada por leitura diagonal, postagem por postagem. A cada nova postagem de instrutor identificada, todas as demais postagens daquele instrutor eram encontradas utilizando o campo ‘forumpostid’, que identifica usuários do fórum.

Tratamento Automático

Antes de efetuar os cálculos de consenso foi efetuada um tratamento de limpeza no campo ‘text’ dos dados **D1**, contendo o texto da postagem propriamente. Inicialmente dividiu-se as frases em *tokens* – tomemos um *token* como qualquer conjunto de caracteres não espaçados. Em seguida eliminou-se *tokens* numéricos, *links*, fórmulas matemáticas e pontuação. O terceiro passo consistiu da remoção de *stop words* – palavras muito frequentes e com pouco valor semântico como ‘de’, ‘para’, etc – a partir da biblioteca Java Opennlp (opennlp.apache.org) da Apache Foundation (www.apache.org). Além disto, para cada discussão, a média e o desvio padrão de convicção, urgência e positividade da Seção 3.2.1 foram calculados para possível utilização na análise estatística e armazenados em campos de mesmo nome com os sufixos ‘_mean’ e ‘_sd’.

3.3 Mensurando Consenso

Estendendo a Definição 3 temos o conceito de *Soft Consensus*, onde o consenso é definido a partir de modelos capazes de assimilar certo grau de incerteza [37]. Tal potencial diz respeito especificamente ao uso de quantificadores linguísticos e Lógica Difusa. A maioria destes mo-

delos segue metodologias rígidas divididas em rodadas de discussão e geralmente moderadas por um agente externo [16], embora existam alternativas mais flexíveis [7]. Como estamos lidando com discussões já finalizadas e agentes dispersos no tempo, além de pouco comprometidos com o debate, mesmo metodologias mais flexíveis ainda necessitam de certos ajustes para que se enquadram ao nosso propósito. Em face a isto, tomamos toda a discussão como sendo uma única rodada de debate e cada postagem como um agente individualizado. Um vez que os dados utilizados já discriminam opiniões, perguntas e respostas – vide Seção 3.2.1 – ainda tomamos as postagens discriminadas como “opinião” como as alternativas disponíveis para consideração. A distância textual entre uma postagem e cada uma das opiniões mede o nível de apoio que aquela postagem oferece àquela opinião.

Métricas de distância textual podem introduzir ruído na análise, e de acordo com [35] existem três classes principais de algoritmos de similaridade textual. Diante disto, analisou-se o consenso em função de três algoritmos diferentes de distância textual – conforme recomendado em [35] – um para cada classe, comparando os resultados. A primeira classe, e mais simplista delas, é a similaridade baseada em caracteres. Optamos por adotar a distância de edição ou distância de Damerau-Levenshtein, por ser a mais difundida. A segunda classe de algoritmos mede a distância semântica com base em Corpus textuais de vários idiomas, dentre eles o Inglês. Para esta, utilizou-se o algoritmo DISCO [49], por se tratar de uma tecnologia bem estabelecida e com implementação estável. A terceira e última classe de algoritmos traz algoritmos baseados em redes semânticas de palavras e expressões. Para esta classe, aplicou-se um algoritmo baseado em Wordnet [62], por dispor de implementação estável.

Nosso processo semi estruturado para cálculo de consenso segue uma sequência simples: Identificam-se as opiniões dentre as postagens de uma discussão; calcula-se a matriz de preferências comparando todas as opiniões duas a duas, conforme descrito na Seção 3.3.2; a partir da matriz de preferências, calculam-se similaridades entre tais preferências e ordenam-se as opiniões tanto no âmbito geral quanto para cada postagem individualizada; e, por fim, verifica-se o nível de consenso.

3.3.1 Distância Textual

Extrair opiniões e preferências a partir de dados textuais é um problema que pode ser abordado com algoritmos de distância textual. A seguir, selecionamos uma abordagem para cada classe de distância textual, conforme [35], onde dividiram-se os algoritmos de distância textual em três classes distintas. Aplicamos três métricas de distância de classes diferentes, isto para mensurar o ruído introduzido por estas na análise.

Damerau-Levenshtein: A distância de Damerau-Levenshtein [27; 54], também conhecida como distância de edição, é a diferença entre duas cadeias de caracteres. Tal diferença é medida em quantos caracteres precisam ser eliminados de ambas para que estas se tornem iguais, conforme Equação 3.1:

$$lev_{a,b}(|a|, |b|) = \begin{cases} 0, & \text{se } a = b, \\ \max(|a|, |b|), & \text{se } \min(|a|, |b|) = 0, \\ \min \begin{cases} lev_{a,b}(|a-1|, |b|) + 1 \\ lev_{a,b}(|a|, |b-1|) + 1 \\ lev_{a,b}(|a-1|, |b-1|) + 1 \end{cases} & \text{caso contrário.} \end{cases} \quad (3.1)$$

Sendo a e b duas cadeias de caracteres e as operações $a - 1$ ou $b - 1$ correspondem à eliminação de um caractere.

DISCO: DISCO [49] pode ser descrita como relação distributiva ou semelhança distributiva, simultaneamente. Duas palavras são distributivamente semelhantes se possuem muitas palavras co-ocorrentes semelhantes nos mesmos papéis sintáticos. Por outro lado, duas palavras são distributivamente relacionadas quando estão imersas no mesmo contexto, ou no mesmo conjunto livre de palavras co-ocorrentes. Neste sentido, DISCO não utiliza ferramentas sintáticas para correlacionar palavras por semelhança distributiva (tal como uma Wordnet), porém faz aproximações livres de contexto de tais relações. Para tanto [49] utilizou métricas para mensurar informação mútua baseadas em [56].

Wordnet: A Wordnet é uma base de dados textual semântico-sintática [62]. Aqui palavras são separadas por grupos sintáticos como verbos, pronomes, advérbios e adjetivos, e agrupadas em conjuntos de sinônimos, os chamado *synsets*. Uma mesma palavra pode constar em mais de um *synset*. *Stop words* não estão incluídas. *Synsets* são conectados uns

aos outros por relacionamentos semânticos dentro de seus grupos sintáticos e a força de relacionamento entre duas palavras (elementos destes *synsets*) é calculada em função destes relacionamentos.

3.3.2 *Soft Consensus*

Em sistemas de apoio ao consenso, sobretudo em ambientes online, opiniões e preferências tendem a apresentar incertezas e imprecisões que tem sido abordadas usando a Lógica Difusa [37; 16]. Porém, diferenciando-se de discussões convencionais, os fóruns online agregam incerteza até mesmo no processo de discussão. Sendo assim, no âmbito deste experimento, é necessário fazer adaptações nos modelos que tratam consenso com incerteza [65]. No cálculo de consenso que propomos, postagens de alunos e instrutores estão definidas em $P = \{p^1, \dots, p^m\}$, sendo m o total de postagens por discussão. Suas respectivas preferências são extraídas da proximidade textual entre cada postagem e cada opinião disponível $X = \{x^1, \dots, x^n\}$ – sendo n o total de opiniões por discussão – seguidas por comparações pareadas entre estas proximidades. Por simplicidade, mas sem perda de generalidade, aqui assumem-se todos os pesos das postagens/opiniões de instrutores e alunos como iguais¹.

Definição 4. Tomando X como o conjunto de opiniões e P como o conjunto de postagens da discussão, a função $\Delta_{P,X}^{-1}$ é definida como uma função de similaridade dada entre uma postagem $p \in P$ e uma opinião $x \in X$, sendo $\Delta_{P,X}^{-1} : P \times X \rightarrow [0, 1]$.

Definição 5. O conjunto de relações de preferência $M = \{\mu^1, \dots, \mu^u\}$, sendo u o total de preferências de uma dada discussão, pertence ao conjunto das relações $\Delta^{-1}P, X \times \Delta^{-1}P, X$ e é definida por $M : \Delta^{-1}P, X \times \Delta^{-1}P, X \rightarrow [0, 1]$.

Assim, podemos entender $\mu_{i,j}^p \in M$ como a relação de preferência/precedência entre as opiniões x^i e x^j em função de p , sendo $\mu_{i,j}^p = \mu(\Delta^{-1}p, x^i, \Delta^{-1}p, x^j) (\forall p \in P, \forall x \in X, \forall i, j \in \{1, \dots, n\})$. Diante disto, para identificar o grau de consenso da discussão, devemos calcular quatro níveis de apoio – relação de preferência – entre o conjunto de postagens e o de opiniões disponíveis: 1) entre pares de opiniões; 2) entre opiniões e postagens; 3) entre postagens e opiniões; e, 4) geral [65].

¹Os detalhes matemáticos para o cálculo de consenso feito nesta pesquisa estão disponíveis em https://github.com/tiago-clementino/consensus_x_quality

- **Nível 1: Par de Alternativas** – Este nível consiste em gerar uma matriz de similaridade para cada par de postagens em relação a cada par de opiniões $SP^{i,j} = (sp_{k,l}^{i,j}), \forall p^i, p^j \in P (i \neq j), \forall x^k, x^l \in X (k \geq l)$, conforme a equação:

$$sp_{k,l}^{i,j} = \frac{1 + (\mu^{p^i} k, l - \mu_{k,l}^{p^j})}{2} \quad (3.2)$$

Forma-se, assim, um tensor de quatro dimensões $SP_{(m \times m) \times (n \times n)}$ comparando as preferências de todas as postagens duas a duas ($m \times m$) em função às preferências daquela postagem com relação à cada par de opiniões comparadas ($n \times n$). Em seguida, com base em uma função de agregação determinada Φ , obtemos duas matrizes de preferência coletiva 1) comparando pares de opiniões agregando $SP_{(m \times m) \times (n \times n)}$ por postagem

$$sp_{k,l} = \Phi^i(\Phi^j(sp_{k,l}^{i,j})) \forall i, j \in \{1, \dots, m\} \mid i \geq j \quad (3.3)$$

e 2) pares de postagens agregando $SP_{(m \times m) \times (n \times n)}$ por opinião

$$sp^{i,j} = \Phi_k(\Phi_l(sp_{k,l}^{i,j})) \forall k, l \in \{1, \dots, n\} \mid k \geq l \quad (3.4)$$

Aqui usaremos como Φ um quantificador linguístico da lógica Fuzzy que prioriza a concordância à discordância, conforme recomenda [37];

- **Nível 2: Opiniões e Postagens** – Aqui é definido o nível de consenso para cada opinião $cx_k, \forall k \in \{1, \dots, n\}$, calculado em função de todas as postagens P (já agregadas no nível anterior):

$$cx_k = \frac{\sum_{l=1, l \neq k}^n sp_{k,l}}{n - 1} \quad (3.5)$$

- **Nível 3: Postagens e Opiniões** – Este nível de consenso é adotado apenas neste trabalho e define o quão balanceada a discussão está, conforme a equação:

$$cp = \frac{\sum_{i=1}^m \sum_{k=1}^n g(\Delta_{p^i, x_k}^{-1})}{mn} \quad (3.6)$$

Quanto mais próximo de 0,5; mais balanceada a discussão estará;

- **Nível 4: Geral** – O nível final de consenso cg é definido como:

$$cg = cx_k : cx_k \geq cx_l \forall l \in \{1, 2, \dots, n\} \quad (3.7)$$

Em tendo realizado os cálculos destes quatro níveis, dois resultados são especialmente úteis aqui: Com a matriz de preferências agregadas por opinião do Nível 2 podemos identificar qual opinião é mais forte; e o nível de consenso geral – Nível 4 – nos dá o resultado final. Para entender melhor cada um destes resultados, observe no código fonte² a rotina que os processa diretamente. Para tal, observe o método ‘calculeConsensusGeneral’, do tipo ‘ForumThread’, no arquivo ‘/consensus/src/main/java/mooc_forum_analytics/ForumThread.java’.

Para concluir a análise matemática, assume-se que a opinião apoiada pela maioria dos instrutores é a mais acertada – supondo que os instrutores contam com mais experiência no tema de estudo. Seja $a_i(x_k)$ uma função cuja imagem está definida no conjunto $\{0, 1\}$ (tomar como binária tal função se justifica pois, geralmente, apenas um instrutor interfere na discussão, portanto apenas a opinião vencedora interessa) e que representa o apoio da i -ésima postagem de um instrutor à opinião x_k , $k \in \{1, 2, \dots, n\}$. Temos então, que a qualidade da opinião x_k é definida por

$$q_{x_k} = \sum_{i=1}^l \frac{a_i(x_k)}{l} \quad (3.8)$$

Na Equação 3.8, l é a quantidade total de postagens de instrutores para uma certa discussão. Assim, quanto mais apoio dos instrutores, mais qualidade $q \in [0, 0; 1, 0]$ agregada. Além disto, o conceito de consenso utilizado para esta análise está definido em $[0, 0; 1, 0]$, porém acaba restrito a $[0, 5; 1, 0]$, sendo 0,5 o grau de consenso aleatório. Deve-se notar que o consenso abaixo de 0,5 para uma dada opinião significa que há outra opinião contrária acima de 0,5. Para simplificar os cálculos, normalizamos o consenso em $[0, 0; 0, 1]$.

Soft Consensus: Exemplo

Para aprofundar o entendimento a respeito do modelo de consenso utilizado, tomaremos como exemplo um cenário simples com seis postagens e apenas duas opiniões. Sendo x^1 e x^2 as opiniões disponíveis e p^1, p^2, p^3, p^4, p^5 e p^6 as postagens, considere x^1 e x^2 correspondendo às postagens p^1 e p^2 , respectivamente.

- Nível 1: *Par de Alternativas* – Utilizando alguma das três métricas de distância textual,

²no endereço já citado https://github.com/tiago-clementino/consensus_x_quality

tomemos a seguinte matriz hipotética de similaridades entre opiniões e postagens

$$P \times X = S1 = \begin{bmatrix} 1 & 0,0239 \\ 0,085 & 0,85 \\ 0,635 & 0,37 \\ 0,755 & 0,22 \\ 0,35 & 0,585 \\ 0,79 & 0,155 \end{bmatrix} \quad (3.9)$$

Perceba que no nosso caso $S1$ é uma matriz de 6×2 ($m \times n$), sendo seis postagens e duas opiniões disponíveis. A partir de $S1$, definiremos a matriz de preferências $M1$ com base em uma relação de preferência / precedência $\mu_{i,j}^p$ que compara as similaridades entre uma postagem p e duas opiniões x^i e x^j . Para nosso exemplo utilizaremos em $\mu_{i,j}^p$ uma equação similar à Equação 3.2, porém em função de $g(\Delta_{p,x_i}^{-1})$ e $g(\Delta_{p,x_j}^{-1})$, veja abaixo:

$$\mu_{i,j}^p = \frac{1 + [g(\Delta^{-1}p, x_i) - g(\Delta^{-1}p, x_j)]}{2} \quad (3.10)$$

Assim, para este exemplo, a matriz $M1$ resultante é a seguinte:

$$\mu_{i,j}^p = M1 = \begin{bmatrix} \mu_{1,2}^p \mu_{2,1}^p \\ 0,98805 & 0,01195 \\ 0,1175 & 0,8825 \\ 0,6325 & 0,3675 \\ 0,7675 & 0,2325 \\ 0,3825 & 0,6175 \\ 0,8175 & 0,1825 \end{bmatrix} \quad (3.11)$$

Com isto podemos agora calcular SP (Equação 3.12), uma matriz que compara todas as preferências μ entre si, conforme Equação 3.2. No caso geral, na verdade, trata-se de um tensor de quatro dimensões, mas para nosso exemplo (com apenas duas opiniões) três dimensões são suficientes para visualizar $SP_{(6 \times 6 \times 2)}$. Isto ocorre pois as terceira e quarta dimensões representam comparações entre opiniões e como não faz sentido comprar duas opiniões iguais, apenas dois pares de comparações interessam (1×2 e 2×1), podendo assim restringir-se a uma única dimensão. Para termos uma visualização plana, ainda devemos utilizar duas visualizações separadas, uma para

$sp_{1,2}^{i,j}$ e outra para $sp_{2,1}^{i,j}$, veja abaixo:

$$\begin{aligned}
 sp^{i,j}{}_{1,2} = & \begin{bmatrix} - & 0,935275 & 0,676525 & 0,610275 & 0,802775 & 0,585275 \\ - & - & 0,2425 & 0,175 & 0,3675 & 0,15 \\ - & - & - & 0,4325 & 0,62625 & 0,4075 \\ - & - & - & - & 0,6925 & 0,475 \\ - & - & - & - & - & 0,2825 \\ - & - & - & - & - & - \end{bmatrix} \\
 sp^{i,j}{}_{2,1} = & \begin{bmatrix} - & 0,064725 & 0,322225 & 0,389725 & 0,197225 & 0,414725 \\ - & - & 0,7575 & 0,825 & 0,6325 & 0,85 \\ - & - & - & 0,5675 & 0,375 & 0,5925 \\ - & - & - & - & 0,3075 & 0,525 \\ - & - & - & - & - & 0,7175 \\ - & - & - & - & - & - \end{bmatrix}
 \end{aligned} \tag{3.12}$$

A partir de SP (Equação 3.12), com base em uma função de agregação Φ – neste exemplo, para simplificar os cálculos, usaremos a média aritmética – geraremos a

matriz de preferências coletivas agregando SP em função das opiniões:

$$\Phi^i(\Phi^j(sp^{i,j}1, 2)) = \begin{bmatrix} - & 0,935275 & 0,676525 & 0,610275 & 0,802775 & 0,585275 \\ - & - & 0,2425 & 0,175 & 0,3675 & 0,15 \\ - & - & - & 0,4325 & 0,62625 & 0,4075 \\ - & - & - & - & 0,6925 & 0,475 \\ - & - & - & - & - & 0,2825 \\ - & - & - & - & - & - \end{bmatrix} =$$

$$\Phi^i \begin{bmatrix} 0,722025 \\ 0,23375 \\ 0,48875 \\ 0,58375 \\ 0,2825 \\ - \end{bmatrix} = 0,462155$$

$$\Phi^i(\Phi^j(sp^{i,j}2, 1)) = \begin{bmatrix} - & 0,064725 & 0,322225 & 0,389725 & 0,197225 & 0,414725 \\ - & - & 0,7575 & 0,825 & 0,6325 & 0,85 \\ - & - & - & 0,5675 & 0,375 & 0,5925 \\ - & - & - & - & 0,3075 & 0,525 \\ - & - & - & - & - & 0,7175 \\ - & - & - & - & - & - \end{bmatrix} =$$

$$\Phi^i \begin{bmatrix} 0,277725 \\ 0,76625 \\ 0,51166666667 \\ 0,41625 \\ 0,7175 \\ - \end{bmatrix} = 0,5378783333$$

(3.13)

Perceba que os resultados ficam bem próximo de 0,5 (aleatório), o que faz sentido visto que os números deste exemplo foram escolhidos a esmo.

- Nível 2: *Opiniões e Postagens* – Aqui é calculado o nível de consenso cx_k para cada

opinião. Para nosso exemplo, temos apenas duas opiniões disponíveis, portanto cx_k fica restrito a aplicar a Equação 3.5 aos resultados acima (Equação 3.13), onde:

$$\begin{aligned} cx_1 &= \frac{\sum_{l=1, l \neq 1}^2 sp_{1,l}}{2-1} = \frac{0,462155}{1} = 0,462155 \\ cx_2 &= \frac{\sum_{l=1, l \neq 2}^2 sp_{2,l}}{2-1} = \frac{0,5378783333}{1} = 0,5378783333 \end{aligned} \quad (3.14)$$

- **Nível 3: Postagens e Opiniões** – Aqui, com base na Equação 3.6, verifica-se se os valores estão aproximadamente balanceados – quanto mais próximo de 0,5; mais balanceado. Veja para o nosso exemplo:

$$cp = \frac{\sum_{i=1}^6 \sum_{k=1}^2 g(\Delta_{p^i, x_k}^{-1})}{6 * 2} = 0,48491 \quad (3.15)$$

- **Nível 4: Geral** – Agora calculemos o nível de consenso geral com base na Equação 3.7:

$$cg = cx_k : cx_k \geq cx_l \forall l \in \{1, 2, \dots, n\} = 0,5378783333 \quad (3.16)$$

O consenso final foi 0,5378783333 (aproximadamente aleatório); o que faz sentido; visto que os números utilizados foram escolhidos ao acaso.

Perceba ainda que até este ponto, ainda não diferenciamos as postagens de instrutores. Como o nível de consenso final é muito baixo – 0,5378783333 – identificar as postagens de instrutores não faria sentido. Porém, por hipótese, considere que a postagem p^4 é de um instrutor. Agora apliquemos a Equação 3.8 às similaridades de p^4 (Equação 3.9) com relação a x^1 e a x^2 :

$$\begin{aligned} q_{x_1} &= \sum_{i=1}^1 \frac{a_i(x_1)}{1} = \frac{a_i(0,755 > 0,5)}{1} = \frac{1}{1} = 1 \\ q_{x_2} &= \sum_{i=1}^1 \frac{a_i(x_2)}{1} = \frac{a_i(0,22 > 0,5)}{1} = \frac{0}{1} = 0 \end{aligned} \quad (3.17)$$

Caso a opinião x^1 saísse vencedora, este exemplo contribuiria para concluirmos que há relação entre a opinião consensual e a opinião dos instrutores, caso x^2 vencesse, este exemplo contribuiria para a conclusão oposta.

3.4 Consenso \times Qualidade

Esta etapa do trabalho, conforme sintetizado na questão de pesquisa **Q2**, tem o objetivo de buscar uma correlação entre o consenso e a qualidade da solução alcançada. Isto configura

um objetivo parcial em busca da resposta para **QP**, visto que – como já introduzido no início deste capítulo – opiniões tendem a implicar intervenção do instrutor com frequência, reduzir ou mesmo priorizar tal intervenção passa por medir a qualidade da opinião dominante na discussão.

Com o intuito de medir o grau de certeza por parte do aluno em cada postagem, utilizamos o nível de convicção (ou, o inverso de confusão) médio das postagens, já computado, como um fator moderador em nossa análise. Além disto, o volume de postagens também pode ser fonte de imprecisão em discussões, por isso agrupamos as discussões em nossa base por número de postagens e analisamos tais grupos também separadamente. Agrupamos 4.671 postagens em discussões com número de postagens entre dois e cinco e 4.080 postagens em discussões com mais de seis postagens. Escolhemos estes números para, assim, dividir a base de dados em duas partes aproximadamente iguais. Por fim, estudamos as influências destes fatores em dois níveis de consenso diferentes: 0,8 (baixo) e 0,9 (alto); por serem mais recorrentes na literatura de consenso.

Para nossa análise estatística, recorreremos a um modelo de regressão logística [34] para investigar o relacionamento entre nível de consenso e apoio dos instrutores à opinião consensual. A partir daí, tomando o consenso como a variável a ser investigada, e para adequar nosso problema a tal modelo, a partir dos dois limiares de consenso definidos ($\delta_1 = 0.8$ e $\delta_2 = 0.9$), traduzimos o consenso em uma variável binária assumindo “0” quando $cg < \delta$ e “1” quando $cg \geq \delta$.

Em regressão logística, fatores com coeficientes positivos (β) estão positivamente correlacionados à variável sob investigação. Tomamos a taxa de apoio dos instrutores ao consenso como fator principal e a convicção como um fator que também poderia influenciar o consenso, nossa variável sob análise. Assim, um aumento na taxa de apoio dos instrutores à opinião majoritária associado à um aumento na frequência com que se alcança o limiar δ de consenso, validará a afirmação que consenso e qualidade, medida em função do apoio dos instrutores à opinião majoritária, estão relacionados.

Além de efeitos fixos (fatores), devemos também levar em consideração os dois efeitos aleatórios: a métrica de distância textual e o total de postagens por discussão. Isto porque devemos assimilar o fato de que a métrica de distância textual é automática e agrega certa imprecisão, especialmente mediante um grande volume de postagens que pode levar à uma

dispersão no foco de discussão e prejudicar o consenso. A partir da Tabela 3.1, podemos ver que ambos os efeitos aleatórios exercem influência no *intercept* da função de regressão com variância s^2 igual à 0,52 e 0,30 para cg_2 ; e s^2 igual à 0,72 e 0,69 para cg_1 . Assim os fatores: “Apoio dos instrutores” e “Convicção média” para cg_2 são menos susceptíveis a efeitos aleatórios. (Na Tabela 3.1, cg_1 e cg_2 – alcançar ou não o consenso com limiares de $\delta_1 = 0.8$ e $\delta_2 = 0.9$, respectivamente; β – coeficiente de regressão; s.e. – erro padrão; s^2 – variância; e, σ – o desvio padrão.)

Tabela 3.1: Regressão logística: apoio dos instrutores e convicção média como fatores de consenso

	-2*Efeito	fixos		aleatórios	
		β	s.e.	s^2	σ
-5* cg_1	Apoio dos instrutores (a)	-0.75**	0.06	-	-
	Convicção média	-1.43·	0.74	-	-
	(<i>intercept</i>)	2.33***	0.88	-	-
	Modelo de distância textual ^a	-	-	0.72	0.85
	Quantidade de postagens ^a	-	-	0.69	0.83
-5* cg_2	Apoio dos instrutores (a)	-1.88***	0.08	-	-
	Convicção média	-3.29***	0.92	-	-
	(<i>intercept</i>)	4.81***	0.78	-	-
	Modelo de distância textual ^a	-	-	0.52	0.72
	Quantidade de postagens ^a	-	-	0.30	0.55

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, · $p < 0.1$

^a Termos tratados como efeitos aleatórios.

Ainda com base na Tabela 3.1, em números absolutos, vemos que todos os coeficiente para cg_2 são bastante significativos, com $p < 0,001$. Porém, para cg_1 , a convicção média parece ser um fator pouco significativo. A variância s^2 para ambos os efeitos aleatórios (distância textual e quantidade de postagens) parece invalidar os resultados para um baixo limiar de consenso $cg_1 \mid \delta_1 = 0.8$, uma vez que neste caso $s_{cg_1}^2 \in \{0,72, 0,69\}$, e $s_{cg_1}^2 \approx |\beta|$, $\beta = -0,75$. Assim, podemos descartar os resultados para baixo consenso. Já em se tratando do alto limiar de consenso, $\delta_2 = 0.9$, a magnitude de β parece sempre significativa com relação aos efeitos aleatórios $s_{cg_2}^2 \in \{0,52, 0,30\}$, assim $s_{cg_2}^2 < |\beta|$, $\beta \in \{-1,88, -3,29\}$.

A partir destes resultados, que rejeitam a hipótese nula para nossa questão de pesquisa **Q2**, podemos concluir que o apoio dos instrutores, tomado como parâmetro de qualidade, está significativamente relacionado ao consenso ($p < 0,001$) a um limiar $\delta = 0,9$, exercendo um influência negativa da ordem de $-0,032$ para cada décimo de consenso, de acordo com a regra da divisão por quatro [34]. Assim, os resultados desta etapa deste trabalho a indicam que em fóruns de discussão educacionais, existe correlação negativa entre o nível de consenso e a qualidade da opinião consensual – **Q2** – aceitando a hipótese alternativa.

3.5 Conclusões, Trabalhos Futuros e Limitações

3.5.1 Limitações

Dentre as limitações que podem enviesar nossos resultados, a mais significativa é a interpretação do apoio dos instrutores a uma opinião como indicativo de qualidade. Por conjectura, como um dos papéis do instrutor em um fóruns educacionais é sanar dúvidas, possíveis discussões onde consenso e qualidade tenham se correlacionado positivamente podem não ter atraído a participação dos instrutores, ficando de fora da nossa análise.

3.5.2 Conclusões e Trabalhos Futuros

Este artigo buscou estudar a correlação entre o consenso em discussões de fóruns MOOC e a qualidade das conclusões – as quais podem agregar valor pedagógico ao curso online de interesse. Resultados mostraram que o consenso, no contexto dos dados considerados, é inversamente proporcional à qualidade em um fórum MOOC – sendo a opinião dos instrutores nosso parâmetro de qualidade. Consequentemente, rejeita-se a hipótese nula da questão de pesquisa **Q2**. Uma alta convicção média, embora possa ser interpretada como indicativo de correteza, pode dificultar o consenso. Por exemplo, duas opiniões opostas e convictas podem levar à confusão.

Estes resultados estão em linha com aqueles de [39] e [41]. Porém, estes afirmam que consenso pode ser visto como indicativo de qualidade apenas quando um rigoroso e “vigilante” processo de debate e argumentação seja seguido, além da heterogeneidade dos interlocutores assegurada. Contudo, além da dificuldade em estruturar discussões online de modo

geral, segundo [14], a maior parte das postagens realizadas em fóruns de MOOC são feitas no início do curso, quando os alunos ainda não tem muita familiaridade com o conteúdo. Com isto, a opinião consensual apresentada por alunos ainda inexperientes tenderia a não ter o apoio dos instrutores. Ainda sim, tal consideração não invalida os resultados alcançados aqui para a educação online.

Diante de tais resultados, parece relevante tornar o instrutor ciente do grau de consenso em cada discussão. No contexto dos MOOC, elevados níveis de consenso tendem a ser mais comuns sobre opiniões de baixa qualidade ou equivocadas, o que configura um risco para o aprendizado. Tal situação requer intervenção mais urgente. Assim no próximo passo deste estudo apresentaremos uma ferramenta gráfica baseada em consenso voltada a dar apoio a tarefa de oferecer suporte educacional à um fórum MOOC.

Capítulo 4

Proposta de Aplicação baseada em Consenso para Apoio à Decisão de Instrutores sobre quando Interferir em Discussões em Fóruns MOOC

Nesta Parte 4 apresentamos o resultado da investigação acerca da principal questão de pesquisa “Como ampliar o potencial de suporte da equipe de instrutores por meio do fórum de discussão em um MOOC?” (QP), que trata de uma ferramenta de apoio à tarefa do instrutor ao priorizar intervenções mais urgentes com o intuito de, principalmente, evitar desinformação no fórum MOOC.

4.1 Objetivo

Com base nos resultados e conclusões alcançadas na Parte 2 e na literatura [2; 6], identificou-se o suporte do instrutor ao fórum como ponto mais relevante de apoio ao aprendizado e a apresentação seletiva de mensagens como ferramenta com boa aceitação por parte dos usuários (vide Parte 2). Assim, diante da questão de pesquisa QP, tomou-se como objetivo três situações no fórum onde um modelo de apresentação seletiva de mensagens baseado em distância textual e consenso é capaz de dar apoio à atuação do instrutor.

- **Identificar discussões duplicadas:** Níveis de similaridade muito elevados entre duas discussões tanto de modo geral como quando comparamos apenas a opinião mais consensual podem indicar duplicidade entre discussões;
- **Identificar discussões contraditórias:** Por outro lado, níveis de similaridade elevados entre duas discussões de modo geral, mas tendo soluções consensuais distintas indicam resultados contraditórios para o mesmo tema;
- **Identificar intervenções necessárias:** Discussões com níveis elevados de consenso, sobretudo quando associados à duplicidade contraditória – item anterior – indicam necessidade de intervenção por parte do instrutor, uma vez que uma opinião incorreta pode estar sendo consolidada

Baixos níveis de consenso levam o estudante a buscar mais informação, já níveis elevados podem o levar a tomar aquela opinião como definitiva, o que é perigoso. Assim, embora a ferramenta aqui proposta não tenha por objetivo afirmar ao instrutor que a discussão x requer intervenção, esta destina-se a apontar a necessidade de intervenção ao afirmar, por exemplo, que a discussão x possui um nível elevado de consenso e é semelhante de modo geral à discussão y , embora tenha chegado à uma conclusão diferente.

4.2 Solução

Para alcançar os objetivos apontados, desenvolveu-se um modelo de agrupamento e apresentação de postagens por similaridade e nível de consenso. O modelo foi usado como base para a construção de uma ferramenta na forma de interface gráfica de agrupamento e apresentação de postagens e discussões composta por elementos visuais familiares aos usuários – semelhantes ao formato de fórum utilizado hoje em dia. Tal ferramenta foi concebida em conjunto com um dos entrevistados e mostrou bons níveis de aprovação por meio de uma Validação de Face (vide Subseção 4.5.3) feita com educadores experientes em plataformas de MOOC.

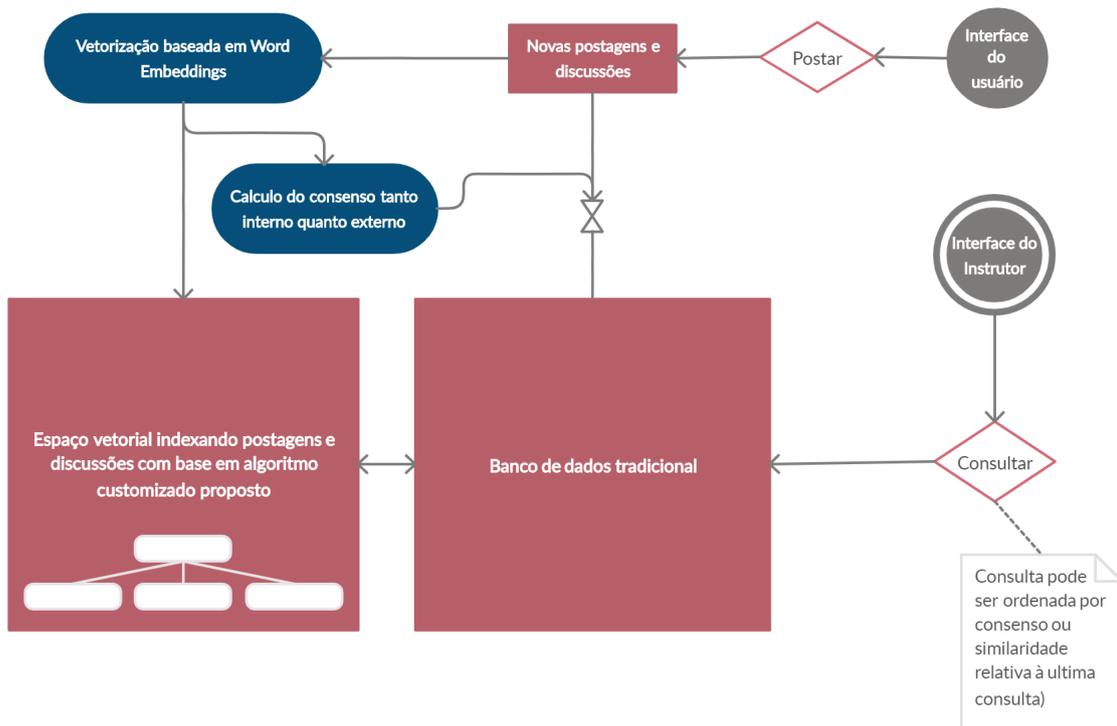


Figura 4.1: Diagrama de comunicação entre módulos do modelo de utilidade proposto.

Com a ferramenta, a tarefa de agrupar postagens por similaridade e consenso é feita em três etapas, a Figura 4.1 traz o diagrama de comunicação entre os módulos responsáveis por cada etapa: 1) Vetorização do texto por *Word Embeddings* [61] – conforme Figura 4.1, recebe novas postagens, as converte em vetor e as envia para o cálculo de consenso e para o índice baseado em Árvores KD; 2) Criação de índices semânticos utilizando uma variação de Árvores KD [71] / Árvores Híbridas [17] – conforme Figura 4.1, recebe as novas postagens na forma de vetores e as indexa em uma estrutura de Árvore; e, 3) Cálculo do consenso com base na Parte 3 deste documento – conforme Figura 4.1, calcula o consenso entre novas postagens a partir da representação vetorial destas postagens. O resultado destas três etapas é exposto ao instrutor por uma interface gráfica definida na Seção 4.4.

4.2.1 Validação da Solução

Aqui foi proposto um protótipo da ferramenta gráfica baseado em *mockups*, que foi utilizado em uma Validação de Face (vide Subseção 4.5.3) para medir o grau de satisfação de

usuários habituados a desempenhar a função de instrutor em fóruns MOOC. Contudo, antes desta etapa final de validação, são realizados dois procedimentos preliminares: 1) Uma verificação do algoritmo de busca pelo vizinho mais próximo, que consiste em uma comparação por Intervalos de Confiança dos resultados de outros algoritmos de vizinho mais próximo mediante alta dimensionalidade (obtidos utilizando Árvores KD e distância Cosseno por força bruta), descrito na Seção 4.4, aos nossos resultados; e, 2) Uma validação baseada em modelos de predição por Aprendizagem de Máquina Supervisionados, tomando o nível de urgência médio daquela discussão – previamente marcado manualmente para cada postagem (campo ‘Urgency’) – como resultado a ser predito y (marcações) e o nível de consenso como ‘feature’ X .

Metodologia de Validação

Assim, a validação do modelo seguiu três passos simples listados abaixo:

1. Verificação da solução utilizando dados gerados por modelos semelhantes para a bateria de testes;
2. Validação da utilidade do modelo em estimar a urgência em intervir. Isto foi feito medindo a correlação entre o consenso utilizado na solução e a necessidade de intervenção urgente por parte do instrutor – informação já contida nos dados;
3. Validação de face utilizada como metodologia de validação do modelo de utilidade propriamente. Aqui instrutores de MOOC foram avaliados ao serem questionados a respeito de *mockups* desenvolvidos com base no modelo.

4.3 Relevância

Embora já existam trabalhos abordando a identificação de postagens e discussões semelhantes [88] e a identificação da necessidade de intervenção do instrutor em fóruns educacionais [6; 79; 18], o presente estudo agrega melhorias que tornam a implantação de tal processo mais simples. Aqui o modelo de aprendizado é não supervisionado, isto significa que não há a necessidade de marcação de dados para treinos periódicos.

Ainda que outros modelos aplicando gamificação [8] e estratégias de recomendação colaborativa [15; 60; 44] tenham bastante repercussão na literatura, no âmbito particular dos MOOC, tais soluções demandam esforço do usuário e desestimulam a participação em um ambiente onde a evasão é um problema significativo. Reduzir custo operacional, automatizar e evitar sobrecarga de trabalho colaborativo por parte dos alunos estão entre as principais contribuições deste trabalho. Abaixo estão elencados os principais argumentos em prol da relevância da ferramenta desenvolvida:

- **Aprendizagem Não Supervisionada:** Já existem trabalhos focados em apontar a necessidade urgente da participação do instrutor em discussões de fóruns MOOC [6; 79; 18], embora todos eles baseiam-se em modelos supervisionados e pré treinados. O presente trabalho apresenta uma solução não supervisionada que não necessita de dados marcados ou retreinamento periódico;
- **Indicativo de Urgência:** Embora este trabalho não tenha o objetivo de apontar diretamente a necessidade de intervenção por parte do instrutor em fóruns MOOC, demonstramos que o nível de consenso em discussões pode ser utilizado pelo instrutor para inferir a necessidade de intervenção;
- **Deteção de Duplicidade:** Nossa solução é capaz de agrupar postagens por similaridade e níveis elevados de similaridade podem significar duplicidade em discussões ou postagens;
- **Deteção de Soluções Divergentes:** Discussões semelhantes dotadas de soluções divergentes indicam a necessidade de intervenção por parte do instrutor;
- **Apresentação Gráfica:** Ao final deste estudo é apresentada uma proposta de interface de gerenciamento de fórum pelo instrutor que da corpo a esta ferramenta.

4.4 O Modelo

Ao lidar com educação em massa, a performance é uma questão computacional a ser levada em consideração. Uma vez que o volume de usuários é indeterminado, é preciso estar preparado para grandes demandas – ainda que nem sempre esta demanda ocorra. Diante disto,

abordou-se o problema de indexar o conjunto de postagens e discussões por similaridade e consenso com foco em eficácia e desempenho. Assim, os seguintes critérios foram usados como norte para esta solução:

- **Atualização:** Os índices da base de postagens devem ser atualizáveis em tempo linearmente proporcional ao conjunto adicionado ou modificado;
- **Consulta:** A consulta pelo vizinho mais próximo útil deve ocorrer com atraso constante;
- **Precisão:** Os resultados, mediante os critérios anteriores, devem ser no mínimo equivalentes às melhores soluções viáveis disponíveis na literatura.

Com base nestes critérios, e em função de resultados anteriores deste estudo e de informações colhidas da literatura, propõe-se a seguinte configuração: Para corresponder ao critério eficácia utilizou-se uma estratégia para medir a distância textual entre as postagens baseada em um método de vetorização com alta dimensionalidade por *Word-Embeddings* aplicado a textos de tamanho variável – vide Seção 4.4.2. Tal abordagem foi escolhida por sua eficácia e por utilizar tecnologias consolidadas [52]. Tendo a representação vetorial de cada sentença, podemos localizar seus vizinhos mais próximos para proceder à indexação. Para tal, foi aplicada a um modelo de similaridade vetorial baseado em Árvores KD – vide Seção 4.4.2 – ou, mais especificamente, Árvores Híbridas. Tal estratégia satisfaz o critério Atualização uma vez que não demanda retrabalho à medida que novas postagens são compartilhadas, sendo necessário apenas a atualização destas novas postagens – para mais detalhes vide Seção 4.1. Conforme será tratado na Seção 4.4.3, a busca em nossa versão de Árvore KD ocorre em tempo constante, embora a alta dimensionalidade do espaço vetorial multiplique este tempo por uma constante elevada. Em função da alta dimensionalidade, soluções baseadas em Árvores KD mostraram-se mais adequadas.

O elemento final para nossa ferramenta funcionar é o módulo de consenso que pode ser dividido em dois: 1) Consenso Interno, onde é calculado o consenso apenas entre postagens da mesma discussão; e, 2) Consenso Externo, onde é calculado o consenso geral para postagens próximas. O consenso em nossa ferramenta é utilizado como filtro para localizar postagens com ideias comuns e divergentes, assim como a similaridade, o que ficará

mais claro na Seção 4.5.3. Estratégias de otimização como caches e listas invertidas foram deixados de lado por simplicidade e por se tratar de uma ferramenta apenas experimental.

4.4.1 Os Dados

Para este estudo foi utilizado a mesma base de dados históricos de postagens em fóruns MOOC **D1**, já descrita nas Seções 2.2.3 e 3.2.1. Como descrito, tais dados já foram pré-tratados e marcados com níveis de urgência (campo ‘Urgency’) e positividade (campo ‘Sentiment’) – dentre outras informações – que variam entre [1, 7], sendo 1 a ausência de urgência e positividade mínima, respectivamente, e 7 a urgência máxima e positividade máxima, respectivamente, conforme [1]. Ainda que a informação contida no campo ‘Sentiment’ possa parecer relevante para esta aplicação, por observação percebeu-se que ao concordarem, duas postagens tendem a ser semelhantes, embora é incomum a discordância mediante mera negação – preservando a semelhança. Assim, ao discordarem, duas postagens tendem a trazer ideias distintas e não meramente opostas.

Tais postagens contam com outros seis campos-chave utilizados neste estudo: 1) O código sequencial de cada postagem ‘forumpostid’; 2) O código da discussão em que a postagem está inserida ‘commentthreadid’; 3) O nome do curso onde aquele texto foi postado ‘coursedisplayname’; 4) A área do conhecimento à qual o curso pertence ‘CourseType’; e, 6) O texto propriamente dito ‘Text’. O código da postagem é usado para identificá-la, enquanto o código da discussão é utilizado para agrupar as postagens por discussão e medir o nível de consenso interno; já os campos ‘coursedisplayname’ e ‘CourseType’ são utilizados para restringir a postagem ao grupo de dados em que ela é útil e assim medir o consenso geral¹.

Tratamentos

O primeiro passo do tratamento dos dados **D1**, de modo análogo ao apresentado na Parte 3 Seção 3.2.1, foi a limpeza do texto – o campo ‘text’ – que foi da divisão do texto em *tokens*, à remoção de *stopwords*. O resultado deste processo foi armazenado no campo ‘cleaned_text’².

¹Para mais informações a respeito dos dados brutos acesse: <https://datastage.stanford.edu/StanfordMoocPosts/>.

²O código fonte utilizado para tal está acessível em: https://github.com/tiago-clementino/consensus_x_relevance/blob/master/data_treatments/clean_and_report.ipynb

No passo seguinte à preparação dos dados **D1**, ocorreu a vetorização do texto tratado – campo ‘cleanned_text’. Para tal, fez-se uso de uma rede semântica pré treinada com base em **Word Embeddings** (vide Seção 4.4.2) a partir da biblioteca Gensim³ escrita em Python. Tal treinamento se deu a partir de uma base de mais de três milhões de palavras e frases coletadas de resultados de buscas na Web feitas por meio do motor de buscas do Google, que serviu de material para criar o modelo de vetorização aplicado à **D1**⁴.

Optou-se por utilizar um modelo pré-treinado com dados não relacionada à **D1** para assegurar o poder de generalidade do algoritmo. Este processo gerou vetores fixos de alta dimensionalidade para cada postagem – 300 dimensões – e o resultado foi armazenado no campo ‘vector’. A alta dimensionalidade é uma necessidade em virtude da definição de preferência para o cálculo de consenso depender de comparações ricamente semânticas. O código fonte utilizado para esta finalidade também foi escrito em Python, utilizou o pacote Doc2Vec da biblioteca Gensim⁵.

4.4.2 Fundamentação

Aqui são apresentadas as tecnologias e métodos utilizados para abordar a questão de pesquisa **QP**.

O Problema da Alta Dimensionalidade

Em nossa árvore de índices, cada índice representa uma postagem na forma de um vetor em uma espaço de 300 dimensões. Tendo em vista que toda a postagem é representada por apenas um vetor, a alta dimensionalidade se faz necessária. Além disto, o método de cálculo do nível de consenso requer um modelo de representação das postagens que descreve com a clareza necessária seu significado. Embora tal dimensionalidade viabilize o método de consenso, traz grandes implicações relacionadas ao desempenho da nossa solução.

Diante da alta dimensionalidade, algumas soluções foram avaliadas, dentre elas a solução escolhida foi a aplicação de um modelo de Árvore KD com algumas adequações. Conforme

³<https://radimrehurek.com/gensim/>

⁴Tal modelo, já pré-treinado, pode ser acessado em: <https://code.google.com/archive/p/word2vec/>.

⁵Este código fonte completo está disponível em: https://github.com/tiago-clementino/consensus_x_relevance/blob/master/data_treatments/vectorize_report.ipynb

descrito na Seção 4.4.3, a Árvore KD aplicada poderia ser descrita como uma árvore híbrida que inclui o conceito de semelhança útil, elemento que torna possível realizar consultas à árvore em tempo constante.

Árvores KD

Um árvore KD é uma generalização do conceito de árvore binária. A raiz de uma árvore KD representa todo o conjunto de pontos sendo particionado. Cada nó é definido como uma partição do conjunto de pontos anterior. Tal partição geralmente representa metade do conjunto total, sob alguma métrica arbitrária. Cada folha da árvore compõe um subconjunto perfeitamente disjunto do conjunto total, em função de alguma métrica de distância [71].

A versão de árvore KD utilizada aqui divide o conjunto de pontos em cada nó em função de uma dimensão específica – dentre as trezentas totais – em negativos e positivos – $d \geq 0,0 || d \leq 0,0$. Necessariamente, nossa árvore KD deve ter $d - 1$ níveis, sendo d o total de dimensões de nosso vetores – $d = 300$.

Busca em Árvores KD

Buscas em árvores KD são operações naturalmente recursivas. A cada nó, o ponto de divisão do conjunto de pontos corrente é avaliado a fim de definir para que direção ou para que nó filho seguir. Na versão utilizada neste trabalho, tal avaliação envolve verificar para um dada dimensão sob análise, se naquele nó é maior que zero $p \geq 0,0$ – sendo p a posição do vetor para aquela dimensão – guiando o algoritmo para o nó filho positivo, ou se é melhor que zero $p \leq 0,0$, guiando o algoritmo para o lado negativo. Ao alcançar um nó folha, todos os pontos daquele conjunto devem ser comparados ao ponto sob análise até encontrar o mais próximo. Caso não seja encontrado um mais próximo satisfatório, deve-se retroceder na árvore até um resultado aceitável. Tal processo é conhecido como *backtracking*.

O *backtracking* é um procedimento necessário para caso o vizinho mais próximo do ponto sob análise esteja em algum dos quadrantes vizinhos e por isto não tenha aparecido na busca convencional. Para evitar a necessidade de *backtracking*, a implementação de Árvore KD utilizada neste trabalho seguiu uma linha diferente no que se refere à busca. Aqui os dois caminhos possíveis do algoritmo em cada nó não são completamente disjuntos. Assim, ao alcançar um dado nó, para decidir qual partição do conjunto restante de pontos seguir é

preciso verificar se $p \geq 0, 0 + \delta$, ou $p \leq 0, 0 - \delta$, caso contrário – sendo δ uma margem de segurança para evitar que o vizinho mais próximo esteja em outro quadrante.

Para esta versão de Árvore KD, o conceito de vizinho mais próximo aceitável é peça chave. O δ deve ser escolhido de modo que o vizinho mais próximo aceitável seja coberto, conforme Figura 4.2. Veja na imagem ‘a’ que o algoritmo retornou o vizinho mais próximo correto graças ao $+\delta$. Já na imagem ‘b’, o algoritmo erra ao retornar o vizinho mais próximo em função de $+\delta$, embora tal erro não seja relevante, visto que o ponto sob análise não possui um vizinho mais próximo aceitável. O δ deve ser equivalente ao raio da circunferência ao redor do ponto, área onde o vizinho mais próximo deve estar. Caso a distância entre o ponto analisado e o seu vizinho mais próximo seja maior que δ , o algoritmo conclui que aquele ponto não tem um vizinho mais próximo útil. De todo modo, veja na imagem ‘c’ que o ponto retornado caso a margem $+\delta$ não fosse utilizada também estaria incorreto.

Em problemas de distância textual entre sentenças podemos entender o conceito de vizinho aceitável como uma sentença de fato semelhante. O vizinho mais próximo inaceitável seria uma sentença menos distante que as demais, mas ainda sim diferente da sentença sob análise.

Consenso

Para o módulo de consenso devemos introduzir dois conceitos distintos: Consenso Interno e Consenso Externo.

- **Consenso Interno:** Nível de consenso entre postagens da mesma discussão
- **Consenso Externo:** Nível de consenso entre postagens de mais de uma discussão calculado em tempo de consulta

Para o módulo de consenso desta ferramenta, o nível de consenso foi calculado tal qual descrito na Seção 3.3.2. Já a distância entre cada postagem foi calculada utilizando duas estratégias distintas, uma para o consenso interno e outra para o consenso externo. O consenso interno foi calculado com base em *Word-Embeddings* e *Wmdistance* para mensurar as preferências entre as opiniões. Já o consenso externo também aplicou *Word-Embeddings*

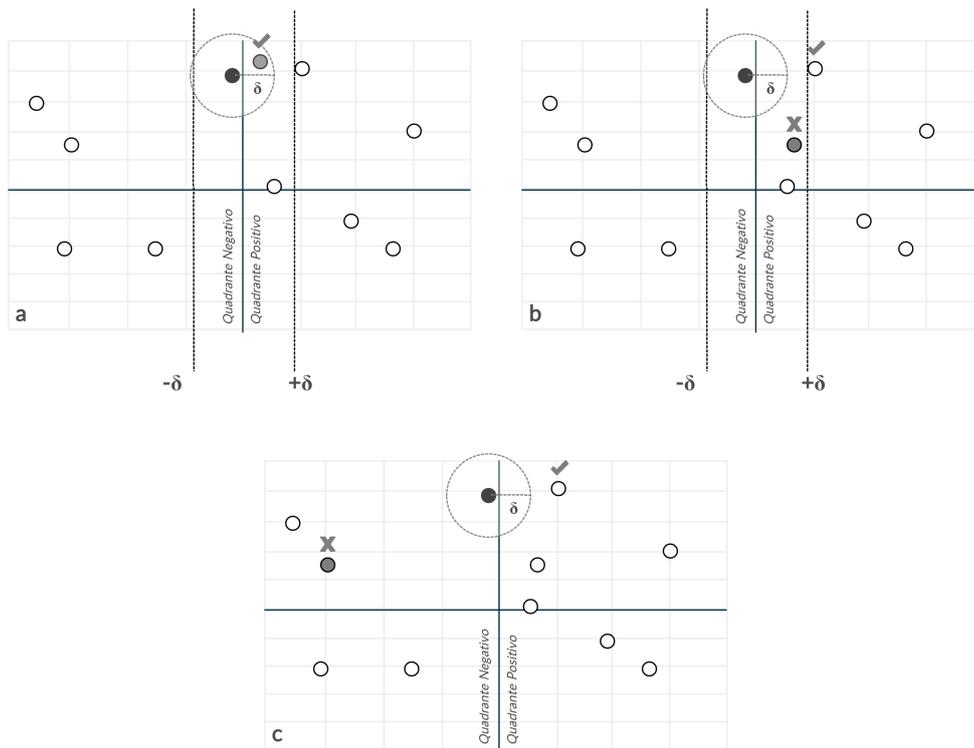


Figura 4.2: Possíveis cenários para identificação de vizinho mais próximo

para vetorização textual, mas utilizou nosso método baseado em Árvores KD para medir preferências entre opiniões em tempo constante.

Word-Embeddings

Word-Embedding é um método utilizado para mapear palavras para vetores de baixa densidade e supera seus concorrentes antecessores em muitos aspectos [61]. Um vetor de baixa densidade é aquele que, embora esteja em um espaço de muitas dimensões, a informação que o define está restrita a poucas [61]. *Word-Embedding* tem sido utilizado com sucesso em diversas áreas relacionadas a este trabalho, como: Análise de Sentimento [89], Recuperação da Informação [48] e até mesmo recomendação de postagens para apreciação urgentes por instrutores de cursos online [36], tema deste trabalho.

Especificamente, neste trabalho foi utilizado uma variação do método de vetorização por *Word-Embedding* para conjuntos de palavras. Descrito em [52], este conceito derivado mede a distância mínima entre os conjuntos de vetores de palavras dos dois textos. Assim, com base no pacote Doc2Vec da biblioteca Gensim para linguagem Python, que implementa o conceito de ‘*Document-Embedding*’, aplicou-se a rotina de inferência – ‘infer_vector’ – que converte um conjunto de vetores de palavras em um único vetor que representa todo o texto (postagem). Tal vetor tem dimensionalidade contante, independente dos diferentes comprimentos textuais das diferentes postagens.

Word-Embedding é um modelo semântico de agregação de informação baseado em predição de vizinhança entre as palavras [61]. Este funciona a partir de uma etapa de treinamento onde o modelo é aplicado a uma base de dados textuais suficientemente grande. A estratégia do modelo consiste em prever a palavra central a partir de suas vizinhas – método *Skip-gram* – ou suas vizinhas a partir da palavra central – método CBOV (do inglês *Continuous Bag of Words*). O resultado é uma malha vetorial onde se pode inferir similaridade a partir de modelos de distância vetorial simples e até realizar operações matemáticas entre vetores-palavra (exp: $REI - HOMEM = RAINHA$).

4.4.3 Busca por Vizinhos Próximos: O Algoritmo

Conforme é possível observar no Código Fonte 4.1 abaixo, o método **Build** cria partições do espaço ou apenas as recupera por meio de uma função Hash. Assim, caso já exista algum vetor na mesma partição do espaço vetorial que o vetor corrente, este é alocado em tempo constante. O código *hash* retornado pela função ‘makeHash’ é uma sequência de símbolos que representam cada passo na árvore para se chegar àquela partição.

Ainda segundo o Código 4.1, o método **Add** constrói a estrutura de uma árvore binária simples que particiona o espaço vetorial a cada nó. Veja que a cada passo o código *hash* daquela partição é decrementado de um símbolo e a ausência de novos símbolos significa o fim da árvore (nó folha).

Por fim, o método **myKdHashs** busca pela partição de um dado vetor postagem. Perceba que seu modus-operandi é semelhante ao do método ‘Add’, tendo o parâmetro ‘edge’ como principal diferença. Tal parâmetro define a margem em que uma partição se sobrepõe à outra para uma dada dimensão.

Código Fonte 4.1: Árvore de Busca por Vizinhos mais Próximos

```

1  '''Este método constrói a nossa adaptação de Árvore KD (no código
    corresponde à classe KdTree'''
2  def build(postings):
3      hashMap = {}
4      #Instância da árvore
5      kdTree = KdTree()
6      for index, row in postings.iterrows():
7          #Converte os valores do vetor em um código hash, o que acelera a
            consulta
8          hash_ = makeHash(row['Vector'])
9          try:
10             '''Cria uma partição para o vetor.
11             Esta partição corresponde ao quadrante onde todos os sinais (
                positivos ou negativos) correspondem aos do vetor em quest
                ão. Caso já exista, apenas a retorna. É claro que podem já
                haver outros vetores neste quadrante.'''
12             partition = hashMap[hash_]
13         except KeyError:
14             partition = None

```

```

15     if(partition is None):
16         partition = Partition(hash_)
17         #Insere nosso vetor na sua partição
18         partition.append(row)
19         #Insere a partição e seu vetor na nossa árvore
20         kdTree.add(row['Vector'], partition)
21         #Array com os hashes das partições ordenados para facilitar a
           busca
22         hashMap[hash_] = partition
23     return (hashMap, kdTree, changedPartitions)
24
25     “Classe de objetos para nossa versão de Árvores Kd”
26     class KdTree:
27
28     “Insere um novo vetor em sua respectiva partição.
29     ‘vector’ seria o vetor contendo um array com todos os 300 valores para as
           300 dimensões utilizadas”
30     def add(self, vector, partition):
31         if(partition is None):
32             return
33         if(vector is None or isEmpty(vector)):
34             self.leaf = partition
35             return
36         value = vector[0]
37         #Caso o valor da próxima dimensão seja maior ou igual a zero,
           segue à direita, caso contrário, à esquerda (árvore binária)
38         if(float(value) < 0):
39             if(self.negativeEP is None):
40                 #todo novo ramo é compreendido como uma nova árvore
41                 self.negativeEP = KdTree()
42                 self.negativeEP.add(vector[1:], partition)
43         else:
44             if(self.positiveEP is None):
45                 self.positiveEP = KdTree()
46                 self.positiveEP.add(vector[1:], partition)
47
48     “Busca-se o ‘vector’ na árvore ‘self’, em função de uma ‘edge’ igual a

```

0.5 por padrão.

49 'edge' é a margem de tolerância para cada dimensão do vetor a ser buscado
. Isto é, se o vetor é positivo na dimensão X, ele deve ser próximo de
todos os demais vetores positivos nesta dimensão e os vetores
negativos com valores iguais ou inferiores a -0.5 '''

```

50 def myKdHashs(self, vector, edge = 0.5):
51     if(vector is None or isEmpty(vector)):
52         if(not self.leaf is None):
53             return [self.leaf]
54         return []
55     hashes = []
56     value = vector[0]
57     if(value < 0 + edge and not self.negativeEP is None):
58         hashes = hashes + self.negativeEP.myKdHashs(vector[1:], edge)
59     if(value >= 0 - edge and not self.positiveEP is None):
60         hashes = hashes + self.positiveEP.myKdHashs(vector[1:], edge)
61     return hashes

```

Análise Assintótica

Dois dos principais requisitos desta ferramenta foram: atualização local, para que novas postagens não exijam reindexação; e, indexação semântica, para garantir o cálculo correto do consenso e busca pelo vizinho mais próximo útil em tempo constante – $O(1)$. Tais requisitos foram atendidos pelo método e ferramental aplicados. A indexação, processo mais custoso, é feita em tempo $O(md)$, sendo m o total de postagens da base e d o total de dimensões do espaço vetorial.

4.5 Verificação e Validação

Aqui apresentaremos com mais detalhes os três passos empreendidos para validação e verificação da ferramenta proposta.

4.5.1 Verificação

A ferramenta apresentada neste trabalho coloca em prática algumas tecnologias já bem fundamentadas. Porém, o algoritmo de busca pelo vizinho mais próximo ‘útil’ é uma versão customizada de Árvores KD, e tal customização demanda verificação – o método de cálculo de consenso também foi adaptado ao cenário dos MOOC, mas este será validado na Seção 4.5.2. Por se tratar de uma proximidade difícil de apontar, marcar dados para teste não foi uma tarefa simples. Deste modo, o conjunto de testes utilizado na verificação foi reunido a partir de resultados alcançados por tecnologias similares já consolidadas.

Embora o algoritmo a ser verificado aqui seja baseado em Árvores KD ou Árvores Híbridas, algumas peculiaridades o fazem apresentar resultados melhores para esta aplicação específica. Para verificar tal afirmação, utilizou-se como ponto de partida o problema simples de encontrar o vizinho mais próximo viável / útil. A partir deste problema, a estratégia de verificação consiste da comparação entre o algoritmo apresentado, Árvores KD puramente e força bruta utilizando a distância Cosseno, tomando os resultados da distância baseada em Word Movers (WMD) [52] como parâmetro de vizinhos mais próximos à postagem sob análise. A escolha destes concorrentes foi definida comparando os resultados visualmente por amostragem. Para tal, tomou-se como base as implementações dos pacotes Kdtree e distances.cosine da biblioteca Scikit-learn e o pacote Doc2Vec da biblioteca Gensim, ambas escritas em Python.

Para a comparação proposta foi aplicado a cada algoritmo a ser comparado o teorema do limite central com base em intervalos de confiança de 95%, por meio do resultado médio da precisão dos resultados dos 50 vizinhos mais próximos à uma dada postagem selecionada aleatoriamente. Tal rotina era repetida 20 vezes para cada média amostral. Este processo foi repetido 100 vezes para gerar um intervalo de confiança⁶.

Definição 6. *A métrica de Precisão utilizada nestes resultados é definida conforme Equação 4.1:*

$$Precisão = \frac{Verdadeiros\ Positivos\ (TP)}{Verdadeiros\ Positivos\ (TP) + Falsos\ Positivos\ (FP)} \quad (4.1)$$

Tal Precisão é o número de postagens classificadas como próximas à postagem sob análise

⁶O código fonte deste processo está disponível na forma de um *Notebook* Jupyter em: https://github.com/tiago-clementino/consensus_x_relevance/blob/master/analysis_v2.ipynb

e que realmente o são (verdadeiros positivos), dividido pela soma entre este número, e o número de postagens classificadas como próximas, mas que não são (falsos positivos).

Por termos limitado o número de postagens mais próximas retornadas, para nossos testes, o total de falsos positivos e falsos negativos é equivalente. Assim, descartando a necessidade de medir o *Recall*.

Distribuição

Antes de verificarmos os resultados, é útil visualizar a distribuição das médias amostrais. Como é possível observar na figura 4.3 ‘a’, os três modelos comparados seguem uma distribuição “aproximadamente” normal, visto que nosso modelo apresenta certa distorção, sem descaracterizá-lo. Já de acordo com as Figuras 4.3 ‘b’, que apresentam a distribuição das diferenças entre as precisões de nosso modelo com relação a Árvores KD e distância Cosseno, percebe-se que além de se tratar de distribuições também normais, ambas têm seus intervalos de confiança totalmente acima de zero. Tal constatação sugere uma diferença significativa entre tais resultados.

Resultados

Conforme Figura 4.4, confirmando o que já havia sido indicado pela distribuição amostral da diferença entre os modelos, nosso algoritmo apresenta resultados melhores quando comparado tanto à Árvores KD quanto à distância Cosseno, para este problema em particular. A parte “a” da figura apresenta o IC (Intervalo de Confiança) da precisão de cada modelo, com a nossa precisão acima das demais. Já a parte “b” da figura apresenta o IC da diferença entre a nossa precisão e a precisão de cada modelo concorrente, sempre acima de zero. Tais conclusões foram tomadas a 95% de confiança.

4.5.2 Consenso × Urgência

Aqui buscamos validar o consenso como ferramenta útil para o reconhecimento da urgência em intervir em uma discussão online. Para tanto, buscou-se identificar uma correlação entre a urgência em dar suporte a uma discussão e o nível de consenso desta discussão. O objetivo aqui é demonstrar que tal correlação existe e o instrutor, ao ter ciência do nível de consenso

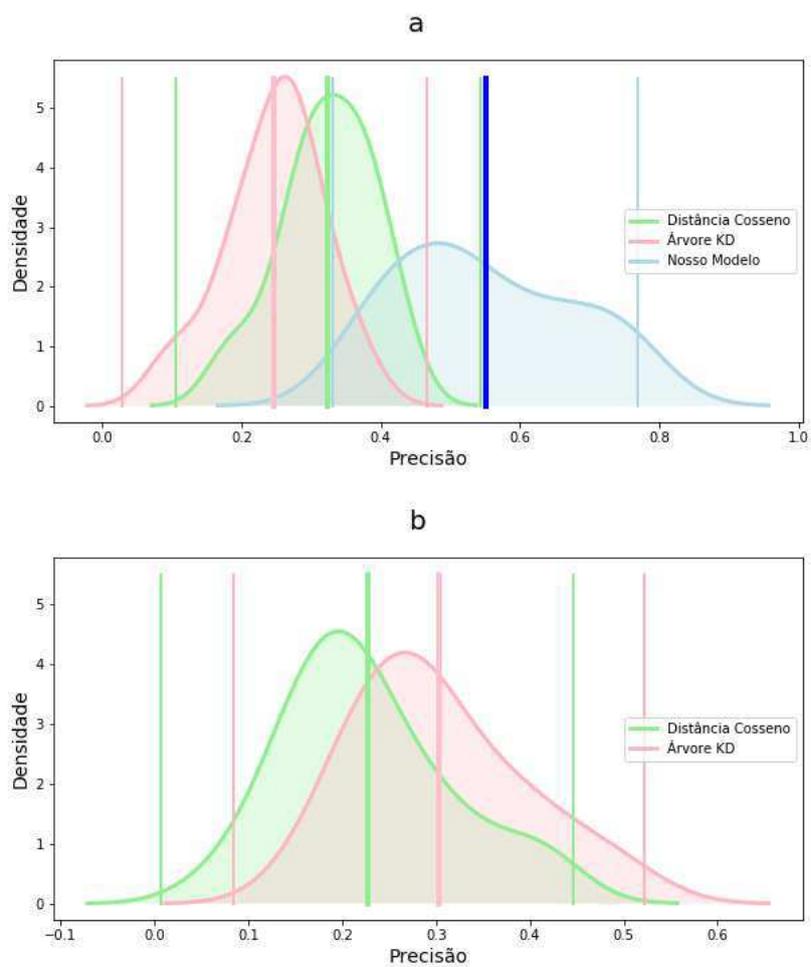


Figura 4.3: Distribuições Amostrais dos Dados de Verificação

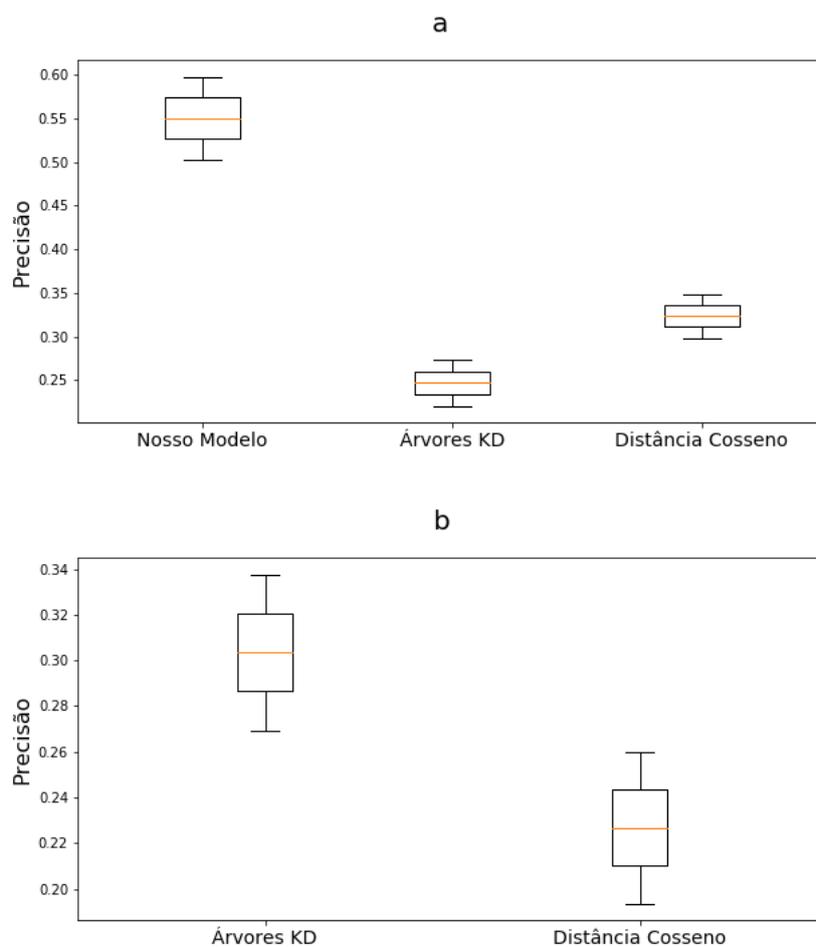


Figura 4.4: Intervalos de Confiança: “a” apresenta a precisão de cada modelo; e, “b” a diferença entre nossa precisão e a de cada modelo concorrente

de uma discussão, será capaz – com suporte da ferramenta gráfica desenvolvido a partir do modelo – de inferir o quão urgente é sua participação naquela discussão, dentre outras coisas. Para calcular tal correlação, utilizaremos modelos e métodos de Aprendizagem de Máquina para prever a urgência com base no nível de consenso, conforme descrito na Seção 4.5.2.

Com os dados já tratados e os níveis de consenso calculados, seguiu-se uma metodologia simples: 1) Utilizando uma técnica de Aprendizagem de Máquina Automática chamada TPOT (do inglês, *Tree-based Pipeline Optimization Tool*) [64], identificar o modelo de predição ideal; 2) Definido o modelo de predição ideal como sendo o GBDT (do inglês, *Gradient Boost Decision Tree*) [33], mediu-se sua precisão conforme descrito na Seção 4.5.2; e por fim, 3) A partir de uma técnica de explicação de modelos de Aprendizagem de Máquina chamada LIME (do inglês, *Local Interpretable Model-agnostic Explanation*) [69], identificaram-se os principais fatores responsáveis pela correlação entre tais grandezas.

Ferramental

Aqui são descritas em poucos detalhes todas as técnicas e ferramentas utilizadas da Seção 4.5.2.

TPOT:

TPOT (do inglês, *Tree-based Pipeline Optimization Tool*) é uma ferramenta de otimização e automatização de soluções em Aprendizagem de Máquina que identifica o modelo de predição ideal para um dado problema e seleciona um conjunto de hiper parâmetros otimizados. Tal procedimento consiste em, com base em um Algoritmo Genético, gerar um conjunto de *pipelines* de algoritmos de aprendizagem que oferece como saída o modelo vencedor, mediante uma configuração específica de seus hiper parâmetros otimizados no formato de código fonte compatível com a biblioteca Scikit-learn de Python e totalmente implementada sobre o Scikit-learn⁷.

Cada instância em um *pipeline* corresponde a uma indivíduo do Algoritmo Genético, conta com uma cópia do conjunto de dados e compete com os demais para ser selecionado. Duas instâncias podem ser compostas por dois algoritmos distintos ou por um único algoritmo com duas configurações diferente. Tal modus-operandi o torna fortemente distribuível.

O Algoritmo Genético de otimização inicialmente cria 100 *pipelines* em árvore aleatori-

⁷Disponível em: <https://github.com/rhiever/tpot>

amente, balanceando e procedendo uma validação cruzada de seus resultados. Para um nova geração o algoritmo seleciona os melhores 20 *pipelines* da população com maiores acurácias e com menos operações. Cada um dos vinte gera 5 cópias. Este processo se repete até ser interrompido ou alcançar um total limite de gerações.

Gradient Boosting:

GBDT é um modelo aditivo de regressão / classificação que consiste de um conjunto de árvores de regressão. Em síntese, a ideia básica deste modelo é combinar uma série de classificadores mais fracos em um forte.

LIME:

LIME é uma técnica voltada a interpretar as previsões de um modelo de Aprendizagem de Máquina *black-box*. A interpretação de um dado modelo f baseia-se em aproximações lineares válidas apenas para uma amostra local q do conjunto total de dados de treinamento Q . Os resultados, diferentemente daqueles do modelo original, são de interpretação tão fácil quanto qualquer modelo linear

O processo de interpretação ocorre introduzindo certas perturbações no conjunto q . Cada instância q^i sofre pequenas modificações em seu conjunto de *features*, resultado em $q^{i'}$, uma versão de q^i com alterações randômicas conhecidas. O LIME então armazena as tuplas $(q^i, q^{i'})$ e $(f(q^i), f(q^{i'}))$, representando as instâncias e suas saídas do respectivo modelo sob análise. Uma função g pondera as diferenças entre q^i e $q^{i'}$ e as relaciona às diferenças entre $f(q^i)$ e $f(q^{i'})$. O resultado disto é uma LWR (do inglês, *Locally Weighted linear Regression*), um modelo linear interpretável, onde seus fatores representam as *features* do modelo original. Nesta etapa de validação da nossa ferramenta foi utilizado uma biblioteca de código aberto⁸ escrita em Python disponibilizada pelos autores da técnica, Ribeiro et al [69].

Procedimento e Resultados

Para esta validação foi utilizado o conjunto de dados **D1**, já descrito na Seção 2.2.3, mediante algum tratamento prévio. Em **D1**, o campo 'Urgency' será especialmente utilizado nesta validação para medir a correlação entre a máxima urgência e o consenso. Tal campo mede a urgência daquela postagem em receber atenção do instrutor. Para distinguir as postagens

⁸ Acessível em: <https://github.com/marcotcr/lime>

mais urgentes, criou-se o campo ‘is_Urgent’ definido em $\{0, 1\}$, sendo que 0 corresponde a pouco urgente e 1 a muito urgente. Além deste campo, criou-se também os campos ‘>0.6’, ‘>0.7’, ‘>0.8’ e ‘>0.9’, também definidos em $\{0, 1\}$ para designar qual o limiar de consenso à qual aquela postagem pertence. O objetivo consiste em medir a correlação entre o limiar de consenso daquela postagem e seu nível de urgência.

Com esta divisão, os dados ficaram significativamente mal distribuídos. Apenas 11% das postagens são muito urgentes. Para contornar este problema, utilizou-se uma estratégia de reamostragem com substituição, nativa da biblioteca Scikit-learn em Python⁹.

Tendo finalizado esta primeira etapa, foi utilizado a técnica de Automatização de Aprendizagem de Máquina TPOT – descrita na Seção 4.5.2 – com o intuito de identificar o modelo de aprendizagem mais eficiente para este procedimento. Conforme é possível observar no Código Fonte 4.2 que apresenta a resposta do método TPOT, GBDT aparece como modelo mais eficiente para o conjunto de dados, além de oferecer todo um conjunto de hiper parâmetros mais adequados.

Para medir o desempenho da GBDT foi utilizado a métrica de precisão **F1**, por ser uma das mais consolidadas em Aprendizagem de Máquina. Tal métrica corresponde à média harmônica entre Precisão – ou *Precision* – e *Recall*. Precisão, conforme Definição 6, mede sobretudo a capacidade do modelo em não apresentar falsos positivos; já *Recall* mede a capacidade do modelo em não apresentar falsos negativos e é definido a seguir

Definição 7. A métrica de *Recall* utilizada nestes resultados é definida conforme Equação 4.2:

$$Recall = \frac{Verdadeiros\ Positivos\ (TP)}{Verdadeiros\ Positivos\ (TP) + Falsos\ Negativos\ (FN)} \quad (4.2)$$

Tal métrica é o número de postagens classificadas como próximas à postagem sob análise e que realmente o são (verdadeiros positivos), dividido pela soma entre este número e o número de postagens classificadas como não próximas, mas que o são (falsos negativos).

Assim, F1 é definida conforme a Equação 4.3

$$F1 = \frac{2 * Precisão * Recall}{Precisão + Recall} \quad (4.3)$$

⁹Os resultados destes tratamentos de dados estão acessíveis em: https://github.com/tiago-clementino/consensus_x_relevance/blob/master/analysis_v3.ipynb

Quanto maior o F1, melhor.

Conforme Tabela 4.1, é possível observar um F1 de 0,67, o que é um valor não muito elevado, mas que indica uma correlação concreta entre consenso e urgência. A diferença exata entre Verdadeiros Positivos, Falsos Positivo, Verdadeiros Negativos e Falsos Negativos é descrita na Figura 4.5, trazendo a matriz de confusão entre estes valores. O nível baixo de correlação não desqualifica nossos resultados. Situações peculiares da operação da interface como a identificação de discussões semelhantes e discordantes, conforme Validação de Face descrita na Seção 4.5.3, maximizam o potencial da solução.

Tabela 4.1: Relatório de Classificação

Resposta correta	Precisão	Recall	F1
Não é urgente	0,72	0,62	0,67
É urgente	0,67	0,77	0,71
acurácia			0,69

Por fim, ainda precisamos averiguar se o modelo de aprendizagem chegou a estes valores em função dos fatores corretos. Para tanto, empregou-se a técnica de explanação de modelos de predição LIME – vide Seção 4.5.2. Conforme é possível observar na Figura 4.6, a principal razão por trás da predição da natureza urgente da postagem sob análise é o fato de seu nível de consenso ser maior que 0,6, 0,7 e 0,9 – o campo “cg_16” = 0 indica que o nível de consenso não está no intervalo [0,6; 0,7]; semelhantemente, o campo “cg_17” = 0 indica que o nível de consenso não está no intervalo [0,7; 0,8]; já o campo “cg_19” = 1 indica que o nível de consenso está no intervalo [0,9; 1,0], para o caso analisado pela ferramenta LIME. Tal postagem foi selecionada aleatoriamente e o LIME apontou que, em função dos fatores indicados, ela tem 58% de chance de ser marcada corretamente como urgente.

Código Fonte 4.2: Configuração do Modelo de classificação Gradient Boosting Recomendada Pelo TPOT

```

1 GradientBoostingClassifier(criterion='friedman_mse', init=None,
    learning_rate=0.1, loss='deviance', max_depth=7, max_features=0.5,
    max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=
    None, min_samples_leaf=19, min_samples_split=3,
    min_weight_fraction_leaf=0.0, n_estimators=100, n_iter_no_change=None,

```

```
presort='auto', random_state=None, subsample=0.6500000000000001, tol
=0.0001, validation_fraction=0.1, verbose=0, warm_start=False)
```

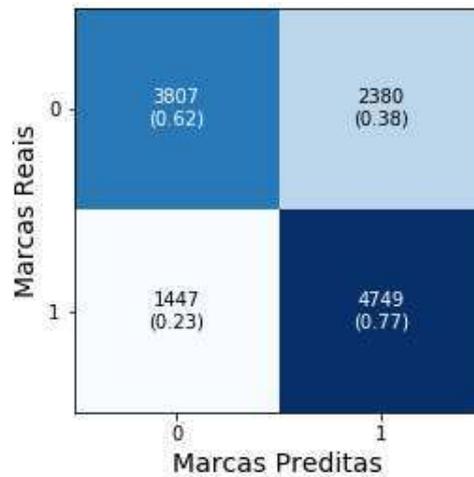


Figura 4.5: Matriz de Confusão da Validação da Relação Consenso \times Urgência

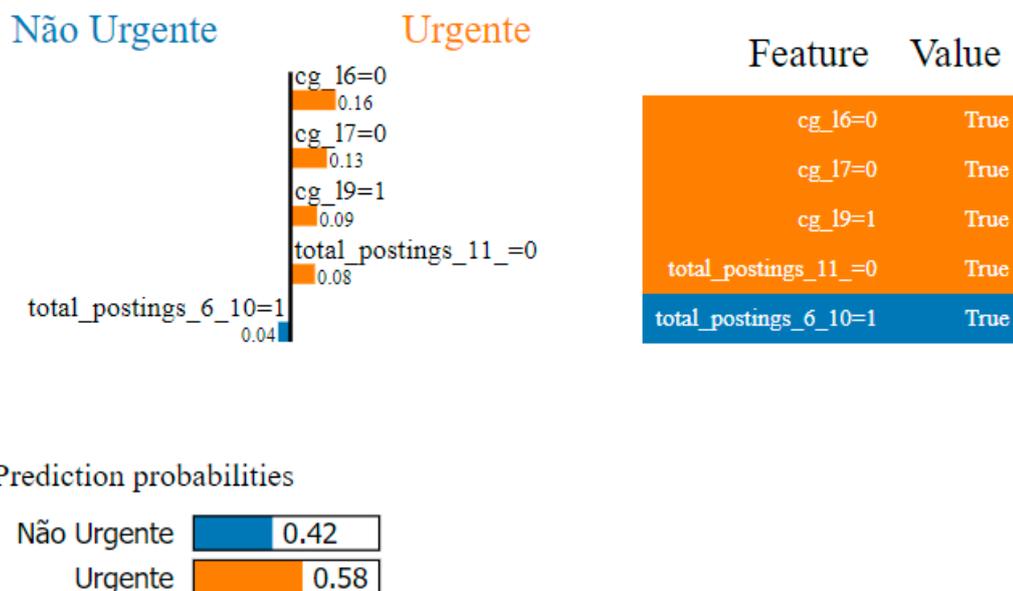


Figura 4.6: Explicação dos fatores da Validação da Relação Consenso \times Urgência a partir da ferramenta LIME

4.5.3 Entrevistas: Validação de Face

Para validar a relevância e aceitação da ferramenta proposta, um grupo de quatro indivíduos com experiência em dar suporte a fóruns MOOC foi entrevistado – tal grupo é remanescente daquele entrevistado na Parte 2 deste trabalho, os quais quando ao final da entrevista demonstraram interesse em participar de uma segunda fase do estudo. Assim, coletaram-se suas impressões a cerca da aprovação ou reprovação de alguns protótipos de interface apresentados – *mockups*. Tal interface visa apresentar as informações úteis ao instrutor, assim como definidas na Seção 4.2. Além da interface proposta, os entrevistados foram apresentados às duas discussões de maior e menor consenso sobre o tema Educação – comum à área de interesse de todos os entrevistados – e ambas envolvendo opiniões ‘questionáveis’. A maioria deles apontou a discussão de maior consenso como aquela que mereceria um suporte mais urgente por parte dos instrutores. Apenas um deles fez ressalvas a respeito dos riscos de ‘priorizar’ a atividade do instrutor e condicioná-lo a não ler todas as discussões. O processo de validação descrito neste parágrafo é denominado ‘Validação de Face’.

A Validação de Face é mais subjetiva que uma validação estatística. De acordo com [40], Validação de Face avalia a aplicabilidade, sensibilidade ou relevância dos itens testados e como eles são vistos pelos entrevistados. Tal metodologia deve ser aplicada em conjunto com formas de validação mais objetivas, assim como a descrita na Seção 4.5.2.

Neste processo de validação em particular, os usuários não utilizaram de fato a ferramenta proposta, mas apenas *mockups* baseados em conteúdo real extraído manualmente dos resultados de tal ferramenta. Sendo assim, foi dado mais ênfase na qualidade dos elementos de apresentação e operação, em detrimento de elementos técnicos como plataforma ou dispositivo alvo. Assim, tal validade é mensurada a partir da rotina adequada de operação da interface, do desempenho do entrevistado ao identificar os elementos gráficos, da relevância do apoio oferecido pela ferramenta à atividade do instrutor, da aceitação e da eficácia percebida.

Mockups

Os *mockups* utilizados como protótipos de interface nas entrevistas foram elaborados com base em critérios de usabilidade e familiaridade às interfaces de ferramentas já utiliza-

das em plataformas como edX (www.edx.org), Coursera (www.coursera.org) e Udemy (www.udemy.com), permitindo que o instrutor navegue de forma gráfica entre as discussões. Para corresponder ao requisito visualidade, a similaridade é apresentada na forma de uma estrutura em árvore – conforme Figura 4.10 – como alternativa à linguagem natural simples. Assim, observe que o protótipo das Figuras 4.8, 4.9, 4.10 e 4.11 foram concebidos com base na sequência de operações descrita no diagrama de atividades da Figura 4.7. Tal sequência de operações foi em conjunto com um dos entrevistados, este tendo experiência no gerenciamento de vários MOOC para o programa de capacitação em saúde do governo brasileiro, o Mais Médicos¹⁰.

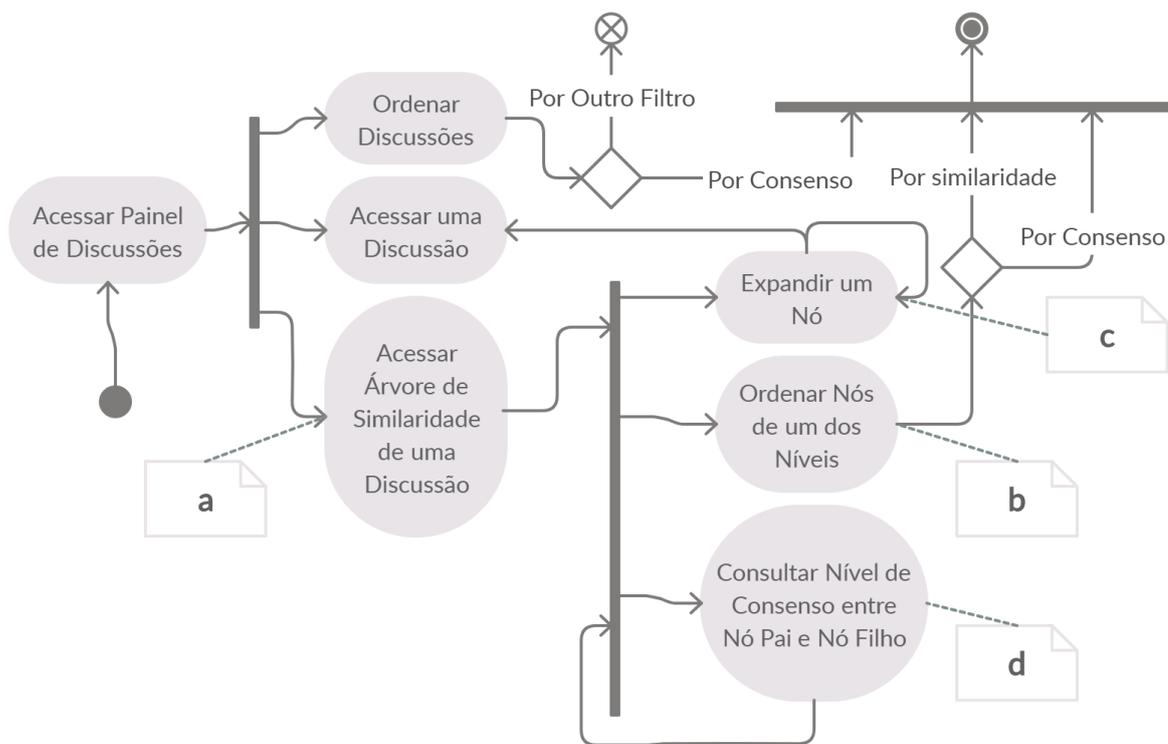


Figura 4.7: Diagrama de Atividades para Validação de Face. Letras são pontos de referência para elementos dos *mockpus*

A Figura 4.7 apresenta o diagrama do protótipo do modelo de utilidade proposto. Cada letra marcada como comentário representa uma tarefa básica: “a” representa a atividade de acessar a árvore de similaridade – a estrutura a direita da listagem de discussões nas

¹⁰<https://www.unasus.gov.br/>

Figuras 4.8, 4.9, 4.10 e 4.11; “b” representa a ação de ordenar os nós de um dos níveis de tal árvore por consenso, por exemplo; “c” representa a atividade de expandir um nó – abrir o próximo nível da árvore contendo as discussões semelhantes àquela selecionada; e, “d” representa a ação de consultar o nível de consenso daquele nó selecionado.

Para verificar a adequação dos *mockups* ao digrama de atividades, observe que tanto no diagrama – vide Figura 4.7 – quanto nos *mockups* – vide Figuras 4.8, 4.9, 4.10 e 4.11 – cada operação em comum está marcada com uma letra. Cada letra traz a referência da tarefa a qual ele está relacionado – conforme Tabela 4.2 – sendo quatro tarefas básicas elencadas com base em operações regularmente realizadas em fóruns de plataformas MOOC: Ordenar, Acessar, Consultar e Expandir discussões, embora tais atividades estejam associadas ao conceito de nível de consenso, além de ganharem uma apresentação gráfica.

Similar to: ...					
Although this course has good intentions, it requires a lot of wallowing in really bad writing, and I just fear that will be damaging. A lot can be learned by reading...	11 posts	64 views	0.73 consensus	staff replied	0.66 Similarity
Hi, I'm confused how to go about finding the 95% confidence interval for an odds ratio. Thanks!	64 posts	442 views	0.7 consensus	pending	0.6 Similarity
Hello, Does anyone know how to calculate interquartile range for binary variables?Thank you. <nameRedac_<anon_screen_name_redacted>>.	36 posts	122 views	0.69 consensus	pending	0.52 Similarity
So would that mean the probability of rolling 1 six on 3 dices or 1 six on 1 die rolled 3 times would be 1/216?	19 posts	26 views	0.59 consensus	pending	0.49 Similarity
How many binary variables? Does this mean the lines that can be designated binary or the individual entries of data? In other words x or 2x	9 posts	32 views	0.56 consensus	pending	0.48 Similarity
Hello can somebody please explain more the difference between mutually exclusive dependent and independent probability .Thank you	5 posts	11 views	0.52 consensus	pending	0.46 Similarity
Can someone explain how we go from a 2.56 risk ratio and interpret that as a 156% increase in depressive disorders?	7 posts	10 views	0.51 consensus	pending	0.42 Similarity
Hi, I'm confused how to go about finding the 95% confidence interval for an odds ratio. Thanks!	1 posts	3 views	- consensus	pending	0.4 Similarity
Hi ! As many I would like to be able to download the video lectures along with the subtitles. Thanks in advance. Cheers !	2 posts	2 views	- consensus	pending	0.39 Similarity
I got 0 marks for lack of uppercase letters or punctuation in quiz 3. How can I correct my answer?	1 posts	1 views	- consensus	pending	0.29 Similarity
I submitted very early and graded immediately four essays, but I don't have a grade yet too :(1 posts	8 views	- consensus	pending	0.25 Similarity

Figura 4.8: Mockup a

Similar to: ...

Although this course has good intentions, it requires a lot of wallowing in really bad writing, and I just fear that will be damaging. A lot can be learned by reading...	11 posts	64 views	0.73 consensus	staff replied	0.66 Similarity
Hi, I'm confused how to go about finding the 95% confidence interval for an odds ratio. Thanks!	64 posts	442 views	0.7 consensus	pending	0.6 Similarity
Hello, Does anyone know how to calculate interquartile range for binary variables?Thank you. <nameRedac_<anon_screen_name_redacted>>.	36 posts	122 views	0.69 consensus	pending	0.52 Similarity
So would that mean the probability of rolling 1 six on 3 dices or 1 six on 1 die rolled 3 times would be 1/216?	19 posts	26 views	0.59 consensus	pending	0.49 Similarity
How many binary variables? Does this mean the lines that can be designated binary or the individual entries of data? In other words x or 2x	9 posts	32 views	0.56 consensus	pending	0.48 Similarity
Hello can somebody please explain more the difference between mutually exclusive dependent and independent probability .Thank you	5 posts	11 views	0.52 consensus	pending	0.46 Similarity
Can someone explain how we go from a 2.56 risk ratio and interpret that as a 156% increase in depressive disorders?	7 posts	10 views	0.51 consensus	pending	0.42 Similarity
Hi, I'm confused how to go about finding the 95% confidence interval for an odds ratio. Thanks!	1 posts	3 views	- consensus	pending	0.4 Similarity
Hi ! As many I would like to be able to download the video lectures along with the subtitles. Thanks in advance. Cheers !	2 posts	2 views	- consensus	pending	0.39 Similarity
I got 0 marks for lack of uppercase letters or punctuation in quiz 3. How can I correct my answer?	1 posts	1 views	- consensus	pending	0.29 Similarity
I submitted very early and graded immediately four essays, but I don't have a grade yet too :(1 posts	8 views	- consensus	pending	0.25 Similarity

Figura 4.9: Mockup b

Similar to: ...

Although this course has good intentions, it requires a lot of wallowing in really bad writing, and I just fear that will be damaging. A lot can be learned by reading...	11 posts	64 views	0.73 consensus	staff replied	0.66 Similarity
Hi, I'm confused how to go about finding the 95% confidence interval for an odds ratio. Thanks!	64 posts	442 views	0.7 consensus	pending	0.6 Similarity
Hello, Does anyone know how to calculate interquartile range for binary variables?Thank you. <nameRedac_<anon_screen_name_redacted>>.	36 posts	122 views	0.69 consensus	pending	0.52 Similarity
So would that mean the probability of rolling 1 six on 3 dices or 1 six on 1 die rolled 3 times would be 1/216?	19 posts	26 views	0.59 consensus	pending	0.49 Similarity
How many binary variables? Does this mean the lines that can be designated binary or the individual entries of data? In other words x or 2x	9 posts	32 views	0.56 consensus	pending	0.48 Similarity
Hello can somebody please explain more the difference between mutually exclusive dependent and independent probability .Thank you	5 posts	11 views	0.52 consensus	pending	0.46 Similarity
Can someone explain how we go from a 2.56 risk ratio and interpret that as a 156% increase in depressive disorders?	7 posts	10 views	0.51 consensus	pending	0.42 Similarity
Hi, I'm confused how to go about finding the 95% confidence interval for an odds ratio. Thanks!	1 posts	3 views	- consensus	pending	0.4 Similarity
Hi ! As many I would like to be able to download the video lectures along with the subtitles. Thanks in advance. Cheers !	2 posts	2 views	- consensus	pending	0.39 Similarity
I got 0 marks for lack of uppercase letters or punctuation in quiz 3. How can I correct my answer?	1 posts	1 views	- consensus	pending	0.29 Similarity
I submitted very early and graded immediately four essays, but I don't have a grade yet too :(1 posts	8 views	- consensus	pending	0.25 Similarity

Figura 4.10: Mockup c

Similar to: ...					
Although this course has good intentions, it requires a lot of wallowing in really bad writing, and I just fear that will be damaging. A lot can be learned by reading...	11 posts	64 views	0.73 consensus	staff replied	0.66 Similarity
Hi, I'm confused how to go about finding the 95% confidence interval for an odds ratio. Thanks!	64 posts	442 views	0.7 consensus	pending	0.6 Similarity
Hello, Does anyone know how to calculate interquartile range for binary variables?Thank you. <nameRedac_<anon_screen_name_redacted>>.	36 posts	122 views	0.69 consensus	pending	0.52 Similarity
So would that mean the probability of rolling 1 six on 3 dices or 1 six on 1 die rolled 3 times would be 1/216?	19 posts	26 views	0.59 consensus	pending	0.49 Similarity
How many binary variables? Does this mean the lines that can be designated binary or the individual entries of data? In other words x or 2x	9 posts	32 views	0.56 consensus	pending	0.48 Similarity
Hello can somebody please explain more the diference between mutually exclusive dependent and independent probability .Thank you	5 posts	11 views	0.52 consensus	pending	0.46 Similarity
Can someone explain how we go from a 2.56 risk ratio and interpret that as a 156% increase in depressive disorders?	7 posts	10 views	0.51 consensus	pending	0.42 Similarity
Hi, I'm confused how to go about finding the 95% confidence interval for an odds ratio. Thanks!	1 posts	3 views	- consensus	pending	0.4 Similarity
Hi ! As many I would like to be able to download the video lectures along with the subtitles. Thanks in advance. Cheers !	2 posts	2 views	- consensus	pending	0.39 Similarity
I got 0 marks for lack of uppercase letters or punctuation in quiz 3. How can I correct my answear?	1 posts	1 views	- consensus	pending	0.29 Similarity
I submitted very early and graded immediatly four essays, but I don't have a grade yet too :(1 posts	8 views	- consensus	pending	0.25 Similarity

Figura 4.11: Mockup d

Entrevistas

O questionário seguiu as perguntas apresentadas na Tabela 4.2, que tinham como resposta cinco alternativas conforme escala Likert – ‘concordo fortemente’, ‘concordo’, ‘neutro’, ‘discordo’ e ‘discordo fortemente’. Embora tais questões tenham resposta objetiva, estas fomentavam uma discussão que levava o entrevistador (o autor) a avaliar o sentimento de cada entrevistado a respeito da ferramenta, conforme Tabela 4.3. Além disto, a discussão posterior a cada pergunta eventualmente acrescentava novos questionamentos que foram incorporados ao questionário. Em função desta dinâmica do questionário, a primeira coluna da Tabela 4.2 traz o total de entrevistados que responderam cada pergunta.

O sentimento percebido do entrevistado para com a ferramenta era avaliado – pelo entrevistador – com base nos seguintes critérios de pontuação:

- **Compreensão:** Cada entrevistado recebia uma breve descrição do que a ferramenta pretendia oferecer. Daí, a cada pergunta relacionada a operação da ferramenta, era

decrementado um ponto de sua nota, partindo de uma nota máxima igual a 10;

- **Aprovação:** Este requisito era avaliado diretamente com base nas reações dos entrevistados e apenas três pontuações eram possíveis: ‘Aprovação Total’, representada pela nota 10; ‘Indeterminado’, representada pela nota 5; e, ‘Reprovação’, representada pela nota 0;
- **Interesse:** Para este critério, a cada pergunta ou sugestão do entrevistado a respeito das possibilidades da ferramenta, era acrescentado dois pontos a sua nota, partindo de uma nota mínima igual a 0.

Tabela 4.2: Questionário para Validação de Face.

	Perguntas	Tela a	Tela b	Tela c	Tela d	Respondentes
1	É relevante para o esforço de suporte ao fórum?	9,07	8,14	9,07	8,14	4
2	A interface apresenta a informação de modo compreensível?	9,07	9,07	9,07	8,14	3
3	Há necessidade de alguma orientação para usar a interface?	8,14	8,14	6,21	6,21	3
4	Com esta ferramenta é possível dar suporte ao fórum em menos tempo?	9,07	9,07	10	9,07	4

Tabela 4.3: Avaliação do Entrevistado pelo Pesquisador.

	Participantes	1	2	3	4
1	Compreensão	7	8	9	6
2	Aprovação	5	5	10	5
3	Interesse	6	6	10	8

Resultados

Como é possível observar na Tabela 4.2, maioria dos entrevistados concordam que a ferramenta é relevante tendo todos reconhecido as tarefas como relevantes, sendo que nenhum dos entrevistados discorda de nenhuma das questões (nenhum avaliação inferior a ‘Neutro’). Tal apoio foi observado tanto a partir do questionário, quanto durante a entrevista.

Com base nas impressões do entrevistador a respeito das reações dos entrevistados – conforme Tabela 4.3 – identificou-se certa confusão ao observar as primeiras telas dos *mockups*. Contudo, por se tratar de uma interface similar àquela já utilizada nos fóruns e razoavelmente intuitiva, a confusão foi apenas inicial. Apenas o entrevistado número quatro teve um pouco mais de dificuldade em operar e compreender a utilidade da ferramenta.

4.6 Resultados e Discussão

Aqui foi apresentado uma ferramenta que recomenda ao instrutor de um fórum educacional discussões que requerem uma intervenção urgente. Diferente de outras alternativas, a ferramenta proposta aqui não recomenda urgência diretamente, mas oferece informações para que o instrutor reconheça a urgência mais facilmente. Isto se dá através de medidas de similaridade entre discussões e exposição de níveis de consenso internos e externos à cada discussão. Assim, eliminando a necessidade de dados marcados para Aprendizagem Supervisionada ou de retreinamento periódico. Deste modo, caso o cenário educacional ao qual esta ferramenta está inserida mude, a ferramenta desenvolvida a partir do modelo não irá deixar de funcionar de forma silenciosa até o próximo retreinamento.

4.6.1 Ameaças à Validade

Por se tratar de um trabalho longo, cheio de etapas, certas ameaças à validade do resultado final podem ser apontadas. Para esta última Parte, a principal ameaça à validade está no método qualitativo de validação por entrevistas semi estruturadas envolvendo um número reduzido de *stakeholders*. Tal ameaça, assim como recomenda [40], foi contornada a partir de outras formas de validação mais objetivas – vide Seção 4.5.2.

Capítulo 5

Resultados, Conclusões e Trabalhos

Futuros

Aqui discutimos conclusões e possibilidades futuras para 1) o uso de *Rationale* como fonte de requisitos para pesquisa e atualização de software em plataformas colaborativas; e, 2) o consenso como indicativo de urgência apontando a necessidade de intervenção para o instrutor.

5.1 *Rationale* como Fonte de Requisitos de Software em Plataformas Colaborativas

No Capítulo 3 deste trabalho foi apresentado um estudo que investiga o potencial da mineração de *User Rationale* espontaneamente compartilhado em fóruns MOOC como fonte de requisitos para pesquisa e manutenção do ferramental de software da plataforma em que estão inseridos. Isto foi feito a partir do estudo do caso onde era feita uma avaliação, com base em *Rationale*, das demandas por novas funcionalidades que pudessem melhorar a experiência em um MOOC. Assim foi definida a seguinte questão de pesquisa para este capítulo **Q1:** *O Rationale subjacente às postagens nos fóruns poderia dar suporte à manutenção, atualização ou inovação em plataformas de software educacional?*

Como os dois principais resultados podemos apontar: o direcionamento para a continuação do trabalho investigando estratégias de apoio ao instrutor por meio de consenso; e,

evidências de que *User Rationale* pode ser utilizado como fonte de diretrizes de pesquisa e requisitos de manutenção de software. Alerta-se para o fato de que foi utilizada uma metodologia exploratória e manual de investigação e seus resultados devem ser vistos como preliminares. Resultados mais definitivos demandam uma análise estatística mais rigorosa.

Como próximos passos para investigação nesta linha do trabalho, um estratégia mais automatizada poderia trazer mais rigor aos resultados. Neste sentido, dois trabalhos podem ser tomados como ponto de partida para se automatizar a extração de *User Rationale* em fóruns educacionais não técnicos a respeito da própria plataforma utilizada. O primeiro [50] apresenta uma metodologia completa em extração e classificação automática de *User Rationale*. Aliado a este trabalho, o segundo artigo [83] explora fóruns de discussão usando um algoritmo de aprendizado de máquina para analisar *User Rationale* em relatos de testes entre desenvolvedores para a plataforma Android¹ por meio da plataforma de fóruns Stackoverflow².

5.2 O Consenso em Fóruns Educacionais

No Capítulo 4 buscou-se estudar a correlação entre o consenso em discussões de fóruns MOOC e a qualidade das conclusões. Assim, definiu-se a questão de pesquisa **Q2**: *Em fóruns de discussão educacionais, existe correlação entre o nível de consenso e a qualidade da opinião consensual?* Resultados mostraram que o consenso em seus extremos, no contexto dos dados considerados (**D1**), é inversamente proporcional à qualidade em um fórum MOOC – sendo a opinião dos instrutores nosso parâmetro de qualidade – confirmando a hipótese alternativa (existe correlação entre consenso e qualidade). Aqui, por se tratar de resultados estatísticos em uma base de dados relativamente diversa – vide Seção 3.2.1 – podemos tomar tais resultados como mais consistentes.

O auto-aprendizado, com o aluno buscando o conhecimento por seus próprios meios, é visto como uma propriedade de um perfil de aluno muito particular [90]. Contudo, algumas ferramentas, dentre elas os fóruns de discussão, apresentam-se como facilitadoras para um certo nível de auto-aprendizado. Neste sentido, simplificar o esforço de buscar conhecimento

¹<https://www.android.com/>

²<https://stackoverflow.com/>

em um fórum, interagir com ele, ou até mesmo moderá-lo, tem sido alvo de bastante atenção da comunidade científica e o consenso é definido como uma ferramenta útil, porém controversa [81]. Sendo assim, este trabalho contribui com uma maior clareza da função ou até da utilidade real de se conhecer o grau de consenso em discussões online. Intuitivamente, reconhecer consenso nos parece algo útil para muitas das atividades de aprendizado online, particularmente em MOOC. Infelizmente, conforme apresentado neste artigo, a relação entre o consenso e a qualidade da discussão pode não ser tão clara, merecendo mais esforços de pesquisa.

Um nível elevado de consenso, embora possa ser interpretada como indicativo de correte, pode gerar desinformação caso não esteja correto. Um estudante navegando em um fórum online corre o risco de interromper sua busca ao encontrar a resposta que satisfaça sua busca, ainda que esta possa não estar necessariamente correta. Assim, se uma discussão que tenha chegado à uma conclusão equivocada com um nível de consenso elevado for visualizada por 90 estudantes em um curso com 600 estudantes – que corresponde a uma taxa habitual – é possível que 15% dos estudantes daquele curso assimilem uma informação errada.

5.3 Uma Ferramenta de Suporte ao Trabalho do Instrutor em Fóruns MOOC

Na literatura até aqui, ferramentas como estas foram apresentadas com base em modelos supervisionados e pré-treinados. Isto é um risco, uma vez que tais modelos não evoluem junto com a plataforma e correm o risco de “envelhecer” caso o cenário mude e os dados usados para o treinamento não mas se apliquem. Tal situação levaria a recomendações incorretas. Já o modelo apresentado aqui dá ao instrutor informações fortemente relacionadas à urgência sem a necessidade de treinamento utilizando dados marcados – supervisionado. Tal estratégia aplicada por meio de uma ferramenta gráfica ajuda a julgar a necessidade de intervenção urgente. Assim, caso o cenário educacional ao qual esta ferramenta está inserida mude, uma ferramenta desenvolvida a partir do modelo não deixará de funcionar de forma silenciosa, e o senso crítico do instrutor poderá identificar isto e atualizações poderão ser feitas. Assim, respondemos a questão principal de pesquisa **QP**: *Como ampliar o potencial de suporte da*

equipe de instrutores por meio do fórum de discussão em um MOOC? Para tal, apontamos o modelo de aplicação proposto no Capítulo 4 como resposta a questão de pesquisa.

Como próximo passo mais direto temos a implementação de uma interface funcional para a ferramenta. Um exemplo de tecnologia que poderia viabilizar tal implementação seria a plataforma Moodle³, ou mesmo código fonte da plataforma OpenedX. Com uma ferramenta funcional seria possível realizar testes práticos. Outra alternativa possível seria desenvolver um modelo que identificasse a desinformação diretamente, sem o consenso como ponto. Embora, neste caso haveriam desafios bem maiores.

³https://moodle.org/?lang=pt_br

Bibliografía

- [1] Akshay Agrawal, Jagadish Venkatraman, Shane Leonard, and Andreas Paepcke. You-edu: addressing confusion in mooc discussion forums by recommending instructional video clips. 2015.
- [2] Carlos Alario-Hoyos, Mar Perez-Sanagustin, Carlos Delgado-Kloos, Mario Munoz-Organero, et al. Delving into participants' profiles and use of social tools in moocs. *IEEE Transactions on Learning Technologies*, 7(3):260–266, 2014.
- [3] Rana Alkadhi, Teodora Lata, Emitza Guzman, and Bernd Bruegge. Rationale in development chat messages: an exploratory study. In *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*, pages 436–446. IEEE, 2017.
- [4] Rana Alkadhi, Manuel Nonnenmacher, Emitza Guzman, and Bernd Bruegge. How do developers discuss rationale? In *2018 IEEE 25th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pages 357–369. IEEE, 2018.
- [5] Omaira Almatrafi and Aditya Johri. Systematic review of discussion forums in massive open online courses (moocs). *IEEE Transactions on Learning Technologies*, 2018.
- [6] Omaira Almatrafi, Aditya Johri, and Huzefa Rangwala. Needle in a haystack: Identifying learner posts that require urgent response in mooc discussion forums. *Computers & Education*, 118:1–9, 2018.
- [7] Sergio Alonso, Ignacio J Pérez, Francisco Javier Cabrerizo, and Enrique Herrera-Viedma. A linguistic consensus model for web 2.0 communities. *Applied Soft Computing*, 13(1):149–157, 2013.

-
- [8] Manuela Aparicio, Tiago Oliveira, Fernando Bacao, and Marco Painho. Gamification: A key determinant of massive open online course (mooc) success. *Information & Management*, 56(1):39–54, 2019.
- [9] Yoav Artzi, Patrick Pantel, and Michael Gamon. Predicting responses to microblog posts. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 602–606. Association for Computational Linguistics, 2012.
- [10] Juliano Assunção and Gustavo Gonzaga. Educação profissional no brasil: inserção e retorno. 2015.
- [11] Bernard M Bass. Amount of participation, coalescence, and profitability of decision making discussions. *The Journal of Abnormal and Social Psychology*, 67(1):92, 1963.
- [12] Kent Beck and Erich Gamma. *Extreme programming explained: embrace change*. addison-wesley professional, 2000.
- [13] Harry N Boone and Deborah A Boone. Analyzing likert data. *Journal of extension*, 50(2):1–5, 2012.
- [14] Christopher G Brinton, Mung Chiang, Shaili Jain, Henry Lam, Zhenming Liu, and Felix Ming Fai Wong. Learning about social learning in “moocs”: From statistical analysis to generative model. *IEEE transactions on Learning Technologies*, 7(4):346–359, 2014.
- [15] Nicholas Buhagiar, Bahram Zahir, and Abdolreza Abhari. Using deep learning to recommend discussion threads to users in an online forum. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.
- [16] Francisco Javier Cabrerizo, Francisco Chiclana, Rami Al-Hmouz, Ali Morfeq, Abdullah Saeed Balamash, and Enrique Herrera-Viedma. Fuzzy decision making and consensus: challenges. *Journal of Intelligent & Fuzzy Systems*, 29(3):1109–1118, 2015.
- [17] Kaushik Chakrabarti and Sharad Mehrotra. The hybrid tree: An index structure for high dimensional feature spaces. In *Proceedings 15th International Conference on Data Engineering (Cat. No. 99CB36337)*, pages 440–447. IEEE, 1999.

- [18] Muthu Kumar Chandrasekaran, Min-Yen Kan, Bernard CY Tan, and Kiruthika Ragu-pathi. Learning instructor intervention from mooc forums: Early results and issues. *arXiv preprint arXiv:1504.07206*, 2015.
- [19] Preetha Chatterjee, Kostadin Damevski, Lori Pollock, Vinay Augustine, and Nicholas A Kraft. Exploratory study of slack q&a chats as a mining source for software engineering tools. In *Proceedings of the 16th International Conference on Mining Software Repositories (MSR'19)*. ACM, 2019.
- [20] Snigdha Chaturvedi, Dan Goldwasser, and Hal Daumé III. Predicting instructor's intervention in mooc forums. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1501–1511, 2014.
- [21] Shaiful Alam Chowdhury and Abram Hindle. Mining stackoverflow to filter out off-topic irc discussion. In *Proceedings of the 12th Working Conference on Mining Software Repositories*, pages 422–425. IEEE Press, 2015.
- [22] T. Clementino. Do suporte ao reconhecimento de consenso em discussões estruturadas: Uma revisão exploratória. Technical report, Universidade Federal de Campina Grande, 2019. Disponível em: https://github.com/tiago-clementino/consensus_x_quality/blob/master/consensus/doc/consenso.pdf.
- [23] T. Clementino. Systematic literature review: Colaborative engagement on moocs. Technical report, Universidade Federal de Campina Grande, 2019. Disponível em: https://github.com/tiago-clementino/rationale-research-introduction/blob/master/doc/mooc_tools_review.pdf.
- [24] Tiago Clementino and Antão Moura. Consenso como indicativo de qualidade em fóruns educacionais: Uma análise no âmbito de mooc. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 30, page 1965, 2019.
- [25] Tiago Clementino and Antão Moura. Rationale in forum discussions as support for software upgrading of online course platforms. In *Brazilian Symposium on Computers*

- in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 30, page 349, 2019.
- [26] Yi Cui and Alyssa Friend Wise. Identifying content-related threads in mooc discussion forums. In *Proceedings of the 2nd ACM Conference on Learning@ Scale*, pages 299–303, 2015.
- [27] Fred J Damerau. A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7(3):171–176, 1964.
- [28] Venkatesh T Dhinakaran, Raseshwari Pulle, Nirav Ajmeri, and Pradeep K Murukanaiyah. App review analysis via active learning: reducing supervision effort without compromising classification accuracy. In *2018 IEEE 26th International Requirements Engineering Conference (RE)*, pages 170–181. IEEE, 2018.
- [29] Tawanna R Dillahunt, Brian Zengguang Wang, and Stephanie Teasley. Democratizing higher education: Exploring mooc use among those who cannot afford a formal education. *The International Review of Research in Open and Distributed Learning*, 15(5), 2014.
- [30] Willem Doise, Gabriel Mugny, A St James, Nicholas Emler, and D Mackie. *The social development of the intellect*, volume 10. Elsevier, 2013.
- [31] Stephen Downes. An introduction to connective knowledge, 2005.
- [32] Dalia Elalfy, Walaa Gad, and Rasha Ismail. A hybrid model to predict best answers in question answering communities. *Egyptian informatics journal*, 19(1):21–31, 2018.
- [33] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- [34] Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2006.
- [35] Wael H Gomaa and Aly A Fahmy. A survey of text similarity approaches. *International Journal of Computer Applications*, 68(13):13–18, 2013.

- [36] Shou Xi Guo, Xia Sun, Shi Xiong Wang, Yi Gao, and Jun Feng. Attention-based character-word hybrid neural networks with semantic and structural information for identifying of urgent posts in mooc discussion forums. *IEEE Access*, 7:120522–120532, 2019.
- [37] Enrique Herrera-Viedma, Francisco Javier Cabrerizo, Janusz Kacprzyk, and Witold Pedrycz. A review of soft consensus models in a fuzzy environment. *Information Fusion*, 17:4–13, 2014.
- [38] Sushant Hiray and Venkatesh Duppada. Agree to disagree: Improving disagreement detection with dual grus. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pages 147–152. IEEE, 2017.
- [39] Randy Y Hirokawa. Consensus group decision-making, quality of decision, and group satisfaction: An attempt to sort “fact” from “fiction”. *Communication Studies*, 33(2):407–415, 1982.
- [40] Ronald R Holden. Face validity. *The corsini encyclopedia of psychology*, pages 1–2, 2010.
- [41] Irving L Janis and Leon Mann. *Decision making: A psychological analysis of conflict, choice, and commitment*. Free press, 1977.
- [42] Nishant Jha and Anas Mahmoud. Mining user requirements from application store reviews using frame semantics. In *International working conference on requirements engineering: Foundation for software quality*, pages 273–287. Springer, 2017.
- [43] David W Johnson and Roger T Johnson. Conflict in the classroom: Controversy and learning. *Review of educational research*, 49(1):51–69, 1979.
- [44] A Kardan, Amir Narimani, and Foozhan Ataiefard. A hybrid approach for thread recommendation in mooc forums. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 11(10):2195–2201, 2017.

- [45] Mohammad Khalil and Martin Ebner. What massive open online course (mooc) stakeholders can learn from learning analytics? *Learning, design, and technology: An international compendium of theory, research, practice, and policy*, pages 1–30, 2016.
- [46] Javed Ali Khan, Lin Liu, Lijie Wen, and Raian Ali. Crowd intelligence in requirements engineering: Current status and future directions. In *International working conference on requirements engineering: Foundation for software quality*, pages 245–261. Springer, 2019.
- [47] Jihie Kim, Jia Li, and Taehwan Kim. Towards identifying unresolved discussions in student online forums. In *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 84–91. Association for Computational Linguistics, 2010.
- [48] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [49] Peter Kolb. Disco: A multilingual database of distributionally similar words. *Proceedings of KONVENS-2008, Berlin*, 156, 2008.
- [50] Zijad Kurtanović and Walid Maalej. Mining user rationale from software reviews. In *2017 IEEE 25th International Requirements Engineering Conference (RE)*, pages 61–70. IEEE, 2017.
- [51] Zijad Kurtanović and Walid Maalej. On user rationale in software engineering. *Requirements Engineering*, 23(3):357–379, 2018.
- [52] Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. From word embeddings to document distances. In *International conference on machine learning*, pages 957–966, 2015.
- [53] Jintae Lee. Design rationale systems: understanding the issues. *IEEE expert*, 12(3):78–85, 1997.
- [54] Vladimir I Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710, 1966.

- [55] Rensis Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.
- [56] Dekang Lin. Automatic retrieval and clustering of similar words. In *COLING 1998 Volume 2: The 17th International Conference on Computational Linguistics*, volume 2, 1998.
- [57] Tharindu Rekha Liyanagunawardena, Andrew Alexandar Adams, and Shirley Ann Williams. Moocs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distributed Learning*, 14(3):202–227, 2013.
- [58] Margaret Mazzolini and Sarah Maddison. When to jump in: The role of the instructor in online discussion forums. *Computers & Education*, 49(2):193–213, 2007.
- [59] Alexander McAuley, Bonnie Stewart, George Siemens, and Dave Cormier. The mooc model for digital practice. 2010.
- [60] Fei Mi and Boi Faltings. Adaptive sequential recommendation for discussion forums on moocs using context trees. In *EDM*, 2017.
- [61] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [62] George Miller. *WordNet: An electronic lexical database*. MIT press, 1998.
- [63] Sunny Miranda, Fernanda Ferreira, Evandro Costa, and Fábio Paraguaçu. Experimento de uma ferramenta de fórum baseado em mapas conceituais: uma nova forma de interação. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 26, page 917, 2015.
- [64] Randal S Olson and Jason H Moore. Tpot: A tree-based pipeline optimization tool for automating machine learning. In *Automated Machine Learning*, pages 151–160. Springer, 2019.
- [65] Ignacio J Pérez, Francisco Javier Cabrerizo, Sergio Alonso, YC Dong, Francisco Chiclana, and Enrique Herrera-Viedma. On dynamic consensus processes in group decision making problems. *Information Sciences*, 459:20–35, 2018.

- [66] Thiago RM Pessoa and Daniel F Pires. Imt: Uma api para suporte a comunicação instantânea em ferramentas de educação a distância.
- [67] Peter Potash and Anna Rumshisky. Towards debate automation: a recurrent model for predicting debate winners. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2465–2475, 2017.
- [68] Minghui Qiu and Jing Jiang. A latent variable model for viewpoint discovery from threaded forum posts. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1031–1040, 2013.
- [69] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144. ACM, 2016.
- [70] Jennifer C Richardson, Adrie A Koehler, Erin D Besser, Secil Caskurlu, JiEun Lim, and Chad M Mueller. Conceptualizing and investigating instructor presence in online learning environments. *The International Review of Research in Open and Distributed Learning*, 16(3), 2015.
- [71] John T Robinson. The kdb-tree: a search structure for large multidimensional dynamic indexes. In *Proceedings of the 1981 ACM SIGMOD international conference on Management of data*, pages 10–18. ACM, 1981.
- [72] Cristóbal Romero and Sebastián Ventura. Educational data science in massive open online courses. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(1):e1187, 2017.
- [73] Sara Rosenthal and Kathy McKeown. I couldn't agree more: The role of conversational structure in agreement and disagreement detection in online discussions. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 168–177, 2015.

- [74] Ken Schwaber. Scrum development process. In *Business object design and implementation*, pages 117–134. Springer, 1997.
- [75] Bakhtawar Seerat and Farouque Azam. Opinion mining: Issues and challenges (a survey). *International Journal of Computer Applications*, 49(9), 2012.
- [76] George Siemens and Peter Tittenberger. *Handbook of emerging technologies for learning*. University of Manitoba Winnipeg, 2009.
- [77] Ian Sommerville. Software engineering 9th edition. *ISBN-10*, 137035152, 2011.
- [78] Thomas Staubitz, Tobias Pfeiffer, Jan Renz, Christian Willems, and Christoph Meinel. Collaborative learning in a “moooc” environment. In *Proceedings of the 8th annual international conference of education, research and innovation*, pages 8237–8246, 2015.
- [79] Xia Sun, Shouxi Guo, Yi Gao, Jun Zhang, Xiaolin Xiao, and Jun Feng. Identification of urgent posts in mooc discussion forums using an improved rcnn. In *2019 IEEE World Conference on Engineering Education (EDUNINE)*, pages 1–5. IEEE, 2019.
- [80] Peter Tolmie, Rob Procter, Mark Rouncefield, Maria Liakata, and Arkaitz Zubiaga. Microblog analysis as a program of work. *ACM Transactions on Social Computing*, 1(1):2, 2018.
- [81] John Trimbur. Consensus and difference in collaborative learning. *College English*, 51(6):602–616, 1989.
- [82] Daan Van Knippenberg, Carsten KW De Dreu, and Astrid C Homan. Work group diversity and group performance: an integrative model and research agenda. *Journal of applied psychology*, 89(6):1008, 2004.
- [83] Isabel K Villanes, Silvia M Ascate, Josias Gomes, and Arilo Claudio Dias-Neto. What are software engineers asking about android testing on stack overflow? In *Proceedings of the 31st Brazilian Symposium on Software Engineering*, pages 104–113. ACM, 2017.
- [84] G Vinodhini and RM Chandrasekaran. Sentiment analysis and opinion mining: a survey. *International Journal*, 2(6):282–292, 2012.

-
- [85] Li Wang, Su Nam Kim, and Timothy Baldwin. The utility of discourse structure in identifying resolved threads in technical user forums. In *Proceedings of COLING 2012*, pages 2739–2756, 2012.
- [86] Grant S Williams. Modeling crowd feedback in the mobile app market. 2019.
- [87] A. F. Wise, Y. Cui, W. Jin, and J. Vytasek. Mining for gold: Identifying content-related “mooc” discussion threads across domains through linguistic modeling. *Internet and Higher Education*, 32:11–28, jan 2017.
- [88] Alyssa Friend Wise, Yi Cui, and Jovita Vytasek. Bringing order to chaos in mooc discussion forums with content-related thread identification. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, pages 188–197. ACM, 2016.
- [89] Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, Jun Zhao, et al. Relation classification via convolutional deep neural network. 2014.
- [90] Barry J Zimmerman and Dale H Schunk. *Self-regulated learning and academic achievement: Theoretical perspectives*. Routledge, 2001.
- [91] Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, and Michal Lukasik. Stance classification in rumours as a sequential task exploiting the tree structure of social media conversations. In *Proceedings of COLING, the International Conference on Computational Linguistics*, pages 2438–2448, 2016.