

Consenso como Indicativo de Qualidade em Fóruns Educaçãoais: Uma Análise no Âmbito de MOOC

Tiago L. P. Clementino¹, José Antão B. Moura¹

¹Departamento de Sistemas e Computação – Universidade Federal de Campina Grande (UFCG)
Caixa Postal 10.106 – 58.429-900 – Campina Grande – PB – Brasil

tiagolucas@copin.ufcg.edu.br, antao@dsc.ufcg.edu.br

Abstract. *Chat discussions and forums in online teaching platforms, such as those for Massive Open Online Courses – MOOC, serve to build consensual conclusions that add pedagogical value. The study of the correlation between consensus and the quality of conclusion in online discussions has received little attention. This article uses “Soft Consensus” – when consensus may be graded, from none to unanimity – to present a statistical study where such correlation is measured from MOOC forum data and instructor support for consensual conclusion is the indicator of quality. Preliminary results point to a negative correlation.*

Resumo. *Discussões em chats e fóruns de plataformas de ensino online, como aquelas para Massive Open Online Courses – MOOC, servem para construir conclusões consensuais que agreguem valor pedagógico. O estudo da correlação entre consenso e a qualidade da conclusão em discussões online tem recebido pouca atenção. Este artigo usa “Soft Consensus” – quando há gradação do consenso, desde nenhum até unanimidade – para apresentar um estudo estatístico onde tal correlação é medida a partir de dados de fóruns de MOOC e o apoio dos instrutores à conclusão consensual é o indicador de qualidade. Resultados preliminares apontam uma correlação negativa.*

1. Introdução

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 consensual ou definitiva. Diante da dificuldade de se alcançarem conclusões definitivas em certos temas de debate, o consenso ganha força como instrumento de avaliação de qualidade da solução proposta. Além disto, a presença de dois ou mais grupos de consenso destoantes em uma mesma discussão pode apontar uma necessidade de intervenção do instrutor. Aqui, “consenso” é definido como segue.

Definição 1: “Consenso” é uma medida de concordância entre as respostas ou opiniões de um grupo de indivíduos sobre determinada questão ou assunto.

A definição 1 permite modelar desde uma concordância parcial até a unanimidade. A definição se baseia naquela de *Soft Consensus* [Herrera-Viedma et al. 2014] que acomoda incertezas.

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 conjectura, 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 (do inglês, *Massive Open Online Courses*), tomando o apoio do corpo de instrutores à conclusão consensual como parâmetro de qualidade.

Este artigo pretende avaliar a correlação entre consenso e qualidade em discussões online, e especificamente no ambiente educacional (MOOC). Para tanto, utilizou-se uma base de dados 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.

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, mas agora no contexto de educação online.

2. 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 [Johnson and Johnson 1979]. A divergência leva à reflexão que promove o entendimento profundo [Doise et al. 2013, Van Knippenberg et al. 2004]. O conflito motiva a colaboração [Doise et al. 2013]. Todavia, tomar uma conclusão consensual como sendo de qualidade pode ser uma decisão precipitada. [Hirokawa 1982] 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 [Janis and Mann 1977], de modo geral, o grupo tende a aceitar melhor decisões consensuais. Já [Bass 1963] mediu se 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 [Cabrerizo et al. 2015]. O advento da Web 2.0 e 3.0 e as comunidades online tornaram o ambiente de discussões ainda mais flexível [Herrera-Viedma et al. 2014, Alonso et al. 2013]. 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 [Pérez et al. 2018]. A análise de discurso tem mostrado resultados significativos em áreas como marketing [Hiray and Duppada 2017, Rosenthal and McKeown 2015, Vinodhini and Chandrasekaran 2012, Seerat and Azam 2012] e política [Tolmie et al. 2018, Potash and Rumshisky 2017, Zubiaga et al. 2016, Qiu and Jiang 2013]. Em análise de discurso, o consenso é frequentemente um importante argumento de relevância [Zubiaga et al. 2016, Rosenthal and McKeown 2015].

[Potash and Rumshisky 2017] traz 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 [Zubiaga et al. 2016] apresenta uma abordagem que agrega opiniões em núcleos de consenso com o intuito de diferenciar rumores de fatos em redes sociais. [Rosenthal and McKeown 2015] 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 [Hirokawa 1982].

Aqui, como será descrito na próxima seção, 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 [Gomaa and Fahmy 2013]. 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 deste 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. Design de Pesquisa

O presente 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, *este estudo objetiva 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 é a seguinte:

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

4. Os Dados

Os dados utilizados nesta pesquisa são fruto de discussões em fóruns de 12 MOOC ofertados em Inglês pela Universidade de Stanford na plataforma edX entre 2013 e 2015. Tais MOOC abrangiam três áreas do conhecimento: Ciências Humanas, Medicina e Educação; totalizando 29.604 postagens agrupadas em 22.804 discussões. As discussões podem ainda ser divididas do seguinte modo: 20.268 tendo apenas uma postagem; 2.208 tendo entre 2 e 5 postagens; 179 tendo entre 6 e 10 postagens; e 48 com mais de 11 postagens. Aqui, apenas as discussões com pelo menos duas postagens foram utilizadas.

4.1. Tratamento Manual

Os dados foram utilizados originalmente por [Agrawal et al. 2015]. 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 demanda urgente de intervenção de um instrutor; “positividade”, o quão afirmativa é a postagem; e, “confusão”, o nível de incerteza associado a afirmação na postagem. As postagens foram ainda classificadas como pergunta $\{0, 1\}$ – não ou sim, respectivamente; resposta $\{0, 1\}$; ou, opinião $\{0, 1\}$. Deve-se notar que o campo “confusão” parece ter sido mensurado em escala invertida, medindo na verdade a “convicção” em cada postagem.

Com os dados já tratados, classificamos manualmente cada postagem como sendo de autoria ou não de um instrutor $\{0, 1\}$. Além disto, urgência, positividade e convicção foram mapeados para a faixa de números decimais $[0, 0; 1, 0]$ a fim de normalizar nossa análise.

4.2. Tratamento Automático

Antes de efetuar os cálculos de consenso, algumas providências foram tomadas. Com o objetivo de melhorar o desempenho dos algoritmos de distância textual, sinais de pontuação foram substituídos por espaços em branco e todos os *tokens* (palavras, numerais, links, etc, desde que delimitados por espaços em branco) com dois caracteres ou menos, foram eliminados. Em seguida foi aplicada uma estratégia de remoção de *stop words* (palavras com pouca informação semântica como “de”, “para”, “por”, 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 4.1 foram calculados para possível utilização na análise estatística.

5. Mensurando Consenso

Como foi adiantado na Definição 1, *Soft Consensus* é um conceito onde o consenso é definido a partir de modelos capazes de assimilar incerteza [Herrera-Viedma et al. 2014]. Contudo, tal potencial diz respeito apenas a quantificadores linguísticos e Lógica Difusa. A maioria destes modelos segue metodologias rígidas divididas em rodadas de discussão e geralmente moderadas por um agente externo [Cabrerizo et al. 2015], embora existam alternativas mais flexíveis [Alonso et al. 2013]. 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. Os dados utilizados já discriminam opiniões, perguntas e respostas. Deste modo, 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 [Gomaa and Fahmy 2013] existem três classes principais de algoritmos de similaridade

textual. Diante disto, analisou-se o consenso em função de três algoritmos diferentes de distância textual, 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 [Kolb 2008], por se tratar de uma tecnologia bem estabelecida e com implementação robusta. 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 [Miller 1998], 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 dois a dois (conforme descrito na Subseção 5.2); 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.

5.1. 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 [Gomaa and Fahmy 2013], 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 [Damerau 1964, Levenshtein 1966], 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.

DISCO: DISCO [Kolb 2008] pode ser descrita como relação distributiva e semelhança distributiva, simultaneamente. Duas palavras são distributivamente semelhantes se possuem muitas palavras co-ocorrentes 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 (tal como uma Wordnet), porém faz aproximações livres de contexto de tais relações. Para tanto [Kolb 2008] utilizou métricas para mensurar informação mútua baseadas em [Lin 1998].

Wordnet: A Wordnet é uma base de dados textual semântico-sintática [Miller 1998]. 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.

5.2. “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 [Herrera-Viedma et al. 2014, Cabrerizo et al. 2015]. Porém, diferenciando-se de discussões convencionais, os fóruns online agregam incerteza até mesmo no processo de discussão. Sendo assim, no âmbito deste estudo, é necessário fazer adaptações nos modelos que tratam consenso com incerteza [Pérez et al. 2018]. Os detalhes do desenvolvimento matemático para o cálculo de consenso feito nesta pesquisa estão disponíveis em https://github.com/tiago-clementino/consensus_x_quality. Tendo realizado os cálculos necessários para identificar o nível de consenso, três resultados são especialmente úteis: i) Uma matriz de preferências agregadas por opinião para cada métrica de proximidade entre cada opinião e cada postagem (i.e., as linhas da matriz são indexadas por opinião). Esta matriz ($n \times m$) – onde n é o total de opiniões identificadas no conjunto de postagens para cada discussão e m , o total de postagens por discussão – é usada para identificar qual opinião é mais forte (maior média harmônica de cada linha); ii) Uma matriz de preferências agregadas por opinião – proximidade entre cada opinião e cada postagem (i.e., as linhas da matriz são indexadas por postagem); e, iii) o nível de geral de consenso cg , definido por

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

Onde cx_k é o nível de consenso da opinião k e igual à média harmônica da linha k em i).

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

$$q_{x_k} = \sum_{i=1}^l \frac{a_i(x_k)}{l} \quad (2)$$

Na Equação 2, 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 $[0, 0; 0, 1]$.

6. Consenso vs. Qualidade

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. Mais detalhes em [https://github.com/tiago-clementino/consensus x quality](https://github.com/tiago-clementino/consensus-x-quality).

Para nossa análise estatística, recorreremos a um modelo de regressão logística [Gelman and Hill 2006] 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, acompanhado do 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, 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, além de que um grande volume de postagens leva à uma dispersão no foco de discussão que pode prejudicar o consenso. A partir da Tabela 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 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 1. Regressão logística: apoio dos instrutores e convicção média como fatores de consenso

	Efeito	fixos		aleatórios	
		β	s.e.	s^2	σ
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
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 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 o estudo 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 QP, 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 [Gelman and Hill 2006].

7. 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.

8. 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 definida. Uma alta convicção média, embora possa ser interpretada como indicativo de correteude, pode dificultar o consenso. Por exemplo, duas opiniões opostas e convictas podem levar à confusão.

Estes resultados estão em linha com aqueles de [Hirokawa 1982] e [Janis and Mann 1977]. 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 [Brinton et al. 2014], 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, por ser apresentada por alunos ainda inexperientes, a opinião consensual tenderia a não ter o apoio dos instrutores. Ainda sim, tal consideração não invalida os resultados alcançados aqui para a educação online.

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 [Zimmerman and Schunk 2001]. 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 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 [Trimbur 1989]. 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. Por exemplo: considerar – além da qualidade da conclusão – satisfação, interesse ou compreensão dos alunos como indicadores adicionais de qualidade.

Referências

- Agrawal, A., Venkatraman, J., Leonard, S., and Paepcke, A. (2015). Youedu: addressing confusion in mooc discussion forums by recommending instructional video clips.
- Alonso, S., Pérez, I. J., Cabrerizo, F. J., and Herrera-Viedma, E. (2013). A linguistic consensus model for web 2.0 communities. *Applied Soft Computing*, 13(1):149–157.
- Bass, B. M. (1963). Amount of participation, coalescence, and profitability of decision making discussions. *The Journal of Abnormal and Social Psychology*, 67(1):92.
- Brinton, C. G., Chiang, M., Jain, S., Lam, H., Liu, Z., and Wong, F. M. F. (2014). Learning about social learning in “moocs”: From statistical analysis to generative model. *IEEE transactions on Learning Technologies*, 7(4):346–359.
- Cabrerizo, F. J., Chiclana, F., Al-Hmouz, R., Morfeq, A., Balamash, A. S., and Herrera-Viedma, E. (2015). Fuzzy decision making and consensus: challenges. *Journal of Intelligent & Fuzzy Systems*, 29(3):1109–1118.
- Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7(3):171–176.
- Doise, W., Mugny, G., James, A. S., Emler, N., and Mackie, D. (2013). *The social development of the intellect*, volume 10. Elsevier.
- Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.
- Gomaa, W. H. and Fahmy, A. A. (2013). A survey of text similarity approaches. *International Journal of Computer Applications*, 68(13):13–18.
- Herrera-Viedma, E., Cabrerizo, F. J., Kacprzyk, J., and Pedrycz, W. (2014). A review of soft consensus models in a fuzzy environment. *Information Fusion*, 17:4–13.
- Hiray, S. and Duppada, V. (2017). 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.
- Hirokawa, R. Y. (1982). Consensus group decision-making, quality of decision, and group satisfaction: An attempt to sort “fact” from “fiction”. *Communication Studies*, 33(2):407–415.
- Janis, I. L. and Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. Free press.

- Johnson, D. W. and Johnson, R. T. (1979). Conflict in the classroom: Controversy and learning. *Review of educational research*, 49(1):51–69.
- Kolb, P. (2008). Disco: A multilingual database of distributionally similar words. *Proceedings of KONVENS-2008, Berlin*, 156.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710.
- Lin, D. (1998). Automatic retrieval and clustering of similar words. In *COLING 1998 Volume 2: The 17th International Conference on Computational Linguistics*, volume 2.
- Miller, G. (1998). *WordNet: An electronic lexical database*. MIT press.
- Pérez, I. J., Cabrerizo, F. J., Alonso, S., Dong, Y., Chiclana, F., and Herrera-Viedma, E. (2018). On dynamic consensus processes in group decision making problems. *Information Sciences*, 459:20–35.
- Potash, P. and Rumshisky, A. (2017). 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.
- Qiu, M. and Jiang, J. (2013). 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.
- Rosenthal, S. and McKeown, K. (2015). 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.
- Seerat, B. and Azam, F. (2012). Opinion mining: Issues and challenges (a survey). *International Journal of Computer Applications*, 49(9).
- Tolmie, P., Procter, R., Rouncefield, M., Liakata, M., and Zubiaga, A. (2018). Microblog analysis as a program of work. *ACM Transactions on Social Computing*, 1(1):2.
- Trimbur, J. (1989). Consensus and difference in collaborative learning. *College English*, 51(6):602–616.
- Van Knippenberg, D., De Dreu, C. K., and Homan, A. C. (2004). Work group diversity and group performance: an integrative model and research agenda. *Journal of applied psychology*, 89(6):1008.
- Vinodhini, G. and Chandrasekaran, R. (2012). Sentiment analysis and opinion mining: a survey. *International Journal*, 2(6):282–292.
- Zimmerman, B. J. and Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*. Routledge.
- Zubiaga, A., Kochkina, E., Liakata, M., Procter, R., and Lukasik, M. (2016). 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.