

Análise de Sentimento na Educação: Um Mapeamento Sistemático da Literatura

Míria Bobó, Fernanda Campos, Victor Ströele, Regina Braga, José Maria N. David

Programa de Pós-graduação em Ciência da Computação
Universidade Federal de Juiz de Fora (UFJF)

Rua José Lourenço Kelmer, s/n-Campus Universitário - 36036-900 - Juiz de Fora - MG

miria.b13@gmail.com, fernanda.campos@ufjf.edu.br,
victor.stroele@ice.ufjf.br, {regina.braga, jose.david}@ufjf.edu.br

***Abstract.** Sentiment Analysis (SA) is the computational study of opinions, feelings and emotions expressed in a text. In educational context, it is used as a method to mitigate deficiencies in current educational systems to obtain student feedback or teacher evaluation. This paper aims to analyze the techniques and methods of Sentiment Analysis, in educational context, with the aim of identifying which sentiment have been considered, which approaches have been used to tackle SA challenges and the origin of the texts analyzed. So, a Systematic Mapping Study was conducted according to the guidelines proposed in the Evidence-Based Software Engineering literature.*

***Resumo.** A Análise de Sentimentos (AS) é o estudo computacional de opiniões, sentimentos e emoções expressas em um texto. Na educação, ela é empregada como método de mitigar deficiências dos sistemas educacionais atuais de ensino para a obtenção do “feedback” dos alunos ou avaliação de professores. Este artigo tem como objetivo analisar as técnicas e métodos de Análise de Sentimentos na Educação para identificar quais sentimentos têm sido considerados e a origem dos textos analisados. Para tanto, um Mapeamento Sistemático da Literatura foi conduzido de acordo com técnicas relacionadas a Engenharia de Software Baseada em Evidências.*

1. Introdução

A Mineração de Dados Educacionais (*Educational Data Mining - EDM*) é uma área da ciência cujo objetivo é monitorar o desempenho dos alunos e entender os seus processos de aprendizagem [Rani and Kumar 2017]. Ela aplica métodos computacionais para modelar os perfis dos estudantes, identificando suas necessidades. Porém, para que a acurácia destes sistemas seja satisfatória, é necessário que informações sobre os alunos sejam coletadas e armazenadas [Mac Kim e Calvo 2010]. Essas informações podem ser de diferentes fontes e formatos. Alguns estudos têm mostrado que informações sobre o estado emocional do aluno agregam valor ao perfil do mesmo [Zhao et al. 2016] [Rani and Kumar 2017], outros se relacionam ao seu grau de satisfação [Preuss et al. 2018]. Como resultado, o desempenho acadêmico pode ser afetado de forma direta. O método computacional usado para obter estas informações é a Análise de Sentimentos (AS).

Análise de Sentimentos (*Sentiment Analysis - SA*) é uma área multidisciplinar que pode aplicar técnicas de Processamento de Linguagem Natural (PLN), Linguística

Computacional (LC) ou Análise de texto, com o propósito de identificar o sentimento do autor em relação ao texto escrito. Ela é aplicada em diferentes domínios com diferentes objetivos [Feldman 2013]. Na Educação, ela é empregada com o propósito de proporcionar um ensino personalizado, tentando mitigar aspectos que influenciem negativamente a motivação e o rendimento dos alunos [Mac Kim and Calvo 2010]. Ela pode focar tanto na subjetividade [Pang and Lee 2008] como na polaridade do texto [Ískender and Bati 2015], extraindo e classificando sentimentos de forma automática.

Por ser interdisciplinar, ela abrange soluções que vão de técnicas simples de PLN (ex.: *segmentation*, *stemming*, *tokenization*, *POST-tagging*) a algoritmos mais sofisticados (ex.: *Naive Bayes*, *Support Vector Machine*, *Maximum Entropy*). Essas soluções podem ser agrupadas em três abordagens: (i) Lexical (usa técnicas de PLN e Análise de texto) [Kumar and Jain 2015]; (ii) Aprendizado de Máquina [Kumar and Jain 2015]; e (iii) Híbrida, que é a combinação das duas abordagens acima citadas. O contexto educacional é composto de características que influenciam a escolha da solução de AS a ser aplicada e dos recursos educacionais utilizados. Por esta razão, verifica-se a necessidade de se identificar abordagens e cenários analisados na literatura, e soluções que têm gerado resultados significantes, entre outras questões.

Assim sendo, um Mapeamento Sistemático da Literatura (MSL) é adequado visto que ele tem como propósito prover a visão geral de uma determinada área de pesquisa [Kitchenham et al. 2011]. MSL é um tipo de estudo secundário que se propõe a identificar e classificar conteúdo relacionado a um determinado tópico de pesquisa. Ele visa prover o estado da arte do tópico em questão e ressaltar quais subtópicos necessitam de mais estudos primários [Kitchenham et al. 2011]. É utilizado usado para estruturar a área de pesquisa identificando tanto outras áreas, que podem ser aprofundadas, como lacunas a serem preenchidas. O objetivo deste trabalho é realizar um MSL capaz de fornecer uma visão geral da Análise de Sentimentos aplicada à Educação.

Este trabalho está organizado da seguinte maneira: a seção 2 detalha o protocolo e as questões do mapeamento. Nas seções 3 e 4 são discutidos os resultados em relação a sentimentos e emoções. As considerações finais são apresentadas na seção 5.

2. Mapeamento Sistemático de Literatura

A necessidade de um mapeamento advém da carência de trabalhos que evidenciam o estado da arte de um tópico de pesquisa. Inicialmente, com o objetivo de recuperar estudos secundários sobre Análise de Sentimentos. Como resultado, oito trabalhos foram obtidos, porém não exploraram as peculiaridades da AS que são relevantes no contexto educacional. Como exemplo, cabe citar: as emoções a serem consideradas nos textos escritos por alunos, a modalidade de ensino mais apropriada o uso de abordagens de AS, a classificação de sentimentos que melhor se enquadra no contexto educacional, e a finalidade da aplicação da AS na área de Educação, entre outras.

Como o emprego da AS na Educação é influenciado pelas características do contexto, um MSL é capaz de examinar a extensão desta área de pesquisa, ao agrupar trabalhos em subtópicos que agreguem evidências para responder as questões de pesquisa.

2.1. Planejamento

Durante o planejamento foi definido o protocolo a ser usado no mapeamento sistemático, de modo a mitigar o viés inerente à pesquisa. As questões de pesquisa, estratégia de busca,

fontes de dados, *string* de busca e critérios de seleção foram definidos e avaliados em um processo cíclico, de modo a garantir a reprodutibilidade deste trabalho.

2.1.1. Questões de pesquisa

As questões de pesquisas devem ser construídas com o foco na extração de dados que possam ser agrupados e analisados. No mapeamento sistemático estas questões tendem a ser mais genéricas, de modo a reunir os estudos em subáreas ou subtópicos [Kitchenham et al. 2011]. Usando o método GQM (*Goal/Question/Metric*), proposto por Basili e Rombach (1988), foi definido o escopo do mapeamento:

Analisar as técnicas, métodos e algoritmos de Análise de Sentimentos com o objetivo de identificar quais sentimentos têm sido considerados e a origem dos textos analisados do ponto de vista do estudante em relação aos seus colegas da turma no contexto da Educação.

Com base nesta definição foram elaboradas as seguintes questões de pesquisa: (QM1) *Como as publicações sobre Análise de Sentimentos na Educação estão distribuídas ao longo do tempo?* (QM2) *Quais são os autores mais ativos na área?* (QM3) *Quais meios/veículos de publicação se interessam pelo assunto?* (QM4) *Qual é o propósito de se empregar a Análise de Sentimentos no contexto educacional?* (QM5) *Em quais modalidades de ensino a Análise de Sentimentos tem sido aplicada?* (QM6) *Qual classificação de sentimentos é mais usada?* (QM7) *Quais tipos de emoções têm sido considerados nos estudos?* (QM8) *Quais fontes de dados têm sido usadas para a AS no contexto educacional?* (QM9) *Qual abordagem tem sido mais empregada no contexto educacional?*

2.1.2. Definição da *string* de busca

Os termos usados na *string* de busca, e seus sinônimos, foram derivados do método PICOC, proposto por Petticrew e Roberts (2008). Numa primeira fase, foram considerados sinônimos de *Sentiment Analysis* os termos: *Opinion Mining*, *Sentiment Mining*, *Emotion Detection*, *Web Opinion*, *Document Classification*, *Polarity Classification*, *Text Clustering*, *Content Mining*. Contudo, constatou-se que estes termos são sinônimos, ou subáreas, da Mineração de Opinião (*Opinion Mining*), e que os artigos retornados não abordavam a Análise de Sentimentos no contexto educacional. Por isso, optou-se pela utilização do termo *Emotion Analysis* que aparecia frequentemente nos artigos de interesse. Abaixo é apresentada a versão final da *string* de busca usada neste mapeamento:

(("Sentiment Analysis" OR "Emotion Analysis") AND (education OR teaching OR "Online Learning" OR "Virtual Learning Environment" OR "Distance Learning" OR "Distance Education")) AND (techniques OR methods OR algorithm OR architecture))

2.1.3. Avaliação do protocolo

As avaliações do protocolo foram feitas através de artigos de controle [Altrabsheh et al. 2014] verificando se eles eram retornados nas buscas efetuadas com a *string* modificada. Especialistas na área de tecnologias educacionais participaram também da revisão dos termos escolhidos para compor a *string*. Além disso, foram feitas reuniões periódicas com revisores, que trabalham com aplicações computacionais no contexto educacional, de forma a discutir soluções e definir as estratégias a serem adotadas.

2.1.4. Critérios de Seleção

Foram definidos alguns critérios de seleção visando analisar a relevância dos artigos recuperados das bases de pesquisa. Os critérios de inclusão (CI) e de exclusão (CE) são apresentados a seguir:

(CI1) Trabalhos que apresentam a Análise de Sentimentos aplicada no contexto educacional; (CI2) Trabalhos escritos em Inglês; (CI3) Trabalhos que apresentam algoritmos ou arquiteturas ou métodos ou técnicas de Análise de Sentimentos aplicados à Educação; (CI4) Trabalhos publicados em Conferências ou Revistas ou Relatórios Técnicos que foram revisados por pares.

(CE1) Trabalhos que não apresentam a Análise de Sentimentos aplicada no contexto educacional; (CE2) Trabalhos não escritos em Inglês; (CE3) Trabalhos que não apresentam algoritmos ou arquiteturas, métodos ou técnicas de Análise de Sentimentos aplicados à Educação; (CE4) Trabalhos publicados em Conferências ou Revistas ou Relatórios Técnicos que não foram revisados por pares; (CE5) Trabalhos que tenham versões mais recentes; (CE6) Trabalhos que o acesso ao texto completo é negado ou pago; (CE7) Trabalhos duplicados.

2.2. Condução

Foram feitas buscas em 6 bibliotecas digitais usando a string de busca, e considerando apenas os campos: título do artigo, resumo e palavras-chave. Como retorno, foram obtidos 345 artigos, dos quais 47 eram provenientes da ACM Digital Library, 96 da Compendex, 74 da IEEE Xplore, 3 da Science Direct, 89 da Scopus e 36 da ISI Web of Knowledge.

O conjunto de artigos foi submetido a uma inspeção para remoção de duplicatas, e um novo conjunto foi gerado com 221 artigos. Cada artigo foi avaliado pelos critérios de seleção aplicados apenas ao título, resumo e palavras-chave, e se obteve 80 artigos. Foi realizada uma segunda avaliação pelos critérios de seleção, agora aplicados ao texto completo. Como resultado, um conjunto de 70 artigos foi obtido. Adicionou-se a esse conjunto 6 artigos provenientes do *snowballing*. A Figura 1 ilustra o processo de aquisição dos artigos.

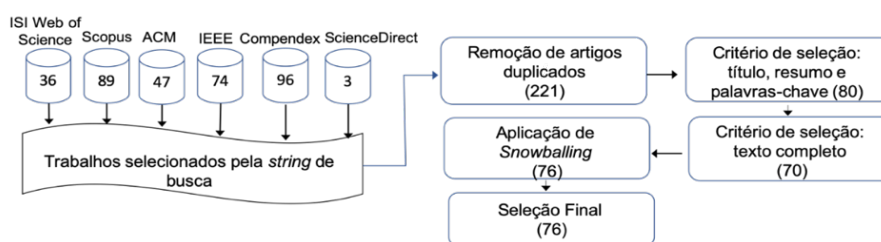


Figura 1. Fases da seleção dos artigos do mapeamento

2.3. Resultados

Os artigos¹ obtidos na seleção final foram lidos buscando responder as questões de pesquisa. A seguir, cada questão é discutida.

QM1: Como as publicações sobre Análise de Sentimentos na Educação estão distribuídas ao longo do tempo?

A primeira publicação aparece no ano de 2007, e o assunto continuou sendo tema de interesse de pesquisadores. O maior número de publicações ocorreu em 2016 (21

¹ <http://www.ufjf.br/nenc/projetos/analise-de-sentimento-na-educacao/mapeamento-sistematico/>

publicações). Também foi possível observar uma tendência de crescimento do interesse dos pesquisadores na área de Análise de Sentimentos.

QM2: Quais são os autores mais ativos na área?

Considerando apenas os 76 artigos usados no mapeamento, os autores que mais publicaram sobre o assunto fazem parte de 3 grupos de pesquisa. O primeiro grupo é formado pelas seguintes pesquisadoras: Nabeela Altrabsheh, Mihaela Cocea e Sanaz Fallahkhair que publicaram 3 artigos em conjunto, sendo que Altrabsheh e Cocea escreveram outro artigo com Mohamed Medhat Gaber, totalizando 5 trabalhos sobre o assunto. O segundo é composto por: Taoufiq Zarra, Raddouane Chiheb, Rdouan Faizi, Abdellatif El Afia que publicaram 3 artigos, e o terceiro por: Michalis Feidakis, Thanasis Daradoumis e Santi Caballe que escreveram 2 artigos em conjunto.

QM3: Quais meios/veículos de publicação se interessam pelo assunto?

Na computação, os veículos de publicação que mais se destacaram, e a quantidade de artigos, foram: (i) *Advances in Intelligent Systems and Computing* (3); (ii) *BRAIN-Broad Research in Artificial Intelligent and Neuroscience* (3); (iii) *CEUR-Workshop Proceedings* (3); (iv) *SIEE-International Symposium on Computer in Education* (3); (v) *Computers in Human Behavior* (2); (vi) *IISA-Conference on Information, Intelligence, Systems and Applications* (2); e (vii) *Journal of Theoretical and Applied Information Technology* (2).

QM4: Qual é o propósito de se empregar a Análise de Sentimentos no contexto educacional?

Foram vários os objetivos encontrados nos artigos, mas como a proposta do mapeamento é agrupar trabalhos em subtópicos, definiu-se 4 objetivos que pareciam se enquadrar melhor nos trabalhos: (i) ensino-aprendizagem; (ii) avaliações; (iii) funções do AVA; e (iv) sistemas de recomendação.

Em relação ao ensino-aprendizagem, aproximadamente 42% dos trabalhos se propõem a aplicar a AS para identificar as dificuldades enfrentadas pelos alunos, e assim recorrer a medidas que melhorem o seu processo de ensino e aprendizagem. Estes trabalhos empregam métodos de obtenção de *feedback*, em tempo real ou não, dos estudantes com o objetivo de proporcionar uma melhor experiência de aprendizagem.

Outro objetivo comum entre os artigos (40%) é o emprego da AS nas avaliações de curso, docente ou instituição de ensino. A maior parte destes trabalhos usam formulários com questões que permitem aos alunos discorrerem sobre a entidade a ser avaliada, e a AS aplicada nestes textos.

Foram também encontrados trabalhos que aplicam a AS para validar melhorias ou funcionalidades novas dos sistemas *e-learning* ou AVAs (Ambientes Virtuais de Aprendizagem). Eles representam 13.8%, e utilizam a AS para obter a opinião dos alunos em relação às novas funcionalidades.

Alguns artigos (4.6%) também usaram a AS em sistemas de recomendação de conteúdos educacionais. Nesses trabalhos, os pesquisadores aplicam a AS tanto para identificar o melhor recurso a ser recomendado como para avaliar a recomendação.

QM5: Em quais modalidades de ensino a Análise de Sentimentos tem sido aplicada?

Nem todos os artigos deixavam clara a modalidade usada como cenário de aplicação da AS e, por isso, consideramos que estes trabalhos não respondiam à questão QM5. Porém, existem alguns trabalhos em que a solução não depende da modalidade, como é o caso dos trabalhos que empregam a AS em avaliações e, nestes casos específicos, foram consideradas as duas modalidades.

Foi observado que o ensino a distância tem sido o cenário mais utilizado pelos pesquisadores. Isso acontece porque esta modalidade de ensino possui um leque de demandas em que a AS pode ser usada para a averiguação de causas ou reconhecimento de soluções, como por exemplo: identificação de unidades de ensino que os alunos têm mais dificuldades, obtenção do *feedback* dos estudantes, redução de abandono do curso por parte dos alunos, avaliação do curso ou da plataforma de ensino, entre outros.

QM6: Qual a classificação de sentimentos é mais usada?

Nesta questão esperava-se identificar apenas duas escalas de sentimento (positivo, negativo e neutro; ou positivo e negativo) porém foi verificado que, apesar de serem as mais usadas, existem trabalhos que utilizam outras escalas. As classes usadas nos artigos são: 2 Sentimentos (67,9%): positivo e negativo; 3 Sentimentos (23,2%): positivo, neutro e negativo; 4 Sentimentos (3,6%): muito positivo, positivo, negativo e muito negativo; 5 Sentimentos (3,6%): muito positivo, positivo, neutro, negativo e muito negativo; 6 Sentimentos (1,8%): brincadeira, positivo, pensativo, neutro, negativo, irritado.

QM7: Quais os tipos de emoções têm sido considerados nos estudos?

A falta de estudos que identificam as emoções relevantes no contexto educacional é apontada por alguns trabalhos como uma deficiência da área, e não houve consenso na lista de emoções usadas pelos pesquisadores. Apenas 11.8% dos artigos mencionam usar emoções nas suas abordagens de AS. As emoções mais comuns são *anger, sadness, fear, joy, confusion, boredom e disgust*, que foram citadas em pelo menos 4 artigos. Verifica-se uma predominância de emoções consideradas negativas. Isso pode ser um indicativo da preocupação dos autores em detectar e sanar problemas ligados à educação.

QM8: Quais fontes de dados têm sido usadas para a AS no contexto educacional?

A grande maioria dos trabalhos usou rede social ou fórum do AVA como fonte de dados para as suas aplicações. Este fato corrobora a ideia de que a AS na educação, é mais aplicada ao contexto de cursos a distância. Alguns trabalhos faziam o uso de redes sociais e fóruns de AVAs para aumentar a acurácia de suas abordagens. Os formulários/questionários foram muito utilizados em trabalhos de avaliação de curso, docente ou instituição de ensino.

QM9: Qual abordagem tem sido mais empregada no contexto educacional?

O mapeamento considera 4 tipos de abordagem de Análise de Sentimento: Lexical, Aprendizado de Máquina, Híbrida e Semântica. Na abordagem “Híbrida” foram considerados todos os trabalhos que usavam mais de uma solução das abordagens Lexical e Aprendizado de Máquina (ex.: uma técnica de Processamento de Linguagem Natural e dois algoritmos de aprendizado de máquina e vice-versa).

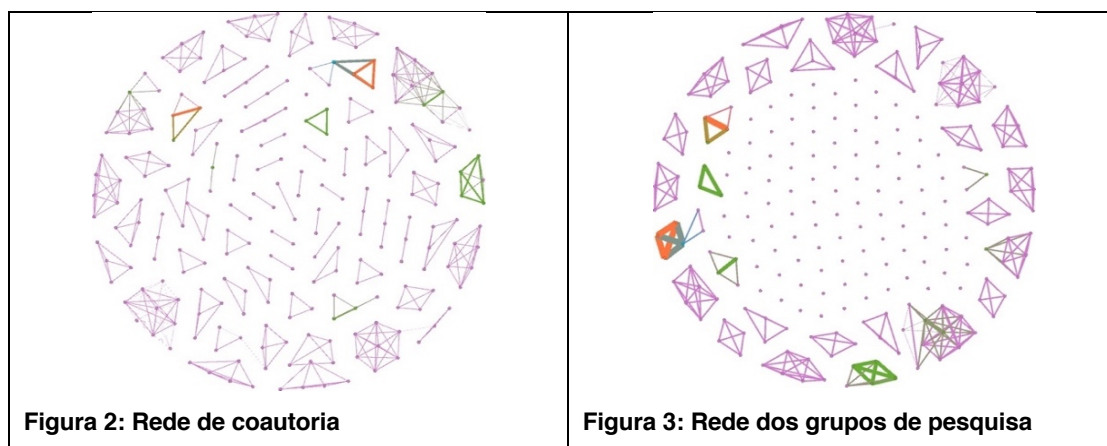
A maior parte dos trabalhos analisados usa a abordagem Lexical (40,7%) como solução. Isso acontece pelo fato de essas abordagens se mostrarem melhores para

problemas de domínio específico. Os dicionários utilizados possuem palavras do contexto a ser analisado, proporcionando assim uma melhor acurácia, em relação as outras abordagens. Outras abordagens foram observadas, tais como: Aprendizado de Máquina (28,8%), Híbrida (28,8%) e; Semântica (1,7%).

3. Análise de coautoria

Para realizar as referidas análises, foi construída uma rede na qual os nós representam os autores e as arestas representam as publicações em comum com eles, ou seja, só existe aresta se dois autores são coautores em, pelo menos, um artigo. Também foi possível representar, graficamente, a relevância do autor para o mapeamento. O número de artigos escritos pelo autor é representado pela cor do nó que se estende até as suas arestas (rosa = 1, verde = 2, laranja = 3 e azul = 4 artigos). A espessura das arestas indica a quantidade de trabalhos entre os autores, arestas mais grossas indicam um maior número de trabalhos em conjunto.

Com a sumarização dos trabalhos foi possível criar uma rede de coautoria dos artigos deste mapeamento. O uso da rede para expressar esta informação permite evidenciar algumas características da área em questão, analisando apenas a sua estrutura. A criação desta rede de coautores tem por objetivo identificar: (i) se a rede de coautoria possui poucas, porém grandes, componentes conexas (rede dividida em poucos subgrupos com muitos nós conectados); (ii) se existe um caminho, e se este caminho é relativamente curto entre os grupos coesos (grupos com mais de uma publicação e com os autores que mais publicam); e (iii) quais nós seriam as pontes entre os grupos coesos. A Figura 2 mostra a rede gerada usando o algoritmo de *Fruchterman Reingold*².



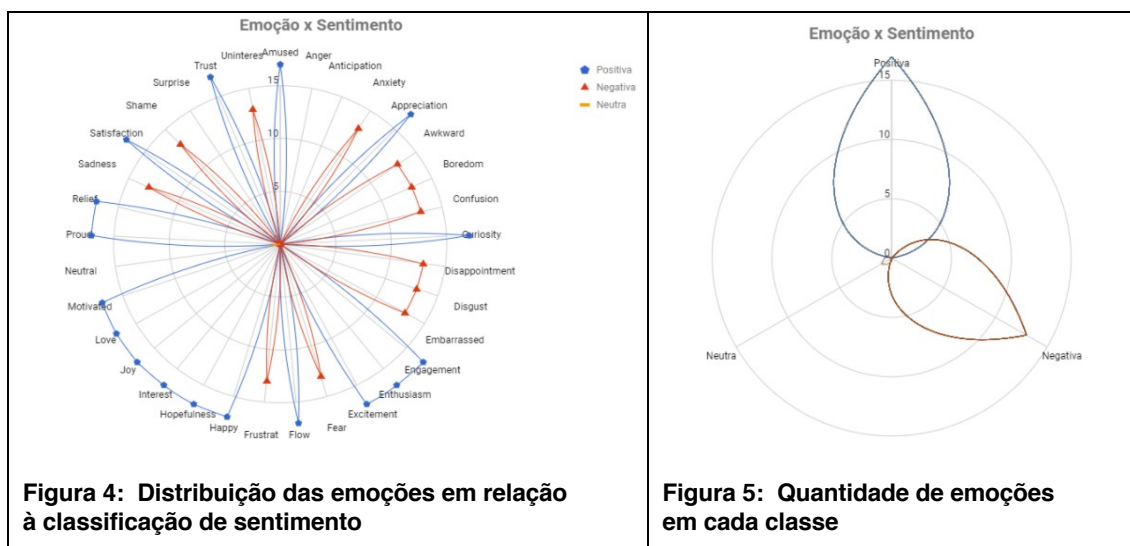
Como pode ser observado pela Figura 2, a rede possui 64 pequenas componentes conexas, que significa que a grande maioria dos autores faz parte de pequenos grupos isolados. Provavelmente, isto se deve ao fato de estes autores não serem pesquisadores do tema propriamente dito, boa parte deles pesquisam sobre AS, e usaram a Educação como domínio de aplicação. O número de componentes da rede pode suscitar a crença de que, por ser um subtópico de pesquisa relativamente recente, a área AS na educação ainda precisa ser explorada, já que temas mais explorados tendem a ter muitos grupos de pesquisa fortes que se conectam formando grandes componentes conexas.

² <https://github.com/gephi/gephi/wiki/Fruchterman-Reingold>

Para facilitar a visualização de propriedades desta rede foi feita a remoção de arestas dos grupos com menor contribuição na área (grupos que publicaram apenas uma vez, que possuem até 3 autores e que não estejam ligados a outros grupos). A Figura 3 apresenta a nova rede gerada. Por ser desconexa, a rede não possui caminho entre grupos coesos e, conseqüentemente, não é possível identificar os nós que seriam pontes entre eles. Porém, uma característica visível destes grupos é o número relativamente pequeno de autores participantes em relação aos tamanhos dos grupos encontrados nesta rede.

4. Análise do conjunto de emoções

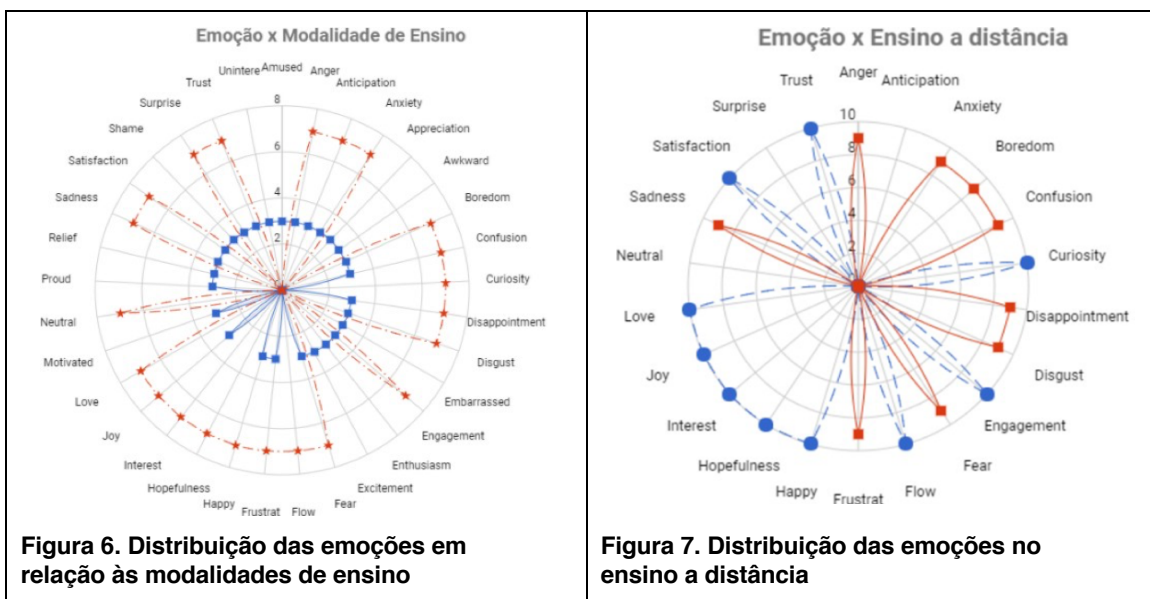
Nove trabalhos citaram o uso de um conjunto de emoções nas suas abordagens de AS. Isso sinaliza que estudos mais aprofundados em relação ao conjunto de emoções devem ser considerados no contexto educacional. Considerações sobre a distribuição deste conjunto em relação a algumas questões de pesquisa são apresentadas a seguir.



Foram identificadas 33 emoções (*Amused, Anger, Anticipation, Anxiety, Appreciation, Awkward, Boredom, Confusion, Curiosity, Disappointment, Disgust, Embarrassed, Engagement, Enthusiasm, Excitement, Fear, Flow, Frustration, Happy, Hopefulness, Interest, Joy, Love, Motivated, Neutral, Proud, Relief, Sadness, Satisfaction, Shame, Surprise, Trust, Uninterested*) porém, os autores não explicitam a relação delas com as classificações de sentimento. A Figura 4 mostra o gráfico radar, para mapear essas emoções em sentimentos positivos, negativos e neutros. Apenas uma emoção neutra foi mapeada como sendo da classe neutra e duas emoções (*Antecipation e Surprise*) ficaram sem identificação de classe, já que ambas dependem fortemente do conteúdo do texto. A escala representa a quantidade de emoções que cada classe de sentimento possui.

A Figura 5 apresenta a quantidade de emoções por classe de sentimento. A diferença entre o conjunto de emoções positivas (15 emoções) e o de emoções negativas (13 emoções) é pequena. Isso se deve, muitas vezes, à limitação inerente ao corpus lexical usado nas abordagens, nas quais é bastante comum encontrar um conjunto maior de palavras que representem emoções ou sentimentos positivos, do que negativos ou neutros. Contudo, as emoções negativas são as mais citadas pelos autores, numa tentativa de identificar dificuldades que os alunos enfrentam no processo de ensino e aprendizagem.

Quanto à modalidade de ensino usada para a extração das emoções, constata-se, pela Figura 6, que: i) os cursos a distância são os que mais se preocupam em identificar as emoções dos estudantes, e ii) o conjunto de emoções positivas continua sendo relativamente maior que o conjunto negativo (Figura 7) porém, as emoções mais citadas, nesta modalidade, também continuam sendo as negativas. A escala na Figura 6 representa a quantidade de emoções em cada modalidade de ensino, e na Figura 7 a quantidade de emoções positivas e negativas.



Nas fontes de dados usadas para a extração das emoções, é possível observar uma relação ou padrão: fontes menos usadas possuem um conjunto menor de emoções (p. ex. diário) e fontes mais usadas, um conjunto maior de emoções (p. ex.: redes sociais).

5. Considerações Finais

Neste trabalho foi apresentado um mapeamento sistemático sobre a Análise de Sentimentos no contexto educacional. Buscou-se analisar as técnicas, métodos, arquiteturas e algoritmos de AS com o objetivo de identificar quais sentimentos têm sido considerados e a origem dos textos analisados do ponto de vista do estudante.

Como contribuição, este mapeamento oferece oportunidades de pesquisas através da utilização das técnicas de AS para a Educação. Foi possível constatar que o assunto é do interesse da comunidade acadêmica, apresentado uma curva de crescimento. A educação a distância foi identificada como sendo a modalidade mais comum da aplicação dessa técnica. Não foi verificado um consenso na lista de emoções a serem consideradas no contexto educacional, o que foi apontado por alguns trabalhos como um tópico que precisa de estudos mais aprofundados. Mas, foi observado que grande parte dos pesquisadores acreditam que redes sociais, fóruns de AVAs ou questionários são fontes ricas de dados para a detecção de sentimento, independente da modalidade de ensino.

A abordagem de solução mais usada na literatura, no contexto educacional, tem sido a Híbrida, na qual técnicas de PLN são usadas para pré-processamento dos dados, e algoritmos de aprendizado de máquina são usados para a detecção de sentimentos. Alguns trabalhos afirmam que essa abordagem garante melhor acurácia nos resultados dos algoritmos de aprendizado de máquina. Sugerem a realização experimentos adicionais

usando mais técnicas da abordagem lexical ou avaliação de novas combinações das abordagens lexical e aprendizado de máquina.

Como trabalhos futuros cabe investigar as técnicas usadas na abordagem híbrida, explorar o conjunto de emoções relevantes no contexto educacional e suas respectivas classificações (*joy* = positiva, *indifference* = neutra, *anxiety* = negativa), e associá-las aos diferentes tipos de atividades, e recomendações, entre outros recursos educacionais.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

6. Referências

- Altrabsheh, N., Cocea, M., Fallahkhalil, S. (2014) “Sentiment analysis: towards a tool for analysing real-time students feedback”, In: Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on. IEEE, p. 419-423.
- Basili, V.R., Rombach, H.D. (1988) “The TAME project: Towards improvement-oriented software environments”, IEEE Trans. on Software Engineering, 14(6), p. 758-773.
- Feldman, R. (2013) “Techniques and applications for sentiment analysis”. Communications of the ACM, v. 56, n. 4, p. 82-89.
- İskender, E., Bati, G. B. (2015) “Comparing Turkish Universities Entrepreneurship and Innovativeness Index's Rankings with Sentiment Analysis Results on Social Media”. Procedia-Social and Behavioral Sciences, v. 195, p. 1543-1552.
- Kitchenham, B. A., Budgen, D., Brereton, O. P. (2011) “Using mapping studies as the basis for further research – a participant-observer case study”, Information and Software Technology, v. 53, n. 6, p. 638-651.
- Kumar, A., Jain, R. (2015) “Sentiment analysis and feedback evaluation”. 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE), p. 433-436.
- Mac Kim, S., Calvo, R. A. (2010) “Sentiment analysis in student experiences of learning”. In: Educational Data Mining (EDM), p. 111-120.
- Pang, B., Lee, L. (2008) “Opinion mining and sentiment analysis”, Foundations and Trends in Information Retrieval, v. 2, n. 1-2, p. 1-135.
- Petticrew, M., Roberts, H. (2008) “Systematic reviews in the social sciences: A practical guide”. John Wiley & Sons, 2008.
- Preuss, L., dos Santos, J., Guedes, G. P. (2018) “VITA - Estimando a satisfação de estudantes por meio da Análise de Sentimentos”, Anais do XXIX Simpósio Brasileiro de Informática na Educação (SBIE 2018), Fortaleza, CE, p. 1143-1152.
- Rani, S., Kumar, P. (2017) “A Sentiment Analysis System to Improve Teaching and Learning. Computer”, v. 50, n. 5, p. 36-43.
- Zhao, H., Ji, X., Zeng, Q., Jiang, S. (2016) “A teaching evaluation method based on sentiment classification”. International Journal of Computing Science and Mathematics, 7(1), p. 54-62.