

Análise de Relevância Temática de Postagens em Fóruns de Discussão em Relação ao uso de Vídeos como Recurso Didático

**Crystiano J. R. Machado¹, Alexandre M. A. Maciel¹, Rodrigo L. Rodrigues²,
Ronaldo P. de Menezes³**

¹Universidade de Pernambuco (UPE) – Recife/PE 50.720-001, Brasil

²Universidade Federal Rural de Pernambuco (UFRPE) – Recife/PE 52.171-900, Brasil

³University of Exeter (UE) – Exeter EX4 4QD, Reino Unido

{cjrm, amam}@ecomp.poli.br, rodrigo.linsrodrigues@ufrpe.br,
r.menezes@exeter.ac.uk

Abstract. *This work proposes the modeling and development of a technological solution that allows the thematic relevance analysis of the postings made by students in discussion forums, in relation to videos indicated by teachers as didactic resource to promote the deepening on the topic of discussion in the forum. For this, it was used the Text Mining for the processing of the texts and extraction of characteristics. The results show that the appropriate application of these techniques can generate potentially useful indicators by teachers in improving pedagogical practice, in order to identify students who may be learning difficulties.*

Resumo. *Este trabalho propõe a modelagem e o desenvolvimento de uma solução tecnológica que possibilite a análise de relevância temática das postagens feitas por estudantes em fóruns de discussão, em relação a vídeos indicados por professores como recurso didático para promover o aprofundamento sobre o tema de discussão no fórum. Para isto, utilizou-se a Mineração Textual para o processamento dos textos e extração de características. Os resultados mostram que a aplicação apropriada destas técnicas podem gerar indicadores potencialmente úteis por professores na melhoria da prática pedagógica, no sentido de identificar estudantes que, possivelmente, possam estar com dificuldades de aprendizado.*

1. Introdução

Os Ambientes Virtuais de Aprendizagem (AVA) oferecem diversos recursos e ferramentas para serem utilizados como instrumento de interação, sendo o fórum de discussão um dos mais populares [Soares et al., 2016]. Estes fóruns representam um ambiente de interação onde, a partir da inserção de um tópico de discussão, os participantes podem interagir com respostas/comentários acerca do tema proposto ou interagir em postagens já inseridas por outro participantes. Diante destas possibilidades, os fóruns de discussão têm sido amplamente adotados pela EAD, estando presente em mais de 72% dos cursos à distância [ABED, 2016].

Atualmente existem diferentes tecnologias suportadas pelos AVA e que são disponibilizadas nos fóruns de discussão como recurso didático. Além do uso tradicional de conteúdos textuais, destaca-se o uso de vídeos como estratégia de ensino na EAD. Neste contexto, o uso de vídeos como instrumento pedagógico é importante no sentido de respeitar as ideias de múltiplos estilos de aprendizagem e de inteligências [Oliveira, 2013]. Através dos vídeos que são inseridos pelo professor no fórum de discussão, a participação dos alunos se dá através de postagens contendo a sua compreensão acerca do conteúdo apresentado.

Desta forma, avaliar o envolvimento dos alunos nos fóruns é uma atividade importante para garantir a efetiva construção do conhecimento. Através da análise de postagens dos discentes, o professor pode diagnosticar possíveis dificuldades de aprendizado e obter informações relevantes sobre os estudantes. Por outro lado, devido ao grande volume de textos gerados diariamente nos fóruns, a análise exclusivamente manual dos conteúdos postados pelos alunos passa a tornar-se inviável [Azevedo, Behar, & Reategui, 2011].

Diante deste contexto, o objetivo deste trabalho é propor a modelagem e o desenvolvimento de uma solução tecnológica que possibilite a análise de relevância temática das postagens feitas por estudantes em fóruns de discussão, em relação a vídeos indicados por professores para promover o aprofundamento sobre o tema de discussão no fórum. Para isto, este trabalho está organizado da seguinte forma: na seção 2 apresenta-se a fundamentação teórica; na seção 3 é demonstrada arquitetura proposta; na seção 4 é feita a caracterização do experimento realizado, resultados e discussões; por fim, na seção 5, são feitas as considerações finais.

2. Fundamentação Teórica

2.1. O Vídeo Como Recurso Didático em EAD

Os vídeos são recursos dinâmicos que, de maneira lúdica, conseguem sensibilizar e manter os alunos motivados durante o processo de ensino-aprendizagem [Dallacosta, 2004]. O uso deste tipo de mídia como estratégia educacional tem sido beneficiado pelo desenvolvimento de repositórios online, disponíveis, em boa parte, de forma gratuita. A diversidade destes vídeos online, combinada com as inúmeras ferramentas presentes nos AVA, acabam por trazer novas oportunidades para integrar o conteúdo multimídia com a EAD [Mattar, 2009].

Uma mudança qualitativa no processo de ensino-aprendizagem acontece quando o professor consegue integrar, sob o aspecto da inovação, diversas tecnologias [Moran, 2006]. Neste contexto, a utilização de vídeos desempenha um papel pedagógico fundamental. No entanto, pelo fato da interpretação do conteúdo através de vídeos ocorrer de forma mais subjetiva se comparado a linguagem puramente textual, as estratégias do professor para avaliar o aprendizado dos alunos em ambientes de EAD devem ser pensadas considerando as particularidades deste novo recurso.

Uma das formas de verificar se o aluno entendeu um conteúdo repassado por meio de vídeo é a qualidade da sua discussão a respeito do mesmo. Geralmente, professores disponibilizam vídeos reflexivos nos fóruns de discussão para gerar um diálogo entre os estudantes sobre o seu conteúdo. É por meio da qualidade de cada uma das postagens que

o professor passa a ter um elemento perceptivo, que faz com que ele visualize o quanto a turma conseguiu absorver do conteúdo.

Neste contexto, o uso de ferramentas tecnológicas que busquem a extração automática de informações a partir de vídeos educacionais pode apresentar uma visão abrangente do conteúdo absorvido pelos estudantes. Por outro lado, embora a quantidade de pesquisas tenha aumentado nos últimos anos através do desenvolvimento de ambientes inteligentes de aprendizado e das ferramentas de *analytics* sobre aprendizado baseado em vídeos [Giannakos, 2013], avaliar de forma eficiente a aprendizagem dos estudantes continua a ser uma tarefa desafiadora. De acordo com Giannakos et al. (2013), um dos maiores desafios nestes ambientes é coletar as experiências de aprendizado em vídeos e contribuir para o desenvolvimento de vídeos educacionais.

Partindo desta problemática, propomos a modelagem e o desenvolvimento de um recurso que quantifique de forma objetiva o aprendizado dos alunos, levando em consideração a relação entre o conteúdo abordado nos vídeos e as contribuições dos estudantes nos fóruns. Para isto, foi utilizada a Mineração Textual.

2.2. Abordagem de Mineração Textual em AVA

Inspirada na mineração de dados, a Mineração Textual (MT) realiza o processo de extrair padrões não-triviais a partir de textos, com o intuito de extrair informações úteis de dados não estruturados ou semiestruturados [Tan, 1999]. Esse processo, no entanto, pode ser realizado por meio de diversas técnicas, das quais podemos destacar a representação dos textos como redes complexas, que é a estratégia de modelagem utilizada neste trabalho.

Nesta abordagem, a rede obtida por meio da mineração textual apresenta uma estrutura formada por nós, que simbolizam as palavras do textos, e arestas, que representam a ligação entre as palavras. O objetivo da modelagem textual como rede complexa é codificar as relações entre os conceitos do texto, que é obtida através de características e propriedades analisadas como medidas estruturais da rede. Estas medidas são frequentemente usadas para capturar informações sobre a importância relativa dos nós e arestas.

Dentre as métricas potencialmente úteis na extração de características em redes formadas a partir de conteúdos textuais, podemos destacar o *betweenness* [Freeman, 1977]. O *betweenness* é uma medida de centralidade que considera que um vértice é importante se este fizer parte de um grande número de caminhos mínimos. Altos valores de *betweenness* indicam que os nós têm alta influência na rede, pelo simples fato de que as informações que circulam na rede tendem a passar por esses nós [Medeiros, 2015]. No contexto das redes que representam conteúdos textuais, as palavras cujos nós da rede possuem altos valores de *betweenness* são aquelas que possuem alta frequência e também aquelas que conectam conceitos relacionados, podendo indicar aspectos sobre o contexto do conteúdo escrito.

Atualmente, diversas pesquisas têm sido desenvolvidas com base na MT e em técnicas de representação textual através de redes complexas em aplicações que estão voltadas ao contexto dos AVA. Dentre estas pesquisas, podemos destacar o trabalho realizado por Lee e Segev (2012), onde foi aplicada a MT através das redes complexas para propor um mecanismo de geração automática de mapas conceituais em e-learning a partir de documentos escritos. Também no contexto educacional, Azevedo (2011)

utilizou a mineração textual através de grafos para desenvolver um recurso para análise qualitativa das contribuições registradas pelos alunos nos fóruns de discussão.

3. Arquitetura Proposta

Esta seção apresenta a arquitetura do Motor de Processamento Textual proposto. Para esta pesquisa, foi utilizado um banco de dados com registros reais de um AVA. Os textos extraídos dos fóruns do AVA foram escritos por alunos e professores, em Português brasileiro. A Figura 1 apresenta a arquitetura geral do motor de processamento, que foi desenvolvido em linguagem Python.

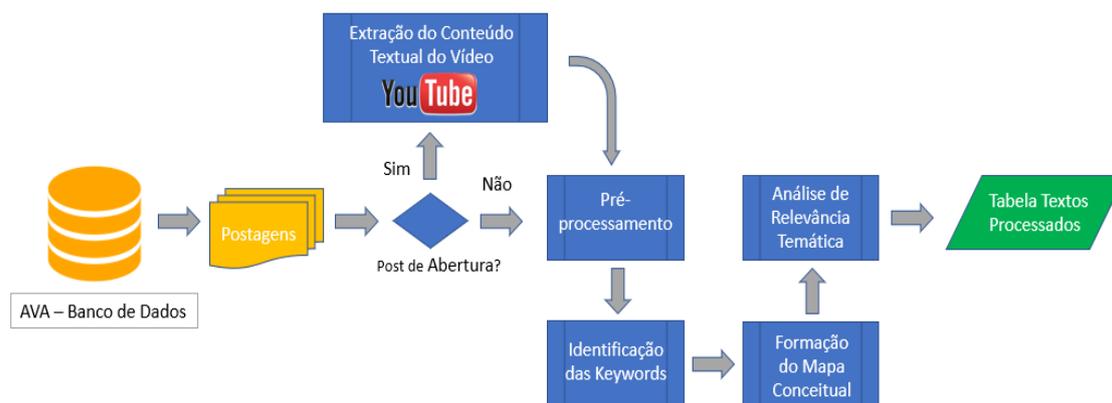


Figura 1. Arquitetura geral do Motor de Processamento Textual.

O motor de processamento é inicializado com a identificação do fórum a ser analisado. Para cada postagem obtida do banco de dados através de consultas SQL, os seguintes módulos são executados sequencialmente: Extração do conteúdo textual do vídeo (para o post de abertura do fórum); Pré-processamento; Identificação das *Keywords*; Formação do Mapa Conceitual; Análise de Relevância Temática. Esses módulos realizam tarefas específicas no processo de extração de conhecimento e são descritos em detalhes a seguir. No final da execução, é obtida uma tabela contendo os dados de todas as postagens processadas.

3.1. Extração do Conteúdo Textual do Vídeo

A execução deste módulo tem por objetivo identificar os *links* dos vídeos, inseridos pelo professor momento de abertura do fórum. Uma vez que os endereços Web dos vídeos foram extraídos do post de problematização, nossa abordagem para a extração do conteúdo textual dos vídeos utilizou o YoutubeDL [Gonzalez, 2018], que é uma biblioteca de vídeo baseada em linha de comando. Este recurso está disponível em linguagem Python e atualmente suporta centenas de sites de vídeo e áudio online.

Através desta biblioteca é possível obter um conjunto de informações associadas ao vídeo. Para a proposta deste trabalho, a informação alvo a ser extraída do vídeo é a legenda, que é um recurso de transcrição do vídeo. Existem, no entanto, duas categorias de legenda: manual e automática. A legenda manual é gerada a partir da ação humana de incluir no vídeo o texto a cada instante de tempo, enquanto que a automática é gerada a partir do processamento computacional do sinal de áudio obtido do vídeo.

O módulo de extração do conteúdo textual busca, inicialmente, extrair a legenda manual do vídeo, uma vez que são menos suscetíveis a erros de tradução, se comparado

às legendas automáticas. Caso a legenda manual não tenha sido inserida no vídeo, utiliza-se, então, a legenda automática. Na etapa seguinte, é feita a transformação do texto para a codificação UTF-8. Após isto, é feita a remoção das *tags* html existentes no texto e, por fim, é feita a padronização dos espaçamentos.

3.2. Pré-processamento

O módulo de pré-processamento é responsável por manipular o conteúdo de uma postagem que foi inserida no fórum. Para isto, recebe como entrada os dados extraídos do banco de dados.

Inicialmente, é feita a remoção dos *links* para páginas Web existentes no texto. Em seguida, o conteúdo da postagem passa por um processo de Lematização que é responsável por transformar as palavras flexionadas em suas respectivas formas canônicas, o que torna possível representar cada palavra do texto em um formato único, independente das variações de gênero e número [Antiqueira, 2007]. Nesta pesquisa, a Lematização foi realizada através do CoGrOO API, que é um verificador gramatical de código aberto, aplicado a língua portuguesa brasileira [Silva, 2013].

Após a Lematização, é feita a separação das sentenças do texto. Na etapa seguinte, as palavras existentes em cada sentença são etiquetadas morfossintaticamente através do CoGrOO API. Por fim, é feita a remoção das *stopwords*, levando-se em consideração a classe morfossintática atribuída a cada palavra. Na abordagem adotada neste trabalho, consideramos que apenas as palavras classificadas como verbo, substantivo, adjetivo ou advérbio devem permanecer no texto. Desta forma, todas as demais palavras são descartadas.

3.3. Identificação das *Keywords*

O módulo responsável por executar a extração de *keywords* recebe como entrada o texto pré-processado e tem como objetivo identificar os principais termos do contexto da argumentação. Inicialmente é feita a indexação das palavras no texto. Esta etapa é caracterizada por identificar unicamente cada palavra no texto, como forma de gerar uma lista em que cada posição corresponde a uma palavra diferente. Em seguida, a matriz de adjacência é gerada. Esta matriz armazena o valor das relações de adjacência para cada par de palavras vizinhas dentro de cada sentença do texto.

Uma vez que a matriz de adjacência tenha sido criada, a rede do texto pode ser construída. Nessa rede, os nós representam as palavras, enquanto as arestas entre os nós indicam que há uma relação de adjacência entre os pares de palavras. Após a formação da rede, o valor do *betweenness* é calculado para cada nó da rede. Essa métrica pressupõe que um vértice é importante se for acessado por um grande número de caminhos mínimos. Com base no contexto do conteúdo escrito no post, através desta medida é possível identificar as palavras com maior importância na estrutura textual.

3.4. Formação do Mapa Conceitual

O módulo de formação do mapa conceitual recebe como parâmetros as sentenças etiquetadas, que representam o texto pré-processado, e a lista contendo as *keywords*. A execução deste módulo busca modelar a rede complexa que representa o mapa de conceitos em relação ao texto da postagem. Para os pares de *keywords*, os relacionamentos são estabelecidos levando em consideração o número de aparições em

uma frase e o número total de palavras na frase. Desta forma, através das arestas, é possível armazenar as informações que identificam a força de relação entre as *keywords*.

Consideramos que a força da relação entre pares de *keywords* é inversamente proporcional ao tamanho das sentenças, com base no trabalho realizado por Lee e Segev (2012). Isso significa que o relacionamento entre duas *keywords* em uma sentença mais curta é mais forte do que em frases mais longas. Assim, para calcular a força de relação entre duas *keywords* em um texto, usamos a Equação (1):

$$R(k_i, k_j) = \sum_{t=1}^n \frac{2}{len(S_t)} \quad (1)$$

onde $R(k_i, k_j)$ é o valor da relação entre o par de *keywords* k_i e k_j , n são as sentenças que contêm estas *keywords* e $len(S_t)$ é o tamanho da sentença S_t .

Após realizar o cálculo da força de relação para todos os pares de *keywords*, a rede complexa que representa o mapa conceitual do texto é criada. Nesta rede, a espessura das arestas reflete diretamente a relação entre as palavras. Da mesma forma, o tamanho dos nós, que é obtido a partir do *betweenness*, caracteriza a importância da palavra para o contexto do conteúdo escrito. A Figura 2 exemplifica o processamento aplicado no texto da postagem de um estudante para a formação do mapa conceitual.

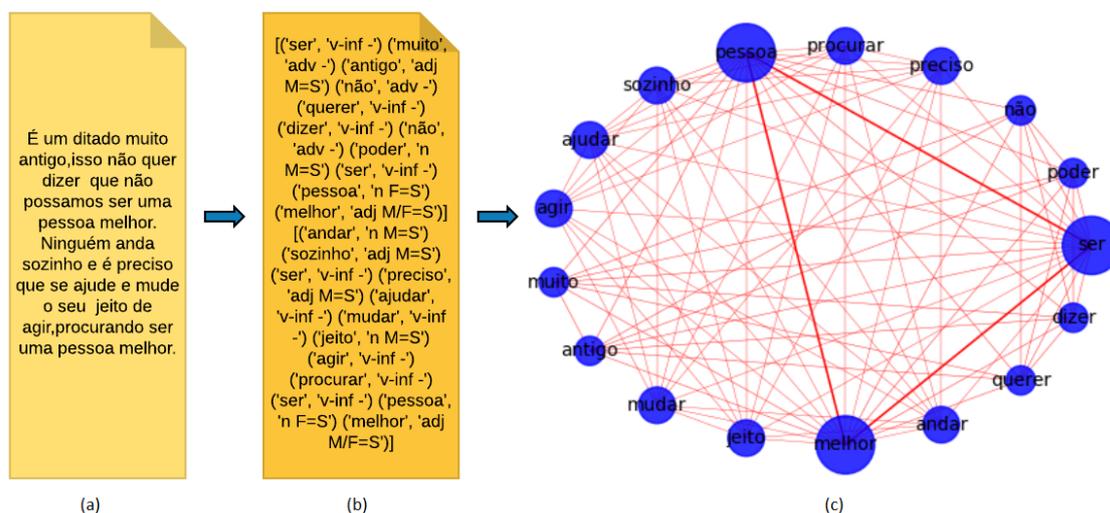


Figura 2. (a) Texto original; (b) Texto pré-processado; (c) Rede do Mapa Conceitual.

3.5. Análise de Relevância Temática

Uma vez que a criação do mapa de conceitual esteja concluída, o módulo de análise de relevância temática é executado. Este módulo recebe como parâmetro as redes complexas para os mapas conceituais, tanto para o texto da postagem que está sendo processado, quanto para o texto da legenda do vídeo. A finalidade desta etapa é identificar se o texto em questão está relacionado ao contexto do tema abordado no conteúdo do vídeo.

Para mensurar a relevância temática de uma contribuição textual, avaliamos as equivalências entre os nós e as arestas das redes formadas. O peso destes elementos, nós e arestas, representa sua importância na rede. Ao identificar um elemento na rede complexa do texto do estudante (N_e) que possua um equivalente na rede do texto do texto

da legenda do vídeo (N_p), o Coeficiente de Equivalência (CE) é calculado de acordo com a Equação (2):

$$CE(e, p) = W_{(p, N_p)} \times 2^{-|W_e - W_p|} \quad (2)$$

sendo $CE(e, p)$ o coeficiente de equivalência entre os elementos e e p , onde $e \in N_e$ e $p \in N_p$. W_e e W_p representa o peso dos elementos e e p , respectivamente. $W_{(p, N_p)}$ denota o peso ponderado do elemento p em relação a todos os elementos de mesma natureza, isto é nós ou arestas, em N_p , o qual é calculado pela Equação (3):

$$W_{(p, N_p)} = \frac{W_p}{\sum_{i \in N_p} W_i} \quad (3)$$

Esta abordagem para o cálculo de CE possibilita com que o peso do elemento influencie no resultado da equivalência. Além disto, outro fator incorporado é a similaridade entre os pesos do par de nós ou arestas equivalentes. Esta característica possibilita com que quanto mais próximos forem os valores dos pesos do par analisado, mais similares eles serão, o que significa que a influência no contexto também é semelhante.

A verificação de equivalência considera que dois nós são equivalentes se eles possuem as mesmas palavras, ou se as palavras são sinônimas entre si, e que pertençam a mesma classe morfossintática. O uso de sinônimos é uma característica importante a ser considerada ao realizar a comparação das redes complexas. Caso a rede da postagem do estudante possua algum nó que seja sinônimo de um nó no texto de problematização, estes serão considerados equivalentes. A análise de sinonímia utilizou o TeP 2.0 [Maziero et al., 2008]. Após a análise da relevância temática, os valores calculados para os atributos CE-Nós e CE-Arestas são inseridos em uma tabela externa, juntamente com o identificador da postagem. Essa tabela, que representa a base de conhecimento obtida como resultado do Motor de Processamento Textual, é responsável por armazenar os dados de todas as postagens no mesmo fórum.

4. Caracterização do Experimento, Resultados e Discussões

Para avaliar a arquitetura proposta neste trabalho e viabilizar a análise dos resultados obtidos, foi utilizado um banco de dados com registros reais do AVA utilizado pelo Núcleo de Educação a Distância (NEAD) da Universidade de Pernambuco. Esta seção descreve o experimento realizado, com base na organização do cenário que foi explorado.

O experimento realizado utilizou o fórum da disciplina “Fundamentos da Educação Especial”, do 5º período do curso de Letras. O fórum selecionado, cujo tema foi “1º fórum temático: sou cego, mas é a escola que não me vê. Sou surdo, mas é a escola que não me escuta...”, possui um total de 94 postagens. O vídeo inserido na abertura do fórum teve como título “Visão histórica da deficiência”, com duração de 9m45s.

Cada postagem realizada pelos alunos no fórum é avaliada pelo professor através de uma nota que varia entre 1 (nota mínima) e 5 (nota máxima). Desta forma, desenvolvemos scripts SQL para extrair a nota de avaliação dada pelo professor a cada postagem do fórum e adicionamos esta informação na tabela com os dados processados do texto. Com isto, a tabela formada pelos dados processados passou a armazenar, além

dos atributos CE-Nós e CE-Arestas, a nota atribuída para cada postagem. A inclusão da nota do avaliador humano possibilitou a avaliação da metodologia proposta, através da classificação das postagens quanto sua relevância temática. Vale ressaltar que em nenhuma das postagens processadas foi atribuída nota mínima (1) pelo professor.

Para realizar a classificação dos textos das postagens, adotamos cinco algoritmos de aprendizagem supervisionada: *K-Nearest Neighbors* (KNN), *Decision Tree* (DT), *Random Forest* (RF), *Support Vector Machine* (SVM) e *Multi-Layer Perceptron* (MLP). Todos os algoritmos testados neste trabalho foram utilizados através do Weka 3.8, de acordo com a configuração padrão dos mesmos fornecida pela ferramenta. Em todos esses algoritmos, foi utilizado o método de validação cruzada *K-fold*, com o valor de *K* igual a 10. Para cada algoritmo, foram realizadas 100 simulações, e o resultado considerado foi calculado pelo valor médio de todos os valores obtidos em cada algoritmo. Para avaliar o desempenho dos classificadores ao processar os dados extraídos da arquitetura proposta, utilizamos as métricas: *Accuracy*, *Precision*, *Recall* e *F-Measure*.

4.1. Cenário de Modelagem e Avaliação

Para a modelagem e avaliação dos dados processados, foi feita a organização das postagens com o objetivo de formar classes com características específicas sob o aspecto da relevância temática. Para isto, definimos que as postagens com nota do professor iguais a 2 ou 3 teriam a nota substituída pelo rotulado “Pouco-relevante”. Por outro lado, as postagens que receberam notas iguais a 4 ou 5, teriam a informação da nota substituída pelo rótulo “Relevante”. Esta manipulação dos dados buscou investigar o comportamento dos algoritmos ao predizerem a classe dos registros (“Pouco-relevante” ou “Relevante”) a partir dos valores dos atributos CE-Nós e CE-Arestas em cada postagem, que foram calculados pelo Motor de Processamento Textual. Após a atribuição dos rótulos das postagens, a disposição quantitativa dos dados das classes ficou apresentada conforme descrito na Tabela 1.

Tabela 1. Disposição quantitativa dos dados das classes.

Classe	Notas consideradas	Total de Postagens
Pouco-relevante	2 e 3	44
Relevante	4 e 5	50

Os resultados mostrados na Figura 3 representam os valores das métricas de desempenho que foram extraídos dos classificadores.

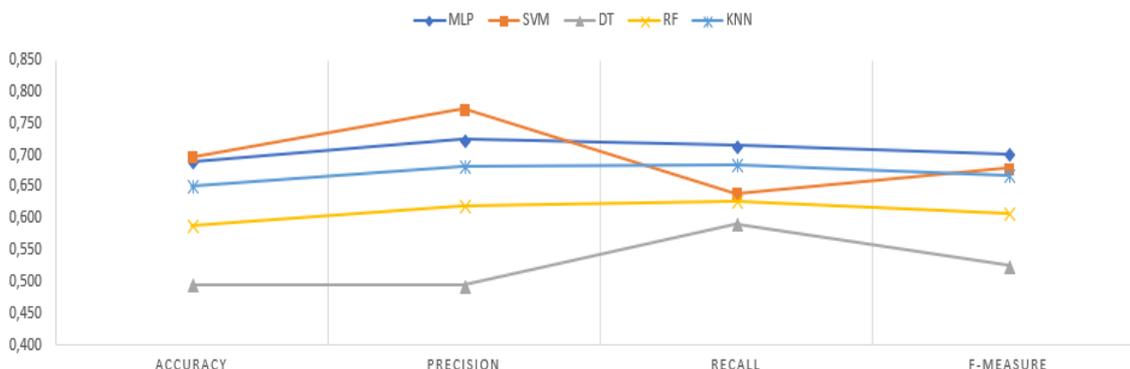


Figura 3. Performance dos classificadores.

Ao observarmos o resultado da classificação em relação a *Acuracy*, percebemos que os melhores resultados foram obtidos pelos algoritmos MLP e SVM, com uma taxa de acerto de aproximadamente 70%. Este resultado aponta para um índice de classificação geral satisfatório, uma vez que esta métrica leva em consideração a taxa de acerto obtida tanto para a classe “Pouco-relevante” quanto para a classe “Relevante”.

Ao avaliarmos os valores obtidos em relação ao *Precision*, percebemos que os algoritmos SVM e MLP também se destacaram ao atingir os melhores resultados, com uma taxa de acerto de 77,3% e 72,5% respectivamente. Com isto, podemos observar que houve um desempenho melhor destes classificadores ao ser considerado o valor preditivo positivo, o que é caracterizado pelo índice de acerto das postagens que foram corretamente classificadas como “Relevante”. Por outro lado, ao analisarmos o comportamento dos algoritmos ao lidarem apenas com postagens da classe “Relevante”, que é representado pelo *Recall*, o melhor resultado foi atingido pelo algoritmo MLP, o qual obteve uma taxa de acerto de 71,5%. Da mesma forma, ao levarmos em consideração simultaneamente as postagens classificadas de maneira incorreta para ambas as classes, o que é obtido pela métrica *F-Measure*, o melhor desempenho permaneceu associado ao algoritmo MLP, com 70,3% de aproveitamento.

5. Considerações Finais

Este trabalho propôs a modelagem e o desenvolvimento de uma solução tecnológica aplicada na análise de relevância temática das postagens feitas por estudantes em fóruns de discussão, em relação a vídeos indicados por professores como recurso didático para promover o aprofundamento sobre o tema de discussão. Para atingir estes objetivos, foram aplicadas técnicas de Mineração Textual e Ciência das Redes para o processamento das postagens. Em seguida, os dados processados foram avaliados em cinco algoritmos de aprendizagem supervisionada.

Os resultados obtidos apontam para um aproveitamento satisfatório da metodologia proposta, refletido em uma performance considerável dos classificadores utilizados. Observa-se que o processamento adequado dos conteúdos gerados pelos estudantes pode oferecer informações potencialmente úteis no processo de ensino-aprendizagem, e que podem fornecer ao professor subsídios importantes a respeito dos alunos para tomada de decisões pedagógicas.

Como trabalhos futuros, planejamos incorporar características que levem em consideração a similaridade entre as postagens dos estudantes. Além disto, pretendemos incluir no motor de processamento a percepção da evolução do contexto da discussão, na medida em que novas postagens relevantes são inseridas. Estas iniciativas buscam a melhoria do desempenho do processamento textual, que é refletido em uma maior taxa de acerto na classificação das postagens quanto a relevância temática.

Agradecimentos

Este trabalho foi apoiado pelo NEAD (Núcleo de Educação a Distância), Departamento de Educação a Distância da Universidade de Pernambuco.

Referências

ABED - Associação Brasileira de Educação a Distância. Censo EaD.br: relatório analítico da aprendizagem a distância no Brasil 2015. Curitiba - Brasil: InterSaberes. 2016.

- Antiqueira, L. Desenvolvimento de Técnicas Baseadas em Redes Complexas para Sumarização Extrativa de Textos. Dissertação de Mestrado. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. São Carlos/SP, Brasil. 2007.
- Azevedo, B. F. MineraFórum: um recurso de apoio para análise qualitativa em fóruns de discussão. Porto Alegre, Brasil: Universidade Federal do Rio Grande do Sul - UFRGS. 2011.
- Azevedo, B. F., Behar, P., & Reategui, E. Qualitative analysis of discussion forums. *International Journal of Computer Information Systems and Industrial Management*, 671-678. 2011.
- Dallacosta, A. Possibilidades educacionais do uso de vídeos anotados no Youtube. Departamento de Educação e Cultura do Exército Brasileiro. 2004.
- Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry*, American Sociological Association, pp. 35-41. 1977.
- Giannakos, M. N. Exploring the video-based learning research: A review of the literature. *British Journal of Educational Technology*, 44(6). pp. 191-195. 2013.
- Giannakos, M. N., Chorianopoulos, K., Ronchetti, M., Szegedi, P., Teasley, S. D. Analytics on video-based learning. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)*. pp. 283-284. 2013.
- Gonzalez, R. G. youtube-dl Download videos from YouTube. <<http://rg3.github.io/youtube-dl/>> Acessado em: 1 jun. 2018.
- Lee, J. H., & Segev, A. Knowledge maps for e-learning. *Computers & Education*. pp. 353-364. 2010.
- Mattar, J. Youtube na Educação: o uso de vídeos em EaD. Em: Congresso da Associação brasileira de educação à distância. São Paulo, 2009. Anais eletrônicos - Resumos. 2009.
- Maziero, E. G., Pardo, T. A., Di Felippo, A., & Dias-da-Silva, B. C. A base de dados lexical e a interface web do TeP 2.0 – Thesaurus Eletrônico para o Português do Brasil. Workshop em Tecnologia da Informação e da Linguagem Humana. pp. 390-392. XIV Simpósio Brasileiro de Sistemas Multimídia e Web. Vila Velha, Brasil. 2008.
- Medeiros, M. A. Reconhecimento de autoria de textos utilizando redes complexas. Recife, Brasil: Universidade de Pernambuco – UPE. 2015.
- Moran, J. M. O vídeo na sala de aula. In *Revista Comunicação & Educação*. São Paulo, ECA-Ed. Moderna. pp. 27-35, jan./abr. de 1995.
- Moran, J. M. A integração das tecnologias na educação. 2006. Disponível em: <http://www.eca.usp.br/prof/moran/site/textos/tecnologias_eduacao/integracao.pdf> Acesso em 15 jun. 2018.
- Oliveira, D. S. de., O Uso do Vídeo em EAD: Desafios no Processo de Ensino Aprendizagem. *Revista Cesuca Virtual: Conhecimento sem Fronteiras - ISSN 2318-4221*, Vol. 1, N. 1, Jul/2013.
- Silva, W. D. Aprimorando o corretor gramatical CoGrOO. Instituto de Matemática e Estatística, Universidade de São Paulo. Brasil. 2013.
- Soares, F. B. M., Machado, C. J. R., Diniz, D., Maciel, A. M. A. Educational Data Mining to support Distance Learning students with difficulties in the Portuguese Grammar. Anais do XXVII Simpósio Brasileiro de Informática na Educação (SBIE 2016). pp. 956-965. Brasil. 2016.
- Tan, A. Text Mining: the state of the art and the challenges. *Workshop on Knowledge Discovery from Advanced Databases*, (pp. 71-76). Pequim. 1999.