Aplicação de Análise de Sentimento em Fóruns Educacionais para Prevenir Evasão

Diogenes Azevedo¹, Rafael Ferreira¹, Valter Mendonca¹, Péricles Miranda¹

¹Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco (UFRPE)

{diogenesloazevedo, rafaelmello, valternetosnt, periclesmiranda}@gmail.com

Abstract. *With the spread of the Internet and distance learning courses, the data generated by teachers and students have grown exponentially in recent years, especially textual data from the educational forums of virtual learning environments. However, even with the success of this method of education, the courses dropouts have proved a major problem to be addressed. The identification of a possible unmotivated student is useful to prevent the dropout. Computational techniques such as data mining could be used to deal with this problem. This paper proposes a sentiment analysis technique to aid the teacher detecting student's motivation based on their posts in educational forums. The experiments shows that the proposal reach to a accuracy of 82% in the polarity classification.*

Resumo. *Com a disseminação da Internet e dos cursos de educação à distância, os dados gerados por professores, alunos e tutores vêm crescendo exponencialmente nos últimos anos, em especial os dados textuais, oriundos dos fóruns educacionais dos ambientes virtuais de aprendizagem. Contudo, mesmo com o sucesso dessa modalidade de ensino, a evasão vem se mostrando um grande problema a ser combatido. A identificação de estudante com possibilidade de desistência é uma das formas encontradas para prevenção evasão. Técnicas computacionais, como mineração de dados, podem ser usadas para lidar com esse problema. Diante deste cenário, este trabalho propõe uma técnica de análise de sentimento que auxilie o professor na detecção da motivação do aluno a partir de postagens de fóruns educacionais, prevendo um possível caso de evasão. Os experimentos mostram que a proposta alcançou uma taxa de acerto de 82% na classificação da motivação dos estudantes.*

1. Introdução

Em 2003, a educação à distância no Brasil já gerava uma receita por volta dos R\$ 350 milhões, com 34 cursos criados [Pereira et al. 2007]. Com a disseminação da Internet e dos próprios ambientes digitais de aprendizagem, em 2014 o Brasil já tinha mais de 900 cursos ofertados à distância [Alves et al. 2015].

Junto com o aumento na quantidade de cursos, também cresceu consideravelmente o índice de evasão [de Almeida Bizarria et al. 2015]. Barroso e Falcão [Barroso and Falcão 2004] afirmam que existem pelo menos três variáveis explicativas para o fenômeno da evasão, são elas: i) econômica – quando o aluno não tem condições socioeconômicas de prosseguir no curso; ii) vocacional – quando não há identificação

do aluno com o curso; iii) institucional – quando a evasão dá-se por reprovações nas disciplinas iniciais, inadequação aos métodos de estudo ou por desmotivação do aluno principalmente quando não consegue acompanhar o ritmo das disciplinas.

Existem várias pesquisas que vêm tratando o problema da evasão [Ferreira and Elia 2013], uma possível solução para esse problema é a aplicação métodos de mineração de dados ou de texto em ambientes virtuais de aprendizagem para prever a eminência de uma evasão [Rigo et al. 2014]. Dentre as ferramentas disponíveis em ambientes educacionais, o fórum proporciona a maior interatividade entre alunos, professores e tutores, pois ele gera a oportunidade de esclarecimento de dúvidas, consultas, e a socialização dos participantes [Bastos et al. 2011].

Contudo, o aumento do número de cursos à distância tem gerado uma grande quantidade de postagens nos fóruns educacionais, o que dificulta o acompanhamento individual do aluno por meio do professor [Bastos et al. 2011]. Sendo assim, técnicas de mineração de texto podem ser utilizadas para extrair informações dos fóruns auxiliando o combate contra evasão [Silva et al. 2015].

Pang e Lee [Pang and Lee 2008] afirmam que uma parte importante da mineração de texto é descobrir o que as pessoas estão pensando e sentindo, e que existem vários recursos ricos em opiniões a serem mineradas. No contexto educacional, a análise de sentimento pode ser usada para mensurar a motivação dos alunos a partir de interações em fóruns educacionais [Azevedo et al. 2010]. Com essa informação é possível propor medidas para diminuir a evasão institucional em cursos a distância.

Diante deste cenário, este trabalho propõe uma técnica de análise de sentimento, para o idioma português, com o objetivo de identificar a motivação do aluno a partir da sua interação no fórum educacional. A partir deste resultado o professor pode buscar alternativas que motive novamente o aluno, na tentativa de evitar sua evasão. A técnica proposta foi avaliada usando dois bancos de dados, o primeiro é formado por postagens geradas artificialmente e o segundo utiliza postagens de diversas turmas de um curso de Ciência da Computação. A proposta atingiu uma acurácia de 71,90% e 82%, ao classificar o aluno como motivado ou não, nos bancos de dados propostos, respectivamente.

O trabalho é dividido da seguinte forma: A Seção 2 e 3 apresentam conceitos básicos sobre fóruns educacionais e os trabalhos relacionados. Na seção 4 são apresentados detalhes sobre o sistema proposto. Na Seção 5 e 6 são apresentados respectivamente os resultados e uma discussão sobre os mesmos. Por fim, na Seção 7 são apresentadas as conclusões e trabalhos futuros.

2. Fórum Educacional

O fórum de discussão é uma ferramenta de comunicação assíncrona, que tem como função principal promover discussões entre os participantes acerca de algum tema. Essa ferramenta é frequentemente utilizada nos ambientes virtuais de aprendizagem, por ser um recurso que proporciona a interação coletiva dos participantes.

Um aspecto importante é que o fórum é considerado uma das ferramentas que permite maior interatividade em ambientes educacionais. Segundo o estudo promovido por [BARROS and CARVALHO 2011] o fórum foi apontado por 69,2% dos alunos como a ferramenta mais interativa. Outras ferramentas que também têm essa característica são:

tarefa (41,0%), *chat* (38,5%) e questionário (20,5%). Isso mostra que a possibilidade de geração de conteúdo no fórum é ainda maior do que as outras ferramentas.

O ganho pedagógico causado pela utilização dos fóruns educacionais é perceptível, mesmo que os alunos envolvidos no processo empreguem pouco tempo para utilizar esta ferramenta [Cheng et al. 2011].

Contudo, com o aumento da interação entre os usuários do fórum, ocorre um aumento também do volume das postagens, tornando-se por vezes inviável o gerenciamento de todas elas pelo professor, demandando muito tempo do mesmo. Por isso faz-se necessária a existência de técnicas automáticas que auxiliem a análise das postagens dos fóruns educacionais, como o objetivo de maximizar o ganho de conhecimento [Azevedo et al. 2010] ou evitar a evasão [Silva et al. 2015].

3. Trabalhos Relacionados

Esta seção apresenta aplicações de técnicas de mineração de texto focadas em prevenir a evasão e utilização de análise de sentimento em postagens de fóruns educacionais.

O trabalho proposto em [Longhi et al. 2009] detalha uma metodologia para realizar análise de sentimento em postagens de fóruns. Ela classifica a postagem em insatisfeito, desanimado, satisfeito e animado. Apesar do trabalho ser promissor, pouco foi detalhada a solução proposta e os experimentos utilizaram apenas 5 postagens.

Seguindo a mesma linha, [Coutinho et al. 2016] propõe a utilização do algoritmo de rede bayesiana para identificação de polaridade (positiva, negativa ou neutra) de mensagens geradas a partir de *chats* educacionais.

Uma proposta de utilização da teoria da semântica lexical computacional para realização de análise de sentimentos para português foi apresentada em [Rigo et al. 2013]. Este trabalho utiliza o SentiLex-PT3, que é um recurso léxico semântico em português, para extrair os sentimentos das postagens. Esse recurso contém 6321 lemas adjetivos e 25406 formas flexionadas que podem ser usadas para determinar a polaridade de um texto.

[Machado et al. 2015] propõe um novo recurso léxico semântico, chamado de *Personalitatem Lexicon*, que pode ser usado para análise de sentimento em postagens. Esse recurso tem como principal função extrair características relacionadas a personalidade dos alunos.

Em [Reis et al. 2010] é proposta uma nova categorização para sentimentos diferente dos trabalhos anteriores. Os autores propõe classes como: amor-afetivo, esperança-pensar crítico, humildade, entre outros. Porém, o trabalho apresenta apenas um estudo qualitativo de análise de sentimento em fóruns, não foi proposta a utilização de nenhuma técnica de mineração de texto.

É importante ressaltar que a análise de sentimento pode ser usada para lidar com o problema da evasão. Contudo, os autores dos trabalhos citados não deixam explícita essa relação. [Silva et al. 2015] propõe a utilização do quantitativo de postagens em fóruns educacionais para prever a evasão. Apesar de o trabalho ter alcançado um resultado interessante, a aplicação de mineração de texto pode melhorar consideravelmente o sistema proposto.

Diante disto, este trabalho propõe um sistema de mineração de texto para identificar a motivação do aluno. As principais novidades do trabalho são: (i) Utilização de uma abordagem que utiliza técnicas estatísticas e aplicação de um léxico semântico; (ii) Tratamento de palavras emocionais e de negação; (iii) Utilização de diferentes corpus de postagens para avaliação; (iv) Propõe o paralelo entre análise de sentimento e evasão escolar.

4. Proposta

A técnica proposta identifica a polaridade da motivação do aluno como positiva ou negativa. Para esse trabalho vamos considerar a polaridade positiva como o aluno motivado e negativa para o desmotivado (que tem risco de evasão). A polaridade é obtida a partir de 4 etapas: pré-processamento, distribuição de pesos, tratamento da negação e classificação. As próximas seções detalham cada uma dessas etapas explicando tanto o funcionamento como as configurações de saída de cada etapa, que serve de entrada para as seguintes.

4.1. Pré-processamento

Esta etapa está baseada na execução de processos de *tokenização* e eliminação de *stopwords* [Baeza-Yates and Ribeiro-Neto 2013]. A *tokenização* consiste em decompor o documento em palavras usando delimitadores. Estes delimitadores podem ser espaços em branco, quebra de linha, pontuação e caracteres especiais, entre outros. A eliminação de *stopwords* remove palavras que agregam pouco ou nenhum significado ao texto, geralmente são artigos, preposições e pronomes. Existem várias listas de *stopwords* disponíveis na Internet nos mais variados idiomas e, para este trabalho, foi utilizada a lista disponível em: <https://gist.github.com/alopes/5358189>.

Após o pré-processamento cada postagem contém apenas palavras que contribuem e incrementam o significado do texto.

4.2. Distribuição de Pesos

Esta etapa recebe como entrada a lista de palavras pré-processadas e tem como objetivo mensurar a importância de cada palavra para as classes positivas e negativas. Para isso, foi realizada uma análise da distribuição de peso de cada palavra de acordo com sua relevância para cada classe utilizando duas técnicas: a implementação do algoritmo tf-idf e o incremento de peso baseado em um dicionário léxico semântico proposto no trabalho de Pasqualotti [Pasqualotti 2008]. Este dicionário foi escolhido porque ele propõe grupos de palavras que demonstram emoção relacionadas a motivação do autor do texto.

O tf-idf, do inglês *term-frequency - inverse document frequency* [Baeza-Yates and Ribeiro-Neto 2013], é uma maneira eficaz de distribuir pesos para as palavras do documento. O *tf* é a probabilidade de uma palavra aparecer no documento, calculada por meio da razão entre a frequência que a palavra aparece e o número total de palavras no documento, fórmula 1.

$$tf_{t,d} = \frac{n_t}{\sum_d n_d} \quad (1)$$

Já o *idf*, a frequência inversa, é obtido pelo logaritmo da razão entre o número total de documentos (N) e a quantidade de documentos contendo a palavra t (N_t), fórmula 2.

O tf-idf é, então, a multiplicação desses dois termos, cada palavra distinta do documento possui um valor tf-idf.

$$idf_t = \log \frac{N}{N_t} \quad (2)$$

Após a aplicação do tf-idf todas as palavras do documento têm um peso associado. Porém, em 2008, Pasqualotti agrupou substantivos e adjetivos em 15 conjuntos de palavras que representam emoção, substantivos e adjetivos que apontam o estado de ânimo do autor. Cada grupo de palavra representa um sentimento: amor, ódio, raiva, entusiasmo, indiferença, etc. Com isso, além da pontuação do tf-idf, todas as palavras desses grupos, sejam elas de sentimento positivo ou negativo, recebem um incremento de uma unidade, fazendo com que o classificador atribua uma maior importância a essas palavras.

Para cada palavra da distribuição foram armazenados os pesos positivos e negativos. Esses pesos foram atribuídos com base nas postagens previamente classificadas. Assim, existem três possibilidades na distribuição de pesos: (i) quando uma palavra só existe na base positiva: nesta condição, o peso negativo da palavra será zero; (ii) quando uma palavra só existe na base negativa: analogamente à situação (i), só que desta vez o peso positivo será zero; (iii) quando uma palavra existe em ambas as bases: nesse caso, ambos os pesos são armazenados.

A saída desta etapa é o conjunto de palavras de todos os textos da coleção de treinamento e seus respectivos pesos.

4.3. Tratamento da Negação

A negação é um dos maiores problemas em mineração de texto pois ela inverte o sentido da opinião [Becker and Tumitan 2013]. Assim, por exemplo, nas frases “A aula do professor é interessante.” e “A aula do professor não é interessante.” a palavra interessante tem sentidos diferentes e caso essa diferença não seja tratada, pode-se gerar erros de classificação, diminuindo a acurácia do sistema.

Para lidar com esse problema o trabalho proposto implementa uma técnica que muda o sentido das duas próximas palavras após a ocorrência de um “não” no documento. Foram feitos testes mudando o sentido de até as próximas 10 palavras depois do “não”, dentre os testes o que mais acrescentou acertos ao classificador foi o que alterou apenas o sentido das 2 próximas palavras após a negação.

4.4. Classificação

Na etapa de classificação, cada frase será vista como um somatório de pesos de suas palavras. Para cada palavra da postagem, é consultado seu maior peso (positivo ou negativo). Nos casos em que uma palavra tem o peso negativo maior, esse peso terá o sinal negado no somatório, diminuindo seu valor final. Desta maneira, ao final da postagem, se o somatório for maior ou igual a zero, a frase é dita como uma postagem positiva e, caso o somatório tenha valor menor que zero, a frase é considerada uma postagem negativa.

5. Resultados

Nesta seção são abordados detalhes sobre as bases de dados utilizada na avaliação da técnica proposta, resultados obtidos e um estudo sobre as palavras que obtiveram os mai-

ores pesos positivos e negativos.

5.1. Bases de Dados

A primeira base de dados (BD1) que serviu como entrada para a ferramenta foi construída durante o semestre 2015.2, fruto da disciplina de Tópicos Avançados em Inteligência Artificial onde o foco principal foi a mineração de texto. Na disciplina, todos os alunos tinham como exercício semanal a tarefa de simular postagens de alunos matriculados em matemática discreta em um fórum educacional. A cada semana os alunos escreviam dez textos de no mínimo vinte palavras com o intuito de retratar a motivação do aluno, ou seja, se em uma semana um grupo de alunos simulavam um aluno decepcionado com a disciplina (sentimento negativo), na semana seguinte este grupo simularia alunos motivados (sentimento positivo).

Após a etapa de escrita foi realizada uma validação das postagens. O principal objetivo é identificar se as postagens estavam dentro da polaridade proposta. Cada aluno analisou entre 30 e 40 postagens por semana, atribuindo a polaridade positiva, negativa ou inconclusiva. As postagens que foram classificadas como inconclusivas por um dos alunos foi automaticamente excluída do banco de dados. Além disto, quando houve alguma divergência entre os alunos ela foi resolvida pelo professor da disciplina.

O total de postagens obtida foi de 460 dividido nas diferentes categorias. Em relação a análise de sentimento, inicialmente foram escritas 230 postagens positivas e 230 negativas. Após a validação das postagens esse número ficou em 107 e 121 para positivas e negativas respectivamente. Para utilização no sistema proposto, foram utilizadas apenas as postagens após a validação.

A segunda base de dados (BD2) avaliada foi retirada de diferentes disciplinas de um curso a distância de ciência da computação. É importante frisar que essas postagens foram extraídas de um curso já realizado, ela não foi feita como a primeira que foi gerada artificialmente. A classificação da polaridade das postagens e validação seguiram os mesmos critérios da base anterior. A base original possuía 450 postagens, mas apenas 100 foram consideradas relevantes para a análise de sentimento, sendo 24 postagens negativas e 76 positivas.

5.2. Comparação de Resultados

Para avaliar o algoritmo proposto foi utilizada a taxa de acerto na classificação (acurácia) e uma metodologia utilizando 10-fold cross-validation para treinar e testar o sistema [Baeza-Yates and Ribeiro-Neto 2013].

A Tabela 1 mostra a avaliação de três configurações diferentes na BD1, são elas: (i) classificação utilizando o peso de cada palavra; (ii) utilização dos pesos e palavras de emoção; (iii) utilização dos pesos, palavras de emoção e tratamento de negação. É possível perceber que incrementar os pesos das palavras com os grupos emoção e com a técnica da negação melhorou em 8,67% a acurácia do sistema.

A primeira abordagem atingiu uma taxa de 66,16%, com a adoção das palavras de emoção obtidas no estudo de Pasqualotti; a taxa de acerto foi elevada de 66,16% para 69,43%. A técnica proposta, utilizando todas as etapas, classificou corretamente 71,90% das instâncias.

Tabela 1. Avaliação BD1

Abordagem	Acurácia
Classificador	66,16%
Classificador + Palavras de emoção	69,43%
Classificador + Palavras de emoção + Tratamento de negação	71,90%

A Tabela 2 apresenta os resultados da proposta aplicada ao BD2. A taxa de acerto do sistema atingiu 82,22%, contudo ela foi alcançada pela configuração utilizando apenas o classificador e as palavras de emoção. O comportamento em relação ao acréscimo das palavras de emoção não foi tão relevante quanto na base anterior porque existiam apenas 4 palavras consideradas de emoção no BD2: atenção, pena, ira e problema. Como consequência o tratamento de palavras negativas piorou bastante os resultados nesse caso. A grande quantidade de postagens positivas do BD2 explica o fato do tratamento de negação não ter alcançado um bom resultado, pois, em geral, as palavras de negação melhoram a classificação de postagem com polaridade negativa.

Tabela 2. Avaliação BD2

Abordagem	Acurácia
Classificador	80,55%
Classificador + Palavras de emoção	82,22%
Classificador + Palavras de emoção + Tratamento de negação	46,66%

5.3. Estudo dos Pesos das Palavras

Devido ao grande número de palavras das bases de dados, a probabilidade de cada palavra ocorrer é extremamente pequena, o que gera valores tf-idf também muito pequenos, todos menores que um. Sendo assim, essa seção apresenta um estudo dos maiores valores tf-idf obtidos tanto nas bases dos alunos motivados quanto dos desmotivados. Vale lembrar que por conta do incremento de peso das palavras de emoção, estas tiveram um peso maior que um. Logo, para melhor estudo da distribuição de peso, são mostradas tabelas com e sem o incremento.

Como visto acima, as palavras dos Grupos de Palavras de Emoção têm seu peso incrementado em uma unidade, o que faz com que sejam as únicas palavras com peso maior que um. Pode-se notar que todas são substantivos ou adjetivos que buscam evidenciar a emoção do autor. A Tabela 3 apresenta uma lista com estas palavras.

Quando não há o incremento de peso, as palavras com maiores tf-idf são palavras relacionadas ao tema do documento. Nota-se que as palavras se repetem em ambas as partes da base de dados (ver Tabela 4), tanto nos motivados quanto nos desmotivados. Isto ocorre porque essas palavras se repetem com maior frequência no decorrer de todo o documento, independente da postagem do aluno ser positiva ou não. Nesse caso, no cálculo do somatório de cada frase leva-se em consideração o maior valor. Por exemplo, a palavra *Assunto* tem a polaridade positiva enquanto *Aula é* tem a polaridade negativa.

Além dos testes realizados, também se testou a hipótese de eliminar as palavras que se repetiram em ambas as classes restando apenas aquelas que só ocorriam em uma

Tabela 3. Pesos das palavras com o incremento das palavras de emoção

Palavra	Peso	Polaridade
Animado	1.0030161196830423	Motivado
Apaixonado	1.0013290647166357	Motivado
Atenção	1.0022496762505524	Motivado
Gosto	1.0013290647166357	Motivado
Arrependido	1.0022496762505524	Motivado
Louco	1.001118850355816	Desmotivado
Medo	1.0019070801782282	Desmotivado
Preocupado	1.001118850355816	Desmotivado
Problema	1.001118850355816	Desmotivado
Terrível	1.0025705193463277	Desmotivado

Tabela 4. Pesos das palavras sem o incremento das palavras de emoção

Palavra	Peso	Polaridade
Assunto	0.010291934754526008	Motivado
Aula	0.009161495175341226	Motivado
Bem	0.011813678242435212	Motivado
Entender	0.011097697742893686	Motivado
Exercícios	0.011294915339531858	Motivado
Assunto	0.00929862018065958	Desmotivado
Aula	0.009300999205031727	Desmotivado
Cadeira	0.008236967202155178	Desmotivado
Difícil	0.008694576491163798	Desmotivado
Entender	0.00927647806044798	Desmotivado

delas, mas os resultados alcançaram resultados piores, o que constata que as palavras que se repetem no quadro acima são importantes para o classificador por representarem o tema e assunto do documento.

6. Discussão

Como apresentado nas seções iniciais, o principal objetivo deste trabalho é oferecer uma técnica automática de análise de sentimento para classificar postagens de fóruns educacionais e identificar se o aluno se encontra motivado ou desmotivado. Esse resultado auxilia professores e tutores de cursos EaD no combate à evasão. Um cálculo simples mostra que, se um professor tem uma turma de 30 alunos e cada um deles posta algo no fórum cinco vezes por semana, ao fim do mês o professor terá 600 postagens para ler.

Diante desse cenário, o professor pode se beneficiar da análise de sentimento para identificar alunos pouco motivados a continuar suas atividades acadêmicas. Dessa forma, a técnica proposta nesse trabalho emite alertas que podem servir como indício que o autor da postagem está com dificuldades no aprendizado (evasão institucional) e o professor pode, por sua vez, direcionar seu tempo para auxiliar aos alunos com potencial chance de evasão.

Outra forma de se melhorar o ensino do curso é dar atenção às palavras com

maior tf-idf (tabelas 3 e 4). Desta maneira, o professor pode identificar o sentimento dos seus alunos visualizando as palavras das tabelas relacionadas ao tema do seu curso que possivelmente indicam a causa de problemas no aprendizado. Vimos que na tabela 4 as palavras “exercícios” e “difícil” aparecem com pesos maiores nas postagens desmotivadas e isso pode ser um indício de que os exercícios propostos pelo professor podem estar com dificuldade elevada, desmotivando alguns alunos. Com isso, o professor também pode identificar os assuntos que têm gerado mais dificuldade para os alunos.

7. Conclusão e Trabalhos Futuros

Este trabalho apresentou uma técnica para identificação da motivação de alunos a partir de suas postagens em fóruns educacionais. O trabalho utiliza uma abordagem mista, que combina técnicas estatísticas e semânticas para atingir o objetivo final. A partir do resultado da proposta, o professor identifica alunos com possibilidade de evasão com antecedência, assim podem ser tomadas medidas afim de evitar a desistência do curso.

Para validar a proposta, foram utilizando diferentes bases de dados que possuem pelo menos 100 postagens divididas nas classes motivado e desmotivado. Os trabalhos anteriores que fazem análise relacionada à proposta utilizam bases consideravelmente menores para avaliar a proposta. O sistema atingiu, no melhor caso, uma acurácia de 82% ao classificar o aluno como motivado ou não.

Por fim, o trabalho apresentou alguns cenários onde o professor pode se beneficiar da identificação de motivação. Os principais listados são: combater a evasão e identificação dos assuntos que estão gerando mais dúvidas entre os alunos.

Como trabalhos futuros espera-se: (i) Aplicar a técnica proposta para fóruns de outras áreas, diferentes da computação; (ii) Analisar outras técnicas para tratamento de navegação; (iii) Relacionar a técnica proposta com outras aplicações automáticas que levem à diminuição da evasão.

Referências

- Alves, T. M. F., Menezes, A. H. N., and Vasconcelo, F. M. d. B. P. (2015). Crescimento da educação a distância e seus desafios: Uma revisão bibliográfica. *Revista de Educação do Vale do São Francisco-REVASF*, 4(6):63–74.
- Azevedo, B. F. T., Behar, P. A., and Reategui, E. B. (2010). Aplicação da mineração de textos em fóruns de discussão. *RENOTE*, 8(3).
- Baeza-Yates, R. and Ribeiro-Neto, B. (2013). *Recuperação de Informação-: Conceitos e Tecnologia das Máquinas de Busca*. Bookman Editora.
- BARROS, M. d. G. and CARVALHO, A. B. G. (2011). As concepções de interatividade nos ambientes virtuais de aprendizagem. *Campina Grande: EDUEPB*.
- Barroso, M. F. and Falcão, E. B. (2004). Evasão universitária: O caso do instituto de física da ufrj. *Encontro Nacional de Pesquisa em Ensino de Física*, 9:1–14.
- Bastos, H. P. P., Bercht, M., and Wives, L. K. (2011). Presença social e pertencimento em fóruns educacionais: manifestação e percepção de afetividade. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 1.

- Becker, K. and Tuminan, D. (2013). Introdução à mineração de opiniões: Conceitos, aplicações e desafios. *Simpósio Brasileiro de Banco de Dados*.
- Cheng, C. K., Paré, D. E., Collimore, L.-M., and Joordens, S. (2011). Assessing the effectiveness of a voluntary online discussion forum on improving students' course performance. *Computers & Education*, 56(1):253–261.
- Coutinho, E., Moreira, L., Paillard, G., and de Lima, E. T. (2016). Análise do sentimento de mensagens de chats em uma turma de graduação de um curso de educação à distância. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 5, page 1019.
- de Almeida Bizarria, F. P., da Silva, M. A., Tassigny, M. M., and Carneiro, T. C. J. (2015). Papel do tutor no combate à evasão na ead: percepções de profissionais de uma instituição de ensino superior. *Educação, Ciência e Cultura*, 20(1):p–85.
- Ferreira, V. d. S. and Elia, M. d. F. (2013). Uma modelagem conceitual para apoiar a identificação das causas da evasão escolar em ead. In *Anais: XIX Workshop de Informática na Escola, Campinas, SP: UNICAMP*.
- Longhi, M. T., Behar, P. A., Bercht, M., and Simonato, G. (2009). Investigando a subjetividade afetiva na comunicação assíncrona de ambientes virtuais de aprendizagem. *Anais do SBIE2009-XX Simpósio Brasileiro de Informática na Educação, Florianópolis, SC. Novembro*.
- Machado, A. A., Longhi, M. T., Nunes, M. A., Pardo, T. A. S., et al. (2015). Personalitatem lexicon: um léxico em português brasileiro para mineração de traços de personalidade em textos. In *Simpósio Brasileiro de Informática na Educação, XXVI. Universidade Federal de Alagoas-UFAL*.
- Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135.
- Pasqualotti, P. R. (2008). *Reconhecimento de expressões de emoções na interação mediada por computador*. PhD thesis, Universidade do Vale do Rio do Sinos.
- Pereira, A., Schmitt, V., and Dias, M. (2007). Ava-ambientes virtuais de aprendizagem em diferentes contextos. *Editora Ciência Moderna, Rio de Janeiro*.
- Reis, E. M., Vasconcelos, F. H. L., and Martins, C. A. (2010). Uma análise da interação em fóruns de discussão em um ambiente virtual de aprendizagem. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 1.
- Rigo, S. J., Alves, I. M., Gazola, O., Belau, F., Barbosa, J. L., and Costa, C. (2013). Abordagem linguística para identificação da dimensão afetiva expressa em textos de ambientes virtuais de aprendizagem—um léxico da emoção. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 24, page 738.
- Rigo, S. J., Cambruzzi, W., Barbosa, J. L., and Cazella, S. C. (2014). Aplicações de mineração de dados educacionais e learning analytics com foco na evasão escolar: oportunidades e desafios. *Revista Brasileira de Informática na Educação*, 22(01):132.
- Silva, F., da Silva, J., Silva, R., and Fonseca, L. C. (2015). Um modelo preditivo para diagnóstico de evasão baseado nas interações de alunos em fóruns de discussão. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 26, page 1187.