

Processamento do Discurso em Textos Dissertativos-Argumentativos: Uma Abordagem Baseada em Mineração de Argumentos e Aprendizado Supervisionado de Máquina

Jonathan Nau, Rudimar Luís Scaranto Dazzi, Aluizio Haendchen Filho, Anita Fernandes

Laboratório de Inteligência Aplicada (LIA)
Universidade do Vale do Itajaí (UNIVALI) – Itajaí, SC – Brasil

{jonathan.nau, aluizio.h.filho}@gmail.com,
{anita.fernandes, rudimar}@univali.br

***Abstract.** Identifying the structure of the argument is useful to analyze student's speech and logical reasoning. The discourse understanding process includes the understanding of morphology, syntax and semantics. In this work, speech processing follows a methodology composed by the following phases: (i) creation of a corpus for training and validation of learning; (ii) implementation of the argument mining approach; and (iii) application of machine learning techniques. The results achieved were among the average found in the literature for works carried out in the English language. The identification of the intervention proposal was the highlight, with an F-value of 0.744.*

***Resumo.** A identificação da estrutura da argumentação é útil para analisar o discurso e o raciocínio lógico do aluno. O processo de entendimento do discurso abrange a compreensão da morfologia, da sintaxe e da semântica. Neste trabalho, o processamento do discurso segue uma metodologia composta das seguintes fases: (i) criação de um corpus para treinamento e validação da aprendizagem; (ii) implementação da abordagem para mineração de argumentos; e (iii) aplicação de técnicas de aprendizagem de máquina. Os resultados alcançados ficaram entre a média encontrada na literatura para trabalhos realizados na língua inglesa. O destaque foi a identificação da proposta de intervenção, com um valor-F de 0,744.*

1. Introdução

De acordo com Russel e Norvig (2004), a inteligência artificial (IA) é o estudo, planejamento e construção de agentes inteligentes. Pode ser classificada em diversos ramos, tais como: representação do conhecimento, planejamento, percepção e aprendizado. O processamento de linguagem natural (PLN) está ligado a cada um desses ramos: (i) representação do conhecimento é essencial para sistemas de PLN, pois fornecem a representação do conhecimento linguístico específico ou fornecem informações extralinguísticas; (ii) planejamento está associado as etapas para atingir o objetivo; (iii) percepção é quando ocorre, por exemplo, a propagação de informações sonoras, em que a mensagem é transmitida por ondas sonoras produzidas pelas cordas vocais; e (iv) aprendizado envolve a construção de uma representação, depois da

máquina examinar dados que podem ou não ter sido analisados anteriormente [Kurdi 2016].

O reconhecimento de voz, processamento sintático, semântico e do discurso são aplicações importantes do PLN. O foco deste trabalho é o entendimento do discurso, que, Segundo Barros e Robin (1996), é uma tarefa complexa para a máquina, por isso o seu processamento automático não é tão simples. O discurso compreende uma produção linguística composta por mais de uma frase. As redações são um tipo de produção textual, em que, o entendimento do discurso é realizado para o texto ser compreendido como um todo, para isso é organizado conforme uma estrutura bem definida.

A redação bem construída utiliza o conceito de textualidade. Val (2009), define textualidade como sendo o conjunto de características que fazem um texto ter uma construção, isto é, que não seja um amontoado de frase soltas sem conexões. Neste trabalho, as redações utilizadas estão escritas conforme modelo do ENEM, definido pelo Ministério da Educação do Brasil. Segundo INEP (2017), para uma boa redação é preciso compor uma tese, uma opinião a respeito do tema proposto. A tese deve ser apoiada em argumentos consistentes, estruturados com coerência e coesão. A construção de componentes argumentativos com coesão e coerência é imprescindível para alcançar uma boa redação. Por fim, é necessário elaborar uma proposta de intervenção social para o problema apresentado no desenvolvimento do texto.

A redação avalia o autor nas habilidades de leitura, interpretação e produção textual. A utilização de redações no processo de ensino-aprendizagem possibilita avaliar processos cognitivos mais elevados, além de treinar e aprimorar a capacidade de interpretação textual, melhorando a qualidade da aprendizagem. Entretanto, o processo de correção da redação demanda tempo, custo e a subjetividade dos avaliadores. Devido a isso, propostas de ferramentas e sistemas para a correção automática de redações buscam automatizar o processo de correção, de modo que os problemas da correção manual possam ser minimizados [Dikli 2006].

Conforme Stede e Schneider (2018), a mineração de argumentos é uma nova área para aplicação de PLN, no qual, o objetivo da mineração é o entendimento do discurso por meio da extração automática da estrutura argumentativa de um texto. Na literatura são encontrados alguns trabalhos, no entanto, todos os trabalhos apresentam soluções para a língua inglesa. O problema é que além de não haver abordagens para o português brasileiro, também há falta de um corpus específico para realizar o aprendizado de máquina desta área.

Dentro deste contexto este trabalho apresenta uma pesquisa que teve por objetivo selecionar e aplicar uma abordagem ponta a ponta, para a mineração de argumentos em redações dissertativas-argumentativas do modelo ENEM. A abordagem ponta a ponta se caracteriza por realizar todas as tarefas designadas para a mineração de argumentos e para isso, utiliza as técnicas de processamento de linguagem natural e aprendizado de máquina. Um conjunto de redações precisa ser anotado com a estrutura argumentativa, esse conjunto é necessário para ocorrer o aprendizado de máquina, esse conjunto é conhecido como corpus. Para validar a abordagem um corpus de 50 redações do português brasileiro foi anotado com sua estrutura argumentativa.

Este trabalho visa contribuir com a área de processamento de linguagem natural, disponibilizando um corpus específico e uma abordagem implementada para o português. O resultado desta pesquisa pode auxiliar os processos de correção automática

da redação, fornecendo recursos dos componentes argumentativos da redação. O processo de ensino-aprendizagem pode ser favorecido, visto que esta pesquisa pretende fornecer informações mais detalhadas sobre a redação, auxiliando professores em suas correções.

2. Trabalhos relacionados

Na literatura são encontrados alguns trabalhos para a mineração de argumentos [Desilia et al. 2017; Gema et al. 2017; Nguyen e Litman 2016; Persing e Ng 2016; Stab e Gurevych 2017, 2014; Tahir 2017]. Entretanto, todos os trabalhos encontrados apresentam soluções apenas para a língua inglesa. O português brasileiro carece de corpora específicos de redações anotadas com a estrutura da argumentação no nível do discurso.

Para a abordagem ponta a ponta, foram selecionados dois trabalhos: (i) End-to-End Argumentation Mining in Student Essays [Persing e Ng 2016]; e (ii) Parsing Argumentation Structures in Persuasive Essays [Stab e Gurevych 2017]. Ambas utilizaram a métrica Valor-F para aferir o resultado dos classificadores. A métrica Valor-F utiliza a média harmônica entre os indicadores de precisão e recall, e é tanto melhor quanto mais próxima de 1.

Os autores Ng e Persing (2016) apresentaram os primeiros resultados sobre argumentação ponta a ponta em redações persuasivas usando uma abordagem de pipeline. Nesta abordagem, a média das tarefas apresentou Valor-F de 0,388. Stab e Gurevych (2017) apresentam uma abordagem que identifica os limites dos componentes, realiza a classificação, e relaciona os componentes. Esta abordagem apresentou Valor-F de 0,762 na média das tarefas.

3. Fundamentação teórica

Este capítulo apresenta a fundamentação teórica dos temas abordados trabalhos. Os conceitos utilizados para auxiliar em sua construção são: (i) mineração de argumentos; (ii) aprendizagem de máquina; e (iii) programação linear.

3.1. Mineração de Argumentos

Conforme Stede e Schneider (2018), a mineração de argumentação é uma aplicação do Processamento de Linguagem Natural (PLN), o qual é composto por um conjunto de tarefas, que são adaptadas conforme o tipo dos textos a serem analisados. No caso das redações, ocorre a extração da tese, dos argumentos e da proposta de intervenção, porém em um artigo científico, é necessário por exemplo, realizar a extração de objetivos, justificativas, trabalhos relacionados e resultados. Portanto, a mineração de argumentação não é um processo unificado e depende exclusivamente da estrutura argumentativa de um determinado texto.

A mineração de argumentos é composta por sete tarefas: (i) identificar o texto argumentativo; (ii) segmentar o texto em unidades discursivas argumentativas; (iii) identificar a afirmação principal (major claim); (iv) identificar o papel/função dos componentes argumentativos; (v) identificar as relações entre os componentes argumentativos; (vi) construir a representação estrutural geral; e (vii) identificar o tipo e a qualidade da argumentação [Stede e Schneider 2018]. A Figura 1 apresenta um diagrama de exemplo, com as unidades argumentativas já separadas.

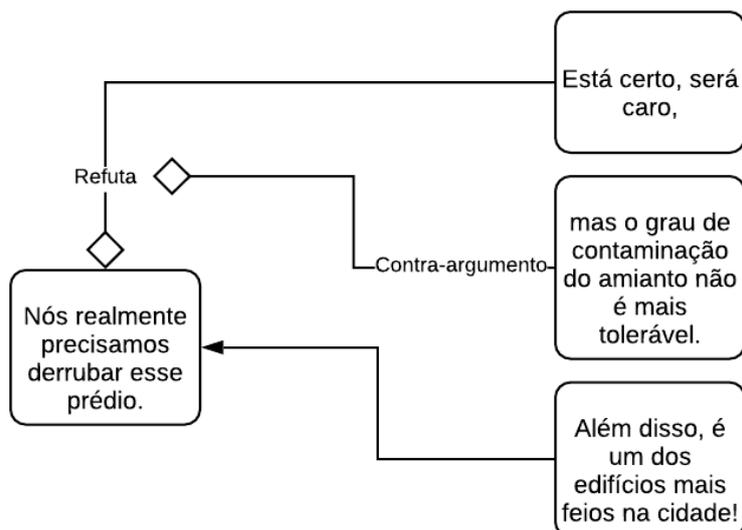


Figura 1. Exemplo de uma estrutura argumentativa

Conforme Figura 1 pode-se observar que o fragmento “Nós realmente precisamos derrubar esse prédio” indica uma afirmação do autor sobre determinado assunto, a proposição “será caro” apresenta uma objeção, porém ocorre um contra-argumento no fragmento, “mas o grau de contaminação não é mais tolerável”. Na proposição “Além disso, é um dos edifícios mais feios na cidade” é adicionado suporte a afirmação principal.

Segundo Walton (2009), na literatura existem diferenças na teoria da argumentação sobre como definir um argumento. A definição mínima diz que, um argumento é um conjunto de proposições, composto de três partes: (i) uma conclusão; (ii) um conjunto de premissas; e (iii) uma inferência das premissas para a conclusão. Além disso, um determinado argumento pode ser suportado ou refutado por outros argumentos.

No Exame Nacional do Ensino Médio (ENEM), a redação exige a produção de um texto do tipo dissertativo-argumentativo, sobre um tema previamente informado de ordem social, científica, cultural ou política. Na escrita da redação, é necessário a apresentação de uma tese – uma opinião a respeito do tema previamente proposto, que por sua vez, deve ser apoiada em argumentos e por fim, para o problema apresentado é necessário elaborar uma proposta de intervenção social, na qual respeite os direitos humanos [INEP 2017]. O ENEM é o principal modelo aplicado no Brasil para avaliar as habilidades de escrita dos alunos do ensino médio.

3.2. Aprendizado de Máquina

O aprendizado de máquina (AM) é capacidade de agentes adquirirem conhecimento por meio da observação de sua interação com o ambiente, e é essencial para um comportamento inteligente. Normalmente, em AM, o aprendizado acontece com a experiência, para isso, empregam a aprendizagem indutiva, na qual a partir de um conjunto de exemplos são geradas conclusões genéricas [Luger 2013; Russell e Norvig 2004].

O aprendizado de máquina é tipicamente classificado em três casos: (i) aprendizagem supervisionada; (ii) aprendizagem não-supervisionada; e (iii)

aprendizagem por reforço. Na aprendizagem supervisionada o aprendizado ocorre por meio da observação de exemplos rotulados, ou seja, a partir de dados de entrada e saída. Na aprendizagem não-supervisionada o algoritmo não recebe valores específicos de saída, apenas de entrada, com isso, a aprendizagem deve ser capaz de identificar padrões e extrair conhecimento. Por fim, na aprendizagem por reforço o algoritmo aprende a partir do reforço.

A mineração de argumentos em redações compreende algumas tarefas que necessitam do aprendizado de máquina. No contexto desta pesquisa, destaca-se o aprendizado supervisionado, com a tarefa de classificação. A identificação de elementos argumentativos e a relação dos argumentos, pode ser visto como uma classificação binária, na qual, uma sentença pode ser argumentativa ou não, e dois argumentos podem ou não, ter relação. Na tarefa da identificação do tipo de argumento, é um problema multiclasse, em que, o algoritmo deve separar os dados em quatro classes: (i) tese; (ii) argumento; (iii) uma proposta de intervenção; e (iv) não argumentativo.

O aprendizado supervisionado com a tarefa de classificação possui diversos algoritmos para realizar no treinamento. Conforme Stab e Gurevych (2017) o *Support Vector Machine* (SVM), apresentou os melhores resultados para as tarefas relacionadas a mineração de argumentos em redações.

3.2.1. Support Vector Machine

Support Vector Machine (SVM) é uma técnica de aprendizado de máquina supervisionado, proposta por Cortes e Vapnik (1995), e pode ser utilizada para problemas de classificação ou regressão linear. O objetivo do SVM é a construção de um hiperplano para separar os exemplos entre positivos e negativos, porém a ideia principal é que a separação seja maximizada.

A teoria do aprendizado estatístico é utilizada no SVM para se aproximar do classificador ideal, ou seja, dado um conjunto de treinamento, a técnica de aprendizagem cria várias hipóteses de modelos de solução e a teoria de aprendizagem estabelece condições matemáticas para a escolha de determinado classificador [Lorena e Carvalho 2007]. Portanto, apesar do SVM não incorporar conhecimento do domínio do problema, é possível alcançar bom desempenho de generalização em problemas de classificação.

3.3. Programação Linear

A Programação Linear (LP, do inglês: *Linear Programming*) foi inicialmente concebida por George em 1947, enquanto desenvolvia uma ferramenta de planejamento mecanizada. O objetivo da programação linear está relacionado com a minimização ou maximização de uma função linear, ou seja, otimizar a função enquanto satisfaz as restrições lineares de igualdade e/ou desigualdade [Bazaraa et al. 2010]. Quando as variáveis do problema são inteiras, a LP também é conhecida por Programação Linear Inteira (ILP, do inglês: *Integer Linear Programming*).

Conforme Stede e Schneider (2018), na mineração de argumentos, a programação linear inteira é utilizada para otimizar a estrutura global dos argumentos, isto é, definir uma função objetivo adequada a ser maximizada para otimizar o resultado dos classificadores de identificação do argumento e das relações. Normalmente, nas

redações as restrições são do tipo árvore, ou seja, ciclos são descartados, todos os nós possuem uma ou nenhuma saída, e o nó raiz não possui saída.

4. Metodologia e desenvolvimento

A solução proposta neste trabalho apresenta uma abordagem ponta a ponta para a mineração de argumentos em redações do português brasileiro. A definição da extração de características e da modelagem da ILP foi desenvolvida conforme proposto por Stab e Gurevych (2017), com adaptações para o português.

A metodologia compreende as seguintes tarefas: (i) preparar um corpus de domínio específico; (ii) separar e classificar as partes argumentativas das não argumentativas; (iii) identificar as relações entre os componentes; (iv) otimizar os resultados com o uso da programação linear inteira.

A Figura 2 apresenta a estrutura da abordagem para mineração de argumentos. A redação é o arquivo de entrada para o processamento do texto. A redação é segmentada em sentenças, e a seguir, de cada sentença é extraído um conjunto de características. As características alimentam os dois classificadores, que realizam a classificação dos componentes e a identificação das relações. O classificador de relação recebe ainda como entrada apenas os componentes que foram classificados como argumentativos. Por fim, a programação linear é aplicada na tentativa de otimizar os resultados. Para isto, são criados um conjunto de restrições na formação da estrutura argumentativa, como, por exemplo, um componente não pode ter duas saídas.

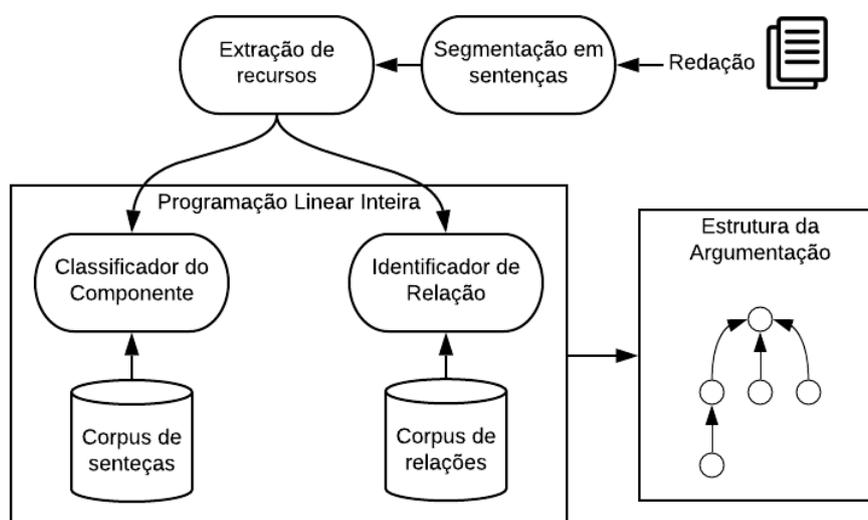


Figura 2: Abordagem para mineração de argumentos

Para o treinamento e validação da abordagem, foi criado um corpus com 50 redações, extraídas do portal Brasil Escola. As redações foram anotadas com a estrutura argumentativa por especialistas em linguística do Programa de Pós-Graduação em Educação (PPGE) da UNIVALI. A tarefa de anotação consistiu em dois especialistas anotarem de forma individual as redações, e um terceiro especialista julgar as divergências encontradas entre os dois primeiros especialistas. O índice de concordância para a anotação dos componentes foi medido pelo coeficiente de Krippendorff [Krippendorff 2004]. A Tabela 1 mostra a correlação entre os anotadores em cada componente da argumentação.

Tabela 1. Concordância entre avaliadores para cada componente.

Componente	α
Tese	0,87
Argumento	0,91
Proposta de Intervenção	0,95
Relações	0,72

A anotação dos componentes teve a média geral de correlação de 0,92. O coeficiente é medido de 0 a 1, no qual 0 indicada nenhuma correlação entre os avaliadores e 1 indica correlação total entre os avaliadores. Dentre os componentes anotados é possível observar a alta concordância entre os anotadores nos componentes do argumento e proposta de intervenção. As anotações das relações apresentaram o menor índice de concordância entre os anotadores, no qual, teve como média geral 0,72 de concordância. Devido a isso, o processo de análise das inconsistências das relações precisou de mais tempo e atenção para avaliar as discrepâncias entre os anotadores por uma terceira pessoa. A Tabela 2 apresenta a estatística do corpus final.

Tabela 2. Estatística do corpus final.

	Total	Média por redação
Sentenças	659	13,18
Palavras (tokens)	20659	413,18
Teses	62	1,24
Argumentos	222	4,44
Propostas de Intervenção	100	2
Não argumentativa	275	5,5
Relações	407	8,14

A tarefa de segmentação foi implementada com a biblioteca do NLTK. Para a extração de características, foram utilizadas as bibliotecas do NLTK, CoGrOO e spaCy. Para implementar os dois classificadores, serão utilizados o SVM. A programação linear foi implementada com o algoritmo PuLP.

5. Resultados

Nessa seção os resultados do trabalho são apresentados e discutidos, com isso, permite avaliar a contribuição do trabalho. Na Seção 5.1 é apresentado os resultados individuais dos dois classificadores para cada *kernel*. A Seção 5.2 apresenta os resultados após a utilização da ILP para otimização da estrutura argumentativa. Na Seção 5.3 é apresentada a análise detalhada dos resultados por classe e a comparação da abordagem com a literatura

5.1 Resultados dos Classificadores

Esta seção apresenta os resultados individuais obtidos para o classificador dos componentes e o classificador das relações. As métricas utilizadas para os testes foram a acurácia, precisão macro, recall macro e valor-F macro. A escolha das métricas foi com base no trabalho de Stab e Gurevych (2017). Os parâmetros de cada classificador foram utilizados com os valores padrão. A Tabela 3 apresenta os resultados de cada métrica para cada classificador e o kernel utilizado.

Tabela 3. Resultados dos classificadores (5-fold)

<i>Kernel</i>	Componente			Relação		
	Valor-F	Precisão	Recall	Valor-F	Precisão	Recall
Linear	0,587	0,610	0,605	0,541	0,543	0,542
Polinomial	0,614	0,645	0,620	0,597	0,588	0,615
RBF	0,637	0,668	0,643	0,494	0,526	0,503
Sigmoide	0,142	0,100	0,250	0,487	0,476	0,500

O kernel RBF apresentou o melhor resultado na classificação dos componentes das redações, com um valor-F de 0,637. Para a classificação das relações o kernel polinomial apresentou o melhor valor-F de 0,597. A Tabela 4 apresenta os resultados com o experimento de 10-fold na validação cruzada e com os mesmos parâmetros mencionados anteriormente.

Tabela 4. Resultados dos classificadores (10-fold)

<i>Kernel</i>	Componente			Relação		
	Valor-F	Precisão	Recall	Valor-F	Precisão	Recall
Linear	0,581	0,622	0,596	0,535	0,539	0,540
Polinomial	0,620	0,670	0,624	0,598	0,589	0,623
RBF	0,655	0,685	0,665	0,487	0,475	0,500
Sigmoide	0,141	0,100	0,250	0,487	0,475	0,500

O kernel RBF apresentou o melhor valor-F na classificação dos componentes, enquanto o kernel polinomial apresentou o melhor resultado na tarefa de classificação das relações. Com 10-fold os resultados da classificação do componente ficaram ligeiramente melhor, porém, na classificação das relações não ocorreu o mesmo e o resultado ficou semelhante ao 5-fold.

5.2 Resultados da Programação Linear Inteira

Esta seção apresenta os resultados individuais obtidos para o classificador dos componentes e o classificador das relações após a aplicação da ILP para otimização. Na ILP os rótulos das sentenças foram alterados conforme solução encontrada, em que, a sentença raiz foi alterada para “tese” e as sentenças que estavam no último parágrafo e tinham uma ligação direta com a tese foram alteradas para “proposta de intervenção”. O uso da ILP para as relações foi pelo fato do classificador individual em alguns casos classificar um número maior de ligações, e em outros casos, não classificava o número de ligações reais das redações.

Tabela 5. Resultados com programação linear inteira (5-fold)

<i>Kernel</i>	Componente			Relação		
	Valor-F	Precisão	Recall	Valor-F	Precisão	Recall
Linear	0,571	0,591	0,595	0,540	0,539	0,543
Polinomial	0,605	0,624	0,622	0,571	0,576	0,569
RBF	0,611	0,620	0,636	0,537	0,535	0,541
Sigmoide	0,150	0,112	0,245	0,496	0,498	0,499

Os resultados da ILP da Tabela 5 são comparados com os resultados apresentados na Tabela 3. Na classificação dos componentes é possível observar que não houve melhoras significativas com a ILP. A classificação das relações com o kernel

RBF teve uma melhora com a implementação da ILP, no qual, o valor-F de 0,494 passou para 0,537. Porém não ultrapassou o melhor resultado de 0,597 do classificador individual. A Tabela 6 apresenta os resultados da programação linear inteira para os testes de validação cruzada de 10-fold.

Tabela 6. Resultados com programação linear inteira (10-fold)

<i>Kernel</i>	Componente			Relação		
	Valor-F	Precisão	Recall	Valor-F	Precisão	Recall
Linear	0,567	0,597	0,589	0,555	0,556	0,557
Polinomial	0,598	0,625	0,615	0,580	0,585	0,578
RBF	0,569	0,577	0,614	0,533	0,529	0,538
Sigmoide	0,150	0,112	0,245	0,496	0,497	0,498

Na comparação dos resultados da Tabela 6 com a Tabela 4 é possível observar que na classificação dos componentes e das relações não houve melhoras significativas após o processamento da ILP.

5.2 Análise dos Resultados

Esta seção apresenta a análise dos resultados com mais detalhes. Para isso, os resultados dos classificadores foram separados por classes. O objetivo dessa análise é verificar se existe alguma classe com dificuldade de aprendizado. A Tabela 7 exibe os resultados (valor-F macro) para as classes do classificador de componente com kernel polinomial, que apresentou a melhor média das tarefas. A proposta de intervenção apresentou o melhor resultado, sendo a classe com a maior facilidade de classificação. Entretanto, na identificação das teses o algoritmo não conseguiu identificar corretamente a maior parte presente na redação.

Tabela 7. Valor-F por classe obtidos pelo *kernel* polinomial para a classificação dos componentes

Abordagem	Não Argumento	Tese	Argumento	Proposta de Intervenção
Sem ILP – 5-fold	0,551	0,491	0,671	0,744
Com ILP – 5-fold	0,551	0,489	0,637	0,744
Sem ILP – 10-fold	0,541	0,518	0,681	0,741
Com ILP – 10-fold	0,541	0,473	0,635	0,741

O motivo para essa ocorrência pode ser o fato da baixa frequência de teses nas redações, enquanto nas redações normalmente encontra-se uma tese, essa mesma redação apresenta pelo menos duas propostas de intervenção. A identificação de sentenças não argumentativas ficou abaixo das sentenças argumentativas, o que mostra a facilidade em classificar as sentenças argumentativas.

Na Tabela 7 é observado uma queda nas classes tese e argumentos, no qual os valores acabaram caindo após a aplicação da ILP. Esse fato foi ocasionado pela troca da classe do componente quando ele foi determinado como raiz da árvore pelo ILP, ou seja, se inicialmente o componente foi classificado corretamente como argumento e na ILP foi considerado como raiz da estrutura, sua classe foi sobreposta pela classe tese. Portanto, a substituição da classe ocasiona a queda nos resultados das duas classes.

A Tabela 8 exibe os resultados (valor-F macro) para as classes do classificador de relações com kernel polinomial, que apresentou a melhor média das tarefas. Os resultados mostram a tendência do classificador em classificar as amostras como “Sem Relação”, um dos motivos está diretamente relacionado ao desbalanceamento das classes, outro motivo pode estar relacionado aos recursos extraídos para representar o que seria uma relação entre duas sentenças.

Tabela 8. Valor-F por classe obtidos pelo *kernel* polinomial para a classificação das relações

Abordagem	Sem Relação	Com relação
Sem ILP – 5-fold	0,954	0,241
Com ILP – 5-fold	0,960	0,182
Sem ILP – 10-fold	0,952	0,243
Com ILP – 10-fold	0,960	0,199

A Tabela 9 apresenta a compilação dos resultados apresentados por outros autores juntamente com nossa abordagem. No resultado médio das tarefas nossa abordagem ficou entre os dois trabalhos encontrados na literatura, com uma média de 0,609 sem ILP e de 0,589 com ILP. Entretanto, uma análise detalhada, com comparação e explanação dos resultados não pode ser realizada pelo fato das abordagens terem aspectos diferentes quanto ao tamanho e idioma do corpus e implementação da abordagem.

Tabela 9. Comparação das abordagens

Abordagem	Componente	Relação	Média
Persing e Ng (2016) com ILP	0,572	0,204	0,388
Stab e Gurevych (2017)	0,794	0,717	0,755
Stab e Gurevych (2017) com ILP	0,826	0,751	0,788
Nossa abordagem	0,620	0,598	0,609
Nossa abordagem com ILP	0,598	0,580	0,589

6. Conclusões

Este trabalho teve como objetivo principal realizar a mineração de argumentos em redações dissertativas-argumentativas do português brasileiro. Visando atender esse objetivo, foram definidos quatro objetivos específicos: (i) criar um corpus público de redações anotadas com a estrutura argumentativa; (ii) selecionar e aplicar uma abordagem para a mineração de argumentos em redações; e (iii) analisar os resultados da abordagem aplicada e validada com o corpus de redação anotado.

O processo de construção do corpus foi uma etapa difícil e que envolveu várias pessoas, principalmente pessoas da área da linguística, no qual foram as responsáveis por anotar as redações de forma fidedigna para que um corpus de 50 redações fosse criado. Os componentes do argumento (tese, argumento e proposta de intervenção) e as relações entre os componentes foram anotados nas redações e as divergências entre os anotadores foram resolvidas.

Os resultados da abordagem implementada ficaram na média dos trabalhos encontrados na literatura. Nossa abordagem não teve resultado superior ou similar ao trabalho utilizado como referência (Stab & Gurevych, 2017), um dos motivos é a diferença do número de redação entre os trabalhos. Ademais, se a análise for realizada

de forma individual para cada classes, as classes de “Proposta de intervenção” e “Sem Relação” conseguiram um bom desempenho, com valor-F superior a 0,7.

A ILP teve como objetivo otimizar a estrutura argumentativa da redação, de forma a montar uma estrutura válida. Os resultados com a implementação da ILP sugerem que, apesar da construção de estruturas válidas, para a redações em português brasileiro a otimização com a implementação da ILP não provoca o efeito alcançando para a língua inglesa.

Portanto, as principais contribuições deste trabalho foram a criação de um corpus de redações anotadas com a estrutura argumentativa¹ e o desenvolvimento de uma abordagem para a mineração de argumentos ponta a ponta em redações do português brasileiro².

Em trabalhos futuros, pretende-se aumentar a quantidade de anotadores e redações anotadas para que o algoritmo tenha mais exemplo para realizar o aprendizado e a validação dos resultados, além disso, analisar a programação linear inteira com o objetivo de melhorar a função objetivo e as restrições para o português brasileiro e realizar testes com outros algoritmos de classificação na tentativa de melhorar os resultados nas tarefas de classificação.

Referências

- Almeida Junior, C. R. C. de, Spalenza, M. A., & Oliveira, E. de. (2017). Proposta de um Sistema de Avaliação Automática de Redações do ENEM Utilizando Técnicas de Aprendizagem de Máquina e Processamento de Linguagem Natural. *Computer on the Beach - COTB*, 8.,.
- Barros, F. de A., & Robin, J. (1996). *Processamento de Linguagem Natural* (p. 61). Sociedade Brasileira de Computação.
- Bazaraa, M. S., Jarvis, J. J., & Sherali, H. D. (2010). *Linear Programming and Network Flows* (4th ed.). Wiley.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Desilia, Y., Utami, V. T., Arta, C., & Suhartono, D. (2017). An attempt to combine features in classifying argument components in persuasive essays. *Workshop on Computational Models of Natural Argument - CEUR*, 18., 2048(July), 71–75.
- Dikli, S. (2006). An Overview of Automated Scoring of Essays. *Journal of Technology Learning and Assessment*, 5(1).
- Gema, A. P., Winton, S., David, T., Suhartono, D., Shodiq, M., & Gazali, W. (2017). It Takes Two to Tango: Modification of Siamese Long Short Term Memory Network with Attention Mechanism in Recognizing Argumentative Relations in Persuasive Essay. *Procedia Computer Science*, 116, 449–459. <https://doi.org/10.1016/j.procs.2017.10.036>

¹ <https://gitlab.com/jonathannau/argument-annotated-essays>

² <https://gitlab.com/jonathannau/argument-mining-essay>

- Haykin, S. (2001). *Redes Neurais: Princípios e Prática* (2nd ed.). Bookman.
- INEP. (2017). *Redação no ENEM 2017: Cartilha do Participante*. http://download.inep.gov.br/educacao_basica/enem/guia_participante/2017/manual_de_redacao_do_enem_2017.pdf
- Krippendorff, K. (2004). Measuring the Reliability of Qualitative Text Analysis Data. *Quality & Quantity*, 38(6), 787–800. <https://doi.org/10.1007/s11135-004-8107-7>
- Kurdi, M. Z. (2016). *Natural Language Processing and Computational Linguistics 1: Speech, Morphology, Syntax* (1st ed.). ISTE - Wiley.
- Lima, F., Haendchen Filho, A., Prado, H. Do, & Ferneda, E. (2018). Automatic Evaluation of Textual Cohesion in Essays. *International Conference on Computational Linguistics and Intelligent Text - CICLing, 19.*,.
- Lorena, ana carolina, & Carvalho, A. c. p. l. f. (2007). Uma Introdução às Support Vector Machines. *Revista de Informática Teórica e Aplicada*, 14(2), 43–67.
- Luger, G. F. (2013). *Inteligência Artificial* (6th ed.). Pearson Education do Brasil.
- Nguyen, H., & Litman, D. (2016). Improving Argument Mining in Student Essays by Learning and Exploiting Argument Indicators versus Essay Topics. *International Florida Artificial Intelligence Research Society Conference - FLAIRS, 29.*, 485–490.
- Passero, G. (2018). Detecção De Fuga Ao Tema Em Redações Na Língua Portuguesa. In *Dissertação de Mestrado*. Universidade do Vale do Itajaí.
- Persing, I., & Ng, V. (2016). End-to-End Argumentation Mining in Student Essays. *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - NAACL-HLT*, 1384–1394. <https://doi.org/10.18653/v1/n16-1164>
- Russell, S., & Norvig, P. (2004). *Inteligência Artificial* (2nd ed.). Elsevier.
- Stab, C., & Gurevych, I. (2017). Parsing Argumentation Structures in Persuasive Essays. *Computational Linguistics*, 43(3), 619–659. https://doi.org/10.1162/COLI_a_00295
- Stab, C., & Gurevych, I. (2014). Identifying Argumentative Discourse Structures in Persuasive Essays. *Conference on Empirical Methods in Natural Language Processing - EMNLP*, 46–56. <https://doi.org/10.3115/v1/d14-1006>
- Stede, M., & Schneider, J. (2018). Argumentation Mining. *Synthesis Lectures on Human Language Technologies*, 11(2), 1–191. <https://doi.org/10.2200/S00883ED1V01Y201811HLT040>
- Tahir, S. B. ud D. (2017). Comparative Analysis of Supervised Learning Approaches for Argument Identification. *IEEE International Multi-Topic Conference - INMIC, 20.*, 20, 1–5.
- Val, M. da G. C. (2009). *Redação e Textualidade* (3rd ed.). Martins Fontes.
- Walton, D. (2009). Argumentation in Artificial Intelligence. In G. Simari & I. Rahwan (Eds.), *Argumentation in Artificial Intelligence*. Springer US. <https://doi.org/10.1007/978-0-387-98197-0>