

Association for Information Systems

AIS Electronic Library (AISeL)

ISLA 2023 Proceedings

Latin America (ISLA)

Fall 8-7-2023

Análise de Sentimento da Poesia de João Cabral de Melo Neto

Patrick Fernandes Rezende Ribeiro

José Marcelo Almeida Prado Cestari Dr.

Follow this and additional works at: <https://aisel.aisnet.org/isla2023>

This material is brought to you by the Latin America (ISLA) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ISLA 2023 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.



Análise de Sentimento da Poesia de João Cabral de Melo Neto

Artigo Completo

**Patrick Fernandes Rezende
Ribeiro**

Aluno de graduação na Universidade
Federal do Paraná
patrick.ribeiro@ufpr.br

**José Marcelo Almeida Prado
Cestari,**

Professor da Universidade Federal do
Paraná
cestari@ufpr.br

Abstract

The purpose of this paper is to analyze if implementing the hermeneutically created negation scope can improve the performance of Sentiment Analysis tools when applied to João Cabral de Melo Neto's poems. To achieve this goal, the paper proposes a hybrid approach that combines lexical analysis based on two dictionaries with a scope of negation created from the author's own poems. Furthermore, some algorithms were used to confirm the negation in the context of Cabralina's poetry. The experiment showed a significant improvement in performance scores when applying one of the dictionaries associated with the scope of negation. This underscores the significance of computational linguists' endeavors to create dictionaries for literary texts and reinforces the need for collaboration between areas of knowledge.

Keywords

Data Analyse, Sentiment Analyse, Lexical Analysis, João Cabral de Melo Neto.

Resumo

Este artigo tem como objetivo analisar se a implementação do escopo de negação criado hermeneuticamente trará um melhor desempenho as ferramentas de Análise de Sentimento quando aplicadas as poesias de João Cabral de Melo Neto. Para isso, propôs-se uma abordagem híbrida que combina a análise lexical com base em dois dicionários em conjunto a um escopo de negação criado a partir das próprias poesias do autor. Além disso, houve a aplicação de alguns algoritmos para validar a negação no contexto da poesia cabralina. Os resultados mostram que o experimento apresentou uma melhoria significativa nos escores de desempenho na aplicação de um dos dicionários associados ao escopo de negação. Isto, destaca a importância de haver um esforço dos linguistas computacionais em desenvolver dicionários específicos para textos literários, assim como reforça a necessidade de colaboração entre áreas de conhecimento.

Palavras-chave

Análise de dados, Análise de Sentimento, Análise Lexical, João Cabral de Melo Neto.

Introdução

João Cabral de Melo Neto é considerado pela crítica como um dos “maiores poetas brasileiros do século XX” (Secchin 2020). O poeta nasceu em 9 de janeiro de 1920, na cidade de Recife, em Pernambuco, e faleceu em 9 de outubro de 1999, no Rio de Janeiro, aos 79 anos. Foi diplomata e escritor, e membro da

Academia Brasileira de Letras. A poesia de João Cabral de Melo Neto tem sido visitada como uma importante fonte de informação em diferentes perspectivas de estudo para a teoria e crítica literária – desde linguísticos, sociais e políticos –, no entanto, os dados de sua poesia apresentam dificuldades para a análise devido à grande variedade e complexidade linguística.

Neste contexto, uma abordagem que traz um potencial significativo de aplicabilidade, é a Análise de Sentimento (Wilson et al. 2006). A aplicação da técnica de análise de sentimentos visa definir métodos capazes de processar informações subjetivas de textos de forma automática determinando sentimentos e emoções expressos em linguagem natural (Duarte 2013).

A Análise de Sentimento da obra de João Cabral de Melo Neto pode verificar o sentimento de cada poesia e de cada conjunto de poesia de forma automática. De forma, que se torna possível verificar se os poemas ou conjunto de poemas expressam um sentimento negativo, neutro ou positivo. Para isso, é imprescindível verificar se há algum método de Análise de Sentimento que consiga trazer essa informação sobre as poesias de forma confiável.

Desse modo, este artigo tem como objeto analisar o sentimento das poesias de João Cabral de Melo Neto mediante uma abordagem híbrida, ou seja, usando tanto a abordagem lexical a partir da aplicação de recursos léxicos de dois dicionários diferentes e a criação de um escopo de negação com base nas poesias, quanto a abordagem de aprendizagem de máquina aplicando sete modelos de algoritmo diferentes. A finalidade desta análise é classificar as poesias em positivo, neutro ou negativo para que essa informação seja utilizada para facilitar pesquisas e estudos literários sobre a poesia de João Cabral de Melo Neto. A hipótese que motivou este artigo é: a criação de um escopo de negação hermeneuticamente criado das poesias de João Cabral de Melo Neto apresentará um melhor desempenho na resultados da análise de sentimento.

Análise da Poesia de João Cabral de Melo Neto

O conjunto da poesia de João Cabral de Melo Neto reside em 20 livros publicados e mais 4 conjuntos de poesias classificadas como póstumas, dispersas e inéditas. Deste conjunto se destacam algumas singularidades linguísticas do autor, como: valorização dos encontros consonantais, preferência por vocabulário controlado pela univocidade, predominância de substantivos concretos, uso de métricas com oito, nove e onze sílabas, obsessão por estrofes configuradas em quadra e uso de linguagem metafórica. A obra de João Cabral de Melo Neto (Secchin 2020) está quase que isolada dentro do panorama literário brasileiro, pois não há uma linguagem tão ostensiva na qual ela possa ser inscrita.

Estas características fazem da poesia de João Cabral de Melo Neto um grande desafio linguístico. Por isso, vários críticos e teóricos literários analisaram a obra cabralina ao longo dos anos, contribuindo para uma melhor compreensão de seus poemas e seu lugar na literatura brasileira. Alguns dos críticos e teóricos mais proeminentes que se dedicaram ao estudo de sua obra são: Antonio Candido que em seu livro “Formação da Literatura Brasileira”, publicado em 1969, destaca a precisão formal e a contenção emocional presente nos poemas; Roberto Schwarz que em seu ensaio “As Ideias Fora do Lugar”, publicado em 1973,

analisa a obra do poeta à luz do contexto social e político brasileiro, explorando temas como a relação entre arte e engajamento político; Antonio Carlos Secchin em seu livro “João Cabral: A Poesia do Menos e outros ensaios cabralinos”, publicado em 1999, explora a construção formal e as temáticas recorrentes na poesia cabralina, como também estuda a valorização do trabalho e a preocupação com a precisão linguística; e Benedito Nunes em seu ensaio “João Cabral e o Poema Dialético”, publicado em 1974, explora a relação entre a obra de Cabral e a poesia dialética, apresentando um poeta preocupado com a investigação do real e a busca de uma compreensão mais profunda da realidade.

Além dessas análises, a poesia de João Cabral de Melo Neto já foi objeto de análise de sentimento de alguns estudos literários. Entre eles temos a análise de Benedito Nunes, em seu livro “João Cabral de Melo Neto”, publicado em 1971, que em um dos capítulos cujo título é “poética negativa” estuda obra “Psicologia da Composição com a Fábula de Anfion e Antiode”(1974) apontando a crise interna do poeta pernambucano como processo criador de sua poesia. Para Benedito Nunes (1971) a obra é negativa não apenas por não se expressar, mas por substituir e controlar o léxico, por exemplo, a substituição de ‘lembança’ por ‘atenção’. Outra análise desse poema, está presente na tese de João Alexandre Barbosa, sob o título “A imitação da forma”, defendida em 1975. Nela, João Alexandre Barbosa (1975) defende que o sentido da negatividade do poema reside na articulação que o poeta faz entre significado e significação na escolha lexicais na composição do poema. Em 1983, como resultado de sua tese, Marta Peixoto publica o livro “Poesia com coisas”, em que defende que há um maior sentimento negativo na poesia cabralina quando analisa o poema “Imagens em Castela”, de “Paisagens com Figuras” de 1956. Em sua análise do poema, a autora demonstra que a frequência e as escolhas lexicais do poeta exemplificam uma poética negativa, por exemplo, no uso dos léxicos ‘vazio’, ‘sem’, ‘escasso’, ‘nada’, ‘morte’ e em suas derivações semânticas. Antônio Carlos Secchin, no livro “João Cabral: a poesia do menos”, de 1985, afirma que as palavras negativas são recursos indispensáveis ao estilo de João Cabral de Melo Neto. Além dessas críticas já consagradas, há a tese defendida em 2018, por Felipe Oliveira de Paula intitulada “A poética negativa de João Cabral de Melo Neto”, em que analisa, exaustivamente, o livro “Os três mal-amados”, de 1943 e o poema Fábula de Anfion, de (1974).

É importante ressaltar que não há nenhum estudo literário de análise de sentimento de todo o conjunto de poesias e obras de João Cabral de Melo Neto. Mesmo os estudos acadêmicos de maior esforço restringem-se a algumas obras e poemas mais lidos. Esta tarefa se realizada manualmente demandaria muito tempo, não sendo viável sua realização, por exemplo, no contexto de uma dissertação ou tese.

O Método de Análise de Sentimento

A Mineração de Textos tem como um dos objetivos extrair conhecimento de um grande volume de dados e, com isso, construir conexões e padrões entre eles fundamentando-se em métricas como temática e frequência dos termos encontrados. O processo de Mineração de Textos (Evangelista et al. 2014) envolve três etapas: o pré-processamento, que está relacionado a limpeza dos dados; a análise dos dados, que se utiliza de alguma técnica computacional para extrair conhecimento; e a avaliação, que verifica os resultados.



Aliado a Mineração de Textos temos o Processamento de Linguagem Natural, utilizado para manipular, processar e compreender a linguagem humana, escrita ou oral. O principal desafio dessa área é ultrapassar a polissemia presente na linguagem humana.

A Análise de Sentimento é uma subárea de investigação do Processamento de Linguagem Natural; ela tem como um de seus objetivos a construção de métodos capazes de extrair informações subjetivas, de modo automatizado, de textos em linguagem natural, de maneira que o seu resultado possa servir de base para conclusões e outras análises. Esta aplicação (Souza et al. 2016), vai além da análise de postagens em redes sociais, ou para saber o sentimento de usuários de determinado produto, ela pode ser aplicada em pesquisas das mais diferentes áreas do conhecimento, como as Ciências Sociais, Direito, Psicologia e Linguística.

Essa área de estudo trata da análise de informações semânticas e sentimento de palavras ou expressões em um texto. Termos isolados, expressões ou todo o texto podem ser classificados como negativos, positivos ou neutros, quando não é possível determinar objetivamente se o sentimento é positivo ou negativo sem um contexto. A Análise de Sentimento busca classificar textos de acordo com seu sentimento, mesmo quando não há uma expressão de sentimento explícita.

A Análise de Sentimento (Funk and Maynard 2011) pode ser dividida em pelo menos duas abordagens: lexical e aprendizagem de máquina. A abordagem lexical recorre ao uso léxicos em um contexto semântico, reunidos em dicionários ou termos pré-compilados com valores atribuídos aos termos segundo seu grau sentimento. Alguns exemplos de léxicos em língua portuguesa são: OpLexicon, SentiLex, Brazilian Portuguese Linguistic Inquiry and Word Count (LIWC), LeIA e Onto.PT. A abordagem de aprendizado de máquina, normalmente, recorre à criação de modelo supervisionados, que utiliza um conjunto de recursos como, por exemplo, os dicionários, treinamento com a aplicação de algoritmos, como: Naïve Bayes, Máquina de Vetores de Suporte (SVM) e Redes Neurais Artificiais, árvores de decisão e outros modelos baseados em regras.

A Análise de Sentimento nos estudos literários foram objeto de análise de Kim e Klinger (2018), em “A Survey on Sentiment and Emotion Analysis for Computational Literary Studies”, um dos primeiros e extensos estudos sobre o tema para estudos literários computacionais, em que se distinguem cinco áreas principais de aplicação da Análise de Sentimento nos textos literários: classificação de textos literários em termos de suas emoções; classificação de gênero e tipologia; modelagem de sentimentos e emoções em textos de séculos anteriores; análise das relações emocionais entre personagens e aplicações diversas. Um estudo crítico atual do uso Análise de Sentimento nos estudos literários é o de Simone Rebora (2023), sob o título “Sentiment Analysis in Literary Studies. A Critical Survey”, faz uma crítica a Análise de Sentimento tanto do ponto vista teórico como computacional, em que conclui que as lacunas e a desconfiança que a ferramenta de Análise de Sentimento apresenta como resultado nos estudos literários podem ser superados na medida que pesquisadores, estudantes e amantes de literatura que anotem textos, testando teorias existentes auxiliem o sonho de construir uma ‘máquina hermenêutica’ compartilhada e orientada para este fim (Rebora 2023).

Entre os estudos que abordam a Análise de Sentimento em texto literários estão, por exemplo, Reed (2018) utiliza o dicionário Vader para explorar a poesia no artigo “Measured Unrest In The Poetry Of The Black Arts Movement”, Vani e Antonucci (2019) no artigo “NOVEL2GRAPH: Visual Summaries of Narrative Text Enhanced by Machine Learning”, apresenta um pipeline que gera resumos visuais das narrativas com abordagem de aprendizado de máquina, Thomson (2017) comparou diferentes edições de “Prelude as Lifespan Gauge”, em que observou a mudança nos aspectos emocionais e, Jacobs (2019) testou o dicionário SentiArt nos romances de Harry Potter, identificando o potencial emocional de trechos na descrição dos personagens.

A Análise de Sentimento da Poesia de João Cabral de Melo Neto

A tarefa de lidar com dados de produção literária – uma vez que se pretende realizar análise de sentimento em poesias de João Cabral de Melo Neto na língua portuguesa tanto por meio da abordagem lexical quanto pela abordagem de aprendizagem de máquina –, exige um esforço extra para realizar o seu pré-processamento, pois mesmo com técnicas bem estabelecidas não funcionam bem, há o desafio de superar a linguagem dentro do contexto da produção literária em língua portuguesa com todas as suas peculiaridades dentro do domínio linguístico brasileiro. Pesquisadores como Gimpel (2009), Hercig e Lenc (2017), por exemplo, debatem que problemas referentes a linguagem figura, uso de informalidades na escrita e fala, regionalismos linguísticos impactam o resultado do pré-processamento e na extração de atributos, por exemplo, com técnicas de tokenização.

Para ultrapassar essa dificuldade, optou-se pelo uso da hermenêutica (Gadamer 2002) para realizar uma normalização lexical dos dados. A hermenêutica foi estabelecida partir da análise do conjunto de 21035 sentenças da própria obra de João Cabral de Melo Neto.

Durante a análise, identificou-se a presença repetitiva de dupla e tripla pontuação, letras isoladas, números isolados, que foram facilmente tratados. Outra situação comum foi o uso de palavras estrangeiras com inglês, francês e com maior frequência o espanhol. Os casos mais difíceis, no entanto, foram os regionalismos tanto como presença fonética, por exemplo: “oitubro”, “tamo”, “experiência”, “fazêndero”, “dinhêro”, “comê”, quanto com presença lexical: ‘barra”, ‘desbanca”, “enfadada”. Essa situação foi tratada através da normalização. No entanto, cabe ressaltar que o uso da hermenêutica apenas deve ser aplicada dentro do próprio contexto do texto, uma vez que depende da interpretação delimitada pelo próprio contexto do texto.

A seguir, usou-se um método de análise de sentimento baseado em léxico e avaliou o uso da hermenêutica e de diferentes modelos para o escopo de negação neste experimento. Os léxicos de sentimento utilizados foram o léxico SentiLex-PT 02, composto por 82.347 formas flexionadas, e o léxico VADER léxicon, composto por 9.000 recursos lexicais.

A modelagem da negação tratou da análise de sentimentos por sentença, ou seja, verso a verso. Esse método, modelou-se por intermédio de um conjunto de indicadores de negação fundamentados nas análises literárias do corpus crítico já estabelecido em Benedito Nunes (1971), João Alexandre Barbosa

(1975), Marta Peixoto (1983), Antônio Carlos Secchin (1985) e Felipe Oliveira de Paula (2018), composto por expressões e palavras que foram analisadas por estes autores como sendo expressões de negação, por exemplo: “sem”, “mesma”, “não”, “nada”, “vazio”, ‘nenhum’ – indicando uma poética de negação intersentenciais e um escopo definido nos versos e nas estrofes da poesia de João Cabral de Melo Neto.

Por fim, o sentimento de uma obra ou poema de João Cabral de Melo Neto é identificado pela soma algébrica dos sentimentos, após a aplicação do tratamento de negação.

Experimento e Resultados da Análise de Sentimento da Poesia de João Cabral de Melo Neto

Nesta seção, apresenta-se o experimento sobre análise de sentimento das poesias de João Cabral de Melo Neto. Primeiro descreve-se o corpus, o léxico e as ferramentas utilizadas, depois apresentam-se os resultados, com as duas abordagens aplicadas.

O Corpus

O corpus usado em neste experimento seguiu a mesma organização proposta por Antonio Carlos Secchin no livro “Poesia completa” (2020) de João Cabral de Melo Neto. O conjunto está composto por poesias publicadas de 1937 até 2014, separados em 20 arquivos referentes as publicações em vida, 2 referentes a publicações póstumas, além de dois arquivos contendo poemas diversos e inéditos até 2014.

Obras	Período	Qtd. Poesias	Qtd. Palavras	Qtd. Sentenças
Totais: 24 arquivos	70 anos de produção	617	95.459	21035

Tabela 1: Síntese Descritiva do Corpus utilizado na Análise de Sentimento

Em síntese, o corpus de análise utilizado no experimento contou com 24 arquivos no formato txt.

Os Léxicos

Utilizou-se neste estudo o SentiLex-PT 02, um léxico de análise de sentimento para língua portuguesa, composto por 7.014 lemas e 82.347 formas flexionadas, sendo: 4.779 (16.863) adjetivos, 1.081 (1.280) substantivos, 489 (29.504) verbos e 666 (34.700) expressões idiomáticas, e o léxico LeIA que é um fork do VADER-léxico para língua portuguesa, composto por 9.000 recursos lexicais polarizados. Além desses léxicos, criou-se um escopo de negação a partir da análise hermenêutica das poesias de João Cabral de Melo Neto que se conformou em um total de 50 palavras e expressões negativas. Cabe ressaltar que a crítica literária da poesia de João Cabral de Melo Neto defende que sua poesia é predominantemente de sentimento negativa: Benedito Nunes (1971), João Alexandre Barbosa (1975), Marta Peixoto (1983), Antônio Carlos Secchin (1985) e Felipe Oliveira de Paula (2018).

Ferramenta

Para realizar a análise de sentimento, optou-se pelo uso da linguagem Python através da implementação dos códigos através do Jupyter Notebook, com auxílio das bibliotecas: NTKL (Natural Language Toolkit), unicodedata e scikit-learn (sklearn).

Resultados

Implementou-se duas abordagens: uma apenas com os léxicos LeIA/VADER e SentiLex, usando métodos com e sem o escopo de negação; e, outra abordagem em que os modelos foram criados com aplicação dos algoritmos: Naïve Bayes, Support Vector Machine (SVM) com os parâmetros kernel: 'linear', 'polynomial', 'RBF' e 'sigmoid', Gradient Boosting e Random Forest. Estes modelos da última abordagem foram implementados em quatro experimentos: no primeiro testamos os modelos com o léxico LeIA/VADER com e sem o escopo de negação e depois o léxico SentiLex com e sem o escopo. Optou-se por esta estrutura de experimento, pois o objetivo foi avaliar o impacto do resultado comparando os modelos implementados com e sem um escopo de negação contextualizada com o texto literário da poesia de João Cabral de Melo Neto.

Os resultados obtidos na primeira abordagem trouxeram a análise de sentimento com o sentimento referente a cada conjunto de poesias com os valores de score e sua classificação de sentimento: neutro, positivo ou negativo. É importante ressaltar que o score foi atribuído em uma escala de ganho de informação, variando de -1 (corresponde a predominância de sentimento negativa) e +1 (corresponde a predominância de sentimento positiva). O resumo dessa abordagem pode ser observado na Tabela 2:

Arquivos	LeIA/VADER	+ Escopo Negativo	SentiLex	+ Escopo Negativo
	Score	Score	Score	Score
ACasaDaFarinha.txt	-0,93	-0,93	-1,00	-1,00
AEducaoPelaPedra.txt	-1,00	-1,00	-1,00	-1,00
AEscolaDasFacas.txt	-1,00	-1,00	-1,00	-1,00
Agrestes.txt	-1,00	-1,00	-1,00	-1,00
AutoDoFrade.txt	-1,00	-1,00	-1,00	-1,00
CrimeNaCalleRelator.txt	-1,00	-1,00	-1,00	-1,00
Dispersos.txt	-1,00	-1,00	-1,00	-1,00
DoisParlamentos.txt	-1,00	-1,00	-1,00	-1,00
IlustracoesFotoDandara.txt	0,86	0,86	1,00	-1,00
Ineditos.txt	-1,00	-1,00	-1,00	-1,00
MorteEVidaSeverina.txt	-1,00	-1,00	-1,00	-1,00
MuseuDeTudo.txt	-1,00	-1,00	-1,00	-1,00
OCaoSemPlumas.txt	-1,00	-1,00	1,00	-1,00
OEngenhero.txt	-0,76	-0,76	-1,00	-1,00
ORio.txt	-1,00	-1,00	-1,00	-1,00
OsTresMalAmados.txt	0,99	0,99	0,00	-1,00
PaisagensComFiguras.txt	-1,00	-1,00	-1,00	-1,00
PedraDoSono.txt	-1,00	-1,00	-1,00	-1,00
PsicologiaDaComposicao.txt	-0,91	-0,91	1,00	-1,00
PrimeirosPoemas.txt	-1,00	-1,00	-1,00	-1,00
Quaderna.txt	-1,00	-1,00	-1,00	-1,00
Serial.txt	-1,00	-1,00	-1,00	-1,00
SevilhaAndando.txt	-1,00	-1,00	-1,00	-1,00
UmaFacaSoLamina.txt	-0,98	-0,98	-1,00	-1,00

Tabela 2: Síntese dos Resultados de cada Conjunto de Poesia

Nota-se que na implementação do método com o léxico LeIA/VADER com o uso conjunto do escopo de negação não alterou em nada o resultado da análise de sentimento. Já na implementação do método com o léxico SentiLex o uso conjunto do escopo de negação apresentou uma alteração significativa no resultado da análise de sentimento. Observa-se, ainda, que na análise de sentimento pelo SentiLex sem o escopo de negação os conjuntos de poemas que formam parte de “Psicologia da Composição” apresentaram na análise pelo SentiLex sentimento positivo as que forma parte de “Os três mal-amados” apresentaram sentimento neutro. Ambas as obras só são analisadas com sentimento negativo após a implementação com o escopo de negação. Já na análise implementada pelo uso LeIA/VADER a obra “Os

três mal-amados “apresentaram sentimento positivo sem e com o escopo de negação. No entanto, cabe ressaltar que estes conjuntos de poemas são analisadas pela crítica literária como obras representativas da poética da negação em João Cabral de Melo Neto, como defendem: Benedito Nunes (1971), João Alexandre Barbosa (1975), Marta Peixoto (1983), Antônio Carlos Secchin (1985) e Felipe Oliveira de Paula (2018) . A síntese do resultado da análise de sentimento das poesias de João Cabral de Melo Neto por estes métodos pode ser observada na Tabela 3:

Sentimentos	Leia/VADER	+ Escopo de Negação	SentiLex	+Escopo de Negação
Neutro	0,00%	0,00%	4,16%	0,00%
Negativo	91,67%	91,67%	83,33%	100,00%
Positivo	8,33%	8,33%	12,51%	0,00%
Totais	100,00%	100,00%	100,00%	100,00%

Tabela 3: Síntese dos Resultados usando Métodos

Como se nota, a aplicação dos métodos demonstra que as poesias de João Cabral de Melo Neto apresentam uma predominância do sentimento negativo independentemente do método aplicado. Também é possível concluir que o uso do léxico SentiLex acrescido do escopo de negação enfatizaram a predominância do sentimento negativo e, por isso, o resultado da análise foi de 100%.

Os resultados obtidos na segunda abordagem – com os modelos criados com aplicação dos algoritmos –, trouxeram os valores de acurácia e medidas F ($\beta = 1$, F1) para cada modelo implementado. O objetivo foi verificar a precisão dos modelos quanto a análise da predominância do sentimento negativo.

Para análise a partir dos modelos realizou-se uma divisão estratificada desse conjunto, pois este apresentou, após a primeira classificação, uma distribuição desigual de categorias de sentimento. O conjunto após a divisão contou com 16.828 sentenças de treinamento e 4.207 de teste, numa relação de proporcionalidade de 80/20 – em que as sentenças foram balanceadas entre duas categorias de sentimento – positivo e negativo. Num segundo momento, foi acrescido a está distribuição o escopo de negação. Esta divisão foi realizada automaticamente mediante um método em Python, em que foi importado a função ‘train_test_split’ do módulo ‘model_selection’ da biblioteca Scikit-learn.

A primeira rodada dos modelos criados com os algoritmos restringiu-se a dois momentos: o primeiro apenas com o léxico LeIA/VADER e o segundo acrescentamos o escopo de negação. A hipótese inicial era que estes modelos não apresentariam resultados diferentes entre si, uma vez que não apresentaram resultados diferentes na primeira abordagem. No entanto, o modelo com o escopo de negação apresentou uma pequena alteração na aplicação dos algoritmos Gradient BoosTing e Random Forest. Como pode ser observado na tabela 4:

Algoritmos	LeIA/VADER			+ Escopo de Negação		
	Precisão	Recall	F1	Precisão	Recall	F1
Naive Bayes	85.89%	83.89%	82.08%	85.89%	83.89%	82.08%
SVM - Linear	94.17%	94.16%	94.12%	94.17%	94.16%	94.12%
SVM - Polynomial	83.09%	78.71%	76.04%	83.09%	78.71%	76.04%
SVM - RBF	92.95%	92.82%	92.72%	92.95%	92.82%	92.72%
SVM - Sigmoid	94.17%	94.16%	94.12%	94.17%	94.16%	94.12%
Gradient Boosting	86.22%	83.97%	82.80%	86.30%	83.97%	82.78%
Random Forest	89.86%	89.40%	89.10%	89.48%	88.81%	88.94%

Tabela 4: Síntese dos Resultados usando Modelos com Algoritmos com LeIA/VADER

Nota-se que os modelos apresentaram em sua maioria resultados que correspondem a hipótese inicial, ou seja, os resultados entre os modelos que não foram acrescidos do escopo de negação foram iguais aos modelos que foram acrescidos. Portanto, acrescentar o escopo de negação não aumentou em nada a precisão que indica uma melhor modelagem da identificação de sentimento negativo. Ao contrário, observando a alteração apresentada na aplicação do algoritmo Gradient Boosting com o uso do escopo de negação verifica-se que o aumento da precisão prejudicou o desempenho do modelo. No caso da aplicação do algoritmo Random Forest o uso do escopo diminuiu a precisão e a recuperação o que fez diminuir o desempenho do modelo. Possivelmente, o motivo para a queda de desempenho nestes modelos reside na pequena representatividade do escopo de negação criado para o experimento.

A segunda rodada dos modelos criados com os algoritmos restringiu-se também a dois momentos: o primeiro apenas com o léxico SentiLex e o segundo acrescentou-se o escopo de negação. A hipótese inicial era que estes modelos apresentariam resultados semelhantes a abordagem anterior, uma vez que houve uma alteração significativa na primeira abordagem. E de fato, a hipótese foi confirmada, houve uma melhora do desempenho do modelo com o acréscimo do escopo de negação. Um único modelo que apresentou um declínio no desempenho foi o que utilizou o algoritmo de Naive Bayes, devido ao aumento da precisão e a diminuição da recuperação. Como pode ser observado na tabela 5:

Modelo	SentiLex			+ Escopo Negativo		
	Precisão	Recall	F1	Precisão	Recall	F1
Naive Bayes	87.36%	86.90%	86.64%	87.87%	85.85%	80.13%
SVM - Linear	88.47%	88.50%	88.46%	93.40%	93.56%	93.09%
SVM - Polynomial	81.73%	77.81%	75.87%	88.64%	86.88%	82.30%
SVM - RBF	89.74%	89.04%	88.80%	91.35%	91.14%	89.75%
SVM - Sigmoid	88.77%	88.77%	88.70%	93.16%	93.33%	92.81%
Gradient Boosting	88.29%	85.83%	85.13%	90.94%	90.56%	88.87%
Random Forest	87.61%	86.36%	85.91%	92.45%	92.29%	91.32%

TABELA 5: Síntese dos Resultados usando Modelos com Algoritmos com SentiLex

Pode-se observar que a implementação dos modelos implementados com o SentiLex acrescidos do escopo de negação teve um efeito positivo para a melhora do desempenho dos modelos. Ao contrário dos modelos com uso do LeIA/VADER que sofreram queda de desempenho em alguns modelos acrescidos do escopo de negação. No entanto, isso é uma surpresa já que o LeIA/VADER é apenas uma adaptação do VADER-léxico para a língua portuguesa, diferentes do SentiLex foi criado para análise em língua portuguesa.

A melhora do desempenho do SentiLex com o escopo de negação na aplicação dos modelos (Tabela 2) deve-se ao fato de que o escopo foi composto por diferentes tipos de palavras e expressões específicas da produção literária de João Cabral de Melo Neto, formada principalmente por substantivos concretos, diferente do escopo do SentiLex que é formado de predicadores em sua maioria adjetivos, classe gramatical pouco usada por João Cabral de Melo Neto.

Conclusão

Nesta pesquisa, fora analisada se a implementação do escopo de negação criado hermeneuticamente traria um melhor desempenho a Análise de Sentimento quando aplicada a textos

literários, neste caso as poesias de João Cabral de Melo Neto. Ao longo do processo de análise, algumas limitações foram identificadas, o que destaca a necessidade de cautela ao interpretar os resultados. Uma das principais limitações foi o tamanho restrito do escopo de negação criado, reflexo da escassez de estudos anteriores sobre esse tema. Além disso, o método utilizado para a criação do escopo ainda pouco explorado também pode ter impactado na qualidade e confiabilidade dos resultados obtidos.

Apesar dessas limitações, este experimento contribuiu para um importante passo para linguística computacional e para os estudos literários, pois mostra aos linguistas computacionais ser possível criar um escopo específico para o trabalho com o texto literário a partir dos próprios estudos acadêmicos do campo dos estudos literários. Para fortalecer a base de conhecimento e melhor compreender a complexidade dos fenômenos estudados, estudos futuros poderiam ser conduzidos com a criação de dicionários específicos e mais representativos para os estudos literários.

Referências

- Balage Filho, P., Pardo, T. A. S., & Aluísio, S., 2013. An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*.
- Barbosa, J. A. 1975. *A imitação da forma: uma leitura de João Cabral del Melo Neto*. No. 4. São Paulo: Livraria Duas Cidades.
- Benevenuto, F., Ribeiro, F., & Araújo, M. 2015. Métodos para análise de sentimentos em mídias sociais. *Sociedade Brasileira de Computação*. (Disponível em: <https://homepages.dcc.ufmg.br/~fabricio/download/webmedia-short-course.pdf> , acessado em 23 de maio 2023).
- Candido, A. 2002. Notas de crítica literária—poesia ao norte. *Textos de intervenção*. São Paulo: Editora 34 pp.135-142.
- Costa, C. H. 2014. *Imaginando João Cabral imaginando*. Editora Unicamp.
- de Paula, F. O. 2018. “A poética negativa de João Cabral de Melo Neto”, Tese de doutorado, Belo Horizonte: Universidade Federal de Minas Gerais.
- Gadamer, H. G. 2002. *Verdade e método: Complementos e índice/trad. de Enio Paulo Giachini. Rev. da trad. de Marcia Sá Cavalcante-Schuback*. Rio de Janeiro: Vozes.
- Grefenstette, G., Qu, Y., Shanahan, J. G., & Evans, D. A. 2004. Coupling niche browsers and affect analysis for an opinion mining application. *Proceedings of Recherche d'Information Assistée par Ordinateur (RIA0)*, April 26-28).
- Kim, E. & Klinger, R. 2018. A Survey on Sentiment and Emotion Analysis for Computational Literary Studies. *ArXiv*. (Disponível em: <https://arxiv.org/abs/1808.03137>, acessado em 31 de julho 2023).
- Melo Neto, J. C. D. 1994. Poesia e composição. *MELO NETO, João Cabral de. Obra completa*. Rio de Janeiro: Nova Aguilar.
- Nunes, B. 1971. *João Cabral de Melo Neto* (Vol. 1). Rio de Janeiro: Vozes.
- Rebora, Simone. 2023. Sentiment Analysis in Literary Studies. A Critical Survey. *DHQ*. (Vol. 17, n. 2) (Disponível em: <http://www.digitalhumanities.org/dhq/vol/17/2/000691/000691.html>, acessado em 31 de julho 2023).
- Rocha, D., Gutierrez, A. A., da Silva Lins, Q., de Moura, Í. L., de Vito Zollner, C., dos Santos Oliveira, A. M. A., & Bianchessi, C. 2020. *CEM ANOS DE JOÃO CABRAL DE MELO NETO (1920-1999): Cidadão do mundo e artesão da palavra*. Editora BAGAI.
- Wilson, T., Wiebe, J., & Hwa, R. 2004. Just how mad are you? Finding strong and weak opinion clauses. In *aaai* (Vol. 4, pp. 761-769) (July).