

Extração de Conhecimento em Texto: Interpretando Opiniões, Percebendo Sentimentos

Carine G. Webber¹, Lauren Girardi Cristofoli², Bryan Zinn²

¹Programa de Pós-Graduação em Ensino de Ciências e Matemática, Área de Conhecimento de Ciências Exatas e Engenharias, Universidade de Caxias do Sul (RS) - Brasil

²Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul, Campus Bento Gonçalves

cgwebber@ucs.br, acid.lauren@gmail.com, bryan.zinn@ifrs.edu.br

Abstract. *This paper presents a case study on Sentiment Analysis in an evaluative institutional process. In order to produce automatic analysis, it has been implemented a software named SentAnalysis. The software analyzes opinions written in the evaluative process, classifying them in positive, negative or neutral feelings. Results obtained from automatic evaluation were compared with those produced by two human evaluators. The comparison allowed us to recognize situations of partial and total convergence between system and evaluators. The evidences show that the developed system presents coherence with human evaluators, observing the fact that around 70% of the cases were of total convergence between human evaluators and system.*

Resumo. *O presente trabalho apresenta um estudo de caso de análise de sentimentos em um conjunto de respostas oriundas de um processo de avaliação institucional. A fim de realizar a análise automática de sentimentos foi desenvolvida uma ferramenta, denominada SentAnalysis. Ela realiza a análise de opiniões escritas no processo avaliativo, classificando-as em positivas, negativas ou neutras. Os resultados do processo de avaliação automática foram comparados com resultados de dois avaliadores humanos. A análise permitiu atestar que houve situações de convergência parcial e total entre sistema e avaliadores. As evidências produzidas permitem concluir que o sistema desenvolvido apresenta coerência com avaliadores humanos, observado o fato de que em cerca de 70% dos casos analisados houve convergência total entre avaliadores e sistema.*

1 Introdução

A mineração de textos é um campo fértil em desafios. Entre seus principais objetivos estão a descoberta, de forma automática, de padrões e anomalias em dados de formato textual. A mineração de textos compreende um conjunto de métodos usados para navegar, organizar, reconhecer e descobrir informações em bases de textos. Uma subárea da Mineração de Textos é a Análise de Sentimentos (LIU, 2010). Ela pode também ser chamada de Mineração de Opiniões. A análise de sentimentos e a mineração de opiniões focam-se principalmente em reconhecer e classificar opiniões,

segundo elas expressem sentimentos positivos ou negativos. Para isso, os indicadores de sentimento mais importantes são as palavras.

O uso da análise de sentimentos tem crescido graças à disponibilidade crescente de textos que emitem opiniões seja em blogs, microblogs, fóruns e nas redes sociais. Observa-se que de fato, a análise de sentimentos ultrapassou as barreiras da computação e foi absorvida pelas áreas de gestão e ciências sociais devido a sua importância para os negócios e para a sociedade como um todo. Por esta razão, a análise de sentimentos se situa ao centro das pesquisas sobre as mídias sociais (LIU, 2012).

As opiniões são centrais a quase todas as atividades humanas porque são influenciadoras de comportamentos. Sempre que se precisa tomar uma decisão, busca-se conhecer opiniões de outras pessoas. No mundo real, as organizações buscam conhecer as opiniões de consumidores sobre seus produtos e serviços. Os consumidores buscam igualmente conhecer as opiniões de usuários de um produto antes de comprá-lo, por exemplo.

Com o crescimento de meios sociais na Web (por exemplo, revisões, discussões do fórum, blogues, microblogues, comentários, e postagens em sites de compartilhamento de informações), os indivíduos e as organizações estão se valendo cada vez mais destas informações para a tomada de decisões. Construir, entretanto, uma visão global sobre as opiniões dispersas dos indivíduos é uma tarefa complexa. Um leitor humano, com conhecimento intermediário, vai fazê-lo mas terá dificuldade em identificar locais relevantes onde consultar, posteriormente extrair e sumarizar as opiniões disponíveis. Neste sentido é que os sistemas automáticos de análise de sentimentos são necessários.

Uma grande questão que permeia pessoas e organizações atualmente diz respeito ao uso de informações textuais para fins estratégicos. Nesta problemática se inserem as avaliações realizadas por organizações com variados fins. Manipular grandes volumes de textos que podem conter opiniões e percepções úteis se torna inviável em um curto espaço de tempo. Além disso, esta tarefa pode ser bem cansativa para um ser humano, que apesar da sua grande capacidade de compreensão, pode facilmente se distrair, não prestando atenção permanente nos detalhes. Neste cenário meios automáticos podem apoiar o ser humano, realizando ao menos uma parte da análise textual, produzindo sumários e antecipando percepções positivas ou negativas.

Um dos fins para o qual a análise de sentimento se aplica é a análise de avaliações textuais. Um tipo de avaliação baseada em texto é a avaliação institucional, prática presente em grande parte dos Institutos e Universidades no Brasil. Ela constitui um instrumento em que alunos, técnicos administrativos, docentes e membros externos têm a oportunidade de avaliar diversos pontos sobre a instituição como, por exemplo, a infraestrutura e os componentes curriculares dos cursos. Usualmente tal avaliação é feita por meio de perguntas objetivas. Porém, ela pode conter campos descritivos, permitindo ao respondente inserir comentários textuais, fazer críticas ou elogios à instituição.

As informações coletadas via questões objetivas são facilmente tabuladas e interpretadas. Contudo, avaliar as respostas descritivas depende da leitura e do tempo de um ser humano. Ele deve neste caso ler, fazer o registro, identificar pontos críticos, emitir julgamentos próprios a fim de lidar com o que está sendo dito em um texto. Tal tarefa parece ser trabalhosa e monótona, porém este tipo de análise é necessária pois as

informações obtidas podem se reverter em melhorias na qualidade dos serviços prestados.

Considerando os aspectos supramencionados, além das necessidades organizações e automatização de tarefas de interesse humano, este trabalho apresenta uma solução baseada na análise de sentimentos aplicada a uma base textual composta pelas respostas descritivas da Avaliação Institucional de uma Instituição de Ensino. Em termos metodológicos seguiu-se os processos típicos da mineração de textos: construção de um conjunto de dados de treinamento contendo opiniões (positivas, negativas ou neutras), aplicação de algoritmos de processamento de linguagem natural e aprendizagem automática, e definição de critérios de análise dos resultados. A fim de apresentar o trabalho desenvolvido, o presente artigo está organizado em 4 seções. A seção 2 traça um panorama dos trabalhos existentes na área de análise de sentimentos. A seção 3 traz aspectos metodológicos e materiais empregados no desenvolvimento do software. A seção 4 aponta resultados observados e a seção 5 conclui o artigo abordando as contribuições do trabalho, bem como os trabalhos futuros.

2 Trabalhos Relacionados

Extrair conhecimento a partir de textos é uma tendência emergente quando se observa o grande volume de dados textuais disponíveis (MATOS, 2010). Conforme Barion e Lago (2008, p.125) denomina-se mineração de textos o processo que visa suprir a necessidade de se descobrir, de forma automática, informações (padrões e anomalias) em textos. A análise textual em geral se baseia na interpretação de textos mas se torna deficiente quando encontra, por exemplo, palavras que podem ter relações difusas e/ou significados ambíguos (Hotho, 2005). A própria identificação de termos mais relevantes em um texto é um trabalho árduo, pois uma palavra não pode ser analisada separadamente. É necessário considerar o contexto no qual ela se encontra e respeitar regras ortográficas e sintáticas que condizem ao idioma do texto que está sendo analisado. Até o momento não existem algoritmos ótimos que consigam lidar com tais dificuldades de forma satisfatória e completa.

Segundo Liu (2012) a análise de sentimentos pode ser feita em três níveis de granularidade: documento, frase ou entidade. Em nível de documento a tarefa consiste em classificar se um documento expressa globalmente um sentimento positivo ou negativo (Pang, Lee e Vaithyanathan, 2002; Turney, 2002). Por exemplo, suponha um comentário enviado ao um site sobre um filme em exibição no cinema. A análise pode determinar se opinião expressa um sentimento positivo ou negativo sobre o filme. Neste nível de análise assume-se que cada documento expressa uma opinião sobre uma entidade ou produto, logo ela não é indicada em casos onde o mesmo documento avalie ou compare múltiplas entidades.

Em nível de frase, a análise vai avaliar se cada frase expressa uma opinião positiva, negativa ou neutra. Por neutro entende-se a ausência de uma opinião. A análise por frases permite distinguir quais frases em um texto emitem uma opinião. Já em nível de entidade a análise busca identificar a que se refere a opinião propriamente dita, buscando reconhecer a qual das entidades mencionadas na frase ela se refere. Por exemplo, suponha a frase: “a companhia aérea presta bom atendimento, porém os banheiros estavam sujos”. Nela são apresentadas opiniões que refletem sentimentos

positivos e negativos. A análise em termos de entidade permite extrair que a opinião positiva se refere a companhia aérea (primeira entidade), enquanto que a opinião negativa se refere aos banheiros disponíveis (segunda entidade). Assim, conclui-se que o sentimento sobre a primeira entidade é positivo e sobre a segunda entidade é negativo.

A análise de sentimentos tem sido utilizada em pesquisas de satisfação como no trabalho de Pang et.al (2002). Da mesma forma, ela é extensivamente aplicada em análises de opinião nas redes sociais, como nos trabalhos de Ferreira (2010) e Duarte (2010). Isto se deve ao fato de que muito conteúdo em formato textual está disponível em ambiente *online*, constituindo assim uma grande fonte de dados (PANG et.al ,2002).

O trabalho de Ferreira (2010) utiliza a Análise de Sentimentos em textos extraídos da rede social Twitter, com o intuito de classificar opiniões (negativas ou positivas) através de algoritmos de extração por passagem. Ainda, o trabalho de Duarte (2013) também utiliza o Twitter para realizar a Análise de Sentimentos, com enfoque em textos na língua portuguesa. Neste trabalho são empregadas diferentes técnicas para extrair informações referentes a diversas entidades em *tweets* em língua portuguesa para depois verificar os sentimentos do público em relação as entidades, mostrando as suas preferências por marcas, empresas e pessoas.

No trabalho de Abbasi et al (2008) é proposto o uso da Análise de Sentimentos para a classificação de opiniões em fóruns da Web em múltiplos idiomas. A pesquisa realizada focou em fóruns extremistas, que disseminavam ódio e em como se classificava os sentimentos presentes nas postagens dos mesmos. O trabalho, utilizando máquinas de suporte vetorial apresentou alta precisão na classificação dos sentimentos descritos nas postagens, demonstrando eficácia na abordagem para o problema proposto. Em outro contexto, a análise de sentimentos foi aplicada à sumarização de pesquisas de satisfação. Isto é possível desde que o questionário permita que as respostas sejam fornecidas em linguagem natural. Trabalhos desenvolvidos por (Oguri, 2007) e (Yi *et al*, 2003) visaram tentar identificar sentimentos negativos ou positivos em textos informados em pesquisas de satisfação de clientes.

3 Materiais e Métodos

A fim de proceder com a análise de sentimentos em um conjunto de dados oriundos de uma avaliação institucional desenvolveu-se uma ferramenta de software denominada SentAnalysis. A ferramenta SentAnalysis foi desenvolvida na linguagem C#. Ela opera em duas etapas e trabalha com arquivos em formato texto na versão corrente. Ela utiliza como dicionário léxico o SentLex PT02, desenvolvido por Carvalho et al (2012). O SentLex é um dicionário léxico de sentimentos em língua portuguesa, sendo constituído por 7.014 lemas, palavras que normalmente se encontram na forma masculina e singular para adjetivos e forma infinitiva para verbos e expressões idiomáticas, e 82.347 formas flexionadas, expressões que não se encontram na forma infinitiva. O léxico descreve 4.779 adjetivos (16.863 na forma flexionada), 1.081 nomes (1.280 na forma flexionada), 489 verbos (29.504 na forma flexionada) e 666 expressões idiomáticas (37.400 na forma flexionada). O SentLex está em formato textual, sendo que a cada termo ele atribui uma polaridade, assumindo valor 1 para positiva, -1 para negativa e 0 para neutra. Essa informação é levada em conta para a análise de sentimentos presentes nas opiniões.

A primeira etapa no uso da ferramenta é de Organização e Pré-processamento dos dados. Essa etapa utiliza o *corpus* do trabalho, o léxico SentiLex PT02 e, ao final da etapa, produz arquivos locais. Por se tratar de uma ferramenta experimental, optou-se por trabalhar com arquivos na camada de persistência. A segunda etapa é a da mineração das opiniões que resultará na análise de sentimentos. Ela faz uso de arquivos locais e, ao final, produz um arquivo global.

Na primeira etapa, a ferramenta deve ler os arquivos que constituem o *corpus* a ser trabalhado. O *corpus* utilizado para o trabalho foi constituído a partir das respostas da avaliação institucional da Instituição de Ensino alvo deste estudo. A avaliação da instituição contém formulários de perguntas de múltiplas escolhas e um campo onde os respondentes podiam deixar algum comentário pertinente. O conteúdo desse campo textual, que é uma resposta descritiva, foi o que constituiu o *corpus* do trabalho. As respostas optativas foram descartadas, por se tratarem de informações estruturadas e não constituírem o foco deste trabalho.

Para cada opinião extraída do banco de dados original da instituição foi gerado um arquivo no formato de texto contendo apenas uma resposta. No total foram gerados 6.878 arquivos. A partir desse momento, não ocorreram mais interações com os bancos de dados, sendo todo o processamento feito via arquivos de texto.

Para a realização da análise de sentimentos não foram considerados todos os arquivos de respostas disponíveis pela avaliação institucional, mas uma amostra de dados do período. Para a definição do tamanho da amostra utilizada no experimento foi considerada a fórmula de Berenson et. al (2000), apresentada na equação 1.

$$n = \frac{N \cdot Z^2 \cdot p(1 - p)}{(N - 1) \cdot e^2 + Z^2 \cdot p(1 - p)}$$

Equação 1: Estimativa de uma proporção
Fonte: Berenson et. al (2001)

A equação calcula o valor do tamanho da amostra (n), considerando:

N: o tamanho do universo

Z: desvio médio que é aceito para alcançar o nível de confiança

e: margem de erro máximo que se quer atingir

p: a proporção que se espera encontrar

Para um universo de 6.878 arquivos de opiniões disponíveis, com uma margem de 5% de erro e um grau de confiança de 92%, a amostra se constituiu de 293 opiniões. O conjunto de opiniões que formam a amostra foi selecionado de modo aleatório.

Depois, para cada arquivo lido, a ferramenta SentAnalysis realiza o pré-processamento. O processo ocorre em quatro passos, assim descritos:

1. Remover as *stopwords* do texto;
2. Separar o texto em *tokens* e calcular a frequência de cada um;
3. Classificar os *tokens* segundo o dicionário léxico;

4. Gerar arquivos de texto para cada opinião avaliada, incluindo os *tokens* e os sentimentos que expressam.

No passo 1 precedente, o sistema percorre o texto e remove suas *stopwords*, tendo em vista que elas não são significantes para o processo de mineração. Considerando como exemplo a frase “A aula de matemática é boa”. Alguns termos (‘A’, ‘de’, ‘é’) não são importantes para a mineração e podem ser removidos sem problema, a fim de se trabalhar apenas com os termos significantes da frase. Removendo-os, o que resta da frase é “aula matemática boa”. Sendo que apenas esses termos serão separados em *tokens* e considerados no processo de análise. A etapa de *stemming* (remoção de radicais) não foi necessária, pois o SentiLex PT02 possui os termos nas formas flexionadas, não necessitando se trabalhar com os radicais das palavras.

No passo 2, os termos que restaram da frase serão separados em *tokens* e sua frequência será calculada. Já no passo 3, os *tokens* são classificados de acordo com a sua polaridade dentro do SentiLex PT02. Para cada *token* gerado, o sistema procura no léxico qual a sua polaridade. Conforme descrito anteriormente, caso o valor seja -1, o termo é considerado negativo; 0 indica termo neutro; e 1 indica termo positivo. Caso o termo não conste no dicionário, ele é considerado neutro, pois não é possível classificá-lo. Durante o processo, a ferramenta contabiliza a quantidade de termos negativos, positivos e neutros.

No quarto e último passo a ferramenta gera os arquivos resultantes. Para cada resposta lida ela gera um arquivo contendo pelo menos três linhas. A primeira linha é composta pela opinião no seu formato original, no caso “A aula de matemática é boa”. A segunda linha apresenta os *tokens* e a sua respectiva frequência (resultantes do passo 2), separados por *pipe* (|), sendo que nesse caso o resultado é: “aula=1|matemática=1|boa=1”. A terceira linha do arquivo apresenta a quantidade de termos positivos, negativos e neutros separados por *pipe*, resultantes do passo 3. No caso, o resultado é: “POS=1|NEG=0|NEU=2”. E, por fim, a ferramenta lista os *tokens* identificados e suas respectivas polaridades, sendo que, em cada linha, é escrito um termo (figura 1).

Figura 1 – Exemplo de resultado da análise de sentimentos

```
A aula de matemática é boa  
aula=1|matemática=1|boa=1  
POS=1|NEG=0|NEU=2  
aula=NEU  
matemática=NEU  
boa=POS
```

Ao fim da primeira etapa o sistema gera, para cada resposta, um arquivo correspondente que, posteriormente, será utilizado para uma classificação geral das opiniões do *corpus*.

Na segunda etapa, o sistema executa a classificação geral/global das opiniões. Sendo essa etapa composta por três passos, explicados a seguir.

No passo 1, o sistema lê os arquivos gerados na etapa 1 e, para cada arquivo lido, a ferramenta seleciona a linha que contém os quantitativos de termos positivos, negativos e neutros, pois eles serão levados em consideração para a classificação das opiniões.

No passo 2 a classificação, ou análise de sentimentos, é feita de fato, levando em consideração a polaridade com maior frequência. Taboada et. al (2011) apresentam como solução uma soma das polaridades dos termos, porque os dicionários utilizados possuem diversos valores para as polaridades dos termos, como por exemplo, -3, -5, -1, 3, 4, entre outros. Tomando como exemplo, uma opinião em que as polaridades sejam 1, -2 e -1 o resultado seria -2, ou seja, negativo.

No caso do presente trabalho, a lógica de somar as polaridades pode ser aproveitada e simplificada. O SentiLex PT02 possui apenas três valores para as polaridades (1, 0, -1), diferente dos outros léxicos, em inglês, utilizados no trabalho de Taboada et al (2011). Nesse caso, todos os termos negativos possuem o mesmo valor, -1, o mesmo acontece com os positivos e neutros. Se, por exemplo, uma frase contém 2 termos positivos e 1 termo negativo, a soma resultaria em 1 ($2+(-1)$), ou seja, uma opinião positiva. Essa lógica, para esse caso específico, é a mesma que quantificar os termos e verificar qual polaridade que aparece mais. No caso apresentado anteriormente, como existem mais termos positivos que negativos, a classificação é positiva.

Como, na fase de pré-processamento o SentAnalysis já realiza a contagem dos termos e a sua respectiva classificação, ou seja, mostra quantos termos positivos e negativos foram identificados, torna-se uma tarefa simples verificar qual deles aparece com mais frequência e realizar a classificação com base nesses valores.

Ao final de todas as classificações o sistema executa o passo 3, que é a criação de um arquivo com indicadores globais das opiniões. Para cada classificação, o sistema gera uma linha no arquivo, contendo um campo identificador (id) para a opinião e a sua classificação, sendo POS para positiva, NEG para negativa e NEU para neutra. Além disso, no final do arquivo o sistema disponibiliza algumas estatísticas: a quantidade de opiniões lidas, a quantidade e percentual de opiniões positivas, a quantidade e percentual de opiniões negativas e a quantidade e percentual de opiniões neutras.

4 Resultados

O experimento realizado se dividiu em duas etapas. Na primeira etapa dois avaliadores humanos foram responsáveis por classificar as opiniões da amostra segundo três categorias: opinião positiva (POS), opinião negativa (NEG) ou opinião neutra (NEU).

O importante em se obter a avaliação de dois especialistas é que a interpretação textual que se pretende realizar depende de percepções individuais e subjetivas, logo é importante ter-se mais de um diagnóstico sobre cada opinião. Na segunda etapa do experimento, as mesmas amostras foram classificadas pela máquina. Sendo assim, ao término das duas etapas, cada opinião contida na amostra possuía três classificações.

Para a análise dos resultados foram considerados quatro cenários possíveis. O primeiro cenário considerado foi onde as três avaliações convergem, ou seja, tanto os

dois avaliadores humanos quanto a máquina classificaram a opinião da mesma forma. Nesse caso assume-se que a máquina teve sucesso na sua tarefa.

O segundo cenário tratou do caso onde a máquina converge com, pelo menos, um avaliador humano (portanto as avaliações humanas divergem). Nesse caso assume-se também que a máquina teve sucesso na sua tarefa.

O terceiro cenário considerou situações onde as três avaliações (duas humanas e a da máquina) divergem entre si. Nesse caso assume-se que houve divergência total entre os avaliadores.

Por fim, no quarto cenário têm-se o caso onde as avaliações humanas convergem entre si, divergindo da máquina. A tabela 1 apresenta exemplos da amostra onde houve convergência e divergência entre avaliadores e software.

Tabela 1 - Comparação entre resultados das avaliações

ID	Avaliador I	Avaliador II	Avaliação Máquina	Resultado
1	NEU	NEU	NEU	ACERTO
2	NEG	NEG	NEG	ACERTO
3	NEU	NEG	NEG	ACERTO
4	NEU	NEU	NEU	ACERTO
5	NEG	NEG	NEG	ACERTO
6	NEG	NEG	NEG	ACERTO
7	POS	POS	POS	ACERTO
8	NEU	NEG	NEU	ACERTO
9	NEG	NEG	NEG	ACERTO
10	NEU	NEU	NEU	ACERTO
11	NEU	NEG	NEG	ACERTO
12	NEU	NEG	NEG	ACERTO
13	NEG	NEG	NEG	ACERTO
14	NEG	NEG	NEU	ERRO
34	NEU	NEG	POS	DIVERGÊNCIA TOTAL
35	NEU	NEU	NEU	ACERTO

Para o *corp us* analisado contendo 293 opiniões, em 222 a máquina

teve sucesso na sua classificação (primeiro e segundo cenário). Em 16 opiniões da amostra ocorreram situações de divergência total (terceiro cenário). Já 56 opiniões da amostra se enquadraram no cenário 4, onde as avaliações humanas convergiram entre si, divergindo da máquina. A Tabela 2 sumariza o resultado das avaliações.

Tabela 2 - Resultado Geral das Avaliações

Resultado	Quantidade	Porcentagem
Sucesso da máquina	222	75,76%
Falha da máquina	56	19,11 %
Divergência total	16	5,40%

Prosseguindo na avaliação, optou-se por considerar apenas as opiniões cujas avaliações dos avaliadores humanos convergiu. Das 293 avaliações, apenas 220 (aproximadamente 75%) tiveram convergência entre os dois avaliadores humanos. Para a próxima análise de dados, essa quantidade será levada em consideração, tendo em

vista que, como ambos os humanos convergiram entre si, a chance da classificação estar correta é maior. Para a exibição dos resultados, foi optado pelo uso de uma matriz de confusão. A Tabela 3 apresenta uma matriz de confusão dos resultados.

Tabela 3 - Matriz de confusão

HUMANOS MÁQUINA	POSITIVO	NEGATIVO	NEUTRO
POSITIVO	59	12	4
NEGATIVO	3	59	5
NEUTRO	13	25	40

A matriz funciona da seguinte forma: as colunas representam as avaliações humanas, enquanto que as linhas representam as avaliações realizadas pela ferramenta.

A posição (1,1) apresenta o total de opiniões onde tanto a ferramenta quanto os seres humanos avaliaram como positivo: 59 no total. A posição (1, 2) apresenta quantas opiniões a máquina avaliou como positivas, enquanto os humanos avaliaram como negativas: 12 no total. A posição 1,3 apresenta quantas opiniões a máquina marcou como positivas, enquanto os humanos classificaram como neutra: 4 no total.

A posição (2,1) apresenta quantas opiniões a ferramenta classificou como negativa enquanto os seres humanos definiram como positivas: 3 no total. A posição (2, 2) apresenta o total de opiniões que tanto a máquina quanto o seres humanos classificaram como negativas: 59 no total. A posição (2,3) apresenta quantas opiniões a máquina classificou como negativa, enquanto os humanos classificaram como neutras: 5 no total.

A posição (3,1) apresenta a quantidade de opiniões que a ferramenta classificou como neutras enquanto os humanos classificaram como positivas: 13 no total. A posição (3,2) apresenta a quantidade de opiniões classificadas pela ferramenta como neutras, mas negativas pelos seres humanos: 25 no total. Por fim, a posição (3,3) apresenta o quantitativo de opiniões que foram classificadas como neutras tanto pela máquina como pelos humanos: 40 no total.

Convém salientar que a diagonal principal da matriz apresenta os verdadeiros positivos para cada classe, ou seja, os valores em que houve convergência entre as avaliações dos seres humanos com as avaliações da máquina. As demais posições apresentam as divergências entre as classificações de ser humano e máquina.

A Tabela 4 apresenta resultados em termos de verdadeiros positivos (convergências) e das divergências.

Tabela 4 - Quantitativo de verdadeiros positivos e divergências

Resultado	Quantidade	Porcentagem
Verdadeiros positivos	158	71,81%
Divergências	62	28,18%

Do total das 293 opiniões, a ferramenta convergiu em 222 com, pelo menos, algum dos humanos, independente se eles convergiram entre si ou não, o que representa aproximadamente 75,76% das respostas. Das convergências humanas, 220 (aproximadamente 75% do total analisado), a ferramenta convergiu em 158 avaliações

com os seres humanos, aproximadamente 71,81% das respostas, entrando no percentual de acerto.

A ferramenta apresentou bons resultados principalmente com dois tipos de opiniões: as respostas curtas e bem objetivas, que continham poucos termos, mas com polaridades bem definidas (positivas ou negativas), como por exemplo: “boa professora” ou “falsa”; e com opiniões mais extensas, principalmente com conteúdo negativo, pois foi uma tendência percebida nas respostas a grande presença de termos negativos, o que tornou as respostas mais objetivas e assim simples de classificar.

Quanto aos erros (divergências com as avaliações humanas), a ferramenta não conseguiu classificar diversas opiniões devido ao contexto que as palavras tomavam nas frases. Isoladamente, a maioria dos termos eram neutros, mas quando dispostos na opinião, assumiam características positivas ou negativas, pelas percepções humanas, sendo que, por não conter algum termo positivo ou negativo, para a lógica da ferramenta, aquilo era considerado neutro.

Outro ponto dizia respeito ao fato de algumas opiniões conterem a mesma quantidade de termos positivos e negativos, sendo que em alguns casos ela classificava como neutro corretamente, pois nenhum termo tinha mais peso que o outro. Entretanto, em outros casos, a opinião não era neutra, considerando a avaliação dos humanos, mas a ferramenta não conseguia determinar quais termos tinham mais peso dentro da frase, se eram os positivos ou negativos, apenas os quantificava e com base nesse quantitativo, ela classificava.

Sumarizando, os casos de divergência entre o diagnóstico humano e de máquina ocorreram devidos aos seguintes fatores:

- a) uso de textos longos contendo comentários positivos e negativos
- b) expressão de uma opinião implícita aos termos empregados
- c) uso de termos em expressões que modificam o seu significado
- d) uso de termos classificados incorretamente no dicionário léxico
- e) erros de digitação (por parte do respondente) que impediram a análise correta
- f) frases curtas que continham termos ausentes do dicionário léxico
- g) uso de termos em sentido diferente do usual, contido no dicionário léxico

Observou-se que quando um texto apresentou opiniões negativas e positivas ao mesmo tempo, os especialistas humanos deram mais importância aos comentários negativos. Nestes casos a opinião foi considerada negativa (mesmo contendo elogios). Contudo, a máquina quantificou os comentários de cada categoria e atribuiu a opinião geral segundo o que foi mais frequente no texto.

Casos onde o dicionário léxico se revelou incompleto também foram identificados. Em uma avaliação, um estudante escreveu: “o professor é absurdo, merece um aumento”. Os avaliadores humanos classificaram a opinião como positiva. Contudo, a máquina utilizou o significado do termo absurdo disponível no dicionário léxico, classificando a opinião como negativa. No dicionário o termo *absurdo* tem como definição: que fala ou age de maneira irracional; estúpido, disparatado, tolo. O que

permite inferir que a opinião é positiva está na segunda parte da sentença: “merece um aumento”. Entretanto, os termos que compõe a frase não permitiram à máquina inferir este novo significado à frase.

Em certos casos os avaliadores humanos julgaram a opinião como sendo neutra. Porém, a presença de termos conduziu a máquina ao erro. Em um exemplo um estudante escreveu: “para quem tem dificuldades de inglês é difícil sem um curso complementar”. Os avaliadores humanos consideraram a opinião neutra. Já a máquina reconheceu os termos “dificuldades” e “difícil” e indicou que a opinião é potencialmente negativa.

Na ausência de termos no dicionário léxico, a máquina encontrou dificuldades em interpretar a opinião do estudante. Por exemplo, um estudante escreveu: “um mito, meu ídolo!”. O termo mito está inserido no dicionário léxico como sendo neutro e o termo ídolo estava ausente do dicionário. Sendo assim, a máquina interpretou a opinião como neutra; os especialistas a consideraram positiva.

De maneira geral considera-se que os resultados obtidos apontam que abordagem é promissora. Entendeu-se por este estudo que a abordagem explorada funciona melhor para textos curtos, contendo palavras utilizadas no sentido literal.

5 Conclusões

O objetivo do trabalho apresentado foi de avaliar a Análise de Sentimentos, usando como exemplo um caso de Avaliação Institucional. A partir da revisão do estado da arte foi possível identificar uma metodologia para conduzir o trabalho. Neste caso, o trabalho foi baseado em um dicionário léxico denominado SentiLex PT02, que busca representar os sentimentos da língua portuguesa. A abordagem baseada em léxico foi selecionada tendo em vista a simplicidade no seu entendimento e implementação, assim como o SentiLex PT02 foi escolhido por estar em português. A implementação do algoritmo de Análise de Sentimentos não se mostrou a parte mais complexa do projeto, mas sim as etapas anteriores a ela de pesquisa, coleta das informações, pré-processamento dos dados, limpeza e formatação das opiniões para que pudessem ser mineradas e analisadas.

Quanto ao processo de análise, convém salientar que diversas respostas da avaliação apresentavam, ao mesmo tempo, pontos positivos e negativos observados pelos alunos, como por exemplo, em uma opinião que um respondente elogiava o esforço e conhecimento de um professor, embora questionasse seus métodos de ensino. Logo, tornava-se difícil concluir se uma opinião assim seria negativa ou positiva, pois ela apresentava ambos os sentimentos. Contudo, a ferramenta desenvolvida tinha que classificá-la mesmo a opinião refletindo os dois sentimentos, sendo que o SentAnalysis considerou como sentimento preponderante aquele que é quantitativamente mais importante. Essa é uma limitação do sistema. Mesmo com limitações, os resultados obtidos foram satisfatórios, tendo a ferramenta obtido mais de 70% de acertos na análise de acordo com avaliações humanas.

Entende-se que as limitações identificadas poderão ser tratadas em um trabalho futuro, em que se possa hibridizar a abordagem aplicada com outra. Espera-se assim considerar o contexto da frase com o léxico adaptado ao universo da avaliação institucional, além de considerar outros fatores, como por exemplo, uma ironia presente

em uma opinião. Outro ponto diz respeito ao uso da ferramenta, que não possui uma interface gráfica para usuário final. Ela pode ser adaptada a fim de disponibilizar uma interface completa para os avaliadores das comissões, sendo aplicado de fato na avaliação, contendo informações mais completas, como uma lista separada por classificação. Com isso, poderia ser possível ao usuário, por exemplo, separar as opiniões negativas a fim de saber quais são as maiores críticas a respeito da instituição. Outro recurso importante que pode ser desenvolvido futuramente é o de associar termos às classificações, por exemplo, identificando os termos mais frequentes presentes em reclamações de respondentes, permitindo ao avaliador identificar rapidamente um possível foco de melhoria para a instituição.

6. Referências

ABASSI, Ahmed; CHEN, Hsinchun; SALEM, Arab. **Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums**. The University of Arizona. 2008.

BERENSON, M. L.; LEVINE, D. M.; STEPHAN, David. **Estatística: Teoria e Aplicações usando Microsoft Excel em Português**. Rio de Janeiro: LTC, 2000.

BARION, E. C. N.; LAGO, D.. **Mineração de Textos**. Revista de Ciências Exatas e Tecnologia, São Paulo, v. 3, n. 3, p. 123-140, 8 Dez. 2008.

CARVALHO, Paula; SARMENTO, Luís; SILVA, Mário J.. "Building a Sentiment Lexicon for Social Judgement Mining". In **Lecture Notes in Computer Science (LNCS)**, International Conference on Computational Processing of the Portuguese Language (PROPOR), Springer, pp. 218-228. 2012.

CRISTOFOLI, Lauren Girardi; WEBBER, Carine, G.; LIMA, Mária de Fátima Webber do Prado. UnderMiner Text Miner – Uma ferramenta para Mineração de Texto para Área Educacional. **RENOTE - Revista Novas Tecnologias na Educação**, vol.11., n.1, 2013.

DUARTE, Eduardo Santos. **Sentiment Analysis on Twitter for the Portuguese Language**. Universidade de Nova Lisboa, Faculdade de Ciência e Tecnologia, 2010.

FERREIRA, Emanuel de Barros Albuquerque. **Análise de Sentimento em redes sociais utilizando influência das palavras**. Universidade Federal de Pernambuco, 2010.

HOTH, Andreas; NÜRNENBERGER, Andreas; PAAß, Gerhard. A Brief Survey of Text Mining. 2005.

LIU, Bing. **Sentiment Analysis and Subjectivity**. In. Handbook of Natural Language Processing. Segunda Edição. 2010.

LIU, Bing. **Sentiment Analysis and Opinion Mining**, Morgan & Claypool Publishers, May 2012.

MICROSOFT. Visual Studio. Disponível em: <<https://www.visualstudio.com/pt-br/downloads/download-visual-studio-vs.aspx>> Acesso em: 02 de out 2015

OGURI, P. **Aprendizado de Máquina para o Problema de Sentiment Classification**, PUC-RIO, 2007. Dissertação de Mestrado. Rio de Janeiro : PUCRio, Departamento de Informática, 2007.

PANG, B.; LEE, L. and VAITHYANATHAN, S. **Thumbs Up? Sentiment classification using machine learning techniques**, In Proc. of the EMNLP, 2002.

TABOADA, Maite et al. **Lexicon-Based Methods for Sentiment Analysis**. Association for Computational Linguistics. 2011.

TURNEY, Peter. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In **ACL**, pages 417–424

YI, J.; NASUKAWA, T.; BUNESCU, R. and NIBLAC, W. **Sentiment analyser: Extracting sentiments about a given topic using natural language processing techniques**, In IEEE Intl. Conf. on Data Mining (ICDM), 2003.