Unisanta Science and Technology, **2019, 3, December**
Published Online 2019 Vol.8 N°2 http://periodicos.unisanta.br/index.php/sat)

UNISANTA
Science and Technology
ISSN 2317-1316

Consistência de Polaridade de Palavras em Aplicações de Análise de Sentimentos

Daniela América da Silva*, Paulo Marcelo Tasinaffo, Johnny Marques, Luiz Alberto Vieira Dias, Adilson Marques da Cunha

ITA-Instituto Tecnológico de Aeronáutica
Praça Marechal Eduardo Gomes, 50 - Vila das Acácias
CEP 12228-900 – São José dos Campos – SP – Brasil.

***E-mail**: damerica@ita.br
Received November, 2019

Resumo: A classificação da polaridade das palavras é importante para aplicações como Mineração de Opiniões e Análise de Sentimentos. E as opiniões expressas em vários sites e mídias (por exemplo, *blogs*, jornais) são um importante critério para o sucesso de um produto ou política do governo. Existem inúmeras obras que, dado um sentimento, analisam a estrutura de uma sentença / documento para inferir sua orientação, o detentor de uma opinião, o sentimento da opinião, entre outros. Entretanto vários dicionários de sentimentos independentes de domínio foram manualmente ou (semi) automaticamente criados, e há inconsistências, pois as polaridades das palavras em um dicionário de sentimentos pode não ser necessariamente consistente (ou correta). Este artigo apresenta um estudo sobre a utilização de um Problema de Satisfabilidade em Lógica Proposicional (SAT) para a verificação de consistência de polaridade de palavras em aplicações de análise de sentimentos.

Palavras chave: Sentimento, Análise, Fraudes, Teste, Polaridade, Dicionário, Satisfabilidade, SAT, NP-Completo, *Fuzzy*, Paraconsistente.

Word Polarity Consistency in Sentiment Analysis Applications

Abstract: Polarity classification of words is important for applications such as Opinion Mining and Sentiment Analysis. And the opinions expressed on various websites and media (e.g. blogs, newspapers) are an important criterion for the success of a government product or policy. There are numerous works that, given a feeling, analyze the structure of a sentence and/or document to infer its orientation, the holder of an opinion, the sentiment in an opinion, among others. However, several domain-independent sentiment dictionaries were manually or (semi) automatically created, and there are inconsistencies, as the polarity of the words in a sentiment dictionary may not necessarily be consistent (or correct). This paper presents a study on the use of propositional satisfiability problem (SAT) to verify word polarity consistency in sentiment analysis applications.

Keywords: Sentiment, Analysis, Fraud, Test, Polarity, Dictionary, Satisfaction, SAT, NP-Complete, Fuzzy, Paraconsistente.

1. Introdução

Este artigo investiga os algoritmos que solucionam o problema de satisfabilidade booleana (SATISFABILIDADE ou SAT) para adquirir os conhecimentos necessários na implementação de testes de dicionários de sentimentos e desenvolvimento de aplicações para a análise de sentimentos.

De acordo com o artigo *Polarity Consistency Checking for Sentiment Dictionaries*, das universidades americanas, University of Purdue, University of Chicago e

Copyright © 2019 ISESC.

University of Binghamton, de 2012 [1], a classificação da polaridade das palavras é importante para aplicações como a Mineração de Opiniões e Análise de sentimentos. Além disto, uma série de dicionários de palavras e/ou sentidos para sentimentos foram manualmente ou (semi) automaticamente construídos recentemente e os dicionários têm imprecisões substanciais. Além de casos óbvios, onde a mesma palavra aparece com diferentes polaridades em diferentes dicionários, os dicionários exibem casos complexos, que não podem ser detectados por mera inspeção manual.

Até a realização do trabalho citado pelo artigo estudado, nenhum dos trabalhos anteriores estudou o problema da verificação de consistência da polaridade para dicionários de sentimentos.

O artigo introduz o conceito de consistência de polaridade de palavras / sentidos nos dicionários de sentimentos.

Dicionários específicos foram selecionados para o experimento e o artigo demonstra que o problema de consistência é NP-completo. Dito isto os autores sugerem utilizar uma solução SAT para detectar inconsistências na polaridade de palavras.

As opiniões expressas em vários sites e mídias (por exemplo, blogs, jornais) são um importante critério para o sucesso de um produto ou política do governo. Por exemplo, é provável que um produto com análises consistentemente boas seja vendido bem. A abordagem geral é resumir a polaridade semântica (ou seja, positivo ou negativo) de frases / documentos pela análise de orientação das palavras individualmente. E os dicionários de sentimentos são utilizados para facilitar o entendimento desta orientação.

2. A História da Satisfabilidade

De acordo com o artigo a História da Satisfação de Franco, J, e Martin, J, de 2007, Universidade de Cincinnati, Estados Unidos [3], pode-se definir a Lógica como:

- A lógica é sobre validade e consistência. As duas idéias são inter definíveis e se fizer uso da negação $\neg$: o argumento de $p_1,..., p_n$ a q é válido se e somente se o conjunto $p_1,..., p_n$ a $\neg q$ é inconsistente. Assim, validade e consistência são realmente duas maneiras de olhar para a mesma coisa e cada uma pode ser descrita em termos de sintaxe ou semântica;
- Entre esses, outro nome para a versão semântica da consistência é a satisfabilidade.

A satisfabilidade também poderá ser analisada através de seus mundos possíveis e a necessidade e possibilidade, conforme descrito a seguir:

- Tradicionalmente, uma frase é considerada necessária (ou necessariamente verdadeira) se for verdadeira em todos os mundos possíveis e, possível (ou possivelmente verdade) se for verdade em pelo menos um mundo possível;
- Se entendermos que um possível mundo seja uma estrutura, a possibilidade acaba sendo apenas outro nome para a satisfabilidade;
- Uma verdade possível é apenas aquela que é satisfatória;
- E em lógica, o nome técnico para uma fórmula necessária é verdade lógica: p é definido como uma verdade lógica (em símbolos, $\models p$ se e somente se, para todos os $\mathcal{A}$, $\mathcal{A} \models p$ (em lógica sentencial, uma verdade lógica é chamada tautologia.);
- Além disso, necessário e possível são predicados da metalinguagem (a linguagem da teoria lógica) porque eles são usados para descrever frases na linguagem "objeto" (a linguagem que se refere a entidades no mundo que são objeto de investigação em teoria lógica).

Há também conceitos centrais sobre a lógica como a sintaxe e a semântica, conforme descrito a seguir:

- O artigo demonstra como a satisfabilidade está intimamente relacionada aos conceitos centrais da lógica;
- De fato, em relação a um sistema axioma completo, a satisfabilidade pode ser usada para definir e pode ser definida por outros conceitos básicos do campo - validade, derivabilidade, consistência, necessidade, possibilidade, verdade lógica, tautologia, e teorema;
- Foram necessários mais de 2000 anos para chegar a definição mencionada neste artigo, sendo claramente enunciado por Tarski na década de 1930 [3]. E portanto, a noção formal de satisfabilidade estava ausente até então, apesar de informalmente compreendido desde Aristóteles.

Ao analisarmos a linha do tempo sobre o estudo da Lógica, temos uma visibilidade do nível de atividade em uma área particular e durante um período particular [3].

De acordo com esta linha do tempo [3], uma noção precisa de satisfabilidade para a lógica de primeira ordem, foi desenvolvida na década de 1930 na obra de Alfred Tarski (1902-1983) [17, 18, 19].

A tarefa de Tarski era definir um conjunto de condições necessárias e suficientes para "p é verdadeiro ", para qualquer fórmula p de sintaxe de primeira ordem. Sua solução não foi definir a idéia em uma única frase aplicável a todas as fórmulas, mas, como Gödel, dar uma definição indutiva, primeiro definindo a verdade para

Copyright © 2019 ISESC

fórmulas básicas e depois estendendo a definição para fórmulas mais complexas. O problema foi complexo, pelo fato de que, diferentemente da lógica sentencial em que o valor de verdade das partes determinam imediatamente o todo (por referência a tabelas verdade), quando toda a expressão é universalmente quantificada, não está claro como sua verdade é determinada pela interpretação de sua parte. Como a "verdade" da fórmula aberta Fx determina que $\forall x: Fx$?

Tarski resolveu o problema em duas etapas:

- Na primeira etapa, ele atribuiu interpretações fixas para as variáveis. Feito isso, é possível dizer quando $\forall x: Fx$ é verdadeiro se soubermos se *Fx* é verdadeiro sob suas várias interpretações. Se *Fx* for verdadeiro em toda interpretação de *x*, então *:x: Fx* também será verdadeiro em cada uma dessas interpretações. Se Fx for falso sob uma única interpretação de *x*, no entanto, $\forall x: Fx$ é falso sob qualquer interpretação das variáveis Tarski cunhou o termo técnico satisfação para se referir à verdade relativa a uma interpretação de variáveis.
- O segundo estágio da definição de Tarski é abstrair uma variável e definir a noção mais simples "*p* é verdadeiro em relação a $\mathcal{A}$". Sua idéia nesta fase, é interpretar uma fórmula aberta *Fx* como verdadeira nesse sentido geral se é "*sempre verdade*" no sentido de estar satisfeito em qualquer interpretação de suas variáveis. Ou seja, ele adota a fórmula simples: *p* é verdadeiro em relação a $\mathcal{A}$ se e somente se, para todas as atribuições de variáveis *s*, *p* é satisfeito em relação a $\mathcal{A}$ e *s*.
- Em notação formal, $\mathcal{A} \models p$ se e somente se, para todos os s, $\mathcal{A} \models_s p$.

Os lógicos adotaram a prática comum de usar o termo "*satisfeito*" em uma estrutura para significar o que Tarski chamou de" *verdadeiro em uma estrutura* ". Assim, é comum dizer que *p* é satisfatório se houver alguma estrutura $\mathcal{A}$ tal que *p* é verdadeiro em $\mathcal{A}$ e que um conjunto de fórmulas *X* é satisfatório se e somente se houver alguma estrutura $\mathcal{A}$ tal que para todas as fórmulas *p* em *X*, *p* seja verdadeira (satisfeita) em $\mathcal{A}$.

3. Análise de Sentimento

A análise do sentimento é uma área da inteligência artificial que se preocupa em ensinar os computadores a reconhecer as emoções humanas presentes no texto. O objetivo é entender a variedade e a força das emoções no texto escrito [4].

Reconhecer emoções é uma tarefa fácil para os humanos. Por exemplo, se alguém disser *"este filme é ótimo"*, saberemos, mesmo sem pensar muito, que a pessoa é (1) positiva e (2) animada com (3) um filme. Mas é mais difícil para as máquinas aprenderem como fazer isso, porque as línguas não são diretas. Alguns exemplos são *"Este filme é melhor do que esse"*, mas não *"esse filme é mais bom que aquele"*, e esse conhecimento vem naturalmente para os humanos entretanto há complexidade para as máquinas.

Além disto, escrever um conjunto de regras preciso e sustentável para entender todas as exceções presentes em um idioma é algo complexo e pior ainda, existem muitas línguas diferentes.

Existem diversas abordagens, porém o desafio é balancear simplicidade e precisão:

- É possível fazer uma lista de palavras positivas (bom, ótimo, excelente) e uma lista de palavras negativas (ruim, terrível, horrível). Depois, analisar um texto, e registrar todas as palavras positivas e negativas que aparecem nele. Se o total for positivo, pode concluir que o texto é positivo, caso contrário, pode concluir que é negativo.
- Pode ainda tornar o processo mais preciso, dando a cada palavra uma "força". Por exemplo, "bom" pode ser 60% positivo, mas "ótimo" soa ainda mais positivo, então pode dar uma pontuação de 80%. Não é tão trabalhoso quanto parece - há listas de emoção de acesso livre e aberto disponíveis *online* (por exemplo, no padrão de bibliotecas do Python) [5].

Porém há alguns problemas com a contagem de palavras:

- Os simples cálculos de palavras muitas vezes não têm a nuance para serem capazes de compreender construções linguísticas mais técnicas. Por exemplo, posso dizer: "Este filme é terrivelmente bom", o que pode gerar uma pontuação neutra ou negativa, dependendo de como a palavra "terrivelmente" é considerada. Embora existam maneiras de contornar esse problema com a contagem de palavras, em geral elas não são sofisticadas o suficiente para entender completamente o problema.
- Uma solução para isso é o uso de algoritmos de aprendizado de máquina, em que o computador aprende a probabilidade de palavras individuais ou pares de palavras ou até mesmo de conjuntos mais longos de texto associados a uma determinada emoção. Entretanto estes algoritmos também apresentam problemas, pois sofrem com a falta de contexto.

Copyright © 2019 ISESC

- Por exemplo, os registros enfatizam demais as palavras individuais, por isso não conseguem entender frases como "bela decepção". Pode-se contornar isso com o algoritmo de aprendizado de máquina Naive Bayes, porque ele pode aprender a associar emoções com pares, trios, ou frases de qualquer tamanho, mas ainda pode se atrapalhar com uma frase "Eu não consigo não gostar de você", se você ainda não mostrou exemplos suficientes desse duplo negativos.
- Além disto há inconsistências na polaridade das palavras de um dicionário de sentimentos, o que causará um interpretação incorreta se um comentário é positivo ou negativo.

O estado da arte atualmente é o aprendizado profundo, que usa técnicas como incorporação de palavras, redes neurais, e mecanismos de atenção para aprender as estruturas complexas presentes na linguagem humana. Mas até mesmo esses modelos podem se confundir com construções linguísticas mais complexas, como a ironia e o sarcasmo.

Além disto até mesmo entre os humanos o nível de concordância sobre exatamente quais emoções estão presentes raramente excede 80\%. De acordo com a JC linguagem e emoção estão unidas, mas a relação nem sempre é clara - até para nós.

Portanto as máquinas só podem fazer o melhor que podemos.

4. Determinação de Inconsistências

Existem inúmeras obras que, dado um sentimento, analisam a estrutura de uma sentença e/ou documento para inferir sua orientação, o detentor de uma opinião, o sentimento da opinião entre outros. Entretanto vários dicionários de sentimentos independentes de domínio foram manualmente ou criado (semi) automaticamente, e há inconsistências, como por exemplo a palavra *barato* pode ser de polaridade positiva em um dicionário e negativa em outro.

As polaridades das palavras em um dicionário de sentimentos pode não ser necessariamente consistente (ou correta).

No artigo da Universidade de Chicago[1], os autores focam na detecção de inconsistências na atribuição de polaridade para as palavras e seus sinônimos e entre dicionários. Desta forma o artigo propõe a resolução do problemas em três passos para a determinação de inconsistências:

- ***A polaridade de uma palavra, de acordo com a distribuição de polaridade de seus sinonimos (Definição 1)***: A distribuição de polaridade de uma palavra é estimada analisando a distribuição de polaridade de seus sinônimos. Definimos a polaridade de uma palavra como: P+, P−, P0 onde eles representam as "probabilidades" que a palavra é positiva, negativa ou neutra, respectivamente. Por exemplo, a palavra barato tem a polaridade distribuição P+ = 0,81, P− = 0,19 e P0 = 0. A distribuição de polaridade de uma palavra é estimada usando as polaridades de seus sinônimos subjacentes. Por exemplo barato tem quatro sentidos, sendo o primeiro sentido positivo e os três últimos sentidos negativos. A probabilidade de a palavra expressar um sentimento negativo é P− = f2 + f3 + f4 / freq (barato) = 0,19, enquanto a probabilidade de a palavra expressar um sentimento positivo é P+ = f1 / freq (barato) = 0,81. P0 = 0 (1- 0.81 - 0.19). Dito isto a polaridade da palavra será a polaridade dominante.
- ***Assinalar a polaridade de uma palavra é um NP-completo***: O adjetivo *cômico* tem polaridade negativa no dicionário 1 (por exemplo por ser considerado um comportamento *ridículo*) e o adjetivo *divertido* tem polaridade positiva no dicionário 2 (por ser considerado um comportamento extrovertido). Através de dedução (ou seja, por aplicações sucessivas da Definição 1), a palavra *engraçado*, que não está presente em um dos dicionários, recebe polaridade negativa atribuída por causa da palavra *cômico* e polaridade positiva atribuída por causa da palavra *engraçado*. Uma palavra tem polaridade *p* se satisfizer a hipótese da Definição 1. A questão a ser respondida é: Dada uma atribuição de polaridades para as palavras, existe uma atribuição de polaridades aos sinônimos que concordam com o das palavras? Em outras palavras, dadas as polaridades de um subconjunto de palavras (por exemplo, a fornecida por um dos três Sinônimos) o problema de encontrar as polaridades é um problema "difícil" (*hard*) [1].
- ***Reduzir o Problema para um CNF-SAT***: O artigo desenvolveu um método de conversão de uma instância do problema de verificação da consistência da polaridade em um instância do CNF-SAT (forma normal conjuntiva). Para cada relação semântica *s* definimos três variáveis booleanas *s-*, *s+* e *s0*, correspondentes às polaridades negativa, positiva e neutra, respectivamente. Nesta seção, usamos -, +, 0 para denotar polaridades negativas, positivas e neutras, respectivamente. E aplicando a transformação CNF-SAT é possível verificar no exemplo 1 como identificar que a determinação de satisfabilidade para as palavras barato, não caro e ordinário é não satisfatível (*cheap* , *inexpensive* e *sleazy*).

Copyright © 2019 ISESC

Quadro 1. Definição de Polaridade.

Definição de acordo com a distribuição de polaridade
Seja *w* uma palavra e S_w suas relações semânticas. Cada relação semântica em S_w tem uma polaridade associada e uma frequência com relação a *w*. *w* tem polaridade *p* e *p* $\in$ {positivo, negativo} se há um subconjunto de relações semânticas $S' \subseteq S_w$ tal que cada relação semântica $s \in S'$ tem polaridade *p* e $\sum_{s \in S'} \frac{f(w,s)}{freq\ (w)} > 0{,}5$. S' é chamado subconjunto de polaridade dominante. Se não há nenhum subconjunto então *w* tem polaridade neutra. $S' \subseteq S_w$ é o mínimo subconjunto dominante de relações semânticas (MDSs - Minimally Dominant Subset of synsets) se a soma da relativa frequência das relações semânticas dominantes em S' é maior que 0,5 e a remoção de qualquer relação semântica *s* de S' fará com que a soma da relativa frequência das relações semânticas em S' - {*s*} menor ou igual a 0,5.

Quadro 2. Polaridade Inconsistente SAT.

SAT aplicado à inconsistência de polaridade
Considere um componente consistindo das palavras *w* = barato (*cheap*), *v* = não caro (*inexpensive*) e *u* = ordinário (*sleazy*). Barato (*cheap*) tem polaridade positiva, enquanto não caro (*inexpensive*) e ordinário (*sleazy)* tem polaridade negativa. As relações semânticas destas palavras são: $\{s^1, s^2, s^3, s^4\}$, $\{s^1\}$, $\{s^3, s^4, s^5\}$ respectivamente, conforme demonstrado nas figuras 1, 2 e, 3 respectivamente (WordNet [2]). A relativa frequência de s^3, s^4 e s^5 com relação a ordinário (*sleazy*) são iguais à ⅓. Temos no total 15 variáveis, 3 por relação semântica, s_-^i, s_+^i, s_0^i, $1 \leq i \leq 5$. O único MDS de barato (*cheap*) é $\{s^1\}$ e coincide com o de não caro (*inexpensive*). Para ordinário (*sleazy*) são $\{s^3, s^4\}$, $\{s^3, s^5\}$ e $\{s^4, s^5\}$. Para cada s^i precisamos de uma cláusula $C(s^i)$. Dito isto, $C(w,+) = s_+^1$, $C(v,-) = s_-^1$ e $C(u,-) = (s_-^3 \wedge s_-^4) \vee (s_-^3 \wedge s_-^5) \vee (s_-^4 \wedge s_-^5)$. Então $\Phi = \wedge_i C(s^i) \wedge [s_+^1 \wedge s_-^1 \wedge ((s_-^3 \wedge s_-^4) \vee (s_-^3 \wedge s_-^5) \vee (s_-^4 \wedge s_-^5))]$. Φ não está na forma normal conjuntiva (CNF) e precisa ser convertido. Para Φ ser verdadeiro, as cláusulas $C(w,+) = s_+^1$ e $C(v,-) = s_-^1$ precisam ser verdadeiras. Entretanto isto fará $C(s^1)$ falso. Portanto, Φ não é satisfatível. As cláusulas $C(w,+) = s_+^1$ e $C(v,-) = s_-^1$ não são satisfatíveis e então as polaridades de barato (*cheap*) e não caro (*inexpensive*) são inconsistentes.

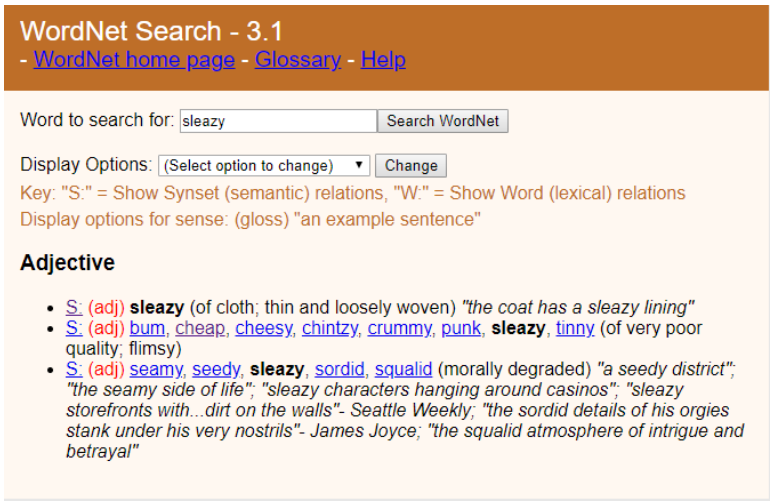


Figura 1. Dicionário WordNet [2], palavra *sleazy*.

Copyright © 2019 ISESC

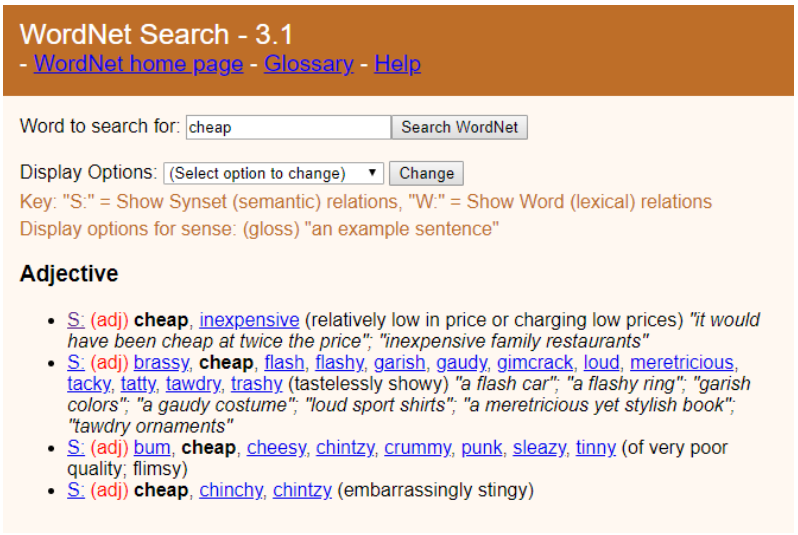


Figura 2. Dicionário WordNet [2], palavra *cheap*.

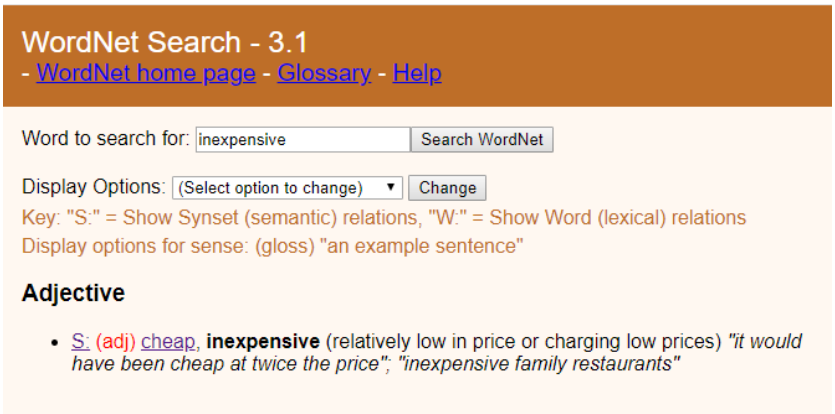


Figura 3. Dicionário WordNet [2], palavra *inexpensive*.

5. Outras Abordagens

5.1 Processamento Natural da Linguagem

Exemplo do Artigo Análise de Sentimento através de Autômata de Estados Finitos (*Sentiment analysis through Finite State Automata*) de 2015 [12], Universidade de Salerno, Itália que utiliza Automata de Estados Finitos para avaliar a composição de uma sentença. Trabalha com o valor de uma sentença (valência). Como exemplificado a seguir, uma negação nem sempre muda a polaridade de uma sentença. Ela pode incrementar ou decrementar o valor da sentença em termos de negação.

- Citroen não produz automóveis eficientes.

  Citroen non[neg] produce auto valide[+2] [-2]

- Os gráficos não são realmente espetaculares.

  Grafica non proprio[neg] spettacolare[+3] [-2]

5.2 Lógica Fuzzy

Há também diversas aplicabilidades de lógica *fuzzy* para análise aprofundada de uma sentença que poderão ser utilizados para identificar o sentimento de uma sentença, conforme demonstrado na tabela 2.

Copyright © 2019 ISESC

Tabela 1. Análise de Sentimento através de Processamento Natural de Linguagem.

Análise morfológica de adjetivos, advérbios, substantivos, "*shift contextual"*, polaridade reversa, negação, intensidade, verbos modais [12]	Automata de Estados Finitos aplicado na análise de sentimento por palavra, sentença, e também do dicionário para expressões complexas, como por exemplo opiniões com alto grau de variações, por exemplo valer a pena (*valerne la pena*). O trabalho mencionado foi realizado para o idioma italiano, entretanto a estrutura da linguagem tem proximidade com o português;
Resolução de Ambiguidade [13]	Exemplo de como identificar de forma apropriada qual a função de uma palavra na gramática de acordo com o contexto. Por exemplo qual o tipo do verbo, presente, passado ou futuro;

Tabela 2. Análise de Sentimento através de Lógica Fuzzy.

Polaridade da sentença através de granularidade [6]	Analisar fuzzy para estimar a granularidade da polaridade (por exemplo, *poor, slight, moderate, very mos* - pobre, leve, moderado, muito mais) da orientação semântica e sua intensidade para sentenças; por exemplo: Moderate: "*occasionally melodramatic*, *it's also extremely effective.*" ("Ocasionalmente melodramático, também é extremamente eficaz".)
Polaridade contextual identificando o sentimento [7]	Analisar polaridade contextual quando um texto pode apresentar diferentes variações de emoções; por exemplo *succumb to temptation* vs. *succumb to disease* (sucumbir à tentação vs. sucumbir à doença), pois são sentimentos diferentes um invoca vergonha e o outro medo.
Polaridade reversa, quando a polaridade da sentença é revertida pela segunda parte da sentença [8]	Identificar e reduzir a polaridade reversa de um texto, por exemplo, Bastante boa atuação, mas no geral um filme decepcionante, a polaridade da primeira parte da sentença é revertida pelo mas;
Análise de Sentimento dual para endereçar problemas de duas negações em uma mesma frase [9][10]	O algoritmo falha ao detectar duas negações em uma frase. Por exemplo: Eu descobri que estes dvds não funcionam bem no meu sistema, são lentos e não confiáveis. Eu não os recomendo. (será identificado como positivo). Ao aplicar a polaridade reversa através de antônimos: Eu descobri que estes dvds funcionam bem no meu sistema, são excelentes e lentos. Eu os recomendo. E dito isto a polaridade original é inconsistente;
Fuzzy Clustering para analisar a parte objetiva e subjetiva de uma sentença [11]	Analisar *fuzzy clustering* para identificar a parte subjetiva e objetiva de uma sentença. Por exemplo: a sentença subjetiva, Esse tipo de coisa é maravilhosa (*This kind of thing is amazing*) expressa uma opinião positiva enquanto a sentença subjetiva O paciente morreu há dois dias (*The patient died in two days*), expressa uma opinião implicita negativa. Dito isto ao analisar um sentimento é necessário avaliar a parte subjetiva e objetiva de uma opinião.

Copyright © 2019 ISESC

5.3 Lógica Paraconsistente

A lógica paraconsistente [20] é uma abordagem adicional para análise de sentimento aplicado à notícias falsas. Conforme descrito no artigo da University College London, intitulado "*How to act on inconsistent news: ignore, resolve or reject*" [16], utilizando lógica paraconsistente para identificar notícias mais inconsistentes do que outras. Apenas com polaridade de palavras não é possível identificar estas inconsistências. Além disto, algumas notícias falsas podem ou não ser significativas.

Para ilustrar a necessidade de avaliar o significado, considere duas reportagens sobre uma partida da Copa do Mundo, onde o primeiro relatório diz que o Brasil venceu a Alemanha por 2 a 0, e o segundo relatório diz que a Alemanha venceu o Brasil por 2 a 0. Esta é claramente uma significativa inconsistência. Agora considere duas notícias no mesmo jogo de futebol, onde o primeiro relatório diz que o árbitro foi Pierluigi Collina e o segundo relatório diz que o árbitro foi Ubaldo Aquino. Essa inconsistência normalmente seria considerada relativamente insignificante.

5.4 Outros Métodos de Aprendizado de Máquina

Adicionalmente outras abordagens de aprendizado de máquina são também possíveis, conforme demonstrado na tabela 3.

6. Conclusões

6.*1 Genéricas*

Estudamos o problema de verificar a consistência da polaridade para dicionários de palavras de sentimentos. Mostramos que esse problema é NP-completo. Mostramos que, na prática, inconsistências de polaridade de palavras tanto em um dicionário quanto em vários dicionários podem ser obtidas usando os solucionadores SAT. E identificar palavras com polaridades inconsistentes permite que os dicionários sejam melhorados

6.*2 Específicas*

Até a realização do trabalho citado pelo artigo estudado, nenhum dos trabalhos anteriores estudou o problema da verificação de consistência da polaridade para dicionários de sentimentos. Este experimento teve como objetivo mostrar que dois dicionários de sentimentos inconsistentes podem fornecer resultados diferentes *quando aplicados à tarefa de análise de sentimento.*

6.*3 Sugestões*

A verificação de consistência de polaridade poderia ser estendida para outras relações semânticas, como hipónimos (uma palavra cujo significado é hierarquicamente mais específico) e hiperónimos (uma palavra cujo significado é hierarquicamente mais abrangente). Por exemplo, rosa ou violeta são hipónimos de flor, e flor é hiperónimo de rosa ou violeta.

Verificar outras técnicas como Processamento de Linguagem Natural utilizando Autômata de Estados Finitos, uma vez que uma negação nem sempre muda a polaridade de um sentença, pois pode incrementar ou decrementar o valor da sentença em termos de negação.

6.*4 Recomendações*

O trabalho sobre subjetividade / sentidos objetivos da palavra mostra que existem situações em que os sentidos subjetivo e objetivo de uma palavra podem conflitar. A implicação dessa descoberta é que algumas relações semânticas podem não ter uma polaridade única, como assumido neste trabalho. Consequentemente, a recomendação é estender a técnica para relações semânticas com múltiplas polaridades.

Tabela 3. Outros métodos de aprendizagem de máquina.

Aprendizado Supervisionado para diversos domínios de mercados e produtos [14]	Revisão dos métodos supervisionados para diversos domínios de mineração de opiniões, como filmes, restaurantes, livros, entre outros.
Figuras de Linguagem como sarcasmo e ironia [15]	Exemplo de *twitter* sarcástico. Exemplo 1: Obrigado Google por dar sempre dois passos para a frente e três para trás. Exemplo 2: Adoro quando meu ônibus está atrasado) (expressão original: *absolutely adore it when my bus is late*).

Copyright © 2019 ISESC

Agradecimentos

Os autores gostariam de agradecer: 1) O Instituto Tecnológico de Aeronáutica (ITA); 2) A Fundação Casimiro Montenegro Filho (FCMF); e 3) A Empresa Ecossistema, pelo suporte no desenvolvimento deste trabalho acadêmico sobre análise de sentimento.

Referências

[1] Dragut WHYCSP E. Polarity Consistency Checking for Sentiment Dictionaries;2012.

[2] WordNet; 2019. Available from: http://wordnetweb.princeton.edu/perl/webwn

[3] A. Biere, M. Heule, H. Van Maaren, T. Walsh, Handbook of Satisfiability: Volume 185 Frontiers in Artificial Intelligence and Applications

[4] Ensinando as máquinas a entenderem emoções através da análise de sentimentos; 2019. Available from: https://www.infoq.com/br/news/2019/09/machines-understand-emotions/?itmsource=presentationsaboutMachineLearningitmmedium=linkitmcampaign=MachineLearning.

[5] Web mining module for Python, with tools for scraping, natural language proces-sing, machine learning, network analysis and visualization; 2019. Available from:https://github.com/clips/pattern.

[6] Appel O, Chiclana F, Carter J, Fujita H. A hybrid approach to the sentiment analysis problemat the sentence level. Knowledge-Based Systems. 2016;108:110–124.

[7] Poria S, Gelbukh A, Cambria E, Hussain A, Huang GB. EmoSenticSpace: A novel framework for affective common-sense reasoning. Knowledge-Based Systems. 2014;69:108–123.

[8] Xia R, Xu F, Yu J, Qi Y, Cambria E. Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis. Information Processing & Manage-ment. 2016;52(1):36–45.

[9] Xia R, Xu F, Zong C, Li Q, Qi Y, Li T. Dual sentiment analysis: Considering two sides of one review. IEEE transactions on knowledge and data engineering. 2015;27(8):2120–2133.

[10] Xia R, Wang C, Dai XY, Li T. Co-training for semi-supervised sentiment classification based on dual-view bags-of-words representation. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers); 2015. p. 1054–1063.22

[11] Wang X, Zhang H, Xu Z. Public sentiments analysis based on fuzzy logic for text. InternationalJournal of Software Engineering and Knowledge Engineering. 2016;26(09n10):1341–1360.

[12] Elia A, Pelosi S, Maisto A, Paolo II VG, Fisciano S. Sentiment analysis through Finite StateAutomata. 2015;.

[13] Voutilainen A, Tapanainen P. Ambiguity Resolution in a Reductionistic Parser. In: Proceedings of the Sixth Conference on European Chapter of the Association for Computational Linguistics. EACL '93. Stroudsburg, PA, USA: Association for Computational Linguistics;1993. p. 394–403.

[14] Lunardi A, Viterbo J, Bernardini F. Um Levantamento do Uso de Algoritmos de Aprendizado Supervisionado em Mineração de Opiniões. Proceedings of XII Encontro Nacional de Inteligência Artificial e Computacional–ENIAC. 2015;p. 262–269.

[15] Silva PDS. Avaliação do Desempenho de Métodos de Análise de Sentimentos na Presença das Figuras de Linguagem Sarcasmo e Ironia;

[16] Hunter A, How to act in inconsistent news: ignore, resolve or reject, 2006, Data and Knowledge Engineering

[17] A. Tarski. The Concept of Truth in Formalized Languages. In Logic Semantics, Metamathematics, A. Tarski, ed., Clarendon Press, Oxford, 1956.

[18] A. Tarski. Truth and Proof. Philosophy and Phenomenological Research, 4:341–75, 1944.

[19] A. Tarski. Contributions to the Theory of Models. Indagationes Mathematicae, 16:572–88, 1954.

[20] da Costa, N., 1974, "On the Theory of Inconsistent Formal Systems", Notre Dame Journal of Formal Logic, 15(4): 497–510.

Copyright © 2019 ISESC