

Análise de Sentimentos

Arthur
Carlos
Deyvirson
Rafael

Agenda

- Definição;
- Motivação;
- Problemas;
- Desafios;
- Objetivos;
- Classificação;
- Ferramentas;

Definição

“Análise de sentimento, também conhecida como mineração de opinião, é o campo de estudo que analisa as opiniões, sentimentos, avaliações, atitudes e emoções das pessoas em relação a entidades como produtos, serviços, organizações, indivíduos, eventos e seus aspectos” (LIU, 2012)

Motivação

Os principais fatores para o aumento das pesquisas foram:

- **Advento de métodos de aprendizado de máquinas;**
- **Disponibilidade de datasets para o treino/aprendizado de algoritmos;**
- **Oportunidade de mercado que a área oferece**

O(s) problema(s) de analisar sentimentos

- **Desafios e definições básicas;**
- Tipos de opiniões;
- Subjetividade e emoções;

Desafios básicos

- Opiniões e sentimentos são **subjetivos**;
- Palavras com sentido adverso dependendo do contexto;
- Diferentes fontes e variados formatos;



Desafios básicos

- 140 caracteres (Quantidade de palavras);
- Linguagem informal, gírias e emoticons;
- Palavras com duplo sentido;



O(s) problema(s) de analisar sentimentos

- Desafios e definições básicas;
- **Tipos de opiniões;**
- Subjetividade e emoções;

Tipos de opiniões

- Regular: uma opinião comum

- **Direta**: é uma opinião que expressa diretamente um sentimento a um determinado alvo.

“A pizza está ótima”

- **Indireta**: é uma opinião que expressa indiretamente um sentimento para um alvo.

“Logo após comer a pizza, passei mal”

Tipos de opiniões

- **Comparativa**: Expressa uma relação de similaridade ou diferença entre dois ou mais alvos.

“A programação do Telecine é melhor de que a da HBO”

O(s) problema(s) de analisar sentimentos

- Desafios e definições básicas;
- Tipos de opiniões;
- **Subjetividade e emoções;**

Subjetividade e emoções

- Subjetividade
 - Uma frase subjetiva expressa sentimentos pessoais
- Emoções
 - São sentimentos e pensamentos
 - Emoções primárias: amor, alegria, surpresa, raiva
 - Cada emoção tem intensidade diferente

Por que saber a opinião das outras pessoas ?

- Em qual candidato votar?
- Qual carro comprar?
- Qual filme assistir?
- Qual livro comprar?
- Como deixar meus clientes mais satisfeitos?

Por que saber a opinião das outras pessoas ?

- **Segundo uma pesquisa feita pelo grupo Kelsey**
 - **81% dos usuários** de internet já fizeram pesquisa online sobre algum produto
 - **Entre 73% e 87%** dos usuários dizem que uma crítica/análise/revisão influenciaram na compra de um produto
 - **20% a 99%** dos consumidores estão dispostos a pagar por um produto 5 estrelas em vez de um 4 estrelas

Análise de Sentimentos

- Visa identificar o sentimento apresentado a respeito de alguma entidade e traduzir as incertezas da emoção;
- Deve ser automatizada;
- Deve prover vantagem competitiva



Objetivos da Análise de Sentimentos

- Analisar produtos, marcas, empresas entre outros;
- Identificar, Interpretar e recomendar ações baseadas em sinais do mercado;
- Interagir com os consumidores;
- Ajudar usuários na tomada de decisão;



Etapas da Análise de Sentimentos

- Coleta de conteúdo (geralmente utilizando APIs de alguma mídia social);
- Classificação;
- Sumarização dos resultados (geralmente utilizando gráficos);
- Gerar conhecimento;

Desafios da Análise de Sentimentos

- Distinção entre opinião e fato;
- Identificação de opiniões de terceiros;
- Gírias, falhas na digitação e “internetês”;
- Sarcasmos e ironia;
- Referência a mais de um item de interesse;
- Propaganda disfarçada.

Recursos Disponíveis

scup 

Zubit *Making sense*

 **SEEKR**
Gestão e Monitoramento de Marcas em Mídias Sociais

livebuzz
social media marketing solution

 **opsocial**

e.life

socialmention*


WHOS TALKIN

KURRENTLY.com

Como classificar sentimentos ?

- **Classificação por polaridade**
- **Classificação por escala**
- **Classificação por análise de elementos e aspecto**
- **Classificação por análise de humor**

Classificação por polaridade

- A forma mais comum de classificação
- Classifica em positivas ou negativas
- Sem opinião explícita é classificada como neutra
- Objetiva e simplificada
- Limitada

Classificação por Escala

- Definida em uma escala 1 a 5 para as publicações
- Posts são considerados muito ou pouco positivo/negativo
- Facilita analisar críticas e comentários

Classificação por análise de elemento e aspecto

- Analisa diferentes opiniões em uma mesma publicação
- Divide a opinião em núcleos
- Análise realizada isoladamente
- Muito usado por marcas e empresas
- Complexo e demorado

Classificação por análise de humor

- Forma mais aprofundada de classificar
- Ir além do “positivo” ou “negativo”
- Pode variar entre termos como “insatisfação”, “frustração”, “revolta”, “satisfação”, “alegria”, “expectativa” entre outros
- Complexa porém útil para compreender reputação (empresas, marcas, produtos e governo)

Afinal, qual é a melhor maneira para classificar ?

- Cada forma de classificação possui particularidades e utilidades específicas;
- As dificuldades podem ainda ser empecilhos em algumas situações;
- É importante observar a necessidade específica e os objetivos de cada caso;

Classificação de sentimentos			
Classificação	Vantagens	Desvantagens	Indicações
Polaridade	<p>agilidade na análise</p> <p>maior consistência na análise manual e automática</p>	<p>pouca profundidade na análise</p>	<p>gestão de crise</p> <p>análise diária constante</p> <p>panoramas gerais</p>
Escala	<p>melhor entendimento de nível de satisfação / Insatisfação</p>	<p>classificação subjetiva</p> <p>Inconsistência na análise</p>	<p>estudos de nível de satisfação com produtos, marcas, governos</p>
Elemento/aspecto	<p>compreensão detalhada da opinião do usuário sobre um elemento</p>	<p>exige maior tempo de dedicação do analista</p>	<p>análise de opinião sobre diferentes características de produtos, serviços, governos</p>
Humor	<p>percepção aprofundada de sentimentos predominantes</p>	<p>classificação subjetiva</p> <p>Inconsistência na análise</p>	<p>análise de opiniões predominantes sobre marcas, empresas, governos</p>

Classificação de sentimentos em nível de documento

- O que é ?
 - Determinar o sentimento geral de um documento como positivo ou negativo;
 - Este documento pode ser um texto, uma postagem, um tweet, uma crítica de produto, etc;
- Dada uma hipótese:
 - O documento somente possui opiniões direcionadas para um único alvo;
 - Esta hipótese é boa para avaliação de produtos e serviços, mas não é boa para Fóruns e Blogs;

Classificação de sentimentos em nível de documento

- Abordagens:
 - Não existe uma bala de prata!
 - Incontáveis abordagens são utilizadas para classificar sentimentos;
 - Duas técnicas que são bastante recorrentes:
 - Parte do discurso (Part of speech - POS);
 - Palavras e frase de sentimentos;

Classificação de sentimentos em nível de documento

- Parte do discurso:
 - Nessa técnica adjetivos são usados como importantes indicadores de opiniões;
 - Advérbios são analisados como intensificadores de adjetivos;
 - Exemplo: "A minha empresa tem um **bom**(adjetivo) ambiente para se trabalhar, mas o salário é **muito**(advérbio) **ruim**(adjetivo)."
 - Identificadas as frases com opiniões, utiliza-se um dicionário de sentimentos;
- Palavras e frase de sentimentos:
 - Identificadas as frases com opiniões, utiliza-se um dicionário de sentimentos;

Classificação de sentimentos em nível de documento

- Problemas de Domínio:
 - A análise de sentimentos é sensível ao domínio estudado;
 - Palavras podem ter sentidos opostos em domínios diferentes;
 - Exemplo de domínios:
 - Livros;
 - Filmes;
 - Política;
 - etc;

Classificação de sentimentos em nível de documento

- Problemas de Idiomas:

- A análise de sentimentos é sensível ao idioma de estudo;
- Ao analisar a orientação semântica do seguinte documento:
 - "Tesla has the best cars ever made"
- Agora ao tentar analisar isso:
 - "Tesla possède les meilleures voitures jamais réalisés";
- Ou até mesmo:
 - 特斯拉有最好的汽车

Classificação de sentimentos em nível de documento

- Resumindo:
 - Extrair opiniões e analisar sentimentos de um documento inteiro possui falhas;
 - O sentimento geral não é detalhado;
 - Exemplos como Blogs e Fóruns não podem ser analisados corretamente;

Classificação de sentimentos em nível de frase

- Não existe uma diferença fundamental entre este nível e o nível de documento;
- Problema a ser resolvido:
 - Dada uma frase, determinar se ela expressa um sentimento positivo, negativo ou neutro;
- Pontos importantes:
 - Saber se uma frase é subjetiva ou não;

Classificação de sentimentos em nível de frase

- Subjetividade em frases
 - Uma frase pode ser classificada em duas classes:
 - Frase objetiva (Expressa fatos);
 - Frase subjetiva (Expressa opiniões);
 - Alguns autores classificam frases buscando a presença de pronomes, adjetivos, números cardinais, advérbios;

Classificação de sentimentos em nível de frase

- Classificação de frases
 - Dada uma hipótese:
 - Uma frase expressa um único sentimento para um único alvo;
 - Dicionários de sentimentos(Wordnet) são utilizados para classificar frases;
 - Uma das técnicas utilizadas é realizar um somatório da pontuação semântica dos sentimentos:
 - + 1 para sentimento positivo;
 - - 1 para sentimento negativo;

Classificação de sentimentos em nível de frase

- Classificação de frases

- Exemplo: "A minha empresa tem um **bom** ambiente para se trabalhar, mas o salário é **muito ruim**."
 - +1 para "bom";
 - -1 para "ruim";
 - Resultado: frase neutra;

Classificação de sentimentos em nível de frase

- Classificação de frases
 - Existe uma abordagem similar que **multiplica** a pontuação semântica dos sentimentos:
 - +1 para um sentimento positivo;
 - -1 para um sentimento negativo;
 - Para o exemplo anterior com a técnica de multiplicação:
 - +1 para "bom";
 - -1 para "ruim";
 - Resultado: frase negativa;

Classificação de sentimentos em nível de frase

- frases condicionais:
 - Uma frase condicional descreve uma implicação ou situação hipotética e suas consequências: causa e efeito;
 - É um problema difícil pois nem sempre frases condicionais expressam opinião;

Classificação de sentimentos em nível de frase

- frases condicionais:
 - Uma frase condicional descreve uma implicação ou situação hipotética e suas consequências: causa e efeito;
 - É um problema, pois nem sempre frases condicionais expressam opinião;
- frases interrogativas:
 - Identificar opiniões em perguntas não é algo popular;
 - Os trabalhos nesse campo de pesquisa são escassos;
 - Exemplo: "Alguém pode me dizer como consertar esse lento celular Samsung?";

Classificação de sentimentos em nível de frase

- Sarcasmo:
 - Se frases condicionais e interrogativas são consideradas difíceis;
 - Sarcasmo é mais fácil ainda!
 - Existem alguns trabalhos na área para identificação de sarcasmo, mas não classificam sentimentos;

Classificação de sentimentos em nível de frase

- Problemas de idiomas:
 - Temos o mesmo problema correlato ao de classificação de sentimentos em nível de documento;
- Estratégias para problema de idiomas:
 - Algumas estratégias são:
 - Traduzir frases do idioma alvo para o idioma fonte e então classificá-las;
 - Traduzir frases do idioma fonte para o idioma alvo e então classificá-las;
 - Traduzir um dicionário de sentimentos do idioma fonte para o idioma alvo e construir um novo classificador;

Classificação de sentimentos em nível de frase

- Resumindo:
 - Para classificar frases é preciso:
 - Diferenciar frases objetivas de subjetivas;
 - Assumir que a frase tenha um único sentimento para um único alvo;
 - Classificar o sentimento de uma frase como positivo, negativo, neutro ou sem opinião;
 - frases condicionais são difíceis;
 - frases interrogativas são mais ainda;
 - Sarcasmo então, nem se fala;

Classificação de Sentimento no Nível de Aspecto

- Motivação

Classificação de Sentimento no Nível de Aspecto

- Etapas necessárias para a classificação no nível de aspecto:
 - Extração de entidades
 - Extração de aspecto
 - Extração do *opinion holder*
 - Extração da hora
 - Classificação do sentimento dos aspectos

Classificação de Sentimento no Nível de Aspecto

- Duas abordagens principais:
 - Aprendizagem de Máquina
 - Baseada em Léxicos

Classificação de Sentimento no Nível de Aspecto

- Aprendizagem de Máquina
 - O alvo da opinião é considerado como *feature* para treinar o classificador
 - Duas abordagens:
 - Geração de um conjunto de *features* que são dependentes da entidade alvo ou aspecto na frase
 - Determinar o escopo da aplicação de cada expressão de sentimento para determinar se ela cobre a entidade alvo ou aspecto na frase

Classificação de Sentimento no Nível de Aspecto

- Aprendizagem de Máquina
 - O alvo da opinião é considerado como *feature* para treinar o classificador
 - Duas abordagens:
 - **Geração de um conjunto de *features* que são dependentes da entidade alvo ou aspecto na frase**
 - Determinar o escopo da aplicação de cada expressão de sentimento para determinar se ela cobre a entidade alvo ou aspecto na frase

Classificação de Sentimento no Nível de Aspecto

- Aprendizagem de Máquina

- Exemplo:
- Jiang et Al. (2011) usaram uma árvore sintática para gerar um conjunto de *features* dependentes do alvo da opinião, que representam alguma relação sintática com a entidade ou aspecto alvo e uma palavra.
- w_i é uma palavra e T a entidade ou aspecto alvo.
- Alguns exemplos de *features* dependentes do alvo usados em Jiang et Al. (2011) são os seguintes: se w_i é um verbo transitivo e T é seu objeto, a *feature* w_i_arg2 é gerada, onde *arg* significa argumento. Se w_i é verbo transitivo e T é seu sujeito, a *feature* w_i_arg1 é gerada.

Classificação de Sentimento no Nível de Aspecto

- Baseado em dicionários léxicos

A classificação baseada em léxico pode ser definida como segue: ela usa (1) um léxico de sentimento incluindo palavras e/ou frases que expressam sentimento, (2) um conjunto de regras para manipular *constructs* de linguagem (e.g. alternadores e cláusulas adversativas) e tipos de frase, (3) Uma função de agregação de sentimento ou um conjunto de relação de alvos e expressões de sentimento derivadas de uma árvore de *parse* para determinar a orientação de sentimento de cada alvo na frase.

Classificação de Sentimento no Nível de Aspecto

Método simples de Ding et. Al. (2008):

1. Marcar expressões de sentimento: (+1; -1)
2. Aplicar alternadores de sentimento.
3. Manipular cláusulas adversativas.
4. Agregar sentimento de *score*: Nesta última etapa, uma função de agregação de sentimento é aplicada aos *scores* de sentimento restantes para determinar a orientação final de sentimento de cada aspecto na frase.

Classificação de Sentimento no Nível de Aspecto

Aprendizagem de Máquina	Léxico
Pode aprender todos os tipos de <i>features</i> através de otimização	Muitas <i>features</i> são difíceis de usar, porque as pessoas não sabem como aplicá-las sem um algoritmo.
Difícil de escalar para um grande número de domínios.	Independente de domínio.
Necessita que o usuário rotule um grande número de dados para cada domínio.	Não requer necessariamente rotulação para cada domínio.
É mais difícil de melhorar porque as frases do problema tem que aparecer várias vezes para que o algoritmo escolha os padrões corretos.	Mais flexível no sentido que o sistema pode ser expandido e melhorado.
	Requer um alto investimento em tempo e esforço para construir uma base inicial de conhecimento do léxico, padrões e regras.

Ferramentas e aplicações - SentiWordNet

- Baseado no dicionário léxico *wordnet*
- É um recurso léxico para mineração de opinião
- Atribui três *scores*: positividade, negatividade, objetividade
- <http://sentiwordnet.isti.cnr.it/>

Ferramentas e aplicações - SentiStrength

- Analisa a força de sentimento positivo ou negativo em textos curtos
- Download: <http://sentistrength.wlv.ac.uk/>



SentiStrength

Quick Tests (English version):

Enter text:

Output: ☒ Dual, ☐ binary, ☐ trinary, ☐ scale

Ferramentas e aplicações - SASA (Sail/Ail Sentiment Analyzer)













- Ferramenta construída com um modelo treinado a partir de *tweets* sobre a eleição presidencial dos EUA em 2012
- <https://pypi.python.org/pypi/sasa/0.1.3>

Ferramentas e aplicações - Globo Esporte: Humor das Torcidas

HUMOR DAS TORCIDAS

Veja quais são as torcidas mais otimistas e pessimistas no Twitter. Os dados referem-se à análise de sentimento das últimas 3 horas. Participe e compartilhe a sua opinião.

Clique no seu time e veja os dados completos.

TIME	AVALIAÇÃO DO TORCEDOR	TWEETS	Como está o seu humor?
1  Chapecoense	83%  17%	72	 
2  Corinthians	82%  18%	3,782	 
3  Ponte Preta	76%  24%	101	 

Ferramentas e aplicações - iFeel 2.0

- Aplicação web que promete detectar sentimento em qualquer forma de texto
- Usa 18 métodos de análise de sentimento estados da arte: sentiwordnet, sentistrength, Sentiment140, StanfordDeep Learning, Emoticons
- <http://blackbird.dcc.ufmg.br:1210/>

Welcome to iFeel 2. 😊

Referências

- LIU, Bing. Sentiment analysis and opinion mining. **Synthesis lectures on human language technologies**, v. 5, n. 1, p. 1-167, 2012.
- LIU, Bing. **Sentiment analysis: Mining opinions, sentiments, and emotions**. Cambridge University Press, 2015.