ANÁLISE DE SENTIMENTOS

Naelson Douglas, Victor Holanda naelson@ic.ufal.br, vglh@ic.ufal.br PPGI - UFAL Sistemas Suportados por ML

SUMÁRIO

- Conceitos Básicos
- Processamento de Linguagem Natural
- Etapas
 - Identificação
 - Classificação
 - Agregação
- Técnicas
 - Léxica
 - Aprendizado de Máquina
 - Estatística
 - Semântica
- Prática
 - Dificuldades
 - Ferramentas
 - Exemplos

- Tomadas de decisão
- Recursos
- Internet, Mídias Sociais
- Análise de Sentimentos
- Novos recursos

- Texto que contenha sentimentos ou opiniões
 - Entidade, Aspecto, Sentimento, Detentor, Data

- Texto que contenha sentimentos ou opiniões
 - Entidade, Aspecto, Sentimento, Detentor, Data
 - Entidade
 - Produto
 - Serviço
 - Tópico
 - Assunto
 - Problema
 - Pessoa
 - Organização
 - Evento

- Texto que contenha sentimentos ou opiniões
 - Entidade, Aspecto, Sentimento, Detentor, Data
 - Aspecto
 - Propriedade ou
 - Característica da entidade em questão
 - Não é obrigatória

- Texto que contenha sentimentos ou opiniões
 - Entidade, Aspecto, Sentimento, Detentor, Data
 - Sentimento
 - Positivo
 - Negativo
 - Neutro
 - Ou ainda: raiva, medo, alegria, ...

- Texto que contenha sentimentos ou opiniões
 - Entidade, Aspecto, Sentimento, Detentor, Data
 - Detentor é a pessoa que escreveu o texto e expressou o sentimento

- Texto que contenha sentimentos ou opiniões
 - Entidade, Aspecto, Sentimento, Detentor, Data
 - Data: Quando o texto foi escrito

- Texto que contenha sentimentos ou opiniões
 - Um único texto pode conter
 - Várias entidades
 - Mais de um aspecto de cada entidade
 - Vários sentimentos

- Texto que contenha sentimentos ou opiniões
 - Um único texto pode conter
 - Várias entidades
 - Mais de um aspecto de cada entidade
 - Vários sentimentos
 - Exemplo:
 - "O Instituto de Computação provê máquinas com ótimo desempenho! A equipe é atenciosa e gentil"

- Texto que contenha sentimentos ou opiniões
 - Um único texto pode conter
 - Várias entidades
 - Mais de um aspecto de cada entidade
 - Vários sentimentos
 - Exemplo:
 - "O Instituto de Computação provê máquinas com ótimo desempenho! A equipe é atenciosa e gentil"
 - Entidade: Instituto de Computação

- Texto que contenha sentimentos ou opiniões
 - Um único texto pode conter
 - Várias entidades
 - Mais de um aspecto de cada entidade
 - Vários sentimentos
 - Exemplo:
 - "O Instituto de Computação provê máquinas com ótimo desempenho! A equipe é atenciosa e gentil"
 - Entidade: Instituto de Computação
 - Aspectos: máquinas; equipe;

- Texto que contenha sentimentos ou opiniões
 - Um único texto pode conter
 - Várias entidades
 - Mais de um aspecto de cada entidade
 - Vários sentimentos
 - Exemplo:
 - "O Instituto de Computação provê máquinas com ótimo desempenho! A equipe é atenciosa e gentil"
 - Entidade: Instituto de Computação
 - Aspectos: máquinas; equipe;
 - Sentimentos positivos: ótimo desempenho; atenciosa; gentil;

- Análise de sentimentos estrutura textual
 - Documento
 - Sentença
 - Entidade e Aspecto

- Análise de sentimentos estrutura textual
 - **Documento**: expressa sentimentos sobre apenas UMA ENTIDADE
 - Exemplo: Revisão de um produto

- Análise de sentimentos estrutura textual
 - **Documento**: expressa sentimentos sobre apenas UMA ENTIDADE
 - Exemplo: Revisão de um produto
 - **Sentenças** ou frases: são analisadas independentemente para identificar possíveis sentimentos (ou ausência de sentimentos)

- Análise de sentimentos estrutura textual
 - **Documento**: expressa sentimentos sobre apenas UMA ENTIDADE
 - Exemplo: Revisão de um produto
 - **Sentenças** ou frases: são analisadas independentemente para identificar possíveis sentimentos (ou ausência de sentimentos)
 - Entidade e Aspecto: não considera a estrutura do texto ou documento. Objetivo de identificar sentimentos sobre entidades ou seus aspectos

- Não existe um padrão ou regra de como medir/representar sentimentos
 - Forma mais conhecida: POLARIDADE
 - Feita de forma discreta: Positiva, Negativa, Neutra. Ou ainda numérica (-1, 1)
 - **EMOÇÕES**: Classificadas com outros modelos
 - Ekman sugere: Surpresa, Alegria, Tristeza, Raiva, Medo e Nojo
 - Plutchik: Ekman + Confiança e Expectativa
 - Parrot: Amor, Alegria, Surpresa, Raiva, Tristeza, Medo
 - Modelo Valence-Arousal-Dominance
 - Três componentes:
 - Valência quanto agradável é um estímulo (infeliz feliz) [0-10]
 - Excitação intensidade [0-10]
 - Dominância poder de um estimulo (fraco/submisso, forte/dominante) [0-10]

- NLP Trata da Manipulação e Transformação do Texto
 - Essencial no **pré-processamento** da Análise de Sentimentos
- Realiza a compreensão da linguagem em diferentes níveis linguísticos
- Processos/Conceitos
 - Tokenização
 - Radical
 - Steam
 - Lema
 - Etiquetação Morfossintática

- NLP está relacionado a compreensão da linguagem (oral ou escrita)
- Classificado em Níveis Linguísticos
 - Fonético, Morfológico, Sintático, Semântico, Pragmático

- Classificado em Níveis Linguísticos
 - Fonético Sons

- Classificado em Níveis Linguísticos
 - Fonético Sons
 - Morfológico Construção / Componentes

- Classificado em Níveis Linguísticos
 - Fonético Sons
 - Morfológico Construção / Componentes
 - Sintático Relações Estruturais

- Classificado em Níveis Linguísticos
 - Fonético Sons
 - Morfológico Construção / Componentes
 - Sintático Relações Estruturais
 - Semântico Significado

- Classificado em Níveis Linguísticos
 - Fonético Sons
 - Morfológico Construção / Componentes
 - Sintático Relações Estruturais
 - Semântico Significado
 - Pragmático Frases e Sentenças
 - Contextos Diferentes
 - Significado de um discurso

PARA ANÁLISE DE SENTIMENTOS OS PRINCIPAIS NÍVEIS SÃO

- Fonético Sons
- Morfológico Construção / Componentes
- Sintático Relações Estruturais
- Semântico Significado
- Pragmático Frases e Sentenças
 - Contextos Diferentes
 - Significado de um discurso

- Técnica de Tokenização
 - Identificação de termos (ou tokens)
 - Remoção de caracteres separadores
 - Espaços em branco
 - Pontuações
 - Quebras de linha
 - Objetivo: Listagem dos tokens que serão processados → Identificar possíveis sentimentos em um texto
 - Exemplo: "Bom dia, o sistema está lento novamente."
 - Tokens: 'bom' 'dia' 'o' 'sistema' 'está' 'lento' 'novamente'

- Técnica de Tokenização
 - Remoção de Stop Words
 - Palavras ou termos irrelevantes para a análise de sentimentos
 - 'bom' 'dia' 'o' 'sistema' 'está' 'lento' 'novamente'

- Técnica de Tokenização
 - Remoção de Stop Words
 - Palavras ou termos irrelevantes para a análise de sentimentos
 - 'bom' 'dia' 'o' 'sistema' 'está' 'lento' 'novamente'

RADICAL – STEMMING – LEMA

- Radical: elemento básico de uma palavra
 - Sem prefixo infixo sufixo
 - Exemplo: igualdade, igualar, desigual
- Stemming: Técnica computacional para identificar o radical de uma palavra
- Lema: modo deflexionado de uma palavra (forma essencial)
 - Exemplo: (Errar Erroneamente Errado) → Erro

- Etiquetação Morfossintática (Part-of-speech ou POS)
 - Identificar classe gramatical de um termo
 - Tokens associado a Classes Gramaticais considerando Contexto
 - Filtrando as que apresentarem SENTIMENTOS
 - Palavras com a mesma GRAFIA podem conter
 - Diferentes classes gramáticas, significados e sentimentos dependendo do contexto
 - Exemplo: Bom dia, o sistema está lento novamente
 - Adjetivos: Bom, Lento
 - Substantivo: Sistema, Dia
 - Adverbio: Novamente
 - Verbo: Está

• Etapas: Identificação – Classificação – Agregação

- Etapas: Identificação Classificação Agregação
 - Identificação: separação de conteúdo subjetivo que apresenta sentimentos sobre assunto ou alvo em um texto

- Etapas: Identificação Classificação Agregação
 - Identificação: separação de conteúdo subjetivo que apresenta sentimentos sobre assunto ou alvo em um texto
 - Classificação: determinar quais sentimentos foram identificados no texto
 - Positivos, Negativos, Neutros, Raiva, Alegria

- Etapas: Identificação Classificação Agregação
 - Identificação: separação de conteúdo subjetivo que apresenta sentimentos sobre assunto ou alvo em um texto
 - Classificação: determinar quais sentimentos foram identificados no texto
 - Positivos, Negativos, Neutros, Raiva, Alegria
 - Agregação: Consolida os sentimentos encontrados afim de encontrar o sentimento final do assunto ou alvo

- Etapa de Identificação (1/3)
 - Entrada: Texto (ou conjunto de textos)
 - Objetivo: Analisar entidades e seus respectivos aspectos
 - Determina previamente as entidades e aspectos possíveis
 - Exemplo:
 - Entidades Times
 - Aspectos Defesa, ataque, vitórias, derrotas, gols, etc.
 - Frase: O Flamengo foi campeão em 2019, com um ótimo ataque, mas com uma defesa satisfatória
 - Entidade: Flamengo
 - Aspectos: Defesa e Ataque
 - Sentimentos: Ótimo, Satisfatória

- Etapa de Identificação (1/3)
 - Esta etapa possibilita separar conteúdos com sentimentos/opiniões daqueles que não apresentam

- Etapa de Identificação (1/3)
 - Esta etapa possibilita separar conteúdos com sentimentos/opiniões daqueles que não apresentam
 - Exemplo: Ontem saímos para o cinema. A temperatura era ótima.

- Etapa de Identificação (1/3)
 - Esta etapa possibilita separar conteúdos com sentimentos/opiniões daqueles que não apresentam
 - Exemplo: Ontem saímos para o cinema. A temperatura era ótima.
 (Sem sentimento) (Com sentimento)

- Etapa de Classificação (2/3)
 - Sentimentos associados a entidades e aspectos
 - Sentimentos podem ser atribuídos a
 - Polaridade
 - Emoção
 - 1. Identificar entidades e aspectos
 - II. Classificação: associa sentimento ou opinião a estas entidades ou aspectos

- Etapa de Classificação (2/3)
 - Exemplo: "O restaurante tinha uma decoração bonita, mas a comida demorou e ainda veio fria"
 - Identificação:
 - Entidade: restaurante
 - Aspectos: decoração, comida
 - Classificação
 - Decoração: bonita polaridade positiva, emoção alegria
 - Comida: Demorou polaridade negativa, emoções raiva e tristeza

Fria – polaridade negativa, emoção tristeza

- Consolidação dos sentimentos
 - Texto Negativo (+,-,-)
 - Emoções Alegria, raiva, tristeza

- Etapa de Agregação (3/3)
 - Sumarização dos sentimentos
 - Etapa responsável por consolidar os sentimentos identificados na etapa de classificação
 - Determina o sentimento final sobre a entidade

Etapa de Agregação (3/3)

- Sumarização dos sentimentos
- Etapa responsável por consolidar os sentimentos identificados na etapa de classificação
- Determing o sentimento final sobre a entidade
- Formas de Expressar o Sentimento
 - Categórico: Negativo, Neutro, Positivo
 - Numérico: -1 0 1

Etapa de Agregação (3/3)

- Sumarização dos sentimentos
- Etapa responsável por consolidar os sentimentos identificados na etapa de classificação
- Determing o sentimento final sobre a entidade
- Formas de Expressar o Sentimento
 - Categórico: Negativo, Neutro, Positivo
 - Numérico: -1 0 1
- Média aritmética das classificações
 - Exemplo: O filme é cansativo. O elenco é péssimo.

Sentimento da frase: (-0.5 - 1) / 2 = -0.75

Etapa de Agregação (3/3)

- Sumarização dos sentimentos
- Etapa responsável por consolidar os sentimentos identificados na etapa de classificação
- Determing o sentimento final sobre a entidade
- Formas de Expressar o Sentimento
 - Categórico: Negativo, Neutro, Positivo
 - Numérico: -1 0 1
- Média aritmética das classificações
 - Exemplo: O filme é cansativo. O elenco é péssimo.

$$(-0,5)$$
 (-1)

Sentimento da frase: (-0.5 - 1) / 2 = -0.75

Negativo

- Etapa de Agregação (3/3)
 - Limites próximos a zero Resultado **Neutro**
 - Exemplo:
 - Considere os limites -0,01 e 0,01 = Neutro
 - Tokens: Legal (0,1), Chato (-0,09)
 - Média Aritmética: (0,1-0,09)/2=0,005
 - Classificação após agregação: NEUTRO

- Etapa de Agregação (3/3)
 - Para sentimentos Discretos
 - Quanto mais vezes a emoção é encontrada mais forte ela é

- Etapa de Agregação (3/3)
 - Para sentimentos Discretos
 - Quanto mais vezes a emoção é encontrada mais forte ela é
 - "O filme é cansativo"
 "As atuações são péssimas"

- Etapa de Agregação (3/3)
 - Para sentimentos Discretos
 - Quanto mais vezes a emoção é encontrada mais forte ela é
 - "O filme é cansativo" "As atuações são péssimas" (Tristeza) (Tristeza + Raiva)

- Etapa de Agregação (3/3)
 - Para sentimentos Discretos
 - Quanto mais vezes a emoção é encontrada mais forte ela é
 - "O filme é cansativo" "As atuações são péssimas" (Tristeza + Raiva)
 - Agregação
 - Considera que Tristeza e Raiva caracterizam o filme.
 - Tristeza é mais forte por aparecer duas vezes

- Léxica
- Aprendizado de Máquina
- Estatística
- Semântica

- Léxica
- Aprendizado de Máquina
- Estatística
- Semântica

MAIS POPULARES

- LÉXICA
 - Não supervisionada
 - Requer dicionários (léxicos)
 - Tokens possuem rótulo/pontuação → Denotam Sentimentos
 - Tokens podem ser palavras ou expressões
 - Para analisar quais sentimentos uma palavra ou expressão apresenta
 - Verificar a presença da mesma no dicionário

- LÉXICA
 - Exemplo de dicionário léxico

Token	Polaridade	Alegria	Surpresa	Raiva	Tristeza
péssimo	-1	não	não	sim	sim
ruim	-0,5	não	não	sim	sim
ótimo	1	sim	não	não	não
rápido	0,25	sim	sim	não	não
muito rápido	0,5	sim	sim	não	não

- Dicionários devem apresentar os tokens em formato de lemas ou radicais
 - Lematização e Stemming

- APRENDIZADO DE MÁQUINA
 - Busca identificar regras gerais em grandes conjunto de dados
 - Gerar Modelo: Extrair automaticamente Informações implícitas
 - Método supervisionado
 - Etapas
 - Aprendizado
 - Aplicação do Modelo

- APRENDIZADO DE MÁQUINA
 - Etapa de Aprendizado:
 - Executa um modelo de classificação usando um conjunto de dados de treino
 - Dados rotulados para treinar o modelo
 - Cada dado para treino representa um texto, seus atributos são:
 - Alvo: Classificação do sentimento
 - Features: tokens que estão sendo classificados, exportados de textos que apresentam sentimentos

- Aprendizado de Máquina
 - Exemplo: "O sistema ficou lento e feio"

DADOS DE TESTE GERADOS:

	Sistema	Lento	Feio	Layout	Não	Agradável	Muito	Prático	Polaridade
1	sim	sim	sim	não	não	não	não	não	Negativo
2									
3									

- Aprendizado de Máquina
 - Exemplo: "O layout do sistema é feio e não é agradável"

DADOS DE TESTE GERADOS:

	Sistema	Lento	Feio	Layout	Não	Agradável	Muito	Prático	Polaridade
1	sim	sim	sim	não	não	não	não	não	Negativo
2	sim	não	sim	sim	sim	sim	não	não	Negativo
3									

- Aprendizado de Máquina
 - Exemplo: "O sistema é muito prático"

DADOS DE TESTE GERADOS:

	Sistema	Lento	Feio	Layout	Não	Agradável	Muito	Prático	Polaridade
1	sim	sim	sim	não	não	não	não	não	Negativo
2	sim	não	sim	sim	sim	sim	não	não	Negativo
3	sim	não	não	não	não	não	sim	sim	Positivo

- Aprendizado de Máquina
 - A escolha de features é essencial para correta classificação
 - Técnicas de Lematização, Stemming e Etiquetação morfossintática normalmente são adotadas
 - É comum restringir a escolha de features a adjetivos e advérbios
 - Representando-os de forma binária
 - A negação também pode ser considerada
 - Outras formas de seleção de features
 - Pesos ao invés de binária
 - Considerar emoticons: © 8 9
 - Palavras alongadas: "Legaaaaallll" "Booommm"
 - Pontuação repetida: "!!!!!" "?????"

- Aprendizado de Máquina
 - Para testar o modelo:
 - Validação Cruzada
 - Treino
 - Teste
 - Etapa de aplicação: Executa o modelo treinado sobre os textos que se deseja classificar
 - Algoritmos de Classificação com Aprendizagem de Máquina
 - Support Vector Machine (SVM)
 - Naïve Bayes
 - Maximum Entropy
 - Redes Neurais

- Estatística e Semântica
 - Menos populares
 - Baseia-se em termos encontrados juntos para detectar sentimentos
 - Um sentimento que um termo expressa pode ser identificado pela ocorrência com outros termos que são conhecidos
 - Abordagem não supervisionada
 - Modelos Estatísticos
 - LDA Latente Dirichlet Allocation
 - LSA Semantic-orientation Latent Semantic Analysis
 - PMI Pointwise Mutual Information

- Dificuldades
- Ferramentas
- Exemplos

- Dificuldades
 - Identificar opiniões implícitas
 - Separar sentimentos de relatos negativos (Noticia de tragédia)
 - Ironia e Negação

- Dificuldade Opinião Implícita
 - A maioria das ferramentas e técnicas consideram apenas Casos Explícitos
 - Casos implícitos necessitam de uma análise complexa

- Dificuldade Separar sentimentos de relatos
 - Texto que são negativos em sua essência
 - Problemas
 - Necessidades
 - Tragédias
 - Não necessariamente expressam sentimentos ou opiniões do escritor

- Dificuldade Ironia
 - O real sentimento expressado é o inverso
 - Representa diferentes emoções do que está escrito
 - Necessário analisar o contexto inserido
 - Exemplo: "Como este político é ótimo"

- Dificuldade Negação
 - Inverter o significado do sentimento
 - Exemplo: "O quadro não é nada bonito"
 - Bonito: sentimento positivo
 - Nada: inverte o sentimento
 - Não: confirma o "nada bonito"
 - Atenuar um sentimento
 - Exemplo: "A comida não é maravilhosa"
 - Não necessariamente a comida é ruim
 - Apesar de n\u00e3o ser maravilhosa, ela pode ser muito boa
 - A negação está atenuando o sentimento

- Dificuldade Entidade e Aspectos variando com o contexto
 - Exemplo: Contexto Futebol
 - Notícias esportivas que utilizem o termo "tricolor"
 - Paulista: "São Paulo"
 - Gaúcha: "Grêmio"
 - Carioca: "Fluminense"

- Dificuldade Termos e sentenças negativas em contexto geral
 - Exemplo:
 - "O arquivo foi executado"
 - A palavra "executado" pode passar um sentimento negativo em um contexto geral, como na frase "homem foi executado", quando na realidade é apenas um termo neutro da Tl.

- Dificuldade Impacto
 - "O jogador de futebol caiu"
 - "O sistema de empresa caiu"
 - O termo "Caiu" pode ser considerado negativo em ambos, porém o impacto é muito maior na empresa quando comparado ao jogador

- Ferramentas
 - Processamento de Linguagem Natural
 - Classificação
 - Análise de Sentimentos em Geral
 - Léxicos

- Ferramentas
 - Processamento de Linguagem Natural e Análise de Sentimentos
 - TreeTagger
 - Lematização e Etiquetação Morfossintática
 - StanfordNLP
 - Etiquetação Morfossintática, Tokenização
 - NLTK
 - Tokenização, Etiquetação Morfossintática, Lemmatization
 - SensiStrenght
 - Análise de Polaridade (-5, 0, 5)
 - WordNet
 - Léxico genérico, Criação e expansão de outros léxicos
 - Classificação

- Ferramentas
 - Textblob
 - VADER

- Exemplos
 - Notebook
 - API Twitter

REFERÊNCIAS

LUMINA – UFRGS: Análise de Sentimentos em Computação. Disponível em: https://lumina.ufrgs.br/ Acesso em 22 de setembro de 2020.

F. Benevenuto. Métodos para Análise de Sentimentos em mídias sociais. Disponível em: https://homepages.dcc.ufmg.br/~fabricio/download/webmedia-short-course.pdf Acesso em 19 de setembro de 2020.