

The background is a light pink color with numerous small, semi-transparent pink hearts scattered across it. A thick black frame surrounds the central text area, with the top-left and bottom-right corners being solid black rectangles.

SENTIMENT ANALYSIS E OS SENTIMENTOS NA LÍNGUA PORTUGUESA

BARBARA CRISTINA RAMOS

Seminário de Tese – 12.06.2019

Sentiment Analysis

(Taboada, 2016)

- A tarefa:
 - ter informação o suficiente (independente da abordagem escolhida) para tornar possível a extração de informação (e, então, detectar a polaridade de cada caso);
- O objetivo geral:
 - Determinar a subjetividade do texto e sua orientação semântica (positiva, negativa ou neutra);
- Os aspectos do processo;
- Principais abordagens: machine learning e baseada em léxico.

Sentiment Analysis em uso



- Análise de **críticas/resenhas** de livros, filmes e produtos:
 - Aplicação direta em mecanismos de busca;
 - Relevante para outros consumidores e também produtores.
- Na **política**:
 - Tweets sobre partidos políticos / candidatos / personagens do cenário político;
 - Indicativos de resultados de eleição;
 - Mais recentemente: características dos autores dos comentários e suas interações online!

Como detectar o sarcasmo no texto escrito?

■ No mercado financeiro:

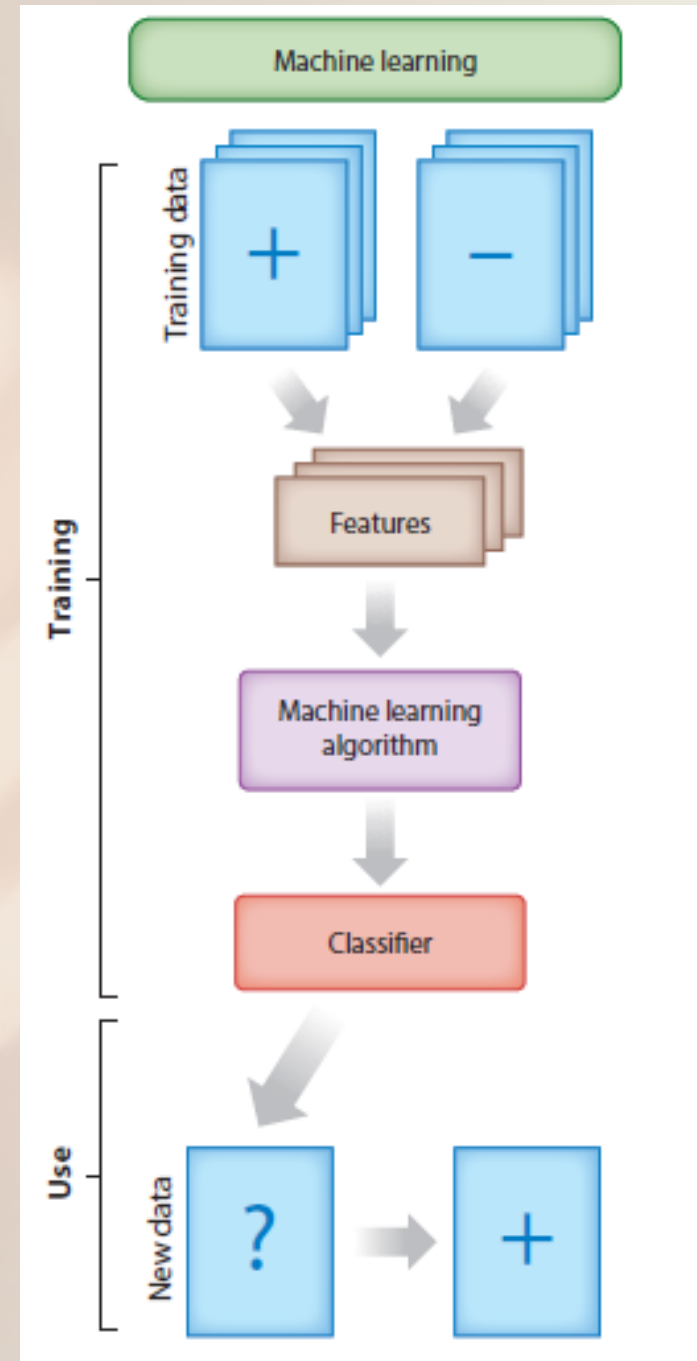
- Sentimentos de investidores em fóruns;
- Acompanhar a evolução de tendências.

■ (Pestian et al., 2012) Análise de **suicide notes** para detectar quais casos poderiam ser reincidentes – *good predictors: abuse, anger, sorrow, forgiveness, love e pride.*

- <https://www.nexojornal.com.br/externo/2019/05/25/Uma-abordagem-inteligente-para-a-sa%C3%BAde-mental> – artigo indicativo do alcance das pesquisas quando áreas diferentes dialogam de forma colaborativa (nesse caso, especialistas de IA, psicologia e, por que não, análise de sentimento?)

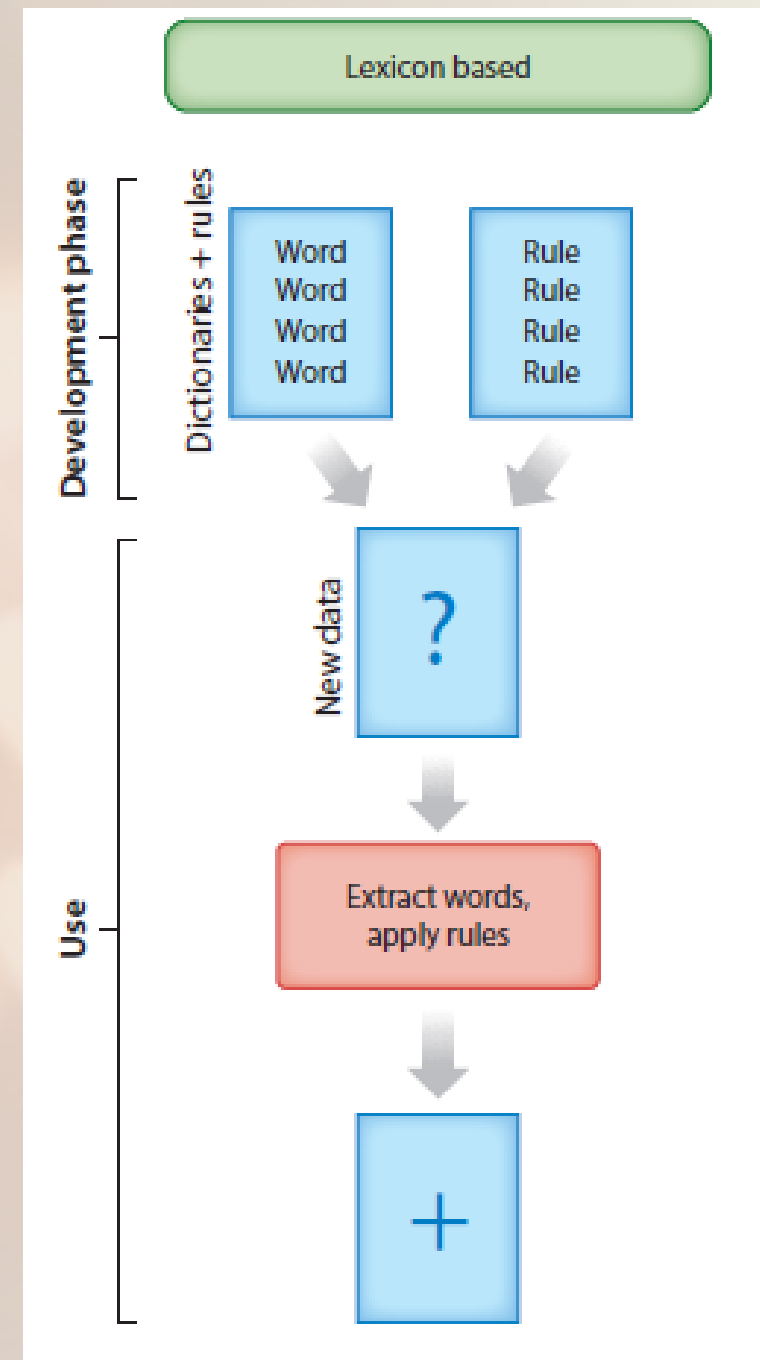
Machine learning approach

- Principal vantagem:
 - Rapidez (com um *dataset* já anotado) 🤔
- Principal desvantagem: São treinados em contextos muito específicos
 - Requer novo material de treino para novos contextos;
 - Depende de nova codificação manual.



Lexicon based approach

- Boa colaboração entre abordagens linguísticas e computacionais;
- Implicações:
 - Determinar a relevância de palavras, expressões e frases;
 - Como extrair padrões do material analisado.



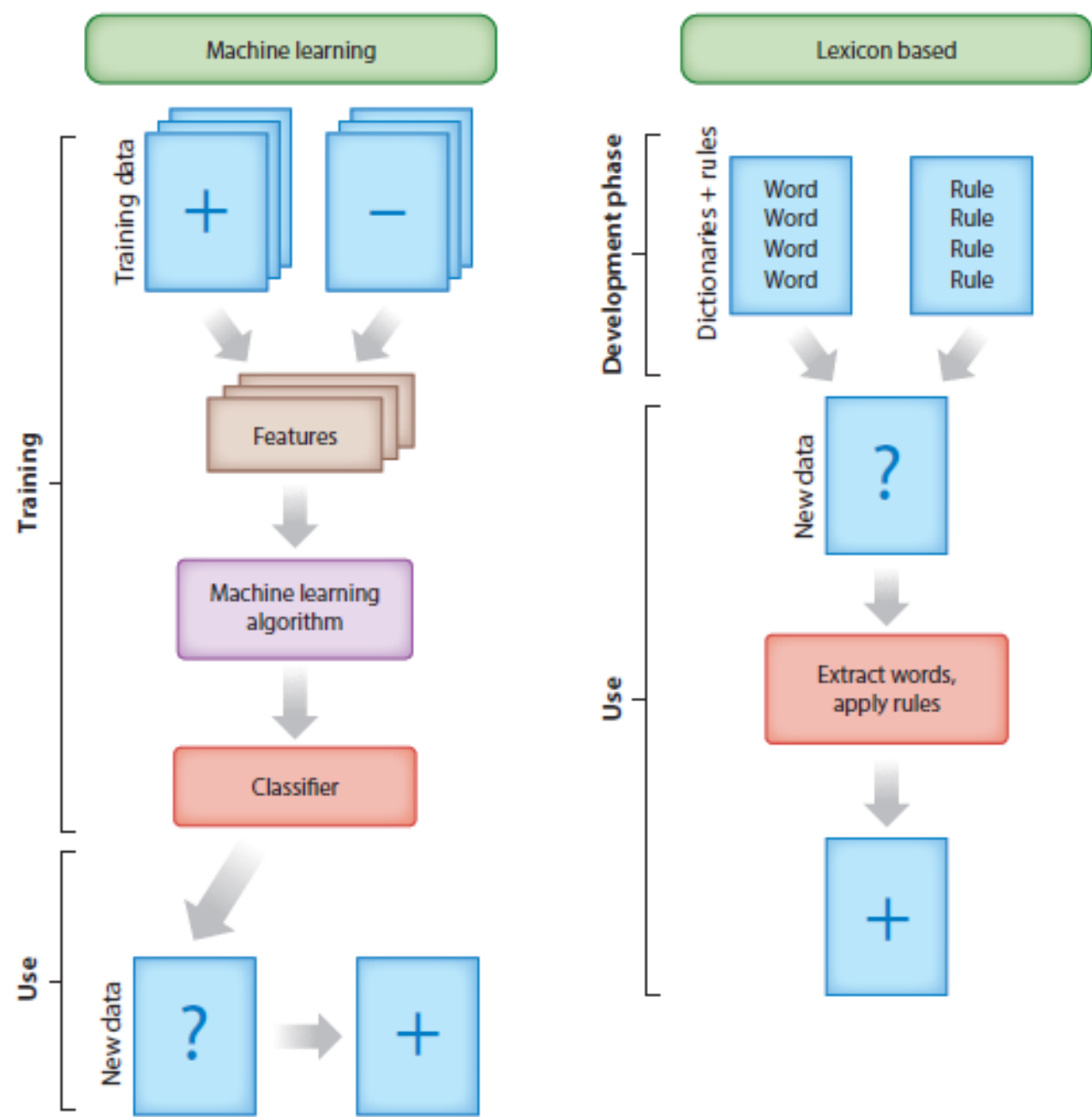


Figure 1

Machine learning and lexicon-based approaches to sentiment analysis.

(Taboada, 2016)

Sentiment Analysis além da língua inglesa (Taboada, 2016):

Pesquisas em Análise de Sentimento em outras línguas:

Alemão;

■ Árabe;

■ Chinês;

■ Espanhol.

■ Francês;

(E português?)

Abordagens possíveis para outras línguas:

- Desenvolvimento de léxicos por pesquisadores falantes nativos;
- Tradução dos textos originais para que a análise seja desenvolvida em língua inglesa. 🤔

Léxicos de sentimento em português

LWIC (Tausczik e Pennebaker, 2010; Balage et al., 2013)

- Abordagem baseada em léxico;
- Como foi elaborada a versão em português?;
- *Affective categories* é uma subcategoria que engloba subcategorias denominadas positive emoticons, negative emoticons, anxiety, anger e sadness.

Multilingualsentiment (Chen e Skiena, 2014)

- Proposta de elaborar um léxico internacional de sentimentos;
- Também contém uma lista com palavras não associadas ao campo semântico de emoção;
- Avalia a polaridade dos termos.

OpLexicon (Souza et al., 2012)

- Léxico de opinião;
- Léxico de orientação semântica de termos não associados ao campo semântico de emoção;

SentiLex (Carvalho e Silva, 2015)

- Idealizado para o português de Portugal;
- Foco em predicadores humanos;
- Léxico elaborado com termos variados e relacionados à polaridade dos predicados e predicadores ;

“Sentimento de N” no AC/DC

- Construção de um léxico de sentimentos a partir do padrão “sentimento de N”
- Pontos positivos:
 - Ir além da orientação semântica;
 - Observar os sentimentos como são descritos/materializados pelos próprios falantes de português (evitando a teoria da universalidade dos sentimentos);
 - Evitar a baixa concordância entre anotadores;

Distribuição das palavras de sentimentos por grupos

- **Grupo 1:** O “N” corresponde a sentimentos que consideramos convencionais, como *culpa*, *medo* etc.

id="Iaiá_Garcia Prosa:romance MdA 1878 romantismo": Luís Garcia não pôde furtar-se a um **sentimento de pena**, ao vê-lo entrar fardado e prestes a seguir para o Sul

- **Grupo 2:** O “N” corresponde a maneiras não convencionais de falar sobre sentimentos.

id="A_semana Prosa:crônica MdA 1892 ": Atentai, mais que tudo, para esse **sentimento de unidade nacional**, que a política pode alterar ou afrouxar, mas que a arte afirma e confirma, sem restrição de espécie alguma, sem desacordos, sem contrastes de opinião

- **Grupo 3:** O “N” não se refere a um sentimento, mas ao possuidor do sentimento.

id="O_Matuto Prosa:romance FT 1878 realismo_ regionalismo_ romantismo": ‘Tais eram as idéias e os **sentimentos de d. Damiana**

- **Grupo 4:** O “N” introduz um modificador do sentimento. Isto é, o “N” não nomeia um sentimento, apenas atribui a ele qualidades.

id="A_Alma_Encantadora_das_Ruas Prosa:crônica JdR 1807 ": Esse **sentimento de natureza** toda íntima não vos seria revelado por mim se não julgasse, e razões não tivesse para julgar, que este amor assim absoluto e assim exagerado é partilhado por todos vós

- **Grupo 5** - O “N” introduz uma mwe iniciada pela palavra “sentimento”.

id="A_Alma_Encantadora_das_Ruas Prosa:crônica JdR 1807 ": Nesse tempo não havia a associação, não havia o **sentimento de classe** e os pobres

Primeiros resultados

- Primeira abordagem – corpus OBras: 120 lemas diferentes;
- Todos os corpora (excluindo-se o OBras): 2060 lemas diferentes:
 - Ponto de corte da análise dos lemas: lemas com frequência acima de 3;
 - 853 lemas diferentes;
 - “Limpeza” dos resultados: erros de digitação, 86 lemas já analisados na primeira abordagem, lemas com variação PT-BR;
 - Total final: 723 lemas para a análise.
- 742 palavras de sentimento (grupos 1 e 2) / 843 palavras:
 - O padrão “sentimento de N” é confiável;
 - 90% são palavras de sentimento.

Primeiros resultados

		Corpus		
		OBras	Todos exceto OBras	Todos
Grupo 1	sema = emo	33 (62%)	76 (35%)	109 (40%)
	sema ≠ emo	20 (38%)	141 (65%)	161 (60%)
Grupo 2	sema = emo	0%	26 (6,1%)	26 (5,5%)
	sema ≠ emo	49 (100%)	398 (93,9%)	446 (94,5%)
Grupo 3		10	58	68
Grupo 4		07	19	26
Grupo 5		01	6	07

A anotação semântica de emoção no AC/DC

- Atualmente: 24 clusters

alívio; admirar; amor; coragem; desejo; desespero; esperança; felicidade; fúria; genérica; gratidão; humildade; infelicidade; ingratidão; insatisfação; inveja; medo; ódio; orgulho; pena; satisfação; saudade; surpresa e vergonha.

Campo lexical 1: lemas referentes à pátria. Sentimento de....

americanismo – anticomunismo – antilhanía – anti-lusitanismo – bairrismo – brasilidade – cidadania – civismo – democracia – germanidade – germanismo – identidade – italianidade – mineiridade – nação – nativismo – nacionalização – nacionalidade – nacionalismo – origem – pátria – patriota – patriotismo – pertença – pertencimento – povo – raça – ufanía – ufanismo – união – xenofobia.

Exemplos:

“É que o sentimento de brasilidade, de patriotismo, nato do brasileiro, impera”

“Eu fui ao diante dele, afirmando que a adoção de uma nacionalidade é ato político, e muita vez pode ser dever humano, que não faz perder o sentimento de origem nem a memória do berço”

“Essa instituição por ele pregada e que se fez realidade entre nós precisa ser conservada sempre, não só por ser ela a defesa do território, como ainda por servir de meio eficaz e rápido para civilizar o caipira, dando-lhe o sentimento de pátria e ensinando-lhe amar e compreender as cores e a história da nossa Bandeira”

Campo lexical 2: lemas referentes à espiritualidade. Sentimento de...

abstenção – adoração – católico – cobiça – comunhão – devoção – doação – fanatismo – fé – idolatria – perdão – religião – religiosidade

Exemplos:

“Helena deixou-se cativar desse sentimento de abstenção e elevação; se alguma dor ou remorso a pungia, esqueceu-os, por um minuto ao menos, entre aquelas paredes desataviadas, diante de um padre, entre uma imagem de Jesus e as obras vivas do Criador”

“Se escrevo sobre o médium, não é por nenhum sentimento de idolatria, mas por reconhecimento ao trabalho de um companheiro que abdicou de si mesmo para servir a causa que abraçou”

Campo lexical 3: lemas referentes ao parentesco. Sentimento de...

avó – irmandade – irmão – mãe – maternidade – orfandade – pai – paternidade
– viúvo

Exemplos:

“Ao contrário do que se pensa, o sentimento de maternidade não é natural, mas sim uma construção social e cultural, como já foi mostrado nos capítulos anteriores”

“Carlos, o marido de Edemeia é a personagem que completa junto com Carvalho, pai da protagonista, o triângulo da contenda e reproduz a figura do ser angustiado e dividido entre a manutenção das funções de vigor e poder reservadas ao sexo masculino, e o sentimento de pai e de esposo”

Próximos passos: como proceder?

- A lista de sentimentos é composta apenas por substantivos – é possível enriquecer a lista com outras classes de palavras?
- Interesses:
 - Considerar outros padrões linguísticos que indiquem a presença de palavras de sentimento; e
 - Relacionar emoções com palavras do corpo humano.
- Estratégias para categorizar uma lista tão grande - como definir o ponto de partida?
- Como reorganizar os clusters já existentes na anotação semântica de emoção do AC/DC?
- Qual o critério para agrupar palavras em clusters de emoção?
- Campos semânticos e riqueza lexical x palavras da lista que se distribuem em outros campos semânticos;



Próximas leituras:

- Benamara F, Cesarano C, Picariello A, Reforgiato D, Subrahmanian V. 2007. **Sentiment analysis: Adjectives and adverbs are better than adjectives alone.** In Proceedings of the 1st International Conference on Weblogs and Social Media (ICWSM).
- Liu B. 2012. **Sentiment Analysis and Opinion Mining.** San Rafael, CA: Morgan & Claypool
- Liu B. 2015. **Sentiment Analysis: Mining Opinions, Sentiments, and Emotions.** Cambridge, UK: Cambridge Univ. Press.
- Pang B, Lee L. 2008. **Opinion mining and sentiment analysis.** Found. Trends Inf. Retr. 2:1-135
- Subrahmanian V, Reforgiato D. 2008. **AVA: adjective-verb-adverb combinations for sentiment analysis.** Intell. Syst. 23:43-50
- Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. 2011. **Lexicon-based methods for sentiment analysis.** Comput. Linguist. 37:267-307
- Wiebe J, Wilson T, Cardie C. 2005. **Annotating expressions of opinions and emotions in language.** Lang. Resour. Eval. 39:165-210

Referências:

- Balage Filho, P.P., Aluísio, S.M. e Pardo, T.A.S. (2013) “An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis”, Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology – STIL, p. 215–219, Fortaleza/Brazil
- Carvalho, P., Silva, M. (2015) “Sentilex-pt: principais características e potencialidades”, In: Linguística, Informática e Tradução: Mundos que se Cruzam, Organizado por Simões, Barreiro, Santos, Sousa-Silva & Tagnin, Oslo Studies in Language 7(1), p. 425–438.
- Chen, Y. e Skiena, S. (2014) “Building Sentiment Lexicons for All Major Languages”, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers), Baltimore, Maryland, USA, p. 383-389.
- Souza, M.; Vieira, R.; Buseti, D.; Chishman, R. E Alves, I. M. (2012) Construction of a Portuguese Opinion Lexicon from multiple resources. In: 8th Brazilian Symposium in Information and Human Language Technology.
- Taboada, M. (2016) Sentiment Analysis: An Overview from Linguistics. In *Annu. Rev. Linguist.* 2:325–347.