

DESCRIÇÃO DO PROJETO

Autor: Lauro C. M. de Paula

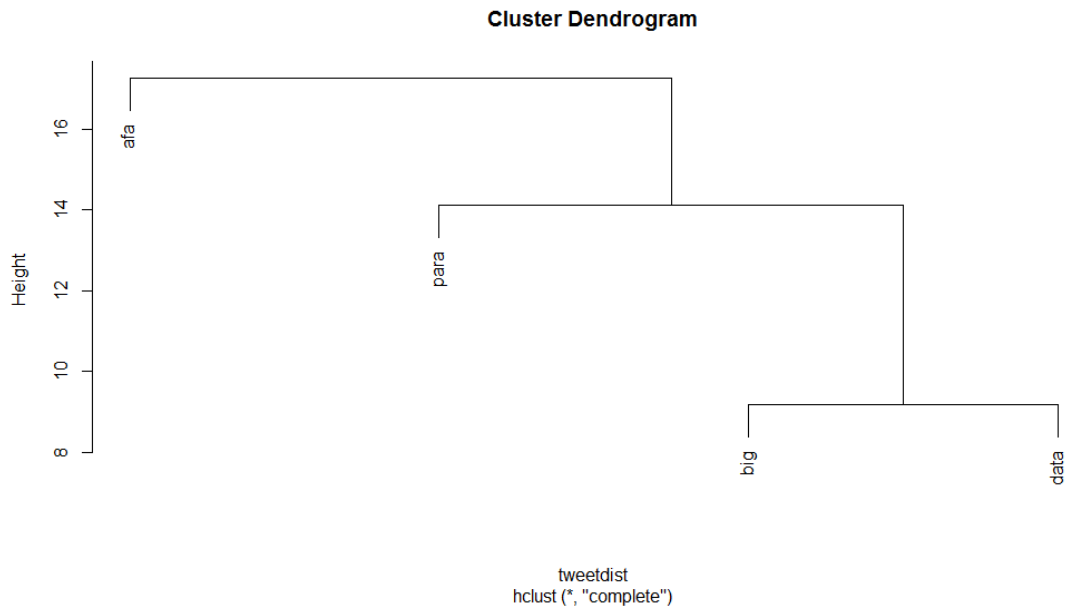
Data: 05/10/2017

Usando o Pacote TwitterR

Após o carregamento de pacotes e autenticação no Twitter, foi realizada uma captura de tweets da Microsoft Br que contêm o termo “big data”. A nuvem de palavras abaixo demonstra que a palavra “afa” (desconhecida) foi a palavra mais encontrada em 100 tweets pesquisados na Microsoft BR. As palavras “big” e “data” são algumas das mais frequentes. Vale ressaltar que, antes da geração da wordcloud, foi realizado todo o tratamento e pré-processamento dos dados coletados.



O dendograma abaixo mostra que as palavras “big” e “data” estão diretamente relacionadas (agrupadas).



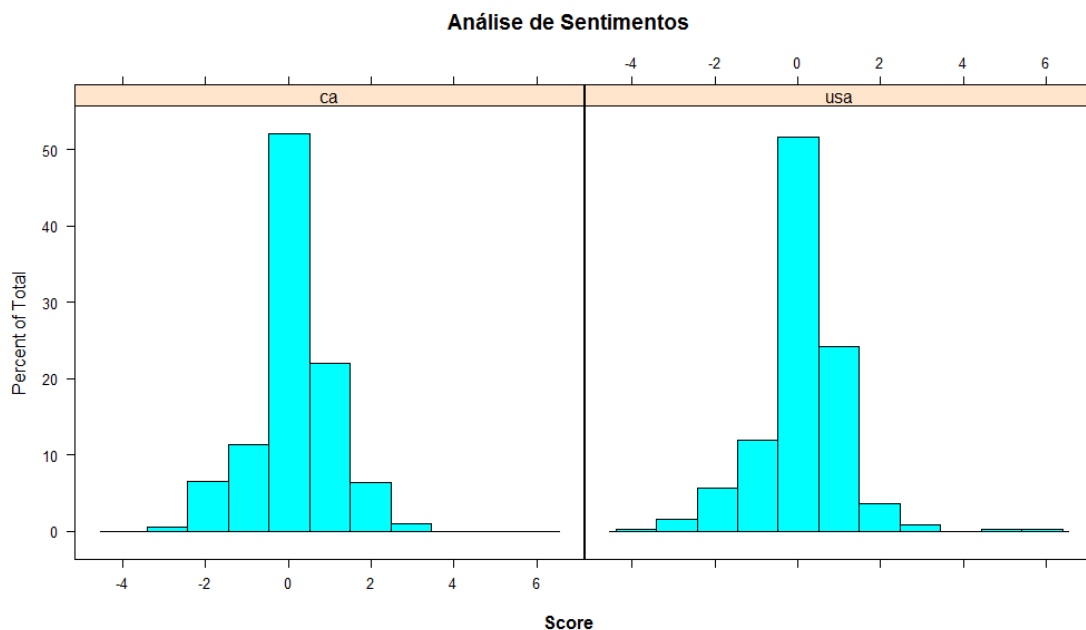
O histograma abaixo mostra uma comparação dos tweets coletados entre dois países: EUA e CAN. Nota-se que, em ambos os países, o número de expressões que não possuem palavras em nossas listas de palavras positivas e negativas (ou que foi encontrada uma palavra negativa e uma positiva em uma mesma sentença) é superior tanto ao número de expressões que possuem palavras com conotação positiva quanto ao número de expressões que possuem palavras com conotação negativa.

Legenda:

0 – neutro

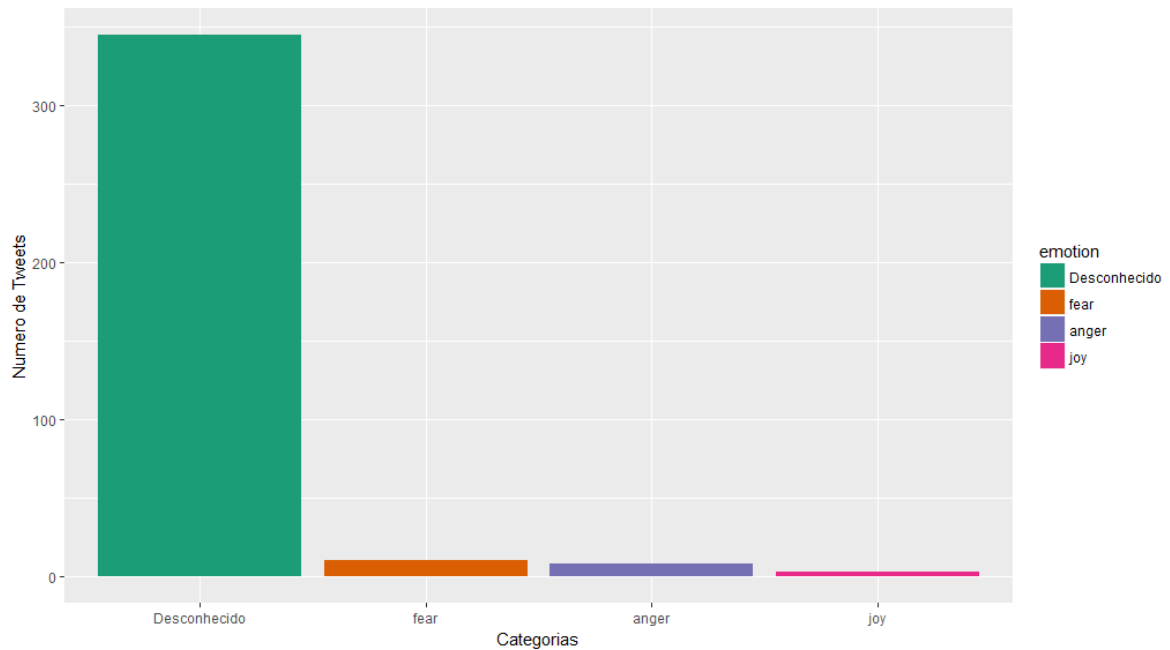
1 – expressão possui palavra com conotação positiva

-1 – expressão possui palavra com conotação negativa



Usando o Classificador Naive Bayes

Ao realizar a análise de sentimentos em uma busca de 1500 tweets contendo a expressão “bigdata” (em português), foi possível obter o gráfico abaixo, que mostra a comparação entre as emoções encontradas. Percebe-se que, como o número de palavras com conotação 0 (neutro) é maior, o número de palavras com emoção desconhecida também é o maior.



Analisando a polaridade das palavras no gráfico abaixo, verifica-se que a categoria de sentimento positivo é maior que a de sentimento negativo.

