ANÁLISE DE SENTIMENTO A COMPANHIAS AÉREAS NORTE AMERICANAS
SENTIMENT ANALYSIS TO NORTH AMERICAN AIR COMPANIES

Marco Alexandre Tomás Tereso

ISLA de Santarém
marco.tereso@islasantarem.pt

Resumo

A disseminação da internet e o crescimento exponencial da sua utilização, tem permitido ao longo dos últimos anos diminuir distâncias entre lugares, comunidades, instituições, organizações e pessoas. O uso recorrente da internet, permite através de diferentes conceitos, traçar o perfil de utilizadores, seus hábitos, gostos, os seus interesses; partilhar exposições de avaliações a determinados produtos e serviços; partilhar conhecimento e experiências na primeira pessoa. Este tipo de dados, quando processados permite obter informações diversas, que podem ser utilizadas para reflexões de análise de sentimentos. Tendo em conta que a utilização das redes sociais se tornou viral, e que muita é a partilha de informação, opiniões e demonstração de sentimentos por parte dos utilizadores, estas tornaram-se uma excelente fonte de dados para a aplicação de técnicas de Processamento de Linguagem Natural. Este estudo tem por base a aplicação de técnicas de análise de sentimento, e a consequente avaliação do serviço prestado por seis companhias Norte Americanas. Este estudo faz uma relação direta com um outro estudo que classifica as companhias através dos atrasos, impedimentos na hora do embarque, extravio de bagagem e reclamações de clientes. Aplicando técnicas de processamento de linguagem natural e análise de sentimento, percebemos que existe alguna relação entre os dados dos dois trabalhos de investigação.

.

Palavras chave: Análise Sentimento; Opinion Mining; PLN; Text Mining.

Abstract

The spread of the internet and the exponential growth of its use have allowed over the last years to reduce distances between places, communities, institutions, organizations and people. The recurring use of the Internet allows, through different concepts, to draw the profile of users, their habits, tastes, their interests; share exposure to certain products and services; share knowledge and experiences in the first person. This type of data, when processed allows to obtain diverse information, that can be used for reflection of feelings analysis. Taking into account that the use of social networks has become viral, and that much is the sharing of information, opinions and demonstration of feelings on the part of users, these have become an excellent source of data for the application of Natural Language Processing techniques. This study is based on an application of techniques of analysis of feelings, and is therefore an evaluation of the service provided by six North American companies. The study is part of a direct evaluation with the other study that classifies as its own words through delays, boarding impediments, lost luggage and customer complaints. Application of natural language processing techniques and meaning analysis, realize that there is a relationship between the data of the research work.

Keywords: Sentiment Analysis; Opinion Mining; PLN; Text Mining.

Em plena era digital, a utilização massiva das redes sociais, tem contribuído para o desenvolvimento de metodologias de análise e interpretação de texto de forma automática. O ramo da inteligência artificial contém algumas áreas que se dedicam esencialmente à implementação de métodos avançados de processamento de texto e consequentemente análise desses mesmos dados. São exemplo, as áreas de Data Mining (Witten, Frank, Hall e Pal, 2016), Text Mining(Pletscher-Frankild, Pallejá, Tsafou, Binder e Jensen, 2015), Opinion

Mining (Balazs e Velásquez, 2016), Sentiment Analysis (Liu, 2010). Cada uma destas áreas da ciência, processam dados textuais com objetivos distintos.

Com a globalização da internet e o consequente crescimento exponencial da sua utilização, surgem novas necessidades e perspetivas de negócio. A divulgação de novos conceitos, empresas, instituições entre outros, permite não só fazer a sua apresentação ao mundo mas também criar opiniões, ideias e sentimentos entre os consumidores dessa informação. A mineração de dados a partir de texto é uma área com bastante aplicabilidade. Este trabalho pretende apresentar os processos para a recolha de dados, provenientes da análise a comentários em serviços online, extraindo informação importante e classificando esse mesmo serviço de forma automática, tendo em conta a natureza do comentário, ou seja, se é positivo, negativo, ou neutro.

O foco deste trabalho é o conceito de Análise de Sentimento, que tem por base obter tweets relacionados com um determinado contexto e extrair a natureza do sentimento, expresso nessa mesma mensagem.

Neste caso em concreto será representado o conjunto de passos necessários para a realização de uma análise deste tipo. Este trabalho de investigação tem como objetivo, avaliar e classificar seis companhias aéreas Norte Americanas, face à percentagem de classificações negativas e positivas por parte dos cibernautas que deixaram uma análise ao serviço prestado pelas mesmas.

Este estudo tem por base um estudo realizado e publicado em 2014, escrito por Brent D. Bowen, da Universidade de Aeronáutica Embry-Riddle, e por Dean E. Headly, da Univesidade de Wichita (2014). O autor criou um ranking baseado em quatro categorias – atrasos, impedimentos na hora do embarque, extravio de bagagem e reclamações de clientes – e foi realizado com os dados das consideradas 15 melhores companhias aéreas norte-americanas. Este estudo apresenta os melhores resultados e a variância de ranking das compañías ao longo de 6 anos. Os dados recolhidos para análise remontan ao ano de 2013, desta forma vamos tentar perceber se das companhias escolhidas existe alguna relação com os resultados de análise de sentimento que vamos obter. Mediante este estudo, e com os dados disponibilizados procura-se fazer uma ponte entre os dados descritos nesse estudo e os resultados que vamos obter.

Este trabalho encontra-se estruturado da seguinte forma: a secção atual é a secção de introdução ao estudo realizado; segue-se a secção de métodos de processamento de texto, onde é feito o enquadramento sobre diferentes áreas de análise de dados textuais, com referência a técnicas de Text Mining e diferentes métodos de processamento e limpeza de dados; de seguida é abordado o conceito de Processamento de Linguagem Natural (PLN), e

referido as suas aplicações práticas; na secção seguinte é abordado o tema de análise de sentimento, que é a base deste estudo; segue-se uma secção dedicada aos métodos de implementação, onde é referido passo a passo o processo que foi implementado na prática para a análise do dataset utilizado; por fim e não menos importante, é apresentada a análise e discussão dos resultados obtidos.

MÉTODOS DE PROCESSAMENTO DE TEXTO

Atualmente existem um conjunto de áreas da ciência que têm como foco a extração de informação a partir de dados textuais. Este tipo de áreas, correlacionam-se com a Inteligência Artificial. De entre as mais conhecidas destacam-se o Text Mining, Data Mining, PLN e Opinion Mining também conhecido por análise de sentimento. Text Mining é uma técnica da Inteligência Artificial, que consiste na extração de informação a partir de texto. Basicamente, a extração de texto consiste em transformar palavras ou frases não estruturadas numa forma adequada para poder aplicar técnicas de Data Mining (Forte, 2015)

Text Mining, ou KDT (Knowledge Discovery from Text) pode ser definido como um processo de conhecimento intensivo, em que um utilizador interage com uma coleção de documentos ao longo do tempo, por meio de um conjunto de ferramentas de análise (Feldman e Sanger, 2007).

A mineração de dados, é o processo de extração ou mineração de conhecimento em grandes quantidades de dados. Este processo resulta da aplicação de técnicas de Text Mining, consiste no processamento dos dados em bruto, através de técnicas de processamento de linguagem natural (PLN) e de métodos analíticos.

O Text Mining pode ser utilizado em áreas diversas e com aplicações diferenciadas. Segundo (Paulraj, 2001), a mineração de dados permite:

- Melhorar a disposição dos produtos, nas prateleiras, mediante o estudo do consumidor
- O departamento de marketing de uma empresa, recorrer ao envio de mensagens promocionais personalizadas, na expectativa de obter melhores retornos
- As empresas preverem a necessidade de reforço de stocks, ao perspetivar picos de vendas
- As agências de viagem aumentar o seu volume de vendas, associando os seus pacotes turísticos ao perfil dos seus clientes

Segundo Dörre, Gerstl e Seiffert (1999), existe ouro escondido nos dados de cada empresa – a extração de conhecimento a partir de texto promete ajudar as organizações a encontrá-lo.

Vamos conhecer em maior detalhe a forma como este tipo de dados são processados.

Processos de Text Mining

O processo de aquisição de informação proveniente de textos, segue um conjunto de processos, de modo a que se torne mais fácil compilar e processar os dados. O processo da extração de informação e apuramento de dados, segue três etapas:

- Análise, procura e seleção de informação (Information Retrieval) - este processo é extremamente importante no desenvolvimento dos restantes. É importante fazer uma boa seleção das fontes de dados (textos, documentos, etc.), para facilitar o processo de recolha de informação de forma automática e no menor período de tempo. Os documentos, dos quais provém a informação, podem ser classificados de: estruturados (quando organizados em tabelas e devidamente organizados); semi-estruturados (quando de certa forma a informação surge classificada por tópicos, ainda que possa surgir desorganizada, como é o exemplo de um jornal, conjunto de artigos e classificados separados na mesma página); ou documentos não estruturados (documentos sem qualquer formatação ou estruturação, como por exemplo um documento de texto corrido ou desorganizado).
- Extração de Informação (Information Extraction) - nesta fase, o mais importante é fazer a filtragem da informação realmente relevante. Mediante aquela que for a seleção que é planeada, este processo tem a finalidade de obter a partir dos documentos, dados específicos, entidades ou relacionamentos.
- Processamento de linguagem Natural (Natural Language Processing, NLP) - por fim, mas não menos importante, o processamento da informação. Nesta fase é importante agrupar dados (provenientes de linguagem falada ou escrita) e classificá-los de uma forma que não seja desconhecida para os programas que manipulam estes mesmos dados. Este processo segue descrito em maior detalhe na secção de pré-processamento, presente neste documento, onde será mais detalhado e exemplificado através de casos concretos.

Pré-processamento

O pré-processamento, consiste na preparação dos dados a serem processados (Silva, 2014). É importante que os dados originais, independentemente da sua proveniência, sejam tratados de forma a serem mais fácil de interpretar para o computador. Esta fase segue um conjunto de etapas, desde o processo de simplificação dos dados recolhidos, limpeza e filtragem dos dados a serem processados. Existem algumas técnicas distintas para realizar o passo a passo deste processo, vamos conhecê-las mais em detalhe.

Separação/segmentação de texto (tokenização)

Os dados obtidos com recurso a técnicas de segmentação de texto, têm o nome de token. É

considerado um token, qualquer palavra constituinte de um texto, ou a um conjunto de n caracteres consecutivos, constituintes de uma palavra, em que n pode receber valores de 1..n, estes tokens têm o nome de n-gram.

O processo de segmentação de frases, obtendo a totalidade de cada palavra, é o mais comum. Este processo consiste na separação de palavras, tendo como referência os espaços entre elas ou os elementos de pontuação (ponto final, virgula, ponto de exclamação, ponto de interrogação, entre outros). A tokenização deve ser ajustada à necessidade de cada problema.

O objetivo deste processo é transformar frases num conjunto de tokens, de forma a poder trabalhar os dados.

Limpeza dos dados

Após o processo de tokenização, existe um conjunto de passos que devem ser seguidos, também eles, ajustados às reais necessidades de cada caso específico.

Na maioria dos casos, a existência de números nos documentos, não acrescentam informação necessária, desta forma, deve-se proceder à remoção dos tokens que contenham numeração.

No processo de limpeza de dados, deve-se proceder à conversão de todos os tokens, para letras minúsculas, evitando assim que palavras iguais mas escritas de forma diferente sejam consideradas diferentes (Ex: Text Mining, TEXT MINING e text mining).

Remoção de stopwords

As palavras identificadas por stopwords, são aquelas que são consideradas que não acrescentam informação de valor. As stopwords mais frequentes são as preposições, artigos e pronomes, como por exemplo "um", "uma", "o", "a", "e", que se tornam irrelevantes para o contexto (Sedbrook e Lightfoot, 2010). Sendo palavras bastante comuns nas frases, e sem que acrescentem valor, a sua eliminação representa um acréscimo na taxa de processamento de dados.

Stemming

Um processo muito comum, é realizarmos uma pesquisa na web, recorrendo a um motor de busca, colocando uma palavra sem nos preocuparmos com o tempo verbal, a pluralidade e as "n" multiplicações que a palavra pode ter. De forma a que a pesquisa seja mais abrangente, os mecanismos de pesquisa encontram o stem da palavra, ou seja, a palavra raiz.

O processo de stemming, permite obter a raiz morfológica de uma palavra, eliminando prefixos e sufixos, para não sobrecarregar a informação gramatical ou lexical da palavra (Moral, de

Antonio, Imbert e Ramirez, 2014). Este processo tem por base fazer uma representação da palavra, excluindo géneros, excluindo termos verbais específicos, e diminutivos (p.ex.: conquistando – conquista, trabalhadora – trabalhador, pequenino – pequeno, pequena, entre outros). Desta forma podemos dizer que:

- Stemming - é a ação de reduzir uma palavra em stems
- Stem - é a parte de uma palavra
- Stemmer - é o artefacto (o programa que executa o processo)

A utilização de stemming por vezes origina erros de análise, esses erros são de dois tipos:

- over stemming - ocorre quando a cadeia de caracteres removida, não é um sufixo mas parte de um stem (p.ex: a palavra "gramática" após processada por um stemmer, é transformada no stem "grama", a sua forma normal seria "gramát")
- under stemming - ocorre quando um sufixo não é removido completamente (por exemplo: a palavra "referência" é transformada no stem "referênc" ao invés de "refer")

Lemmatization

O processo de Lemmatization, consiste num processo capaz de moldar palavras de forma a que, retire a conjugação verbal, caso se trate de um verbo, altere os substantivos e os adjetivos para o singular masculino, colocando a palavra na sua forma de dicionário (por exemplo: amigo – amigo, amiga, amigão; gato – gato, gata, gatos, gatas; ter – tinha, tenho, tiver, tem). Do vocabulário em torno do conceito de Lemmatization temos que:

- Lemmatization - é a ação de reduzir uma palavra em Lemmas
- Lemma - forma básica de uma palavra
- Lemmatizer - é o artefacto (o programa que faz o processamento da ação)

PROCESSAMENTO DE LÍNGUA NATURAL (PLN)

O processamento de Linguagem Natural (PLN), é uma área da ciência que se relaciona com a Inteligência Artificial (IA). O foco de PLN é estudar os problemas de compreensão automática de linguagens naturais humanas. Um dos seus desafios é a compreensão da língua humana e fazer com que computadores consigam interpretar essa mesma linguagem desencadeando funções especificas através do reconhecimento da mesma.

A aprendizagem automática, em todas as áreas, incide sobre a análise de exemplos típicos do mundo real. No caso da PLN, esta mesma análise é feita sobre um corpus, que é um conjunto de documentos ou frases individuais, que foram registados com os valores corretos a serem aprendidos.

Existe um conjunto de algoritmos, de naturezas diversas, tais como, árvores de decisão e modelos estatísticos. Os modelos estatísticos são, nos dias de hoje, mais utilizados, estes

modelos têm a capacidade de tomar decisões flexíveis e probabilísticas, atribuindo um peso a cada característica de entrada. Estes sistemas têm a vantagem de procurar as respostas mais assertivas em cada caso específico.

Aplicações de PLN

A área de PLN é bastante abrangente, e tem imensa aplicabilidade em áreas concretas. Vejamos alguns exemplos da sua aplicabilidade.

- Sumarização Automática: ideal para fazer um resumo sobre notícias, ou conteúdo de uma página
- Tradução de textos: a tradução de elementos entre línguas diferentes, é possível aplicando algoritmos de PLN
- Reconhecimento Óptico de Caracteres (OCR): capaz de fazer um reconhecimento de caracteres através de leitura óptica, e processar esses mesmos dados
- Respostas a Perguntas: um exemplo da sua utilização é quando fazemos uma pesquisa num motor de busca, colocando uma questão (por exemplo: Em que cidade de realizou a EXPO 98?) e obtemos a resposta
- Extração de Relacionamento: permite fazer associações através de critérios de pesquisa em textos (por exemplo: Quem era a mulher de D. Afonso V?)
- Reconhecimento de voz: tem como finalidade fazer o reconhecimento da fala e fazer a sua tradução para escrita.
- Análise de subjetividade (Opinion Mining ou Sentiment Analysis): utiliza a web de modo a recolher informações sobre a opinião pública relativamente a um determinado tema. Este conceito está relacionado com o foco deste trabalho.

Existem muitos outros que poderiam ser descritos e aprofundados. Mas os que aqui foram enumerados, permitem ter uma perceção da sua importância e aplicabilidade.

CONCEITO DE OPINION MINING OU SENTIMENT ANALYSIS

A base deste trabalho é a requisição e análise de comentários a um determinado serviço, relativamente aos sentimentos demonstrados pelo público, na divulgação da sua opinião em relação aos mesmos.

Nas técnicas de PLN, existem subáreas para a análise deste mesmo tipo de comentários, esses conceitos são o Opinion Mining ou Sentiment Analysis. O foco destes conceitos, prende-se com a análise de sentimento e opinião, expressa pelo público, sobre um determinado assunto, conceito ou serviço.

Uma tarefa básica neste conceito, é classificar um determinado texto, como sendo positivo, negativo ou neutro. O conceito de Data Mining, também ele com muitos aspetos em comum,

baseia-se muitas vezes no conceito de Opinion Mining, para poder fazer análises de aproximação em áreas como a gestão e a área das ciências sociais. (Sneka e Vidhya, 2016).

A análise de sentimentos, é aplicada em todos os domínios empresariais, pois as opiniões fazem parte da realidade humana e influenciam os nossos comportamentos (Liu, 2012). Se pensarmos, sempre que temos uma dúvida ou um problema, procuramos muitas vezes a opinião de outras pessoas, no intuito de tentar chegar a uma conclusão/solução o mais assertiva possível. Esta é uma realidade bastante presente e que demonstra a importância da análise de sentimentos.

MÉTODO DE IMPLEMENTAÇÃO

A análise de sentimento, é um processo que permite através da análise de texto perceber opiniões positivas, negativas ou neutras. Até chegar aos resultados pretendidos, os textos em análise passam por um conjunto de processos. É fundamental proceder a uma limpeza dos dados, este processo consiste em remover do texto todas as palavras e caracteres que não acrescentem qualquer informação relevante. Este processo trata essencialmente da remoção de sinais de pontuação, caracteres soltos, algarismos, links de páginas web e remoção de stopwords(palavras que não acrescentam qualquer valor e que são identificadas para cada língua). Posteriormente é necessário fazer a comparação das palavras resultantes com a lista de palavras positivas e negativas de cada língua. Este trabalho foi desenvolvido através da criação de scripts em linguagem R. R, é uma linguagem capaz de manipular grandes quantidades de dados, efetuar cálculos e fazer recriações gráficas por exemplo (Venables, Smith e Team, 2018).

A metodologia de desenvolvimento contemplou um conjunto de etapas, desde a aquisição dos dados; aplicação de processos de limpeza dos dados em análise; separação dos tokens de cada comentario; avaliação dos comentarios, numa escala de positivo, negativo ou neutro, comparando cada token com as listas de *positivewords* e *negativewords*; apresentação de resultados. As próximas secções apresentam em detalhe a metodología aplicada nesta investigação

Aquisição de dados

Para a realização deste trabalho prático foi necessário recorrer à aquisição de dados. Apesar de existirem alguns repositórios online que disponibilizam dados de datasets, a aquisição de comentários torna-se por vezes um processo mais difícil, tendo em conta que nem todos os datasets os disponibilizam. Para a realização deste trabalho utilizou-se um dataset disponível em[1]. Após o download dos dados procedeu-se à separação do ficheiro .csv em diferentes

[1] Disponível para download em: https://www.kaggle.com/crowdower/twitter-airline-sentiment

ficheiros, organizados por companhias diferentes, para facilitar posteriormente a sua utilização.

Limpeza de dados

O processo de limpeza de dados é bastante importante, não só porque permite simplificar a amostra de dados a processar, mas também porque permite apurar apenas o conteúdo essencial. Para a realização deste processo desenvolveu-se um script em R, este processo tem por base fazer a remoção de links, algarismos, sinais de pontuação, acentos, remoção de espaços em branco, remoção de caracteres repetidos, conversão para caracteres minúsculos e remoção de stopwords. O processo de remoção de stopwords deve se ter em conta o idioma utilizado nos comentários, por vezes pode surgir a necessidade de fazer uma limpeza de *stopwords* de outras línguas. Neste caso foram aplicados os processos de limpeza de dados tendo em conta a língua inglesa, visto que os comentários analisados são em inglês.

Análise de sentimento

Após a primeira fase, de limpeza dos dados, aplicando todos os processos anteriormente referidos, os dados obtêm um estado aceitável para a aplicação de métodos de análise de sentimento. Para obter informação precisa sobre a natureza dos comentarios, é necessário proceder à análise de sentimento dos comentarios adquiridos. Neste contexto foi necessário proceder à recolha de conjuntos de palavras positivas e negativas associadas ao idioma de inglês, utilizado neste trabalho. Os documentos de palavras positivas e negativas da língua inglesa, utilizados neste trabalho, resultam de um trabalho de investigação (Minqing e Liu, 2004). Convém referir que, para além das palavras identificadas nos ficheiros enumerados, foram adicionadas as palavras ‘cancel’ e ‘cancelled’ ao ficheiro das palavras negativas, tendo em conta que estas são duas palavras que não constavam na lista mas que são importantes no caso em análise. Para esta análise desenvolveu-se um script em linguagem R, que devolve o número de comentários positivos e o número de comentários negativos, fazendo a diferença e apresentando os valores resultantes sobre o contexto geral.

A Figura1 ilustra o resultado apresentado pelo compilador da linguagem R, aplicando o processo de deteção de palavras positivas no conjunto de comentários. Para uma melhor percepção procedeu-se à identificação dos tokens positivos, assinalados a verde na Figura 1, para uma melhor percepção. As representações numéricas identificam o número do token pela orden do ficheiro criado, enquanto que a representação ‘NA’ representa tokens negativos ou neutros. O mesmo proceso será repetido para encontrar os comentarios negativos, permitindo assim fazer o ranking sobre o sentimento expresso pelos clientes de cada companhia.

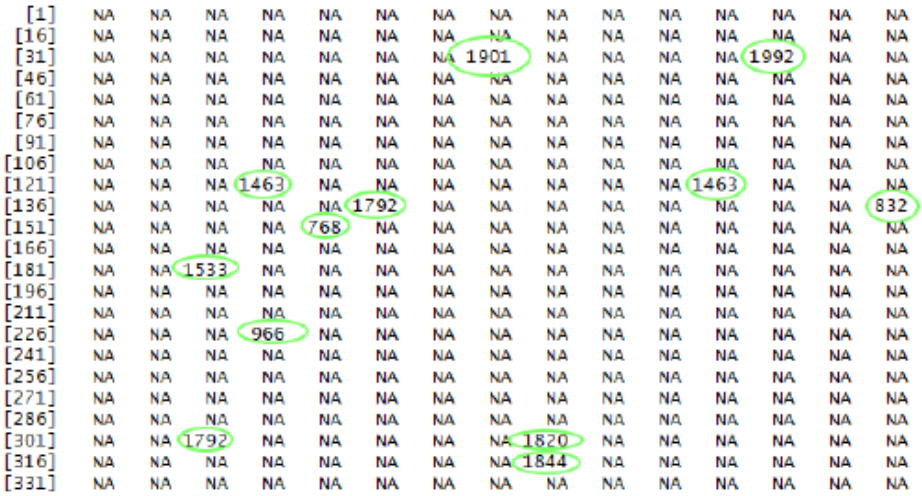

Figura 1. Exemplo da identificação de comentários positivos à companhia aérea JetBlue

Ainda no mesmo script foi gerada uma nuvem de palavras, constituída pelas palavras que se repetem mais vezes nos comentários sobre cada uma das companhias aéreas. A Figura 2 ilustra um dos exemplos representativos, tendo por base o ficheiro de dados da companhia AmericanAir. Este tipo de representação é bastante interessante para analisar as palavras que mais frequentemente são utilizadas para descrever o tópico em análise.

Figura 2. Nuvem de palavras da companhia AmericanAir

Posteriormente, procedeu-se à análise total do corpus de dados, e recriou-se uma nuvem de palavras relativamente às 6 companhias aéreas. A Figura 3 ilustra a nuvem de palavras relativamente às 6 companhias analisadas.

Figura 3. Nuvem de palavras total do corpus em análise

DISCUSSÃO DE RESULTADOS

Tendo por base a atividade experimental desenvolvida, foram analisados os comentários a 6 companhias aéreas, num total de 14608 registos. Dividindo-se por 2755 para a companhia American, 2219 para a companhia Jet Blue, 2415 da companhia Southwest Air, 3806 da United, 2910 da US Airways e 503 da Virgin America. Para cada uma das companhias procedeu-se a uma análise de dados independente e posteriormente uma reunião de dados e comparação entre todas.

A Tabela 1 ilustra os resultados obtidos, sendo que das 6 companhias analisadas, 3 possuem uma diferença positiva e outras 3 uma diferença negativa. Pode-se também analisar que os valores negativos, das companhias American, United e US Airways têm valores bastante negativos, posicionando-se com diferenças acima dos 500 comentários. Por sua vez as companhias com avaliações positivas, não obtêm diferenças tão significativas, sendo a Jet Blue a companhia a conseguir uma maior diferença, situando-se nos 180 comentários.

Tabela 1. Tabela comparativa de análise de sentimento das 6 companhias aéreas

Companhia	**Total**	**Positivos**	**Negativos**	**Diferença**
American	2755	932	1442	-510
JetBlue	2219	895	715	+180
Southwest Air	2415	979	875	+104
United	3806	1412	2069	-657
US Airways	2910	975	1667	-692
Virgin America	503	227	155	+72

A Figura 4 apresenta os resultados apresentados no estudo de Bowen e Headley (2014), e permite fazer algunas considerações tendo em conta os dados a que chegámos.

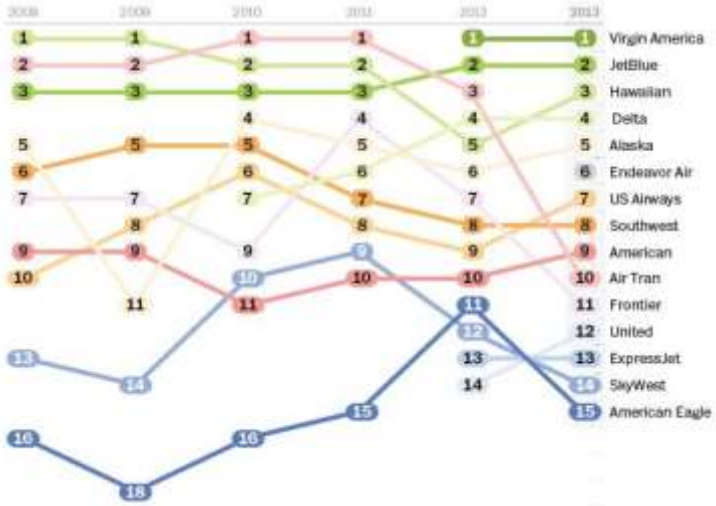

Figura 4. Ranking de classificação melhores compañías aéreas Norte Americanas 2014, Bowen e Headley (2014)

Com base no ranking apresentado na Figura 4, verificamos que as companhias aéreas com análise positiva ocupan o primeiro, segundo e oitavo lugar do ranking; enquanto que as companhias aéreas Norte Americanas com sentimento negativo ocupam as posições sete, nove e doze. Curiosamente apenas existe uma contrariedade entre os resultados, mas que fácilmente pode ser explicada, tendo em conta que ao longo dos tempos a companhia Southwest esteve sempre melhor cotada do que a US Airways, apenas no ano de 2013 houve troca de posições entre ambas. O facto mais curioso é a companhia Virgin America ser cotada como a melhor companhia entre os clientes, e ter menor número de comentarios nas redes sociais. Este facto contribui também para que a sua diferença entre comentários positivos e negativos não possa ser maior.

Este trabalho teve como finalidade aplicar as técnicas de análise de sentimento na prática e fazer uma demonstração dos dados obtidos.

Este é um método que pode ser implementado para fins distintos, com o objetivo de obter uma avaliação pública sobre a classificação de pessoas, serviços ou produtos.

REFERÊNCIAS BIBILIOGRÁFICA

Balazs, J. A., & Velásquez, J. D. (2016). Opinion mining and information fusion: a survey. Information Fusion, 27, 95-110.

Bowen, B. D., Headley, D. E. (2014). Airline Quality Rating 2014 Abstract.

Dörre, J., Gerstl, P., & Seiffert, R. (1999, August). Text mining: finding nuggets in mountains of textual data. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 398-401). ACM.

Feldman, R., & Sanger, J. (2007). The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge university press.

Forte, A. C. B. (2015). Análise de comentários de clientes com o auxílio a técnicas de Text Mining para determinar o nível de (in) satisfação.

Liu, B. (2010). Sentiment Analysis and Subjectivity. Handbook of natural language processing, 2(2010), 627-666.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

Minqing Hu, Liu, B. (2004) "Mining and Summarizing Customer Reviews."Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Seattle, Washington, USA, Aug 22-25

Moral, C., de Antonio, A., Imbert, R., & Ramírez, J. (2014). A survey of stemming algorithms in information retrieval. Information Research: An International Electronic Journal, 19(1), n1.

Paulraj, P. (2001). Data Warehousing fundamentals: A comprehensive guide for IT Professionals. John Willey Interscience Publication.

Pletscher-Frankild, S., Pallejà, A., Tsafou, K., Binder, J. X., & Jensen, L. J. (2015). DISEASES: Text mining and data integration of disease–gene associations. Methods, 74, 83-89.

Sedbrook, T., & Lightfoot, J. M. (2010). Dear: a new technique for information extraction and context-dependent text mining. Communications of the IIMA, 10(3), 3.

Silva, M. A.(2014). O Pré-Processamento em Mineração de Dados como método de suporte à modelagem algorítmica. Dissertação.

Sneka, G., & Vidhya, C. T. (2016). Algorithms for Opinion Mining and Sentiment Analysis: An Overview." International Journal of Advanced Research in Computer Science and Software Engineering 6 (2)

Venables, W. N., Smith, D. M., Team, R. C. (2018). An introduction to R-Notes on R: A programming environment for data analysis and graphics.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

PERFIL ACADÉMICO E PROFISSIONAL DO AUTOR

Marco Tereso - Docente no ISLA Santarém desde 2017. Licenciado em Engenharia Informática – Tecnologias da Informação e Multimédia pela Escola Superior de Tecnologia e Gestão de Oliveira do Hospital; MestrE em Informática e Sistemas – Desenvolvimento de Software pelo Instituto Superior de

Engenharia de Coimbra; Doutorando em Informática na Universidade de Évora. As suas principais áreas de investigação são: Business Intelligence; IoT; Visão Computacional; Machine Learning.

Endereço postal
ISLA Santarém
Largo Cândido dos Reis
2000-241 Santarém
Portugal