**A Importância dos Idiomas na Análise de Sentimento**

**Resumo:** A internet veio revolucionar o mundo e está a mudar a nossa forma de viver, estar e pensar. A análise de sentimento, que classifica opiniões de textos em linguagem natural de forma rápida e precisa, tornou-se muito útil, por exemplo, para as empresas, porque descobrem num curto período de tempo se ocorreu algum problema. A facilidade na partilha de opiniões nas redes sociais, nomeadamente, no Twitter, faz com que as opiniões cheguem a qualquer parte do mundo instantaneamente e, muitas vezes, em idiomas distintos. A avaliação de textos exprimidos noutras línguas para além do inglês, tem algumas limitações, principalmente, por falta de recursos. Embora já existam algumas pesquisas neste campo e com resultados positivos e disponíveis para trabalhos futuros, é necessário um maior empenho e dedicação da parte de novos investigadores.

1. **Introdução**

O aparecimento da internet veio revolucionar o mundo como o víamos antes. É como uma tempestade resistente que não vai passar. Viver sem utilizar os meios informáticos para trabalhar, fazer estudos de mercado, compras e vendas online é impensável. A dependência é de tal forma extrema, que não fazemos nada sem consultar, por exemplo, as avaliações de um produto, filme ou restaurante na internet.

Com a evolução da tecnologia e, principalmente, com a chegada das redes sociais, tornou-se simples tanto a publicação de opiniões como a propagação das mesmas.As opiniões, sendo positivas, negativas ou até mesmo neutras, podem ter um grande impacto na sociedade e, nomeadamente, no futuro de uma marca ou empresa. Estas, se forem cativantes, atrativas e inspirarem confiança, obtém boas classificações e avaliações de clientes de qualquer parte do mundo. Ao serem analisadas e filtradas as classificações menos relevantes, são obtidas opiniões mais gerais. Assim, possíveis problemas são detetados e corrigidos rapidamente. Esta técnica de analisar e quantificar opiniões a partir de textos, é designada por análise de sentimento.

Existem diversas técnicas para analisar sentimento em textos: baseadas em léxicos, *machine learning* ou um sistema híbrido que combina os dois [1]. A abordagem baseada em léxicos, é sustentada por um conjunto de frases e palavras previamente compiladas, que estão relacionadas com sentimentos e que estão presentes em dicionários ou corpus linguísticos. A técnica utilizando *machine learning* é suportada por dados, que fazem com que a máquina aprenda e se torne mais autónoma. Os resultados da análise de sentimento têm sido positivos. Contudo, ainda existem algumas lacunas a serem resolvidas. O facto de os léxicos dependerem do idioma em que o texto está escrito e da maioria das pesquisas se focar maioritariamente na língua inglesa, por ser a língua dominante na internet, são dos maiores problemas a resolver.

1. **Estado da arte**

A Inteligência artificial (IA) é um dos temas mais desafiantes da atualidade. Um dos ramos desta área é a análise de sentimento (*Sentiment Analysis*), que investiga a interação a nível emocional entre um computador e um texto em linguagem natural. Esta técnica, para além de eficaz e organizada, é também relevante no controlo e classificação de emoções e opiniões sobre determinados assuntos. Mas como é que um computador avalia os sentimentos? Máquinas conseguem sentir? Na realidade, são seguidos padrões, regras e algoritmos que auxiliam a avaliação de opiniões. A contagem do número de ocorrências de palavras positivas e negativas, é um dos métodos usados na análise de sentimento. Todavia, as opiniões qualitativas não são fáceis de se quantificar e, por essa razão, existem falhas e limitações neste método, sobretudo na automatização das máquinas.

Ultimamente, foram elaborados vários estudos e artigos sobre análise de sentimento [1-6], particularmente, no contexto das redes sociais como, por exemplo, no Twitter [1,2,4,6], onde é possível partilhar todo o tipo de opiniões para um imenso número de pessoas instantaneamente. Agarwal et al. [2] adquiriram manualmente dados do Twitter (tweets), criaram modelos e disponibilizaram dois dicionários. Um dos dicionários contém *emoticons* e o outro acrónimos e abreviações de palavras com traduções em inglês. Estes dicionários são importantes, porque é frequente encontrar dados com essas características. Além dessas características, muitas vezes, os textos são apresentados noutros idiomas, o que dificulta a sua classificação, pelo facto de alguns idiomas não terem grande suporte de análise de sentimento. Williams et al. [3] coletaram centenas de expressões, palavras e frases associadas a sentimentos em diferentes línguas e implementaram um conjunto de gramáticas que ajudam a reconhecer o idioma num texto. Com esta pesquisa, puderam concluir que a precisão na análise de dados aumenta drasticamente quando esses recursos estão presentes. No entanto, as gramáticas e os dicionários têm um número finito de palavras e, por esse motivo, a análise falha em algumas situações por falta de recursos.

Noutros estudos [4,5], foi sugerido um sistema *multilingual*. Balahur et al. [4], recorreram à tradução automática de textos em quatro idiomas, e verificaram que a qualidade da análise de sentimento relativamente a abordagens e pesquisas anteriores melhorou. Bal D. et al. [5] traduziram manualmente textos em inglês e holandês, e criaram um *pipeline*. Ao avaliar as pontuações de sentimento, notaram que a semântica, a forma como o sentimento é transmitido e certas expressões são muito diferentes entre ambos os idiomas, tornando a sua comparação muito complexa. Esta observação, também é válida para outras línguas.

Em Portugal, ainda existem poucos estudos e recursos sobre análise de sentimento, sendo a maioria dos estudos em inglês. Recentemente, foi realizado um pequeno projeto [6], cujo objetivo foi avaliar o estado emocional dos portugueses durante a pandemia. Utilizando bibliotecas do Python, analisaram e traduziram para inglês, dados de notícias de jornais portugueses e tweets sobre a Covid-19 em Portugal. A partir de estatísticas traduzidas em gráficos, constataram que a classificação média de sentimento foi negativa. Quando foi declarado o desconfinamento, o sentimento melhorou um pouco e, por isso, averiguaram que a população portuguesa está desejosa de voltar à vida que tinha antes da pandemia.

1. **Conclusões**

Com o avanço tecnológico, deparamo-nos com máquinas cada vez mais potentes e capazes de executar tarefas complexas com maior eficiência. A análise de sentimento tem evoluído muito nos últimos tempos e tem-se tornado muito útil na gestão de opiniões em diferentes áreas. Apesar de existirem inúmeras pesquisas sobre este tópico, ainda são poucas as que estão relacionadas com a importância dos idiomas na análise de sentimento.

Ainda que o inglês seja a língua mais utilizada na internet e nas redes sociais, também há um número significativo de pessoas que partilha opiniões noutros idiomas. A forma como se exprime o sentimento varia, porque há diferenças na gramática e em algumas expressões características de um idioma. Parte das pesquisas feitas nesse âmbito, propõem a criação de léxicos para aumentar os recursos. Noutras pesquisas, selecionam textos em diferentes idiomas, traduzem automaticamente para inglês e, em seguida, utilizam métodos desenvolvidos apenas para a língua inglesa. Os resultados têm sido pertinentes e satisfatórios, mas é realmente importante investir em novas pesquisas.

As máquinas são capazes de se automatizar, ou seja, de aprenderem com os dados sem serem programadas manualmente para uma tarefa em específico, mas não há nenhuma máquina até hoje, que seja cem por cento automatizada e eficiente. Portanto, pesquisadores futuros, devem apoiar-se em dados e estatísticas de pesquisas credíveis feitas anteriormente e, assim, desenvolver novos métodos ou aperfeiçoar e melhorar o desempenho de técnicas já existentes.

**Referências**

1. Emeka, O. &. (2019). Sentiment Analysis of Tweet Data. *International Journal of Scientific Research, 4*, 10-24.
2. Agarwal, A. a. (2011). Sentiment analysis of twitter data. Em *Proceedings of the workshop on language in social media (LSM 2011)* (pp. 30-38).
3. Lowri Williams, C. B.-A. (2015). The role of idioms in sentiment analysis. *Expert Systems with Applications*, 7375-7385.
4. Balahur, A. a. (2013). Improving sentiment analysis in twitter using multilingual machine translated data. Em *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013* (pp. 49-55).
5. Bal, D. a. (2011). Sentiment analysis with a multilingual pipeline. Em *International Conference on Web Information Systems Engineering* (pp. 129-142).
6. Ana Frias, J. A. (2020). Análise de sentimento ao longo do período pandémico. *A pandemia nos nossos sentimentos*. Obtido de https://www.dssg.pt/a-pandemia-nos-nossos-sentimentos/