

10 Challenges of sentiment analysis and how to overcome them Part 1

AUTHORS Paul Simmering, Thomas Perry

26 January 2023

A guide for evaluation of sentiment analysis solutions for market research:
What to watch out for when choosing sentiment analysis software

Analytics

6 min read

Share



ARTICLE SERIES

10 Challenges of sentiment analysis

Sentiment analysis: what is it?

Sentiment analysis aims to extract people's opinions about products, brands, issues, or concepts from texts, such as reviews, news texts or social media posts. Sentiment analysis can be done in several ways. One way is to analyse texts manually. This is usually a laborious task. It doesn't scale well.

Since the early 2000s, researchers and industry users have developed various approaches for automating sentiment analysis. Automated sentiment analysis can run on a large scale and provide insights for market and social research. The goal is to measure sentiments about a given research object correctly, fast and at low costs.

Sentiment analysis: what are the challenges?

Sentiment analysis is a feature of many analytics apps, e.g., social media monitoring. But many solutions on the market oversimplify and return answers that are too vague or unreliable to be used as a basis for business decisions.

The closer you look the uglier it gets. We came to this conclusion based on our experience using various software libraries for R and Python, cloud platform APIs and models embedded in social media monitoring tools. We tested their validity for automated sentiment analysis in small experimental projects and on the enormous datasets, our specialised social media monitoring tool Cosmention (www.cosmention.com) generates.

This series of four articles aims to make you aware of the difficult challenges a valid sentiment analysis must meet. It's vital to look closely at what is being sold as sentiment analysis and demand proof of the validity.

Hence, it is a guide for market researchers and market research clients: *What to watch out for when choosing a sentiment analysis software or provider.*

We won't discuss the general requirements for data analysis and social media monitoring, such as defining a clear analysis scope and removing irrelevant, duplicate and spam records from datasets. However, we'd like to point out that many projects already struggle with these requirements.

The 10 challenges of sentiment analysis- General aspects

1. Unclear target: what is a sentiment expression referring to?

To understand the meaning of a sentiment expression (i.e., what sentiment a series of words is expressing), you need to know its target. The sentence *"I like the product"* isn't useful if we don't know what "the product" is.

When analysing reviews, the target product name (the name of the product we wish to understand the sentiment about) is part of the review's metadata. This also includes the review author's name, the website where the review was published and the time of publishing.

But social media posts or other texts don't have metadata that clearly identifies a target. A text can be about many distinct products, brands, people and issues. And a sentiment expressed in it could refer to any of these.

Not having a clear target means that the expressions' meaning is unclear. The fact that a keyword (a word we want to understand the sentiment behind) was used in a text also doesn't necessarily indicate that one or more sentiment expressions in the text refer to that keyword. They may instead refer to a different target in the same text. One can often see from simply reading a few posts that many sentiment expressions don't refer to a keyword that happens to be elsewhere in the text. Such errors are more likely in long texts covering many topics.

Hence, clearly identifying the connection between sentiment expression and the target is essential for any sentiment analysis. The solution is to make the target explicit. This can be done by combining named entity recognition (NER) and dependency parsing.

NER is a machine learning method that highlights entities like products or brands in a text. It works best when the entities have unique names, like "MAC Cosmetics", but it is not well suited to identifying more general issues in discussion. The next step is dependency parsing, which analyses the grammar of a sentence and reveals how sentiment expressions relate to the identified entities.



Dependency-based context example parsed from a customer review of a lipstick. The blue words are targets, and the green and red words are sentiment expressions. Dependency parsing of words is shown by the arrows. The grammatical role of each word is labelled below. Overall, this yields a clean association of sentiment expressions with their targets. (Illustration by Kim Karle).

Another issue is the possibility that the target of a sentiment expression does not have to appear by name in the same sentence as the expression. For example, a person can be referred to by their job (e.g., "the manager") rather than their name. The sentiment model must be smart enough to bundle multiple references to the entity into one sentiment target, e.g., "the manager" and "Jane Doe" must be recognised as being the same person.

This issue appears when working with multiple texts and within texts. In another sentence, Jane Doe may just be referred to as "she", so the sentiment model needs to use a technique called coreference resolution to understand who "she" is.

It becomes even harder when the target is entirely implicit. This means that it doesn't even show up anywhere in the text. For example, the statement "X is the only trustworthy candidate" is an implicit negative statement towards all other candidates. Augenstein et al. (2016) analysed tweets from the 2016 US election with this approach.

2. Unclear opinion holder: are you analysing the right people?

Just like we need to know what is discussed (the target), we need to know *who* is discussing it. A sentiment analysis needs to show the whole sequence: Opinion holder → Sentiment Expression → Target. See Liu (2020) for a full explanation of the sentiment model.

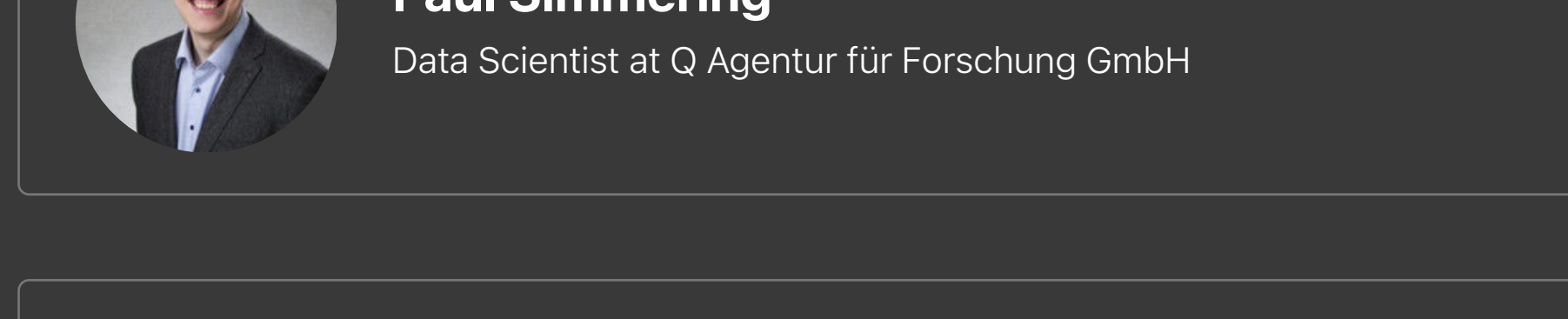
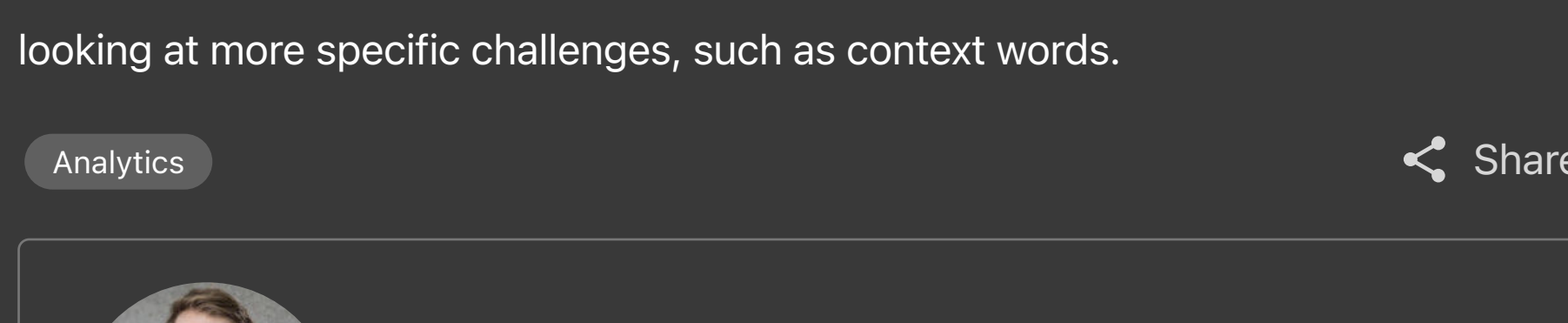
One way to know the opinion holder is to pre-select texts with a clear opinion holder, e.g., customer reviews. However, on social media, the opinion-holder question is much harder to answer.

A post may be written by a genuine customer, an influencer who received money or free samples, a brand, a journalist, or even a spam bot. It may be a retweet or a citation. Solutions for this problem cannot be found in sentiment analysis. Instead, opinion holders can be categorised manually, inferred from the context or metadata (e.g., reviews, customer service tickets, authors of news articles) or predicted by a text classification model based on their texts and author profiles.

In part 2, we'll help you become aware of the danger of lacking nuance and start looking at more specific challenges, such as context words.

Analytics

Share

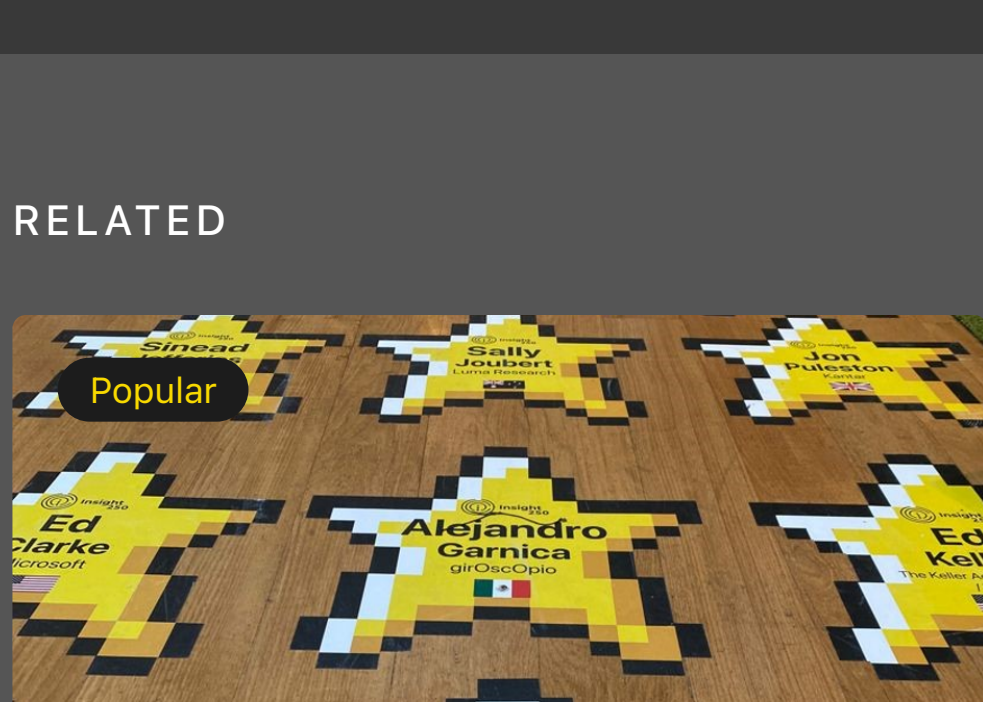


ARTICLE SERIES

10 Challenges of sentiment analysis

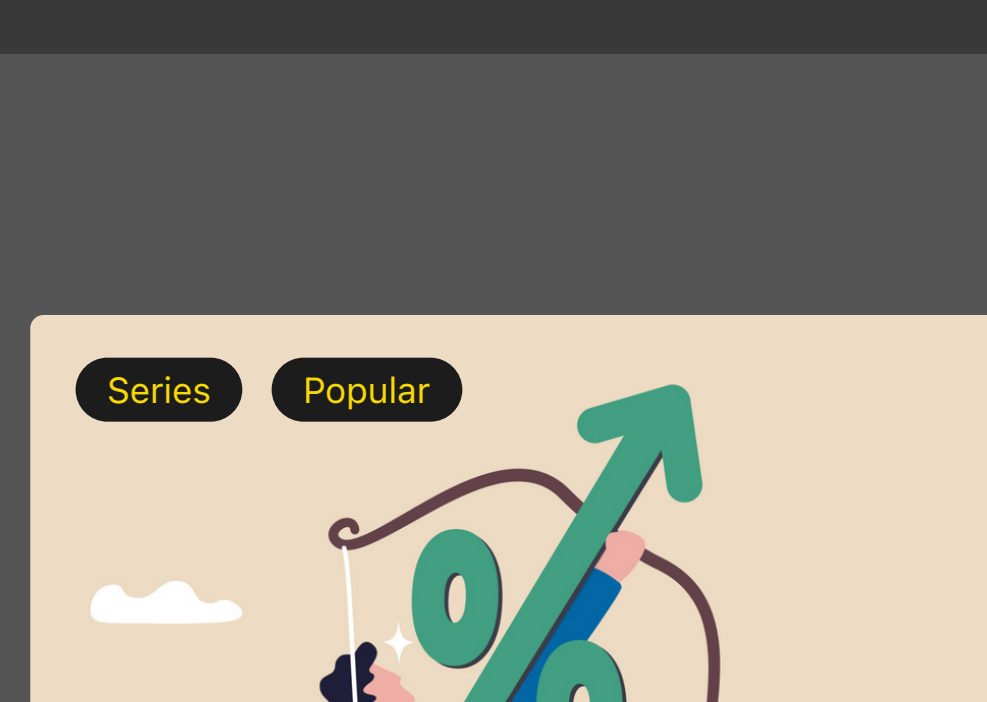
- 10 Challenges of sentiment analysis and how to overcome them Part 1
- **10 Challenges of sentiment analysis and how to overcome them Part 2**
- **10 Challenges of sentiment analysis and how to overcome them Part 3**
- **10 Challenges of sentiment analysis and how to overcome them Part 4**

RELATED



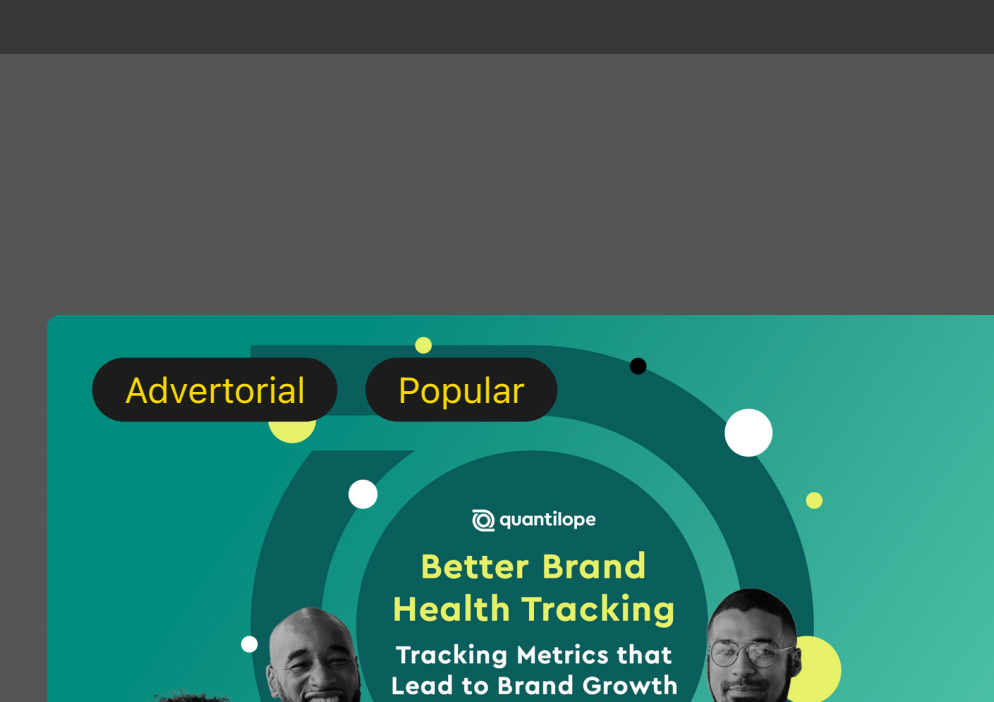
10 October 2023
by Crispin Reale

The inaugural Insight250 legends



10 November 2023
by Lilas Ajaluni

Has inflation undermined Europe's insights industry?



3 June
by Madita Brandhorst

Better Brand Health Tracking: Tracking Metrics...

ESOMAR NETWORK

