New Insights from Written Big Data

Text Mining for Marketing
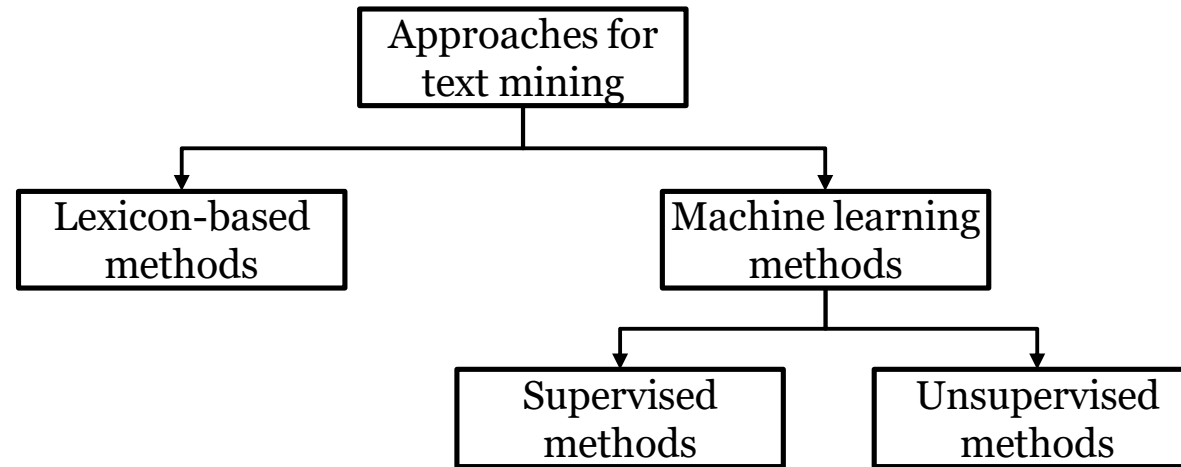
Dr. Evert de Haan

Table of contents

Summary

What is text mining?

› Text mining can be defined as "the discovery by a computer of new, previously unknown information, by automatically extracting information from different written resources". [1]

› Instead of manually going through and classifying written pieces of information, e.g. customer reviews and social media discussions, you let an algorithm do this.

› Advantages are that algorithms are objective (e.g. a piece of text will every time be classified the same way), they can handle much information in a fraction of the time humans can do so, and it is in general much cheaper.

› Disadvantages are that algorithms are not well able to really understand a text; they do not know the context and do not understand sarcasm and cynicism.

› In this report text mining methods, opportunities and challenges in marketing will be discussed.

› [1] http://people.ischool.berkeley.edu/~hearst/text-mining.html

Example text mining approaches

Task	Lexicon-based methods	Supervised methods	Unsupervised methods
Text classification			
› Sentiment	√	√	X
› Content (with custom classes)	X	√	X
Topic modeling	X	X	√
Exemplary methods	LIWC	SVM	LDA

What does ... mean?

› **Latent Dirichlet Allocation (LDA)**: A machine learning technique used to discover and classify the underlying topics of a written piece of text.

› **Linguistic Inquiry and Word Count (LIWC)**: A popular program for text analysis, used for counting words and (basic) sentiment analysis.

› **Machine learning (ML)**: A subset of artificial intelligence. ML methods use a training dataset in order to learn to recognize patterns in the dataset, which can then be used to make (automated) predictions. This can be:

- **Supervised ML**: ML algorithms get training data (e.g. pieces of text) together with indicated and desired outputs (e.g. hand coded sentiment).

- **Unsupervised ML**: ML algorithms recognize patterns by themselves without pre-defined output, e.g. ML algorithms learns by themself that documents which contain many similar words belong to the same topic.

What does ... mean?

› **Natural Language Processing (NLP)**: A field concerned with the interactions between computers and human (natural) languages.

› **Polarity:** The degree to which an expressed opinion in a document or piece of text is positive, negative, mixed, or neutral.

› **Support Vector Machines (SVM)**: A (supervised) machine learning method which can be used to classify data (e.g. for sentiment or content).

› **Sentiment analysis:** A set of NLP techniques that try to find out what the polarity or other emotion is of a (written) piece of information.

› **Topic analysis:** An approach to discover and classify the underlying topics of a collection of texts or documents. LDA is an example of a topic analysis technique.

How popular is text mining?

› The graph below shows the relative worldwide amount of searches for "**natural language processing**" at Google; especially between 2016 and 2018 this topic has become (and stayed) popular.

› The Marketing Science Institute, which connects marketing science with practice, has this as one of their top research priorities for 2018-2020: *"What approaches exist to capture and analyze nonstructured data such as (...) text in order to improve firm communications and the customer experience?"*

› **https://www.msi.org/research/2018-2020-research-priorities/Capturing-Information-to-Fuel-Growth/4.4-approaches-to-ingesting-and-analyzing-data-to-drive-marketing-insights/**

Why is text mining important for managers?

› Firms are facing growing amounts of (customer) data, much of which are unstructured, written data; customer e-mails, comments on social media, reviews, open answers in satisfaction surveys, etc..

› Next to this, firms are trying to become more customer centric, for which firms need to understand the needs, desires and thoughts of their customers.

› Text mining can help to connect these two challenges. With the right text mining procedures, managers can create a dashboard that provides a 360 degree view of their customers.

› In this report, text mining tools will be discussed which can help firms to obtain insights from their written customer feedback (and other written data), and examples will be given on how firms can use these data to improve performance.

Overview of (new) insights on text mining

Research on text mining has shown that:

› There are many tools available for text mining which are valuable for marketing research and marketing practice (Berger et al., 2019; Hartmann et al., 2019; Kübler et al., 2020).

› The sentiment of online written comments on social media is an even better predictor than (survey based) customer satisfaction of future firm performance (e.g. market share, revenue, profit, abnormal stock return) (De Haan, 2020).

› The sentiment and topic of written feedback, e.g. after a service encounter, can help to improve predictions of future customer behavior such as churn (De Haan and Menichelli, 2020).

Recommendations for marketing managers (1/2)

› There is much valuable data available in terms of written feedback, e.g. via correspondence with customers or on social media. These data are typically not systematically analyzed.

› To obtain insights, I recommend:

1. Select the most important written sources that can be used to receive customer insights (e.g. customer correspondence, social media channels, review sites).
2. Periodically (e.g. daily, weekly, monthly) collect all data from these sources.
3. Systematically analyze these data using sentiment and/or topic analyses.

Recommendations for marketing managers (2/2)

› To obtain insights, I recommend (continued):

4. Link the results to one or more KPI's (e.g. customer retention, market share, profit).

5. Monitor and improve the performance of the customer base using the insights from the written data; getting insights is not a goal in itself, the goal is to improve the value to and from the customer base!

› Throughout this report practical and theoretical examples are given, including the (software) tools to do all of this.