

Sentiment Analysis (SA) – Is it all the same?

By Marc Barrachin, CFA and Saeed Shoaraee

This article is the second in a series on Artificial Intelligence (AI), and follows “Demystifying AI”,¹ which was released in April. Sentiment Analysis (SA) – also commonly referred to as Opinion Extraction, Opinion Mining, Sentiment Mining, and Subjectivity Analysis – looks at the use of natural language processing (NLP)² and text analysis techniques to systematically identify, extract, and quantify subjective information and attitudes from different sources. Sentiment can reveal a lot, whether analyzing what a CEO says during quarterly earnings calls, what the social media footprint of a private entity identifies as trends, or the views of employees about a specific company.

We hear a great deal about SA these days, but it is important to understand that it is a broad area and includes many different approaches depending on what you are trying to accomplish, use cases, available data sources, and more. Before getting started, there are a number of important questions that need to be addressed. For example:

- **What are you focused on – i.e., sentiment about what?** Is it financial health, credit risk, business dealings, growth expectations, or employee satisfaction?
- **What is the corpus?** Will it be text, voice, body/facial expressions, or all of these?
- **What document sources will be relevant?** Should it include company sources only (such as press releases, regulatory filings, and investor presentations), or should it also include third-party sources (such as analyst reports, news, and social media)?
- **How will you classify sentiment by its polarity:** Will you label sentiments as positive, negative, or neutral signals, or perhaps some variation of this?

Answers to these questions will help you both frame and scope your SA. There are several other early-stage considerations, as well:

- Are your exposures to public or private companies?
- How much information do you currently have available and how timely is it? Given this, how helpful could SA be for you?
- What geographies are you covering? If you are a large corporation with a global footprint, you will need to deal with the often challenging puzzle of data sources and timing.

¹ “Demystifying Artificial Intelligence (AI): A Framework to Get it Right”, S&P Global Market Intelligence, April 1, 2019, <https://www.spglobal.com/marketintelligence/en/news-insights/research/demystifying-artificial-intelligence-ai-a-framework-to-get-it-right>.

² NLP deals with the interaction of human and machine languages, computer vision, and speech recognition that leverage deep learning techniques.

A Framework for Working with SA

In the Demystifying AI article, we identified six steps to consider when looking at a potential implementation of AI. These steps are also relevant for SA, so we apply the same framework here.

1. Define your problem
2. Define success
3. Determine data needs
4. Evaluate different techniques
5. Assess delivery options
6. Consider maintenance issues

We look at each of these in turn.

1. Define your problem

The first step is to clearly identify what you are attempting to capture, that is, sentiment regarding what? If you are trying to identify risk signals, for example, you will be looking for signs that indicate a potential problem. If a company launches a new distribution partnership, but the CEO's verbiage and phrasing is negative, NLP may pick this up as a risk signal.

Here are several examples of situations where firms may benefit from SA to augment their analysis:

a) Investment Research

- Analysis: Signals that are not necessarily represented in financial statements can help support fact finding. Earnings calls, for example, may help you identify the sentiment of the CEO and CFO during these events by assessing the language or voice tone to determine if they are confident or hesitant. This may be especially important if the changes in these signals can be analyzed over time.

b) Lending or Insurance

- Underwriting: Although there may be a lot of information available for public companies at the time of underwriting (such as financial statements and bank statements), you may want to capture additional insights from management or external news about how the business could evolve in the next few months.
- Portfolio management: Once an insurance policy or a loan has been underwritten, you will need to monitor companies in your portfolio to identify any business challenges they may be facing, or opportunities for further growth. Since financial information can be sparse between releases, news pieces may prove useful.

c) Vendor or Customer Risk Management

- Buyer risk: If you are engaging with customers, you will want to monitor conditions of their businesses and flag any potential risks that could impact cash flow. News may help here, too.
- Supplier risk: Similarly, if you have critical vendors/providers, you will want to monitor their ability to deliver purchased products and services in accordance with agreed terms.

2. Define success

It is important to define criteria that can help increase the likelihood that a potential solution will be successful. For example:

- **Transparency:** Does your use case need to be easily understood by stakeholders, and do you need to explain internally or externally how your SA solution maps input sources to sentiment that may be positive, negative, or neutral?
- **Performance:** What metrics do you need to evaluate the performance of your SA solution? What is the baseline performance for such things as:
 - **Precision** that measures the exactness of a classifier, where higher precision means fewer false positives.
 - **Recall** that measures how many documents with sentiment are rated as sentimental, where higher recall means fewer false negatives.
- **Speed:** Do you need immediate, real-time capture of sentiment, or will periodic updates be adequate?
- **Scalability:** How fast do you need to be able to scale to other languages or use cases? Some methods require feature engineering or manually-crafted rules, which makes them less scalable. In addition, some approaches are language specific, which makes them less scalable to other languages. In Chinese, for example, context and the sequence of words have strong implications for the tone of the text.

3. Determine data needs

There are multiple potential sources of data available for your analysis. They have various characteristics in terms of sources (primary versus secondary), timing (real-time feeds versus quarterly analyst reports), historical availability, and formats. Different data sources will have potential strengths and weaknesses depending on your objectives. You might consider leveraging various sources and creating signals for each, which can be drawn together to help provide a more complete picture. Here are a few examples:

- **Earnings calls, transcripts:** These are primary sources covering U.S. and many global publicly-traded companies, including questions and answers between management and experts. Although the financial data presented will be the same as what is generally available, there is value in understanding the context and management views.
- **News:** The focus should be on reputable news outlets, whether large broad publications or industry-specific. Depending on the source, coverage will tend to center on larger companies, but can also include private companies specializing in a certain industry. Availability tends to be around earnings releases or company announcements, but the goal would be to capture any interim news.
- **Announcements:** Companies announce different events for different reasons. Many large companies announce financial results, investor presentations, marketing events, and mergers and acquisitions. Some of the announcements are mandatory, but you need to be aware of which announcements the company is expected to make and when they will be available.

4. Evaluate different techniques

This step involves trying to identify appropriate techniques for your SA project for the areas outlined below.

- **Pre-processing:** Before you run a sentiment model, you need to prepare the data, which might require many potential techniques, from optical character recognition (i.e., the electronic conversion of images of text into machine-encoded text) to web sourcing (i.e., the extraction of data from websites). Pre-processing is a critical step, especially if your source data is unstructured or is not yet mapped to a company.
- **Labeling:** Labeling is a text classification exercise, where data is tagged with one or more meaningful labels. After obtaining a labeled data set, NLP can be applied so that new unlabeled data can be presented to the model and a likely label can be predicted. In NLP, like other areas of AI, there are two main methods to teach systems to generalize and make predictions:
 - **Supervised learning**, where the system learns how to map input to the output from a set of manually-mapped input/output pairs prepared by humans.
 - **Unsupervised/self-supervised learning**, where the system learns how to map input to the output without access to manually-mapped input/output pairs prepared by humans. Self-supervised learning has been very successful in NLP. For example, in language modeling, one way that a system learns the structure of human language is through following a task: give a sequence of words to the system, mask out 15% of the words in the sequence, and ask the system to predict the masked words.³

A state-of-the-art SA system leverages a combination of transfer learning and task-specific learning. First, some initial understanding of the human language obtained normally through unsupervised/self-supervised learning is “transferred” to the system, which provides a helpful starting point. Following this, the system is trained to combine this initial knowledge with additional lower-level abstraction layers (i.e., the lower the level, the more detail is available) for the specific task of mapping input text to output sentiment labels.

AI systems are only as good as the data set that has been used to train them, and building a reliable label set to train an AI system is critical for its usefulness in downstream applications. Designing how to gather a representative sample of textual input data to send to chosen human annotators and explain what labeling process and rules they have to follow requires planning and resources. In addition, you need to reach a reasonable “inter-annotator agreement” among these individuals⁴ to have their work fit the purpose of the labeling exercise.

Because of the challenges involved in building reliable labeled sets, general application of SA is currently limited. This is especially the case for finance-related applications, where “natural labels”, like star ratings and film reviews used in the movie industry, are not available. For finance-related applications, labels often need to be defined and created manually. This critical step in the success of an SA solution is often overlooked in the industry, with training sets being loosely created for building and assessing the performance of sentiment classification methods. Challenges are even stronger for cases where input

³ To check details of one successful implementation of this method refer to: “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, Cornell University, May 24, 2019, <https://arxiv.org/pdf/1810.04805.pdf>.

⁴ Inter-annotator agreement between labelers is often measured by Cohen’s kappa, Fleiss’ kappa, or Krippendorff’s alpha. For more information refer to: “Inter-Annotator Agreement in Sentiment Analysis: Machine Learning Perspective”, Bobicev, Victoria and Marina Sokolova, RANLP 2017.

corpus contains multiple sentiment labels, for example, in the case of news articles that cover multiple entities.

- **Feature mapping and sentiment classification techniques**

Three broad categories of sentiment classification techniques are: ruled-based sentiment classifiers, machine learning-based sentiment classifiers, and a mixture of the two. Rules-based techniques normally rely on a pre-built dictionary of words and their polarities. Sentiment of the text is then defined by counting how many positive and negative words are appearing in the text.

Machine learning- based classifiers, on the other hand, do not rely on deterministic rules. They transfer text to numerical representations that is understandable by computers (feature mapping) and learn how to map those features to sentiment labels (sentiment classification).

Rules-based approaches are deterministic, easy to define, and easy to implement. However, they can be costly to scale and maintain, as existing rules have to be modified to support specific contexts, and new rules have to be added to support new expressions and vocabulary. In addition, controlling how predefined rules are interacting with each can quickly become a complex task.

Machine-learning approaches are probabilistic and harder to define and implement, but they are more scalable. They can also encode how the tone of a sentence changes based on the order of words, which is a particularly critical feature for some languages and mostly absent in rules-based systems.

Common machine-learning approaches to sentiment classification include mapping text to numerical feature vectors (feature mapping), and then training a classifier (sentiment classification). For sentiment classification, Naïve Bayes, Support Vector Machines, Random Forest, or Deep Learning⁵ classifiers are often being used. For feature mapping, typical methods include:

Bag of Words (BoW): In this method, text is represented as a bag of words (unigrams) or a collection of words (n-grams), without considering their semantics. At its most basic form, each word is mapped to a frequency of its appearance in the document, and those frequencies are used as a feature in the sentiment classification model. Since words that are systematically frequent across all documents have less information value, an enhancement to this basic approach is to penalize words that are appearing frequently in all documents, to identify words that are frequent as well as distinctive (what is known as the Term Frequency-Inverse Document Frequency (TF-IDF) approach).

Word embedding: One main shortcoming of the BoW approach is that it does not capture the meaning of the words. Word embedding is a technique that maps each word, or collection of words, to a vector. The mapping is done in a way where semantically similar words are positioned

⁵ For general overview of Naïve Bayes refer to “1.9 Naïve Bayes”, Scikit Learn, https://scikit-learn.org/stable/modules/naive_bayes.html; Support Vector Machines refer to “1.4 Support Vector Machines”, Scikit Learn, <https://scikit-learn.org/stable/modules/svm.html>; Random Forest refer to “1.11 Ensemble Methods”, Scikit Learn, <https://scikit-learn.org/stable/modules/ensemble.html>; and Deep Learning refer to “Introduction to Neural Networks”, Analytics Vidhya, October 2018, <https://www.analyticsvidhya.com/blog/2018/10/introduction-neural-networks-deep-learning/>.

closer to each other in the vector space, effectively providing the ability to capture semantic information about words and their relationship to one another.

Since their introduction in 2013, word embeddings have been highly leveraged in different NLP tasks. They are trained on large amounts of unlabeled data (unsupervised/self-supervised learning) via algorithms, such as word2vec⁶ or GloVe,⁷ and have been used to initialize the first layer of neural network in a deep learning-based classifier (effectively, transferring the learning obtained in a self-supervised manner to a task, such as SA).

Language modeling: Using word embeddings to initialize a first layer of neural network provides a shallow representation of the language. The rest of the “deep” neural network still needs to be trained for the specific task. For SA, the network should still learn how to drive meaning from a sequence of words and relate them to sentiment labels through layers that come after the initial embedding layer. For that, the network needs access to a relatively large number of labeled documents to achieve reasonable performance.

Since 2018, a series of new approaches⁸ have been introduced by NLP researchers that, instead of just initializing the first layer of neural network, the entire deep pre-trained network is transferred to a downstream task. These deep neural networks have been trained for language modeling tasks, which mask a percentage of words in a sequence and ask the network to predict the masked words. A model trained for a language modeling task has been proven to be useful for a diverse range of other NLP tasks, from SA to question answering, co-reference resolution, and more.

5. Assess delivery options

Whether you are looking to consume the signals within another application (e.g., your origination or portfolio management system) alongside other data sets via your standard workflow platform, screens, API, or Excel[®] spreadsheets, you will need to explain your process. Documentation will be important, but you may also consider leveraging visualization techniques to assist in explaining your model and its results. This is especially important if you are leveraging a machine learning-based approach, as opposed to a rules-based approach for your solution. Good visualization provides transparency to your solution and helps users understand part of the input source, such as sentences or paragraphs, which are main drivers of your model to assign specific labels to the input text. Even for system-to-system applications, transparency to your solution will be in high demand. For example, some clients expect to receive sentiment classification outputs, as well as “snippets” from the inputs that are the main drivers of the sentiment classifier via their APIs to augment their own or existing systems.

6. Consider maintenance issues

Traditionally, standard models have to be recalibrated on a regular basis. With sentiment, the data itself might change over time: types of announcements are created or taken away, languages change over time, or sources dry up or new ones become available. You need to continually monitor the data and output of your model.

⁶ A group of related models that are used to produce word embeddings.

⁷ An unsupervised learning algorithm for obtaining vector representations for words.

⁸ For further details refer to “Deep contextualized word representations (ELMO)”, Allen Institute for Artificial Intelligence, 2018, <https://arxiv.org/abs/1802.05365>; “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, 2019, <https://arxiv.org/pdf/1810.04805.pdf>; and “Universal Language Model Fine-tuning for Text Classification (ULMFit)”, 2018, <https://arxiv.org/abs/1801.06146>.

Conclusion

As discussed in Demystifying AI, you need to be mindful of business risks associated with cyber security, regulation, data privacy, global and region-specific policy issues, as they are important inputs for your solution. It is also important to remember that not all SA is the same. It is a very broad area and there is much to consider. SA initiatives underway at S&P Global Market Intelligence all have different characteristics that reflect their overall use case, available data sources, and chosen methodologies. Following the six-step approach laid out in this article may assist you in framing and scoping any SA you are considering.

About the Authors

Marc Barrachin, CFA

Head of Product Research & Innovation, Risk Services, S&P Global Market Intelligence

Marc Barrachin, CFA, is Head of Product Research and Innovation for the Risk Services group at S&P Global Market Intelligence. Marc searches and analyzes future trends and their potential impact on the group's product offering, and identifies areas of opportunities to build new capabilities. He currently leads projects at the junction of risk, fintech, and regulations. These include cyber risk and the development of a 360 degree view of a company's risk, which encompasses private companies and companies in China. This requires partnering with stakeholders across S&P Market Intelligence segments (e.g., commercial banks, insurance companies, and corporations) and developing new models that leverage a wide array of alternative data and modeling techniques, such as machine learning.

Previously, Marc was Managing Director at IHS Markit. He managed the index business from product to commercial policy, and directed the growth of the credit business through the crisis and subsequent market restructuring and regulatory evolution. Marc also spent over 12 years at Interactive Data in a variety of roles.

Saeed Shoaraee

Data Scientist, Product Research & Innovation, Risk Services, S&P Global Market Intelligence

Saeed is a Data Scientist on the Product Research and Innovation team for the Risk Services group at S&P Global Market Intelligence. Saeed is focused on applying machine learning techniques to extract risk signals from alternative datasets; in particular applying deep learning to natural language processing.

Previously, Saeed held senior quantitative positions at BMO Financial Group, R2 Financial Technologies, and BBVA Corporate & Investment Banking, where he worked on validating credit risk models; developing interest rate and FX derivative pricing, and risk management solutions; and structuring Fixed Income products.

About S&P Global Market Intelligence

At S&P Global Market Intelligence, we know that not all information is important—some of it is vital. Accurate, deep, and insightful. We integrate financial and industry data, research and news into tools that help track performance, generate alpha, identify investment ideas, understand competitive and industry dynamics, perform valuation, and assess credit risk. Investment professionals, government agencies, corporations, and universities globally can gain the intelligence essential to making business and financial decisions with conviction.

S&P Global Market Intelligence is a division of S&P Global (NYSE: SPGI), which provides essential intelligence for individuals, companies, and governments to make decisions with confidence. For more information, visit www.spglobal.com/marketintelligence.

Contact Us

Americas

+1 877 863 1306

Europe, Middle East
and Africa

+44 (0)20 7176 1234

Asia Pacific

+852 2533 3565

credit_analytics
@spglobal.com

www.spglobal.com/marketintelligence

Copyright © 2019 by S&P Global Market Intelligence, a division of S&P Global Inc. All rights reserved.

These materials have been prepared solely for information purposes based upon information generally available to the public and from sources believed to be reliable. No content (including index data, ratings, credit-related analyses and data, research, model, software or other application or output therefrom) or any part thereof (Content) may be modified, reverse engineered, reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the prior written permission of S&P Global Market Intelligence or its affiliates (collectively, S&P Global). The Content shall not be used for any unlawful or unauthorized purposes. S&P Global and any third-party providers, (collectively S&P Global Parties) do not guarantee the accuracy, completeness, timeliness or availability of the Content. S&P Global Parties are not responsible for any errors or omissions, regardless of the cause, for the results obtained from the use of the Content. THE CONTENT IS PROVIDED ON "AS IS" BASIS. S&P GLOBAL PARTIES DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR USE, FREEDOM FROM BUGS, SOFTWARE ERRORS OR DEFECTS, THAT THE CONTENT'S FUNCTIONING WILL BE UNINTERRUPTED OR THAT THE CONTENT WILL OPERATE WITH ANY SOFTWARE OR HARDWARE CONFIGURATION. In no event shall S&P Global Parties be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including, without limitation, lost income or lost profits and opportunity costs or losses caused by negligence) in connection with any use of the Content even if advised of the possibility of such damages.

S&P Global Market Intelligence's opinions, quotes and credit-related and other analyses are statements of opinion as of the date they are expressed and not statements of fact or recommendations to purchase, hold, or sell any securities or to make any investment decisions, and do not address the suitability of any security. S&P Global Market Intelligence assumes no obligation to update the Content following publication in any form or format. The Content should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. S&P Global Market Intelligence does not act as a fiduciary or an investment advisor except where registered as such. S&P Global keeps certain activities of its divisions separate from each other in order to preserve the independence and objectivity of their respective activities. As a result, certain divisions of S&P Global may have information that is not available to other S&P Global divisions. S&P Global has established policies and procedures to maintain the confidentiality of certain non-public information received in connection with each analytical process.

S&P Global Ratings does not contribute to or participate in the creation of credit scores generated by S&P Global Market Intelligence. Lowercase nomenclature is used to differentiate S&P Global Market Intelligence PD credit model scores from the credit ratings issued by S&P Global Ratings.

S&P Global may receive compensation for its ratings and certain analyses, normally from issuers or underwriters of securities or from obligors. S&P Global reserves the right to disseminate its opinions and analyses. S&P Global's public ratings and analyses are made available on its Web sites, www.standardandpoors.com (free of charge) and www.ratingsdirect.com (subscription), and may be distributed through other means, including via S&P Global publications and third-party redistributors. Additional information about our ratings fees is available at www.standardandpoors.com/usratingsfees.