ARX Business Model Investment Series (2022), Vol.1, No. 1, pp. 29-38

Asia-Pacific Research Exchange

Business Model Investment Series

Natural Language Processing in Banking: Current Uses

Kevin Pereira & Rajendra Shroff

Abstract

This article will introduce a branch of Artificial Intelligence called Natural Language Processing (NLP) and dive into two popular NLP models, Word2vec and Bag of Words. Subsequently, the article will examine NLP applications in banking use cases like document search, investment analysis, and customer service. Finally, we will end with some key takeaways for banking executives.

Keywords: Artificial Intelligence, Natural Language Processing, NLP, Business model, business strategy, sector analysis, stakeholder analysis.

Banks are using a branch of Artificial Intelligence called natural language processing (NLP) to automate certain document processing, analysis, and customer service activities. Three applications include:

- **Intelligent document search**: finding relevant information in large volumes of scanned documents.
- **Investment analysis**: automating routine analysis of earnings reports and news so that analysts can focus on alpha generation.
- **Customer service & insights**: deploying chatbots to answer customer queries and understand customer needs.

We will cover real-life examples of what banks are doing in these areas. First, let's go over what natural language processing is capable of.

Introduction to Natural Language Processing

Natural Language Processing (NLP) is a branch of Artificial Intelligence that enables computers to understand human language and respond in kind. This involves training computers to process text and speech and interpret the meaning of words, sentences, and paragraphs in context.

Human-Computer Interactions

Human-computer 'conversations' can be broken down as follows (we'll get to the specific AI methods a little later):

1. We provide text or speech input (e.g. typing into a chatbot interface or talking to a smart speaker).
2. The computer converts the text/speech into a format it understands (e.g. converts speech to text and words to vectors). This helps computers cluster and classify different words.
3. The computer figures out meaning and context using its own data sets.
4. The computer determines an appropriate response and converts it to text or speech that we understand and responds to us.

We interact with apps that use natural language processing every day:

- **Google Translate**: we input text and speech that Google translates for us.
- **Gmail Smart Compose**: You might notice that Gmail suggests the rest of the sentence that you've begun to type. This feature uses the email subject and previous emails to suggest relevant text. Kind of scary, yet kind of cool.
- **Grammarly**: the popular grammar checker that you use because it is so much better than Microsoft Word's spell check.
- **Smart speakers**: No, your conversation with Alexa isn't magic.

Understanding, Processing and Generating Language

Natural language processing is actually an umbrella term that includes two related methods: **Natural Language Understanding** and **Natural Language Generation**.

Natural language understanding (NLU) figures out the meaning behind text and speech. Think of this as reading or listening. This involves taking unstructured text and speech input from humans and converting it to structured formats that computers understand. When you ask Alexa for a weather report, for example, it uses natural language understanding to figure out what you're saying.

Natural language generation (NLG) refers to computer-generated text and speech. NLG turns structured data into text and speech that humans understand. Continuing our previous example, Alexa uses natural language generation when it responds 'It is sunny today. Would you like to place an order for sunglasses?'

AI Methods Used for Natural Language Processing

Natural language processing is often used with other AI methods such as **neural networks**, **deep learning** and **optical character recognition**. Two popular natural language models are **Word2vec** and **Bag of Words.**

Without getting technical, neural networks are a subset of machine learning. When used for natural language processing, they can process text, classify words, cluster similar words, and associate words and phrases with meanings. Deep learning methods (i.e. neural networks with many layers) such as Recurrent Neural Networks are also used.

Optical character recognition (OCR) enables computers to recognize text in scanned documents. OCR can be used with natural language processing to analyze scanned documents or handwritten text.

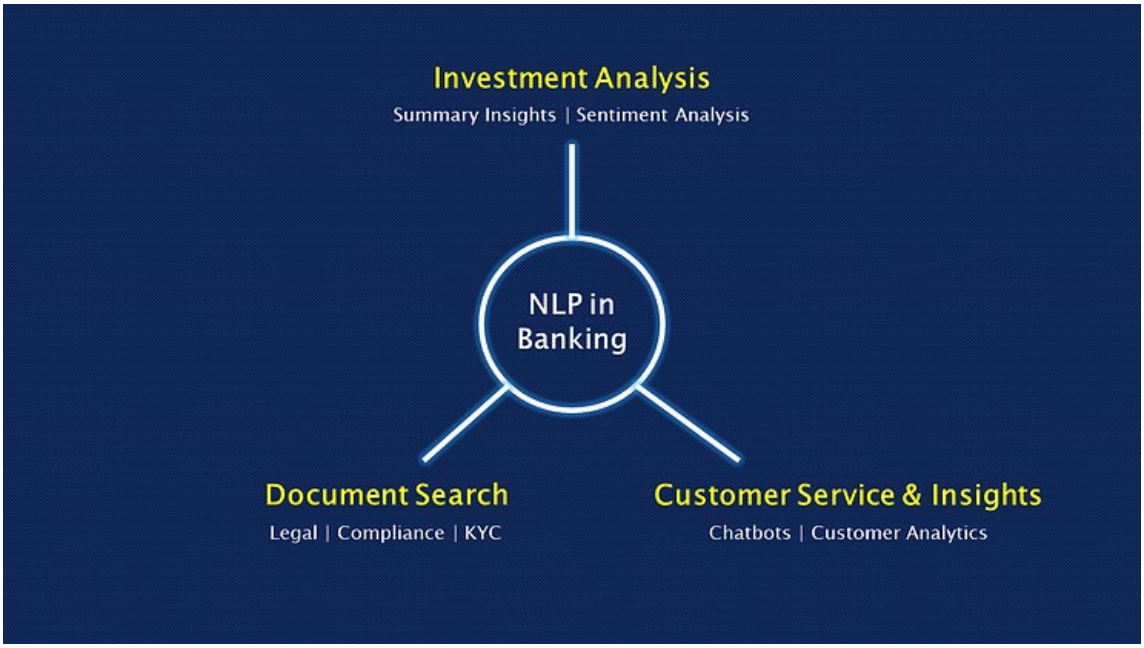
Various natural language techniques are used to determine grammar rules and word meanings. **Syntax analysis** involves determining grammar rules for words and cluster them according to similarity. **Semantic analysis** involves deriving meaning and is used to generate human language. Semantic analysis is challenging because human language rules are complex. Words and phrases take on different meanings in different contexts. Colloquialisms, idioms and sarcasm further complicate matters.

Bag of Words and related algorithms are popular natural language techniques that classify phrases and document by category or type. Bag of Words simply counts how often each word appears in a document (a tally). The algorithm then compares documents and determine the topic of each document. This can be used to train neural networks. Gmail's Smart Compose (mentioned earlier) uses Bag of Words and Recurrent Neural Network models according to Google. Search engines also use these techniques.

Word2vec is another popular natural language model. It is a two-layer neural network that classifies text to determine meaning. It converts words to mathematical 'vectors' that computers can understand. Vector conversion is required because neural networks work better with numerical inputs.

Given a large enough data set, vectors representing similar words are grouped together — similar words are mathematically detected. Properly deployed, Word2vec can infer word meanings with high accuracy based on past appearances. This is useful for document search, sentiment analysis, and even recommendations of which words should come next to complete a sentence.

How Banks Use Natural Language Processing



Banks can apply natural language processing to large volumes of text and speech data to extract information, gain insights, and streamline manual tasks. While time and cost savings are obvious benefits, the ability to identify key information (the proverbial needle in the haystack) can be a competitive difference maker.

Here are three areas where banks are applying natural language processing.

Intelligent Document Search

JP Morgan Chase's COIN (Contract Intelligence) software uses natural language processing to help the bank's legal team search and review large volumes of legal documents.

COIN can reportedly save the bank's legal team 360,000 hours, or 15,000 days, of document search tasks per year. It can extract key data and clauses to help loan officers review commercial loan agreements, for example.

COIN is apparently trained to recognize key information (attributes) within documents that the bank's legal team flags as important. This enables the software to extract key information from documents that are structured differently. The bank claims it extracted 150 relevant attributes from 12,000 commercial credit agreements *in seconds*.

The software's workings are not public since it is used internally. We speculate that it could be powered by natural language processing (to search within documents), optical character recognition (to recognize characters in scanned documents), and machine learning (to classify & cluster data within documents and to improve search algorithms over time).

These methods can be applied to other banking activities. It can help banks to extract various types of customer data that they don't have time to track. This data could help predict customer needs and identify cross-selling opportunities. It can also speed up Know Your Customer (KYC) processes that require document analysis, thereby making customer onboarding easier.

Investment Analysis

Securities research desks at banks are using natural language processing to find valuable insights within mountains of company reports and conference calls.

Banks previously hired armies of analysts to comb through earnings reports and other filings and enter pertinent data into databases and valuation models.

Now, banks are using natural language processing tools that 'read' hundreds of documents at a time and summarize key information for human analysts. Speech analysis tools can 'listen' to analyst conference calls to determine the tone and sentiment behind what company management is saying, which provides insight for equity analysis. These tools are huge time savers and allow analysts to focus on alpha generation.

Banks also use natural language processing for **sentiment analysis**. These tools analyze large volumes of news and social media posts to extract key insights, determine how a company is perceived, or track market reaction to significant events. These timely insights can inform analyst recommendations.

Banks either use tools developed internally or by vendors. One vendor, Dataminr, claims to analyze social media and financial news to identify relevant information including unexpected news, emerging trends, or risks.

On the sell side, natural language generation tools automatically generate reports based on earnings reports and news, for example.

Customer Service & Insights

Major banks are introducing some level of **customer service** automation through chatbots. In early 2019, Bank of America launched Erica, a mobile virtual assistant, which soon amassed over one million users through the BofA mobile app.

Erica accepts voice and text commands and combines predictive analytics with natural language processing to help customers:

- Check balances and transfer money
- Searching for past transactions and account info on demand
- Track spending habits *(probably using predictive analytics, which is a value-add that encourages more chatbot usage)*
- Help customers manage recurring or late payments

Chatbots let customers access account information and perform basic transactions on their phone instead using internet banking or visiting their local branch. Executing transactions through a clean chatbot interface may also take less time.

A bigger win for banks will be using natural language processing for **customer insights.** Using methods related to intelligent document search and sentiment analysis described above, banks can better understand and predict customer needs and pain points.

Sentiment analysis tools can monitor social media to see what people are saying about the bank. Document search tools can analyze feedback forms and customer information to respond to issues, offer tailored products, and increase customer retention.

Takeaways for Banking Executives

Banking leaders realize that natural language processing can automate routine document analysis, research, and customer service.

Cost savings are just the tip of the iceberg. By analyzing text and speech data more quickly and extracting more actionable insights on customers and the market, banks can serve customers better and make better investments. The potential for greater market share and income are the real difference makers.

While we have not covered all possible use cases, banks can apply natural language techniques to any function that processes large volumes of text or speech data. There are numerous applications in compliance, risk management, or order execution, for example.

Key considerations include whether to build AI and natural language processing tools in-house or to license software from an AI vendor. Building in-house requires data scientists, developers, and an organizational AI strategy. While this takes time, internally developed solutions might meet the bank's needs better than a vendor product. In addition, data quality and availability across departments also have to be addressed.

Given the broad range of banking activities that natural language processing can be applied to, banks that apply these solutions across departments are likely to see far greater returns on investment.

References

https://medium.com/analytics-vidhya/understanding-neural-networks-from-neuron-to-rnn-cnn-and-deep-learning-cd88e90e0a90

https://towardsdatascience.com/whatnlpscientistsdo-905aa987c5c0

https://wiki.pathmind.com/bagofwords-tf-idf

https://ai.googleblog.com/2018/05/smart-compose-using-neural-networks-to.html

https://wiki.pathmind.com/word2vec

https://emerj.com/ai-sector-overviews/natural-language-processing-banking-current-applications/

https://promotions.bankofamerica.com/digitalbanking/mobilebanking/erica

https://netfinance.wbresearch.com/blog/bank-of-america-ai-powered-chatbot-strategy

About the Authors

Kevin Pereira is the Managing Director at Blu Artificial Intelligence, a consulting firm specializing in AI consulting and AI education. He also teaches classes on Artificial Intelligence at University of Hong Kong and the Hong Kong University of Science and Technology.

Rajendra Shroff is the Head of Applied AI Research at Blu Artificial Intelligence, a consulting firm specializing in AI consulting and AI education. He is also a Visiting Scholar at the Hong Kong University of Science and Technology and the Hong Kong Metropolitan University, where he teaches classes on Big Data and Artificial Intelligence.