BofA SECURITIES

BofA GLOBAL RESEARCH

Predictive Analytics

Now you're speaking my language: NLP primer with 10 financial markets use cases

Primer | 09 March 2021 | Predictive Analytics | Global

Key takeaways

- NLP can help investors assess risk by either detecting changes in company language or in social media.
- Better NLP tools matter as companies are avoiding traditional financial language in transcripts and filings.
- We illustrate key NLP concepts through 10 financial markets use cases using sentiment analysis, text differencing, etc.

BofA Data Analytics

Exhibit 1: Wordcloud on earnings transcripts

BofA SECURITIES

BofA GLOBAL RESEARCH

Source: BofA Global Research

BofA GLOBAL RESEARCH

Why NLP matters for investors

Today investors are hit with a blizzard of information where traditionally they focused their attention on numeric data for fundamental or quantitative analysis. However, given the emergence of alternative data (https://rsch.baml.com/r?q=cFh9xrqnoW193pFcjkxVYQ), the data is increasingly arriving in unstructured forms such as text, images, audio, etc. By using Natural Language Processing, investors can detect if management is not answering questions on earnings transcripts adequately or how often management is changing their language in 10-K/10-Q SEC filings, both of which could be signs of instability leading to financial underperformance. From a macro perspective, NLP can be used for quantifying sentiment in petrochemical industry publications, finding key inflection points in Central Bank Statements (CB), or with the recent retail trading frenzy, investors can assess what are the key topics trending on Reddit in order to gauge any percolating frenzy versus normal discussion.

What is NLP?

Natural Language Processing (NLP) is one of the most pervasive technologies since the internet age with applications ranging from web search, emails, language translation, virtual agents, artificial intelligence, etc. NLP is the process of turning unstructured text into structured insight (i.e. sentiment) or artificial intelligence (i.e. generating automated text). Financial markets examples typically involve sentiment, textual differencing, topic modeling across text data including news, earnings transcripts, SEC filings, CB statements and social media (i.e. twitter, Instagram, Reddit, etc).

How to talk when a Machine is Listening?

Traditional word lexicons (e.g. dictionaries) have started to lose their value as companies can learn which negative words are the ones to avoid. This creates a feedback effect where companies can adjust their corporate disclosures by adjusting their language as noted by S. Cao, W. Jiang[(*)]. This suggests the need for more sophisticated tools in NLP by utilizing machine learning and deep learning over static word based lexicons.

10 NLP financial market use cases

We present 10 NLP financial market use cases that allow us to bring examples to life. They include: 1) Measuring Responsiveness using Q and A on transcripts, 2) Predicting High Yield Defaults using transcripts, 3) Retail Frenzy vs Normal Discussion on Reddit, 4) Measuring changes in SEC 10-Qs, 5) Petrochemical sentiment on key IHS market commentary, 6) BofA Analyst Tone on equity fundamental reports, 7) Learning from Wikipedia to enhance sentiment, 8) Several Central Bank sentiment indicators, 9) Categorizing Future vs Past Tense on transcripts, 10) LIBOR replacement language on bond disclosures. We use a range of NLP techniques including Support Vector Machines, TF-IDF, K-Means clustering to utilizing the latest cutting edge NLP technologies including Bert, ULMFIT, LSTM and Lexicon enhanced by a Word Embedding.

Summary of 10 NLP Use Cases

Exhibit 2: Summary of 10 Financial Markets NLP Use Cases

Use Case	Tools	Why it's important
Measuring Management Responsiveness to Analyst Questions	Word2Vec, TF-IDF, Cosine Similarity	Assessing management's directness with answering analyst questions to fundamental performance.
Predicting High Yield Defaults using earnings transcripts	Support Vector Machine, TF-IDF	Using earnings transcripts outside of the traditional macro factors offe independent edge and nontrivial prediction.
Reddit Retail Frenzy vs Normal Discussion to assess sentiment	Topic Latent Dirichlet Allocation	Gauging the overall market sentiment with the recent Reddit frenzy di help investors protect portfolios against future asset price bubbles.
Avoid financial disclosure changers on SEC filings	Cosine similarity	Companies who tend to change their 10-Q/10-K more frequently have have lower overall financial performance than those who have more co language from quarter to quarter.
Libor replacement language identifying which MBS are exposed	K-Means clustering, LDA	Identify which legacy MBS deals with outstanding bonds contain Libor order to assess overall market impact with Libor replacement.
Estimating BofA Analyst Tone on Fundamental Equity reports	Long-Short Term Model, Lexicon	Systematic way of extracting sentiment from BofA Analysts fundamen reports offers a quantamental approach to trade sector and industry gr strategies.
Predicting IMDB sentiment with the latest NLP technology	Universal Language Model Fine Tuning	Even though this is our one non-financial example, it is important to sh NLP technology beats all existing models in terms of sentiment. If inve sentiment models this can assist with adequately assessing asset price
Emerging Market Central Bank Sentiment	Market informed custom lexicons	We translate over several Central Bank Statements to assess overall ma which is powerful for trading Emerging Markets FX.
Management discussions on Future relative to past	Dependency Parsing	Management discussion that contains more future tense relative to pa seen to contain more leading information.
Market sentiment with petrochemicals	Word2Vec enhanced lexicon	Our sentiment indicator offers leading information on key inflection pc petrochemical sector

Source: BofA Global Research

Bof

What is NLP?

Natural Language Processing (NLP) is one of the most pervasive technologies since the internet age with applications ranging from web search, emails, language translation, virtual agents, artificial intelligence, etc. Applications with respect to financial markets primarily involve sentiment analysis, measuring responsiveness, predicting high yield bond defaults, topic modelling across wide range of sources of text including news, earnings calls transcripts, SEC filings, social media (Twitter, Instagram, Reddit, etc), job postings and Glassdoor.

Latest NLP technology trends

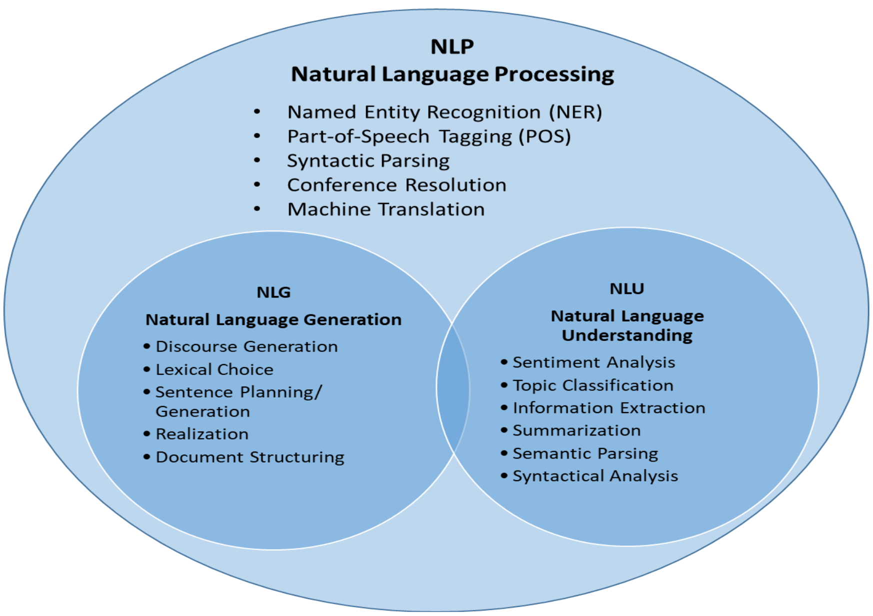
In 2018, NLP experienced a strong year with breakthroughs in Transfer Learning where language models(*) can be trained on large corpus of texts, such as all of Wikipedia, in order to transfer this initial learning to a wide ranged range of other NLP oriented tasks such as sentiment, language translation, sematic textual differencing, etc. One model in particular that has been at the center of these breakthroughs is called Bert (Bidirectional Encoder

Representations from Transformers) that broke the majority of NLP accuracy records across a wide range of NLP problems. However, for sentiment classification for example, Bert models are now already underperforming marginally on the standard NLP datasets (such as IMDB) by a newer model called XLNet[*] that builds in an autoregressive features.

NLP Overview

Natural Language Processing is the process of trying to turn unstructured text data into structured. Structured is established through NLP techniques such as Named Entity Recognition where language tags are established by pre-defined categories such as person name, location, organization, etc as well as through reducing words to their root form as done with an NLP approach called lemmatization.

Exhibit 3: Natural Language Processing Overview



Source: BofA Global Research

BofA GLOBAL RESEARCH

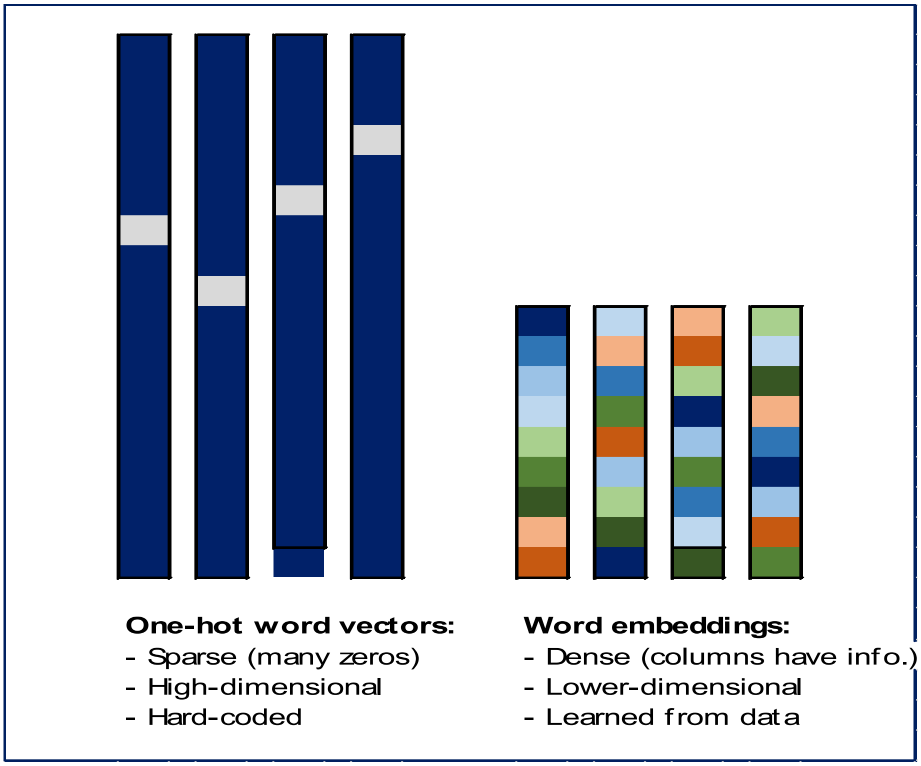
Natural Language Understanding is perhaps the most prevalent with financial markets as approaches such as sentiment analysis, topic classification and semantic textual differences are among the most popular. We will demonstrate these concepts in depth throughout our ten financial markets use cases.

Natural Language Generation tends to be more associated with enabling computers to write text with the latest Artificial Intelligence[*], where the writing is near indistinguishable from humans. Or enabling chat bots through reading comprehension or realization. There tends to be less financial markets applications with NLG. However, language understanding has started to empower sentiment analysis (7 - Learning from Wikipedia) with techniques known as Transfer Learning and with empowering NLP models such as Word Embeddings with sentence context (Word Embeddings are a low dimensional representation of the word but will pull a list of common words that are associated with it as depicted in Exhibit 4. From a practicality standpoint, Word Embeddings can typically have up to 250-500 columns that contain rich information about what a word represents as opposed to a one-hot encoding that can be up to 20k columns.).

What is a Word Embedding?

Historically in NLP problems it was typical to first convert a word into a binary list where it was mostly zeros but only equal to one when the word was present and the full length of the list equaled the total size of the vocabulary (called one-hot word vectors). The issue with this is that this makes the data very sparse (as mostly zeros) and very highly dimensional (since doing this for every word in the vocabulary). An alternative to this approach is a Word Embedding matrix, where we can represent a word with a list of similar words in a lower dimension (i.e. less columns).

Exhibit 4: Better representation of words



Source: BofA Global Research

BofA GLOBAL RESEARCH

Word Embeddings are a low dimensional representation of the word but will pull a list of common words that are associated with it as depicted in Exhibit 4. From a practicality standpoint, Word Embeddings can typically have up to 250-500 columns that contain rich information about what a word represents as opposed to a one-hot encoding that can be up to 20k columns.[*]

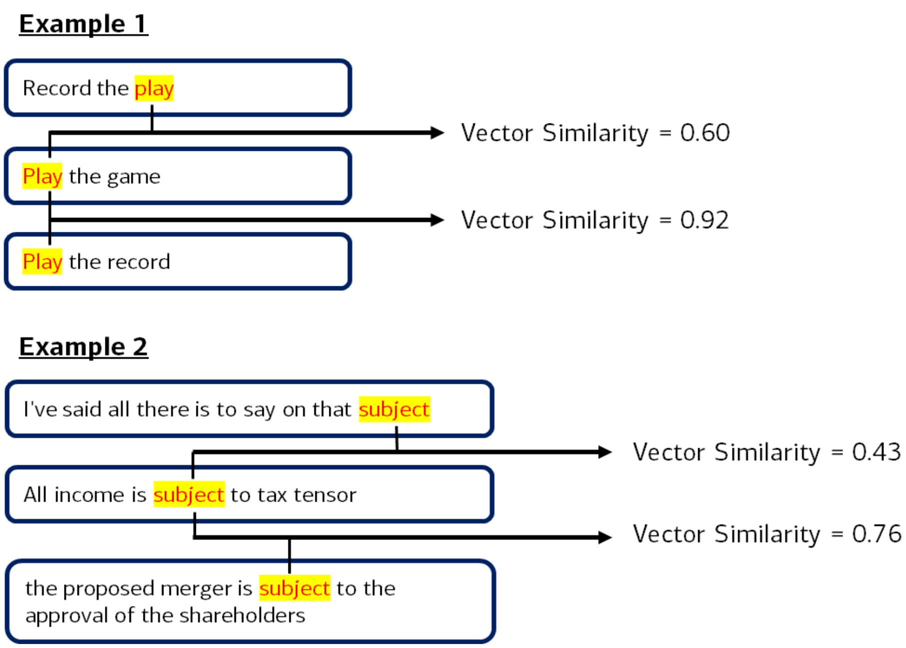
Context matters (Bert)

BERT Embeddings with key examples

In 2018, NLP experienced a phenomenal year with breakthroughs in Transfer Learning where language models[*] can be trained on large corpus of texts, such as all of Wikipedia, in order to transfer this initial learning to a wide ranged range of other NLP oriented tasks such as sentiment, language translation, sematic textual differencing, etc. One model in particular that has been at the center of these breakthroughs is called Bert (Bidirectional Encoder Representations from Transformers) that broke the majority of NLP accuracy records across a wide range of NLP problems. These language models can then be fine-tuned for a relevant text corpus such as in Finance.

In the below examples, we can see how with the word '**play**', Bert embeddings is able to identify that 'Record the **play**' is quite different from '**Play** the game' versus '**Play** the game' and '**Play** the record'. Vector similarity is running the cosine similarity between both but we can essentially think of this as a correlation metric.(*)

Exhibit 5: Context matters: BERT Embeddings



Source: BofA Global Research

BofA GLOBAL RESEARCH

Financial Markets NLP Use Cases

1 - Measuring responsiveness

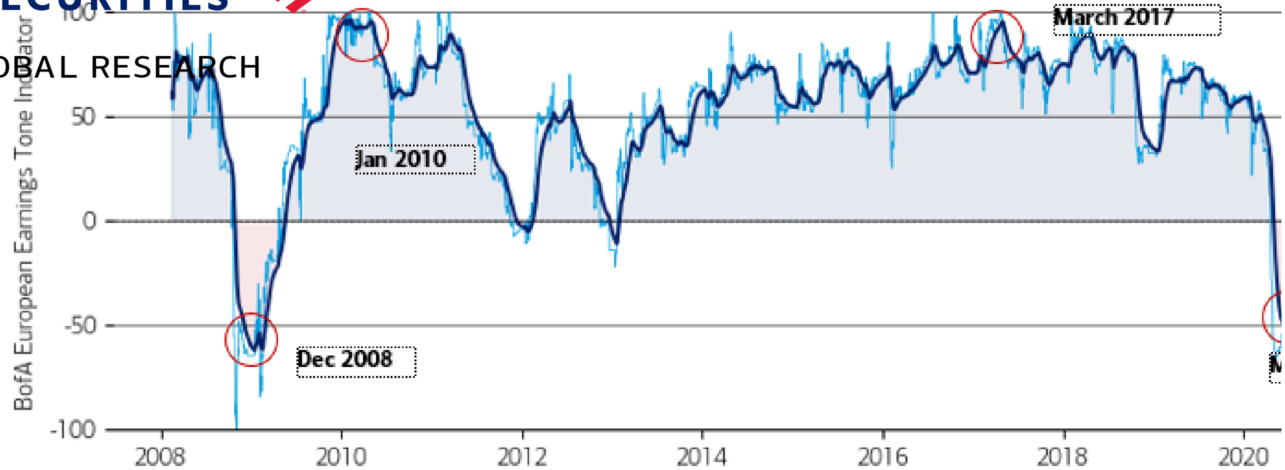
Estimating cosine similarities on Word2Vec using earnings transcripts

Understanding the sematic meaning of words has always been a challenge in machine learning based algorithms. Working with Especially in Finance where words like 'SEC' would mostly refer to regulators instead of "seconds" in time. For example, in our recent NLP work in collaboration with our European Quant team (BofA European Earnings Tone Indicator (https://rsch.baml.com/r?q=GfP7okTuVu-qWxvPrcRRZA)) we measured the responsiveness of management to an analyst question along with sentiment. Off the shelf trained models might not be able to capture such subtleties and there are multiple approaches which can be leveraged to convert the text into numbers that would retain the relationship between words. Retaining such relationships is also important for further systematic analysis involving cosine similarity (See What is a Cosine Similarity?) to quantify the contextual similarity between two text.

Exhibit 6: Our BofA European Earnings Tone Indicator rebounded sharply after May '20 lows and is now firmly in positive territories
BofA European Earnings Tone Indicator

BofA SECURITIES

BofA GLOBAL RESEARCH

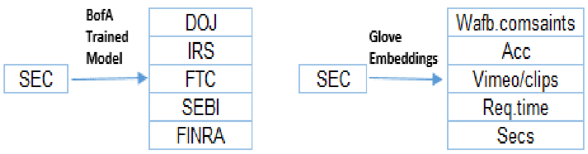


Source: European Equity Quant Strategy, Predictive Analytics, FactSet Earnings Transcript Data, Bloomberg. The light blue line shows the exponential moving average of stock-level Earnings Tone life of 30 calendar days. The BofA European Earnings Tone indicator (dark blue line) is obtained by double exponential smoothing of stock level Earning Tones scores where the trend smoothing f of 15 days. The indicator identified as the BofA European Earnings Tone Indicator above is intended to be an indicative metric only and may not be used for reference purposes or as a measure o financial instrument or contract, or otherwise relied upon by third parties for any other purpose, without the prior written consent of BofA Global Research. This indicator was not created to act a

Bof

These techniques are advancement to frequency base metrics. These tagging and dependency analysis comes handy when we want to analyze the structure of the sentence and the dependence among them. Sometimes they are helpful in focusing on the sub part of the sentence. A model capable of understanding the financial jargon and context was critical because if an analyst asks a question around regulators like SEC then mention of other regulators in answer should also be considered relevant to discussion. There are pre-trained embeddings available but we trained our proprietary word2vec model which is better able to capture such subtleties. We restricted the underlying training corpus to financial documents only instead of generic text. This helped in getting the word embedding that captures the usage of words in financial context. Our model is able to understand that term "SEC" is used to identify the regulators rather than "seconds" in time. Exhibit 8 shows most similar words from BofA trained model and pre-trained Glove embeddings.

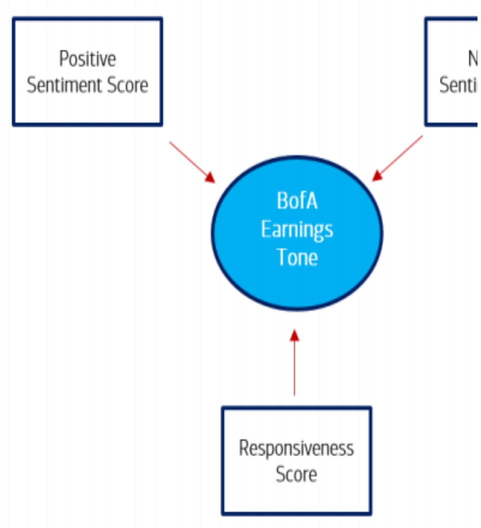
Exhibit 7: Importance of training a Finance specific Word Embeddings



Source: BofA Global Research

BofA GLOBAL RESEARCH

Exhibit 8: The BofA Earnings Tone is an equal-weigh Responsiveness, Positive Sentiment Score and Nega Sentiment Score



Source: BofA Global Research

Bof

Word2Vec model:

> **Word2vec** is a group of related models that are used to produce word embeddings. These models are generally two-layer neural networks that are trained to reconstruct linguistic contexts of words. Word2vec takes as its input a large corpus of text and produces a vector space, typically of several hundred dimensions, with each unique word in the corpus being assigned a corresponding vector in the space. Word vectors are positioned in the vector space such that words that share common contexts in the corpus are located close to one another in the space.

2 - Predicting HY Bond defaults

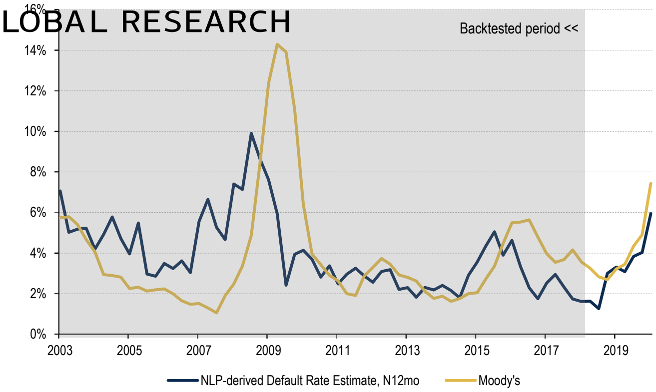
Support Vector Machines using earnings transcripts

Working with our High Yield credit strategists, we developed an NLP model using earnings transcripts to predict High Yield bond default rates (AI meets HY (https://rsch.baml.com/r?q=PzKqXnQz87DoPOIEeIXTbA)). We used a Machine Learning model called Support Vector Machine(*) (SVM) in order to detect key language, if mentioned on the earnings call, leads a higher likelihood of debt default over the next 12 months. The NLP engine parses thousands of key phrases such as 'cost cutting', 'asset sales' and 'cash burn' to establish the linkages for defaults occurring.

BofA SECURITIES

BofA GLOBAL RESEARCH

Exhibit 9: Regression estimate based on the BofA NLP-driven US HY Default Rate Indicator



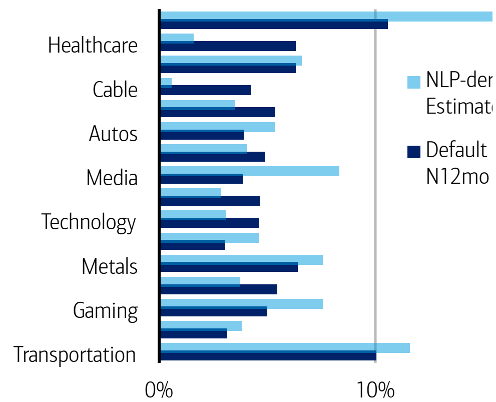
Source: BofA Global Research

Backtested period: from Jan 2003 to Dec 2017.

This performance is back-tested and does not represent the actual performance of any account or fund. Back-tested performance depicts the theoretical (not actual) performance of a particular strategy over the time period indicated. No representation is being made that any actual portfolio is likely to have achieved returns similar to those shown herein.

BofA GLOBAL RESEARCH

Exhibit 10: Sector-level estimate based on the BofA US HY Default Rate Indicator



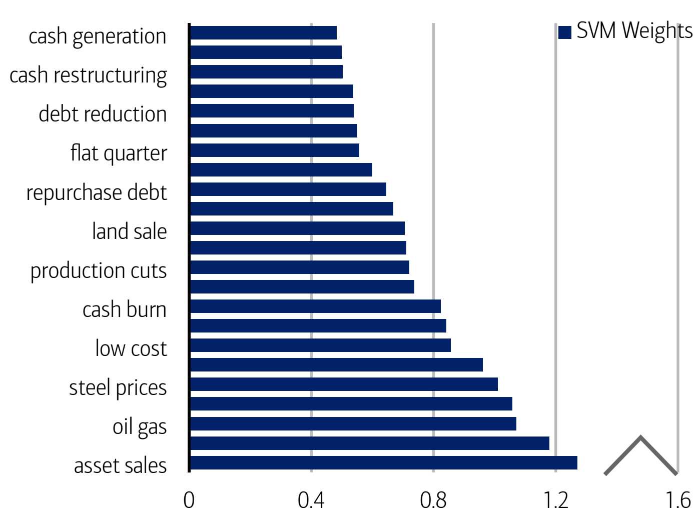
Source: BofA Global Research

Bof

While we have too many phrases to list off as part of our proprietary NLP High Yield default model, the below chart highlights some of the top phrases associated with defaults. Intuitively, when either any of the phrases like "cost cutting", "cash burn", "production cuts" make their appearance on the call, chances are the management is not discussing a particularly strong quarter; our NLP engine confirms this intuition by linking these and other terms directly to subsequent default probabilities.

Exhibit 11: Examples of key phrases from earnings calls associated with High Yield defaults

The SVM weights are the estimated coefficients as a result of training the Support Vector Machine model as described in the methodology section. The higher the weight indicate a higher importance for that phrase. We only show a sample of thousands of phrases to illustrate the mechanics of the model.



Source: BofA Global Research

BofA GLOBAL RESEARCH

We first process the text only from question and answer section while filtering out the prepared management section from the transcripts. Each earning transcript text was then cleansed by removing the stopwords (e.g. the, at, a, an etc.) and then broken into key phrases (also known as bi-grams that creates combinations of two words). An example of creating a bi-gram from a sentence "Our target is cost reduction" that we would use inside the model would be "cost reduction". For this analysis, we chose bigrams because it helped us look at the word combinations which have more semantic value when appeared together. For example, "cost reduction" provides more context than "cost" and "reduction" individually.

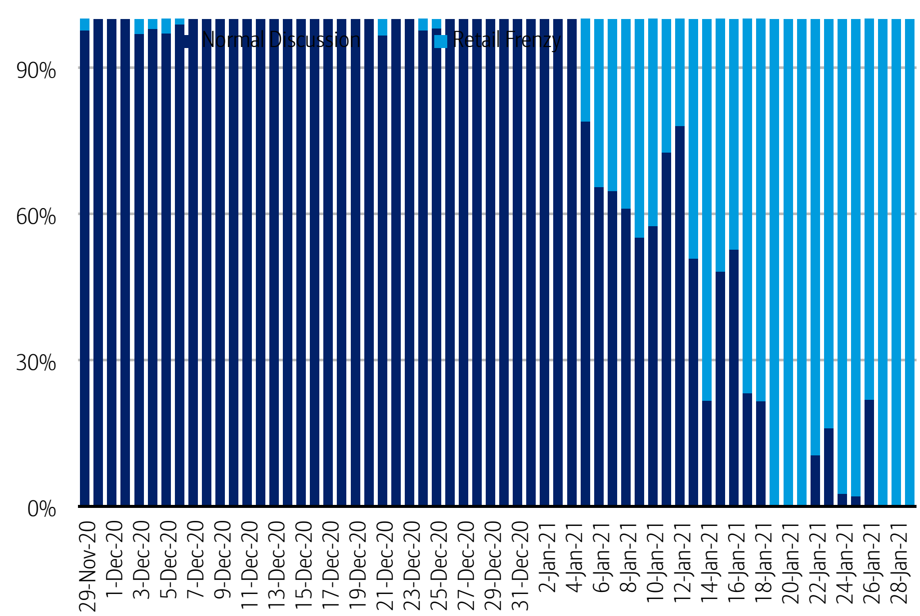
Next, to convert the document into numbers, each bigram was given a TF-IDF[(*)] score which assigns higher scores to the bigrams having more importance in the document than the bigrams which appears repeatedly. The idea behind doing this is we want the model to pick on important phrases that might provide clues for potential financial and liquidity constraints.

3 - Retail frenzy vs normal discussion

Topic Latent Dirichlet Allocation (LDA) on Reddit Wallstreetbets forums

In simple terms, Topic Modeling is an unsupervised learning method to identify the underlying topics in a document. For example, a news article or a blog could be talking about economy, asset classes and geo politics etc. Topic modeling would help to identify these underlying topics and classify other articles into such topics. Latent Dirichlet Allocation (LDA) is a popular technique to fit topic models. It considers each document as mixture of topics and each topic as mixture of words. It allows documents to overlap in terms of content and topics. Topic models could also be helpful in understanding the shift in the focus of text from one topic to other over time.

Exhibit 12: Emergence of Retail Frenzy topic in WSB posts
Running LDA model on text data from Reddit on wallstreetbets forums



Source: BofA Global Research, ListenFrist

BofA GLOBAL RESEARCH

For example, we looked at the discussion in the recently famous Wallstreetbets group on Reddit and saw a shift in the topics of discussion. Before the GameStop frenzy, the group was mostly focused on discussing stocks, their earnings and can be assumed to have one broader topic considering the nature of the group. However, around the GameStop incident a new topic emerged and in a short time took over the whole discussion. Our LDA model was able to identify 2 broad topics in the WSB posts from Dec-2020 to Jan 2021. Topic 1 is dominated by words like "company", "option", "earning", "bear" which we can be seen as normal discussion around stocks. However, Topic 2 is dominated by word like "attack", "retail", "halt", "squeeze", "ape", etc. Even though some of the words in Topic-2 are commonly used in finance but model was able to identify that usage of these words is emerging from a different underlying topic. Note that with LDA topic models they have Coherence tests to pick number of topics, however in practice having human intuition is sometimes better to ensure the topics make sense.

4 - Lazy disclosures - Avoid changers

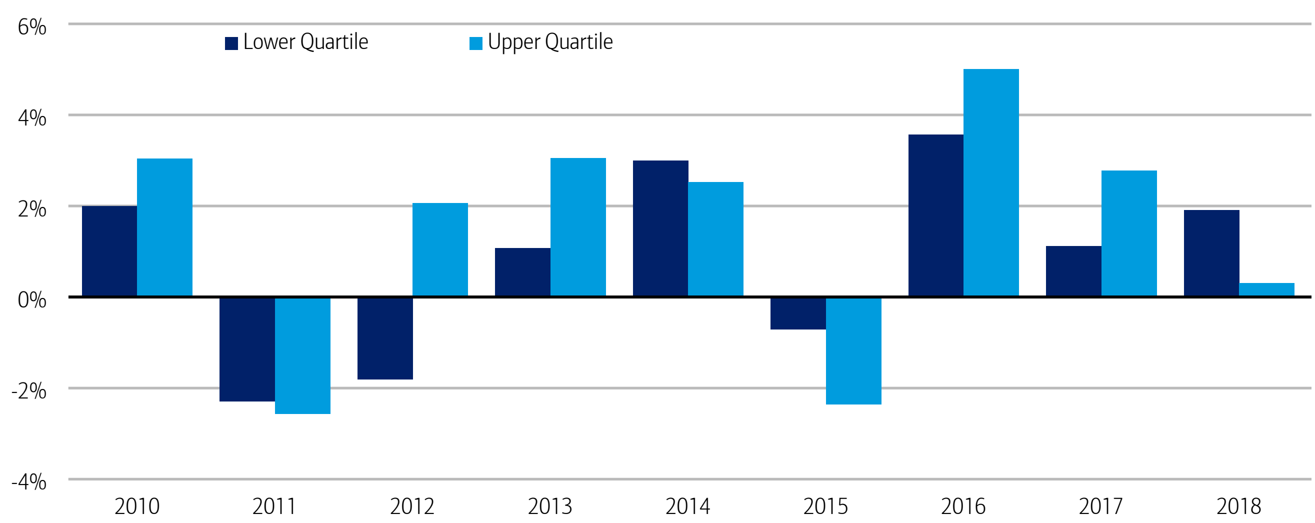
Cosine Similarities on 10-K filings

Firms that are very active in changing their SEC financial filings (i.e. 10-Q/10-Ks) have been shown to have poor subsequent performance on average[(*)]. The intuition is that firms actively changing their disclosures have items they are forced to disclose but the non-changers demonstrate a sense of stability so minimal updates are required.

One common metric for analyzing textual semantic differences is called Cosine Similarity (see appendix for full details). As an illustration, we calculate the cosine similarity of S&P 500 stocks quarter on quarter to score if any changes to their disclosures were made. We find that when bucketing companies into the lower cosine similarity quartile versus the highest cosine similarity quartile, companies with the least of amount of changes tend to outperform those that are changing more.

Exhibit 13: Companies with more changes in their 10-Qs underperform those with the least

S&P 500 average monthly return between companies in the lower quartile of cosine similarity versus those in the highest quartile of cosine sir



Source: BofA Global Research

Bof

Note that the analysis above was done on the full text from a 10-Q, where we can replicate based on key relevant sections in the 10-Q, such as the Management Discussion and Analysis section and/or the Risk sections. However, this comes with tremendous difficultly of fine-tuning the Regular expressions as at times companies will change where they put sections or not report on them at times.

5 - Libor replacement language

Using document term matrix and K-means clustering on legacy non-agency MBS

According to MBS strategists NLP Libor research (https://rsch.baml.com/r?q=0kMAnNo0zxpzPUHB2ayafQ), today there are over 6,500 pre-crisis legacy non-agency RMBS deals that include outstanding bonds indexed to LIBOR. Investors need to prepare for migration away from LIBOR by understanding their exposure, risks and hedging options. Understanding the nature of the fallback language in the legacy deals is an important step in this process.

Due to the large number of issuers and the variation of language and contract formats across various LIBOR deals, manually identifying fallback language in legacy RMBS deals is a nearly intractable effort. The use of artificial intelligence (AI) -- natural language processing (NLP) and machine learning (ML) as an augmented analytics decision-making process enabled us to significantly accelerate the task of finding the LIBOR-replacement clause for each LIBOR deal.

Our process involved the following steps:

- We identified legacy deals that included bonds based on LIBOR.
- We extracted the LIBOR replacement clauses through the corpus of the text of these documents. This task was performed by programmatically breaking the documents into sections and extracting only those sections that are relevant to
- We converted the extracted LIBOR determination sections to document-term matrix (DTM) and ran K-means clustering algorithm to group the extracted language into similar language clusters. K-means clusters the language into pre-specified number of clusters.
- For robustness, we ran the Latent Dirichlet Allocation (LDA) algorithm. Unlike Kmeans, LDA is a fuzzy clustering technique that outputs the probability that the document belongs to a given cluster.
- Our final classification method was semi-supervised learning. We first discerned the patterns and categories in the text through iteration of artificial intelligence and expert judgment. Subsequent clustering took the previous expert-based categorization judgment into account.
- Having pre-grouped the replacement language into clusters, we were able to review and inspect the actual replacement language of each bond. This process was facilitated through an augmented analytics decision-making toolset that we developed. This toolset facilitated manual audit and review over the entire dataset of LIBOR deals by providing the syntax highlighting as well as the programmatically determined categorization. This tooling allowed us to efficiently review and verify that each LIBOR deal categorization was robust, comprehensive and accurate.

6 - BofA Analyst Tone (Sentiment)

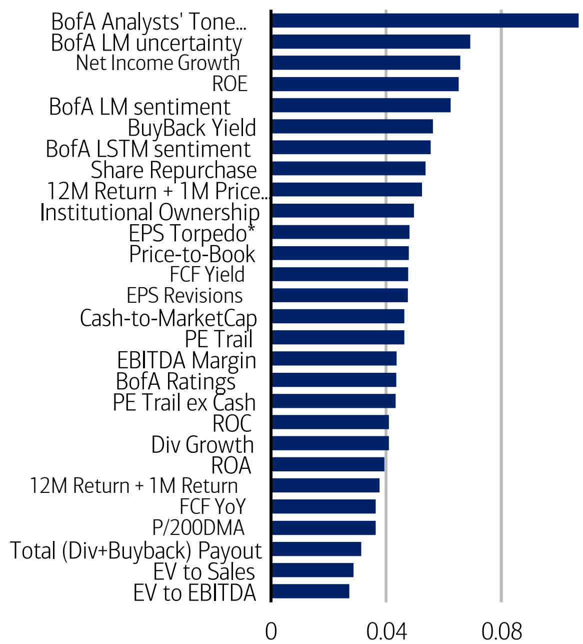
Long Short-Term Memory (LSTM) on equity fundamental reports

We applied three NLP models to BofA US stock research reports since 2010, processing over 30mn words (equivalent to 50 copies of *War and Peace*) to determine analysts' tone on stocks, industry groups and sectors (BofA Analyst Tone (https://rsch.baml.com/r?q=oWOY131Z6Y4aRNAWgXDFSw)). Our analysis suggests that tone is as predictive as fundamentals, and when aggregated to sectors, can beat any other fundamental factor we track (Exhibit 14).

The BofA Analyst Tone factor can be turned into a sector rotation strategy where it bought the five sectors with the highest BofA Analyst Tone and sold the five sectors with the lowest BofA Analyst Tone. BofA Analyst Tone is based on a blend of three NLP approaches applied to US equity analysts' report text data plus BofA ratings.

Exhibit 14: Average Sector Rank IC for best performing factors from 2010 to June 2019**

Note we take the absolute value of any correlation under the assumption one can short that factor if a negative correlation is present



Source: BofA US Equity & Quant Strategy, Factset

* Earnings Torpedo: I/B/E/S FY2 estimate less latest actual annual EPS divided by month-end price.

This performance is back-tested and does not represent the actual performance of any account or fund. Back-tested performance depicts the theoretical (not actual) performance of a particular strategy over the time period indicated. No representation is being made that any actual portfolio is likely to have achieved returns similar to those shown herein.

BofA GLOBAL RESEARCH

This strategy averages three natural language processing (NLP) approaches plus our analysts' ratings. The NLP approaches are: 1) Loughran-McDonald (LM) bag of words (i.e. adding up the positive and negative words); 2) Loughran-McDonald Uncertainty bag of words (i.e. total amount of uncertain words used); and 3) Bidirectional Long Short Term Memory (LSTM) Deep Learning approach trained to predict rating changes and price objective changes which was used to construct sentiment scores (see Machine Learning Primer for investors (https://rsch.baml.com/r?q=-gK0Jh8GNhmlSUrsuZtPtA)). The analysts' ratings are converted into numerical data that gets utilized inside the strategy.

For the LSTM model, we ran the model on documents where the report category is equal to Rating Change or Price Objective Change in order to get labeling. For Rating Change, any positive rating change is labeled as 1 and any negative rating change is labeled 0. Similarly, for Price Objective Change, any positive price objective change is labeled 1 where any negative change is 0. This is in total approximately 16k observations (about 2k were

Rating changes and 14k from Price Objective Changes) from 2010 to 2015 in sample period. Once we had an

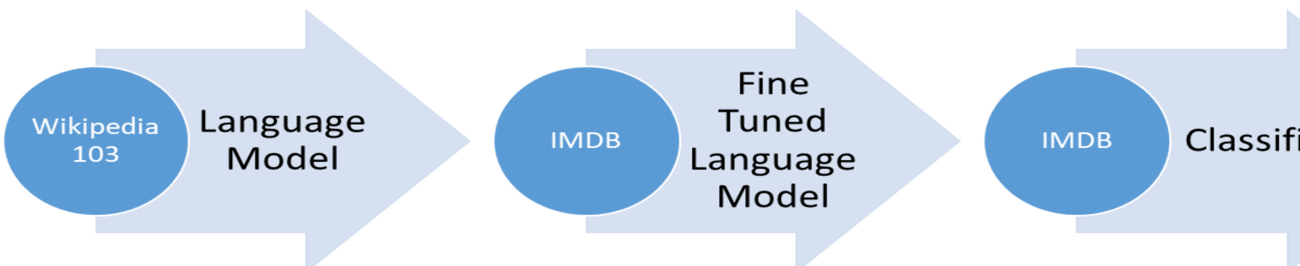
LSTM model we applied it to the rest of the research reports (~90k) to extract predicted probabilities for sentiment. We note there is a slight class imbalance of about 60 percent 1's and 40 percent 0's(*). See Refresher on LSTM in appendix for more details.

7 - Learning from Wikipedia

Universal Language Model Fine-Tuning (ULMFIT) on IMDB

Used the latest cutting edge NLP breakthroughs with techniques called Transfer Learning. The model learns initially on 103 million words from Wikipedia that gets fine-tuned for relevant text dataset. NLP recently experienced breakthroughs with meaningful increase in accuracy rates through an approach called transfer learning. A similar idea occurred with image classification a few years back when Google was able to train on over a billion images constructing the Imagenet dataset which can be used to initialize image classification models. In the NLP space, we focus on a transfer learning model called Universal Language Model Fine-tuning for Text Classification (ULMFiT) from Fast.AI which has achieved significant improvements on most the standard NLP datasets.

Exhibit 15: How Transfer Learning works



Source: BofA Global Research, Universal Language Model Fine-tuning for Text Classification by Ruder, Howard

Bof

There are three phases when estimating an ULMFiT model. They are:

1. Construct a language model using a general language text data set. They use 28,595 preprocessed Wikipedia articles and 103 million words on to build a language model. A language model is an NLP model that learns to predict the next word in a sentence (Think of when you are typing something into Google). The advantage of language models is that they have an abundance of labels as it focuses on predicting the next word in a sentence.

2. A second language model is estimated that is fine-tuned using 1) above to initialize but makes the language for a custom dataset. In this case, the dataset is the IMDB dataset for illustrative purposes but can quite easily be extended to financial corpus text datasets.

3. The fine-tuned language model is eventually fed into a sentiment classification model. The central idea is that it takes more than just a few thousand observations to learn the English language and hence we can incorporate nuanced information that both language models are able to learn.

For 1) above, this does not have to be estimated but one can pre-downloaded Wikipedia weights to initialize the fine-tuned language model. Please see paper by Howard and Ruder called "Universal Language Model Fine-tuning for Text Classification" for further details.

8 - Listening to Central Banks (sentiment)

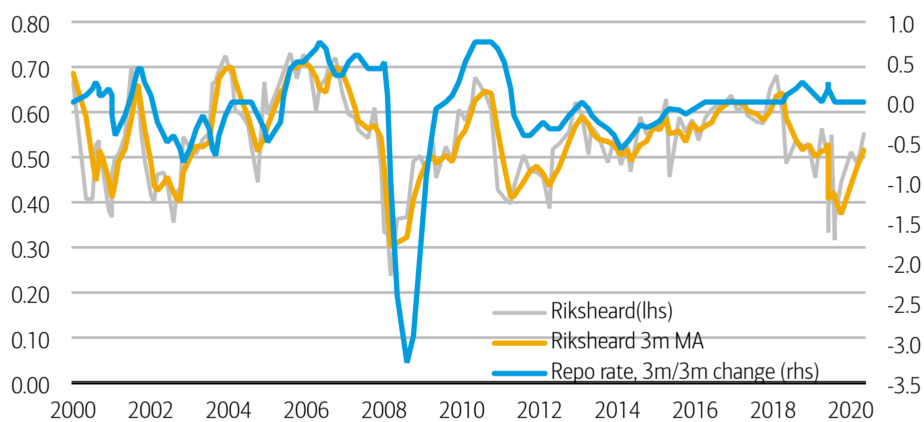
Quantifying the degree of 'hawkishness'

Working with our economists and FX strategists, we developed NLP models to quantify the degree of hawkishness in central banks' statements - BoEMI (https://rsch.baml.com/r?q=0WNoLtobcdynTqBfEVmSTw) (Bank of England Mood Indicator), Riksheard (https://rsch.baml.com/r?q=qV7zIYis!t9sRlcRUpt3Ow), NORBI (https://rsch.baml.com/r?q=rZ7VDpqfKHzGOJHftCcKKg) and EMMI (https://rsch.baml.com/r?q=hoV-E6iMvURreVtQo1vX2g) (Emerging Markets Mood Indicator). In addition, our US economics team decoded the Beige Book (https://rsch.baml.com/r?q=9ZAov5U8KIuTPIMHqZE6iA), produced by the Federal Reserve, which gathers anecdotal information on current economic conditions from its regional Federal Reserve Banks through business contacts, experts and other relevant sources, in order to better harness the information content from the Beige Book.

The 'mood' indicators have provided a leading signal for changes in central banks' policies. In order to compile the dictionary of hawkish/dovish words & phrases, we started with the list from a widely cited ECB paper(*). We added terms that we believe represent the specific way central banks communicate their policy views. Resulting indicators have proven to have significant leading relationship with the changes in central banks' policies.

Exhibit 16: BofA proprietary Riksbank mood indicator

Riksbank shifting neutral



Note: The indicator identified as the BofA Riksbank mood indicator is intended to be an indicative metric only and may not be used for reference purposes or as a measure of performance for any financial instrument or contract, or otherwise relied upon by third parties for any other purpose without the prior written consent of BofA Global Research. This indicator was not created to act as a benchmark.

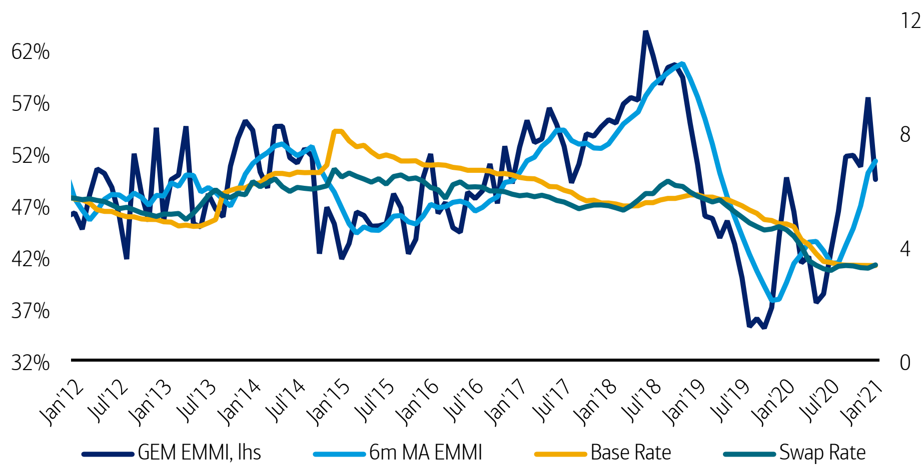
Source: BofA Global Research

BofA GLOBAL RESEARCH

To account for the context, we created a list of contravening words whose presence in a clause flips it into the opposite category. For example, "increase interest rates" vs. "increase quantitative easing". The presence of "quantitative easing" flips the original hawkish nature of the word "increase" into the dovish category. We calculate a hawkishness score between 0 and 1, representing the fraction of clauses that our algorithm determines are 'hawkish'. We can represent this with the following formula:

Exhibit 17: Our Emerging Monetary Mood Indicator measures the hawkishness of eleven central banks

Central banks are becoming more hawkish lately, similar to the 2017 level

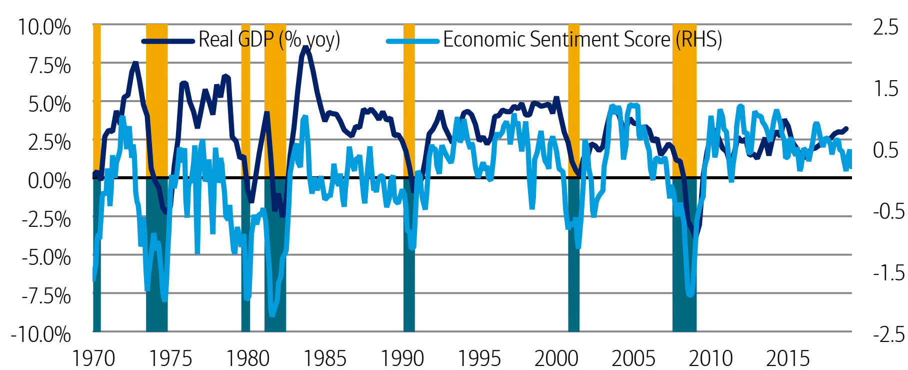


Note: The indicator identified as the BofA Emerging Monetary Mood Indicator is intended to be an indicative metric only and may not be used for reference purposes or as a measure of performance for any financial instrument or contract, or otherwise relied upon by third parties for any other purpose without the prior written consent of BofA Global Research. This indicator was not created to act as a benchmark. **Source:** Bloomberg, Haver, BofA Global Research

BofA GLOBAL RESEARCH

While the economic sentiment score has limited leading features, we see value in monitoring the series. Fed officials consult the Beige Book to look for incremental information on the economy that may not be captured in the official data. Recently, the report has garnered increased attention because it shows heightened concerns around trade policy from business contacts. Indeed, the economic sentiment score turned down indicating that sentiment on the economy softened this year as it fell to 0.18 in today's July Beige Book release compared to 0.61 in last year's July report. A simple word count of trade related terms shows that the weakening in sentiment likely owes heightened concerns around trade negotiations and higher tariffs on imported goods.

Exhibit 18: The economic sentiment score based on the Beige Book generally tracks GDP growth



Note: The indicator identified as the economic sentiment score is intended to be an indicative metric only and may not be used for reference purposes or as a measure of performance for any financial instrument or contract, or otherwise relied upon by third parties for any other purpose without the prior written consent of BofA Global Research. This indicator was not created to act as a benchmark. The economic sentiment score is new and the back-tested performance presented is hypothetical in nature and reflects application of the economic sentiment score prior to its inception date as if the model had been in existence at that time. It is not intended to be an indicative of actual or future performance. The actual performance of the economic sentiment score may vary significantly from the back-tested performance. **Source:** BofA Global Research, BEA, Federal Reserve

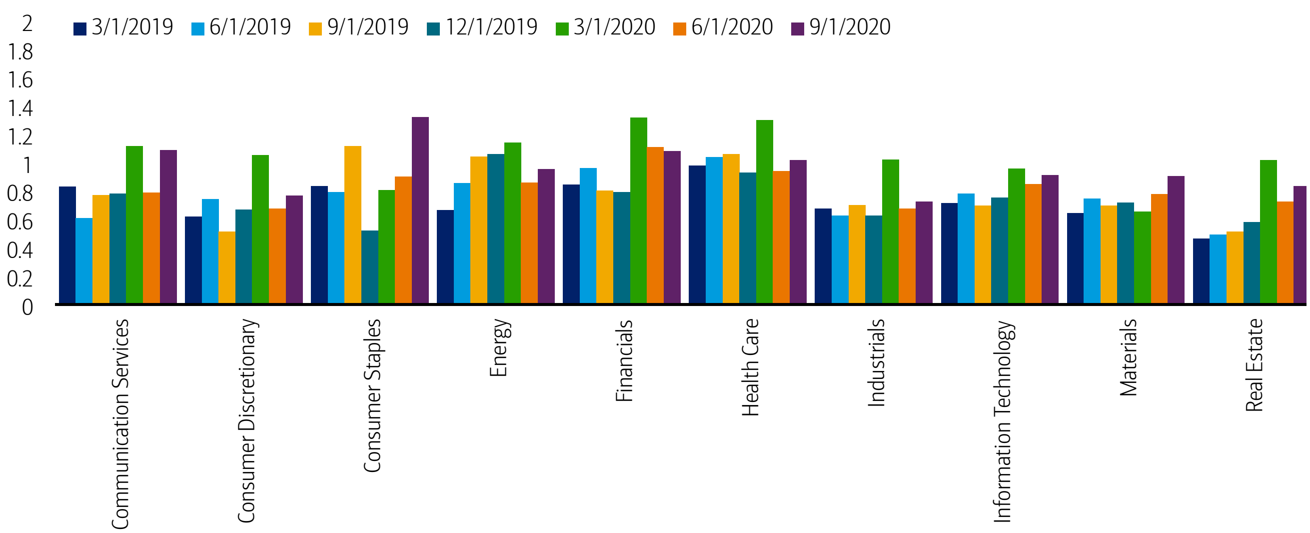
BofA GLOBAL RESEARCH

9 - Management discussions on future relative to past

Dependency parsing on earnings transcripts

As an illustrative example, we extract all language that is considered 'future' tense versus 'past' tense on earnings transcripts as CEO/CFOs speaking more about the future as opposed to the past is a sign of stability. We utilize a traditional NLP technique called Dependency Parsing, which is based on the fact that each sentence is about something and usually contains a subject (the doer), a verb (what is being done) and an object (to whom something is being done). We show an example below of the S&P 500 sectors below:

Exhibit 19: During the initial COVID-19 outbreak saw a spike across most sectors speaking about the future relative to the past
Percentage of words in future tense divided by the percentage of words in the past tense



Source: BofA Global Research

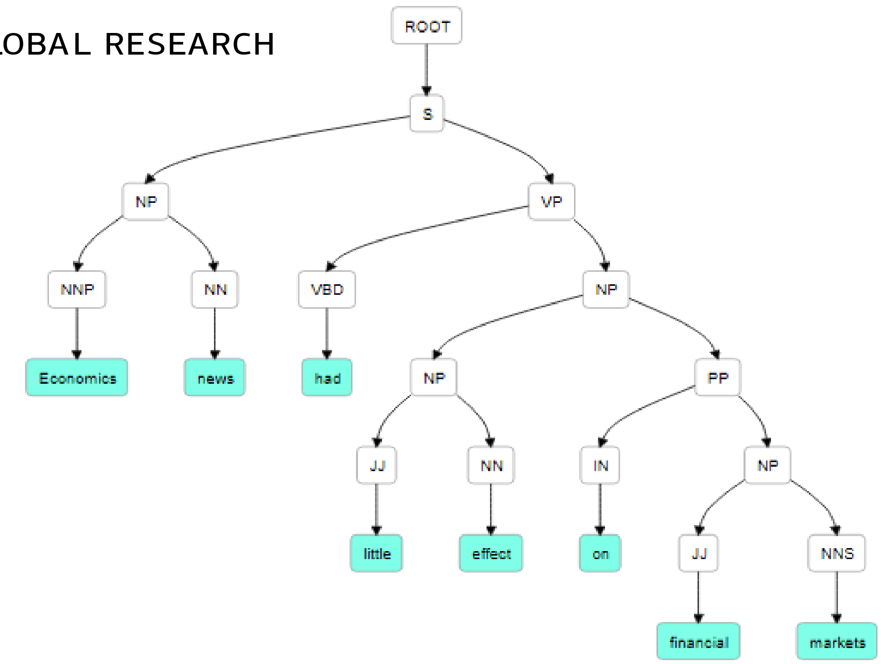
Bof

The Subject-Verb-Object (SVO) is basic word order in present day. In dependency parsing we start with the root of the sentence which is often a verb. The root word is also called as the head of the sentence since it describes what the sentence is about. All other dependent on the root word. Let's consider an example sentence.

Sentence:

"Economics news had little effect on financial markets"

Exhibit 20: Dependency Tree Example



Source: BofA Global Research

In order to understand the Tense in which CFOs or CEOs speak during the earning calls, we applied rules created by combining Dependency Parsing and POS Tag. We considered earning call transcripts of SP500 companies from 2003 to till date for our analysis. We created rules to classify sentence within transcript as Past or Future.

Rules we used:

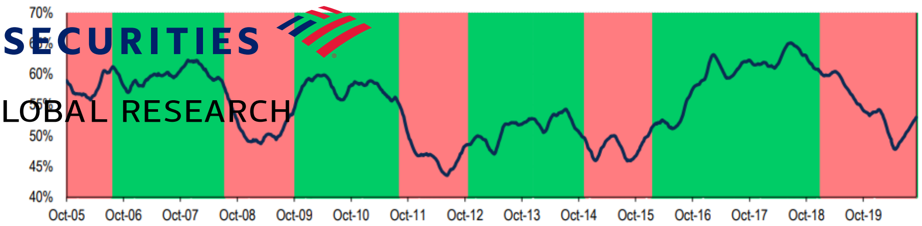
- Past tense can be easily identified using root Verb.
- Past tense can also be identified by checking the a tag of word and its dependent
- For Future tense by checking root is not Past Tense Verb and combining word like will, shall etc.

10 - Petrochemical sentiment

Word2Vec enriched lexicons on market newsletters

Recently, traditional financial dictionaries have begun to lose their relevance and efficiency in capturing the true sentiment in financial text. Especially when the management speakers are prepped to avoid using the words defined as negative in these dictionaries. Secondly, generic dictionaries are not efficient and often fail to capture the sentiment of sector specific text. For example, words such as over-inventory, overproduction and overbuilding are considered negative in chemical industries to imply the imbalance between demand and supply; such words are not part of traditional dictionaries and to calculate sentiment in such cases often requires dictionary building from scratch.

Exhibit 21: : 12MMA of the Petchem Sentiment Indicator: Green indicates time between bullish & bearish signals, red is time between bearish & bullish signals



Source: S&P Global Platts, BofA Global Research

BofA GLOBAL RESEARCH

Models like word2vec are useful to expand or customize dictionaries to find the negative and positive words which are used in similar context. We leveraged our trained word2vec model to build a customized dictionary starting with very few positive and negative words and then iteratively find similar words from text specific to the sector. The customized dictionary was then used to create Petrochemical Sentiment Indicator (https://research1.ml.com/Archive/12181148.pdf?q=4!2YwSjSVlIP9qkhjnQtpg&__gda__=1614777197_bdaba2acf5b14fbeef5582893217a933) using text from S&P Global Platts' Polymerscan reports, a leading publication on global plastic and resins including polyethylene, polypropylene, and polyvinyl chloride (PVC). It provides timely updates on broad market trends, which are often difficult to track, and could help explain stock price movements. The sentiment indicator built using custom dictionary has recently risen sharply since turning bullish in late September, coinciding with a strong rise in global petrochemical stocks prices.

Appendix

Key NLP terms

- Tokenization: Given a chunk of text/a wave of sound, break it into distinct tokens such as phonemes, words, characters, symbols, sentences, or other semantic units.
- Lemmatization: Group together all like terms that share a same lemma such that they are classified as a single item. For example, the lemma for 'walk', 'walked', 'walks', and 'walking' is 'walk'.
- Stemming: Reduce an inflected or derived word into its word stem, base, or root form by cutting off the prefixes and/or suffixes of words.
- Part of Speech (POS): Given a sentence, determine the part of speech for each word. Many words, especially common ones, can serve as multiple parts of speech. For example, "book" can be a noun ("the book on the table") or verb ("to book a flight"); "set" can be a noun, verb or adjective; and "out" can be any of at least five different parts of speech.
- Chunking (also called shallow parsing): Link constituent part-of-speech tagged tokens to higher order phrases/groups that have discrete grammatical meanings. For example, a noun phrase may comprise an adjective sequence followed by a noun.
- Bag-of-words (or Bag-of-n-grams): Describe the occurrence of each word (or n-grams) within a document. The order or semantic structure of the words is ignored. The challenge of bag-of-words is

sparse vector issue, which can be partially reduced by text preprocessing.

- Term Frequency - Inverse Document Frequency (TF-IDF): A statistical measure to measure the importance of a word to a document in a corpus. The TF-IDF penalizes words that are frequent across all the documents since these frequent words may not contain meaningful information.

- Regular expression: A sequence of characters that define a search pattern. Regular expression plays an important role in text preprocessing and feature extraction, relying on developers' business and linguistic knowledge.

Below are POS tag that are commonly used.

- ADJ: adjective
- ADP: adposition
- ADV: adverb
- AUX: auxiliary
- CCONJ: coordinating conjunction
- DET: determiner
- INTJ: interjection
- NOUN: noun
- NUM: numeral
- PART: particle
- PRON: pronoun
- PROPN: proper noun
- PUNCT: punctuation
- SCONJ: subordinating conjunction
- SYM: symbol
- VERB: verb
- X: other

POS Tagging

Parts of speech (POS) are useful information to sentence structure and meaning. A word in a sentence be tagged as noun, verb, adjective, adverb etc depending upon the role it plays in a sentence. Assigning correct POS tags helps us to better understand the intended meaning of the sentence or a phrase. Word classes do have semantic tendencies-adjectives, for example, often describe properties and nouns people- parts of speech are defined instead based on their grammatical relationship with neighboring words or the morphological properties about their affixes.

Consider an example from an Earning Call Transcript:

SENTENCE:

"As Mike said, we're also doubling down on areas such as the new product"

Exhibit 22: Parts of Speech (POS) Tagging example

as *IN* Mike *NP* said *VVD* we *PP* 're *VBP* also *RB* doubling *VVG* down *QQ* on *IN* areas *NNS* such *JJ* as *I*
new *JJ* products *NNS*

Source: BofA Global Research

Bof

There are different approached to do POS tagging to a given sentence or phrase like

- **Lexicon - based:** This approach uses simple statistical algorithm where for each word , it assigns the POS tag that most frequently occurs for that word in a training corpus
- **Rule - based**: This approach uses grammatical rules to assign POS tag. Like all words ending with "ing" are verb
- **Probabilistic technique:** This technique assign POS tags to a word by considering the preceding POS tags. Its assign probability values to the concern word and the previous words and select the combination of with highest Probability. This Method works on the Hidden Markov Models(HMM) for assigning the tags

POS tagging which is a shallow parsing method to understand the structure of sentence is not able to check the grammatical structure of sentence or understand dependencies between words in sentence. To solve the challenge we require deeper parsing techniques like **Dependency Parsing.** POS tags helps us to understand linguistic role of the word in a sentence, it wouldn't enable us to understand how these words are related to each other in a sentence.

What is a Cosine Similarity?

Cosine Similarity is a useful metric with text analytics in order to estimate the similarity between two documents or passages of text (i.e. in the case of comparing questions and answers in earnings transcripts or are SEC filings from different periods). In mathematical terms, it measures the angle between two vectors in a multi-dimensional space. The smaller the angle the more similar two passages of text are.

What is a Support Vector Machine?

Support Vector Machines (SVM) are supervised learning models used for classification and regression analysis. Given a set of training data (observations for which the categorization is known in advance), an SVM builds a model that assigns a new observations to one of the pre-established categories. In our case for predicting defaults, we are using the classification set up.

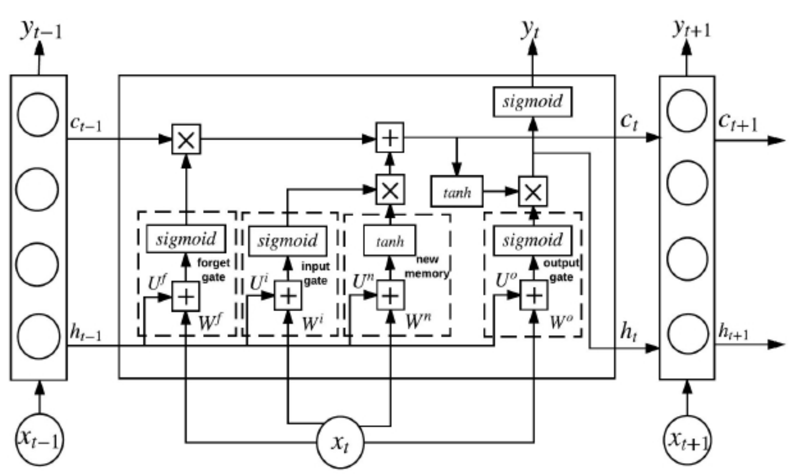
Given a particular set of training data, the algorithm tries to place a line or plane through the data in such a way that data points belonging to one category are on one side of the plane, while data points belonging to the other category are on the other side of the plane. More specifically, this classifier is a so-called hyperplane, i.e. a subspace whose dimension is one less than that of the space it is in. In 2-dimensional space (think of a standard graph with horizontal and vertical axes), the hyperplane is simply a line. In 3-dimensional space, the hyperplane would be a 2-dimensional plane, and so on. Once this hyperplane is established, classification of new data points is given simply by the side of the hyperplane that the data point falls on. Please refer to our Machine Learning primer (https://rsch.baml.com/r?q=-gK0Jh8GNhngF92vUs1BXw) and FX Machine Learning primer for extra details on this model.

Refresher on LSTM

The Long Short Term Memory (LSTM) Deep Learning model is subset of a Recurrent Neural Network (RNN) that is capable of learning long-term dependencies (i.e. word ordering). An LSTM is a more complicated than a traditional Neural Network as it contains four neural layers that collaborate with one another with having two states (hidden and cell).

Exhibit 23 highlights the inner workings of a network with four gates that control the cascade of information being passed around. We defer the reader to view the Deep Learning for Sentiment Analysis paper by Zhang, Wang and Liu for a more detailed description as we leave out for brevity.

Exhibit 23: Long Short Term Memory Network



Source: Deep Learning for Sentiment Analysis paper by Zhang, Wang and Liu

BofA GLOBAL RESEARCH

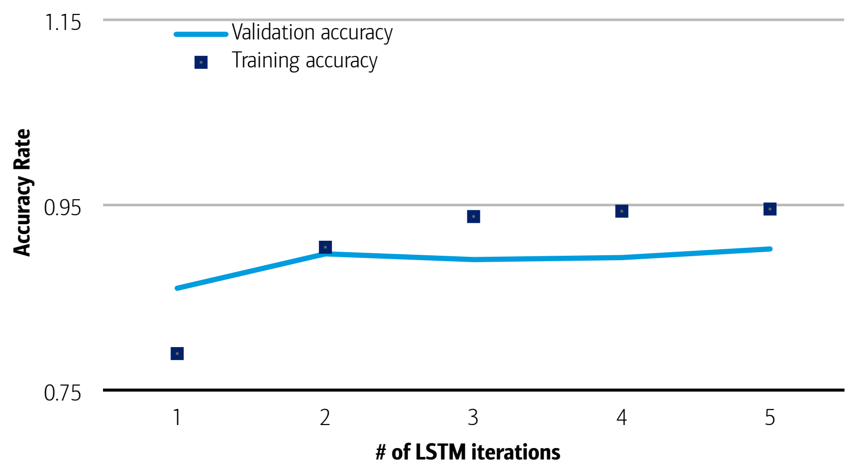
The BofA Analyst Tone product utilizes an LSTM model as one of the key NLP models. Exhibit 24 shows the accuracy per each epoch (i. e. each iteration the model is run in order to find the optimal parameters). There is a balancing act between training and validation accuracy rates and after about 5 epochs the LSTM model finds the best fit. Given that our validation and training accuracies are quite close together this suggests that there is no overfitting present for the model and it will generalize well on new data it has not seen (see Machine Learning Primer for investors (https://rsch.baml.com/r?q=-gK0Jh8GNhmISUrsuZtPtA) for more background on best practices to prevent overfitting).

The hyperparameters of the model include a dropout rate of 10% (controls for overfitting), 500 words for padding (looks at first 500 words in document), Adam for optimization (adaptive learning algorithm), 32 neural node units and Bidirectional (this reverses the order of the sequence to get information for both past and the future). See Refresher on LSTM section for extra details

We note we achieve validation accuracy rates of 90% on our validation dataset (20% of the dataset not included in training the model). In essence, our model is 90% accurate at predicted the tonal sentiment of a BofA research US equity fundamental reports. For more details on LSTM models please see Appendix Refresher on LSTM section.

Exhibit 24: Deep Learning LSTM Model accuracy results

Shows the training of 5 epochs (i. e. each iteration the model is run in order to find the optimal parameters) as a check to make sure the validation and training accuracy broadly match in order to ensure no overfitting has occurred.



Source: BofA US Equity & Quant Strategy

BofA GLOBAL RESEARCH

>> Employed by a non-US affiliate of BofAS and is not registered/qualified as a research analyst under the FINRA rules.
Refer to "Other Important Disclosures" for information on certain BofA Securities entities that take responsibility for the information herein in particular jurisdictions.

Click for important disclosures.

Disclosures

BofA SECURITIES

BofA GLOBAL RESEARCH

Trending

VideoPDF

[The RIC Report (https://rsch.baml.com/r?q=hQshcabjN2YTI1z8yB99Pw&e=mihail_turlakov%40sberbank-cib.ru&h=hPoi!Q)](https://rsch.baml.com/r?q=hQshcabjN2YTI1z8yB99Pw&e=mihail_turlakov%40sberbank-cib.ru&h=hPoi!Q)

The Fiscal Liquidity Trap

Research Investment Committee 2021-Mar-8

Timestamp: 09 March 2021 05:00AM EST

prd - apac - node1 - r7