Narrative Modeling: A Key Component of Narrative Economics

Ali Ozgur Cetinok, Advisor: Dr. Rashed Iqbal
UCLA, Department of Economics
Los Angeles, California, United States

Abstract

Economics is a social science concerned with the description of production, distribution and consumption of goods and services. There are numerous theories, principles and practices in Economics. In addition to the quantitative methods in Economics Theory, it is essential to measure the quantitative impact of the narratives on Economic variables. Narrative Modeling is a significant way to support the foundations of Classical Economics. Individuals are effected by narratives surrounding them, and their decisions are effected by those narratives. Narrative Modeling aims to model how narratives spread and effect individuals' decisions, which has a huge impact on Economic variables. A Narrative Model based on Wikipedia is implemented to classify narratives according to their topics and emotions.

Keywords: Economics, Narrative Modeling, Natural Language Understanding, Data Science, Wikipedia, Semantic Analysis

---

June 21, 2018

Contents

1. Introduction

Narrative Economics is the study of the spread and dynamics of popular narratives, the stories, particularly those of human interest and emotion, and how these change through time to understand economic fluctuations.[1] According to Professor Robert Shiller, Nobel Laurate Economist, narratives behave like living organisms. They appear, spread like a virus and eventually die. Narratives carry ideas and emotions. From these ideas and emotions, individuals' decisions are effected. Their actions follow the narratives that are popular at that time. For instance, if there is a prominent narrative about a possible financial crisis, people will act accordingly, no matter it is factual or not. Narrative Modeling aims to understand the effects of the narratives on human beings to come up with the implications on economic variables. To model the effect of narratives on economic variables, an approach to model narratives should be developed.

Narrative Modeling can be divided into three parts:

- Topic Modeling
- Emotion Modeling
- Call to Action Modeling

After modeling narratives, we can analyze the popularity of each narrative across the world at a specific time. For example, if there is a fear about Stock Markets, we can analyze the sentiment in the market and update our forecasts accordingly. Individuals are not rational, and this irrationality can be captured by evaluating narratives. Combination of Classical Economics Theories and Narrative Economics foundations could lead to more accurate forecasts and analysis about economy.

2. Literature Review in Natural Language Processing

Narrative Modeling, which combines Topic, Emotion and Call to Action Models can be developed by interpreting the principles of Natural Language Processing and Understanding.

2.1. Topic Modeling

To understand how narratives spread and which narratives are prominent, a topic model is essential. Human beings are not able to process large amount of data like computers, therefore identifying popular narratives and trends in Economics is not possible for a human being. With the help of the computational power of computers today, we are able to analyze the text using different algorithms. "Natural Language Processing" (NLP) models are used in Topic Modeling to identify the content of the narrative. NLP focuses on topic modeling with various different algorithms. Here is a summary of methods that have been popular in NLP[2]:

Name	**Methods**
Latent Semantic Analysis	Gets the most frequent words
Probabilistic Latent Semantic Analysis	Generates a word for a topic
Latent Dirichlet Allocation	Probabilistic Models
Non-Negative Matrix Factorization	Liner Algebra
Correlated Topic Model	Logistic Normal Distribution
Natural Language Toolkit	Grammar Rules

Table 1: Methods used in NLP

Natural Language Toolkit (NLTK) is a library in Python that contains the grammar rules of languages and vocabulary information. However, there is not a topic model widely used in NLTK library. I will summarize the most popular algorithms that are used in topic generation.

2.1.1. Latent Semantic Analysis

Latent Semantic Analysis (LSA) is a statistical approach in Natural Language Processing. Counting the appearance of words across various documents and finding the most frequent word is possible by creating a Term-Frequency Matrix. In addition, TF-IDF, which is Term Frequency i-Inverse Document Frequency Matrix, is a numerical statistic that is intended to reflect how important a word is to a document in a collection of documents.

These widely used two methods developed a ground for future work in text mining.

2.1.2. Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a statistical model that allows to discover the topics of documents and sentences.

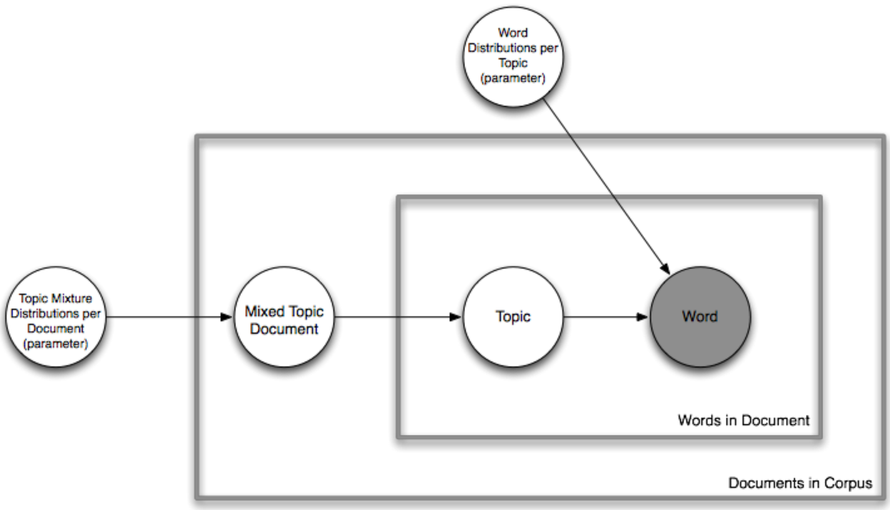

Figure 1: LDA Model

Above, there is a simple representation of LDA Model. In LDA, a large matrix of words and documents is created. Then, a different matrix which contains documents and topics are constructed. By the word distribution in each document, topics are defined using a probabilistic model.

The major drawbacks of LDA are:

1. A large corpus of documents are needed
2. Number of available topics have to be defined beforehand
3. The order and the meaning of the words do not matter

Therefore, in a Narrative Model, LDA is not sufficient to gather insights from the articles.

2.1.3. Non-Negative Matrix Factorization

Non-Negative Matrix Factorization(NMF) is an algorithm based on Liner Algebra to analyze the topic from text.

W is a generated feature matrix and H is a coefficient matrix. With the multiplication of these matrices, V, which is a word-document matrix is created. We can consider each original document in our example as being built from a small set of hidden features. NMF generates these features.

NMF and LDA have same measures to predict the topic of the narrative, therefore their drawbacks are identical. Number of features have to be defined

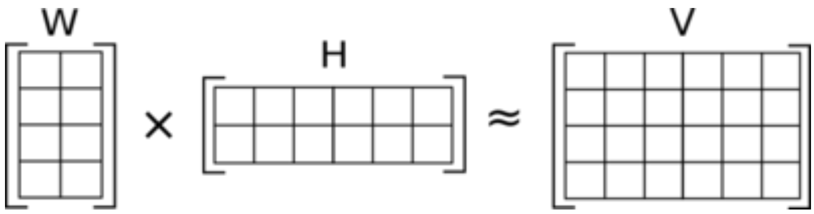

Figure 2: NMF Model

and large number of documents are needed for NMF as well.

2.2. Emotion Modeling

Emotion Modeling aims to gather the overall sentiment and emotions of a narrative. Numerous research is done in Emotion Modeling Field, but the progress is not significant. Natural Language Understanding (NLU) algorithms are present to figure out the sentiments of the text. IBM Watson NLU, Google NLU and Amazon Comprehend are the most popular emotion modeling services that can be found online, however their algorithms can not model complex sentences, which contain ambiguous words and mixed emotions. There are mainly 4 different text-based emotion models[3]:

- Keyword Spotting: Text documents are tokenized and emotion keywords are detected. It is one of the simplest methods to implement. The keyword spotting method is generally adopted sentence by sentence, they analyze the article by summing up the emotions in sentences.

- Lexical Affinity: Words are assigned probabilities with respect to the emotion they may contain. By this way, if there is no other word having the similar probability in a sentence, wrong emotion detection is prevented. Words have couple of meanings, so this method is an extension to keyword spotting, including linguistics.

- Learning Based Approaches: They are mostly supervised learning models, which has lots of text input with emotion labels, and Support Vector Machines,Artificial Neural Networks and Random Forest are used to train the data.

- Hybrid Based Approaches: Hybrid Models combine the 3 different methods mentioned above, to aim higher accuracy.

As mentioned, these widely used methods in emotion recognition are not successful at analyzing sentences without specific keywords.

2.3. Call to Action Modeling

To figure out the economic fluctuations, the aim of the narrative needs to be determined. For example, if an article is about the situation of the exchange rate Dollar/Euro, what is the call to action of it? Will Dollar depreciate or appreciate? With Call to Action modeling, we can identify the ideas that motivate us to act. Call to Action modeling is the most challenging area in narrative modeling. In the literature research, no significant progress is found for Call to Action Modeling.

3. Narrative Model

Implementing a Narrative Model involves three steps as mentioned : Topic Modeling, Emotion Modeling and Call to Action Modeling. Before implementing the model, the text data has to be preprocessed in order to gather the words in the document correctly.

Python programming language is used for model implementation.

3.1. Preprocessing Text Data

For preprocessing the raw data, "regexp" library is used. Unnecessary punctuations and HTML codes are removed from the text data. Then, stop words are removed using the Natural Language Toolkit library.

For the model, words, capital lettered words and 2-grams are gathered from text. After collecting each word, capital letters are removed to prevent double counting of a word.

3.2. Tokenization of Words

To analyze the document, every word and 2-gram are tokenized and represented in number format. Their occurrences are counted in the article and dictionaries that are holding the information are generated.

3.3. Topic Model

Majority of the topic models in sentiment analysis only focus on the words in the document and do not consider the meaning of the words. To prevent this and build a more human-like machine that can extract topic from an article, there are two prerequisites:

- Past Information
- Common Sense

Adequate representation of natural language semantics requires access to vast amounts of common sense and domain-specific world knowledge, as stated above.[4] At the moment, there is no possible way to satisfy the common sense requirement. One of these three prerequisites can be satisfied to get close to a human-like machine right now. The way to extract past information is to find a source that contains every single information, such as an encyclopedia. As we all know, Wikipedia is an online encyclopedia updated by the community and contains a lot of information. Therefore, Wikipedia is used for an information source in the Narrative Model.

In Wikipedia, each page belongs to one or more categories. These categories represent the page's meaning and helps the algorithm to cluster numerous words in a single category. In addition to this, ambiguous words can be classified by its' possible categories. Extracting the category of each word and finding the most frequent category in a document will give us an insight about the article.

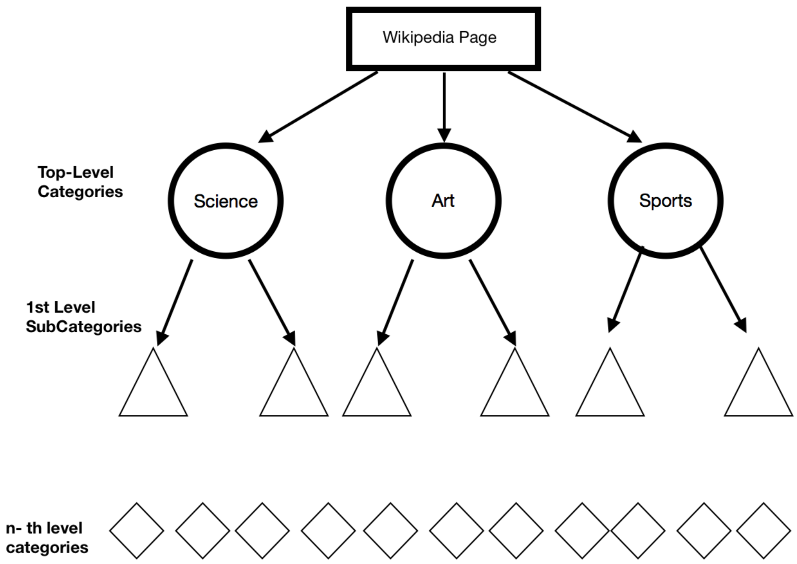

Figure 3: Wikipedia Category Architecture

In addition to categories, Wikipedia pages have anchor texts, which links to the Wikipedia page of anchor texts' article. By gathering these anchor texts, we can analyze the relationship between words. A huge word interaction matrix is created to analyze the most frequent meaning in the text, not the synonyms.

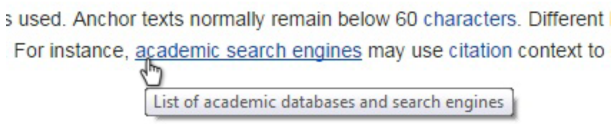

Figure 4: Wikipedia Anchor Texts

Here is the summary of Wikipedia implementation in the topic model:

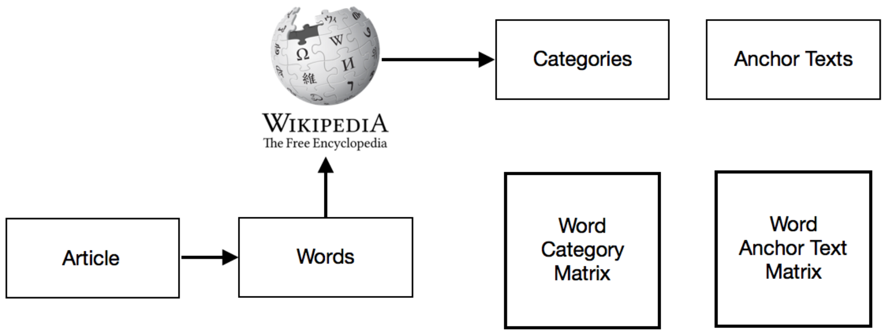

Figure 5: Wikipedia Implementation in Topic Model

After creating two interaction matrices, most frequent categories and anchor texts can be found. By having most important categories, anchor texts and most frequent words, we can analyze the importance of each sentence in the article. To achieve this, each sentence is tokenized in the article and given weights, according to the frequent words, categories and anchor texts it contains. Then, most important sentences can be provided as an output.

The output of the topic model is as follows:

- General Categories
- Related Words (Anchor Texts)

- Most Frequent Grams
- Most Important Sentences

Below, there is a sample output, which analyzes the New York Times article "How the Ice Age Shaped New York"[5]

Topic Model

General Categories: ['fluvial landforms', 'slope landforms', 'boroughs of new york city', 'glaciology', 'coastal and oceanic landforms', 'geomorphology']

Related Words: ['united states', 'earth', 'water', 'latin', 'glacier', 'new york city']

Most Frequent Grams: ['mr. horenstein', 'new york', 'horenstein said', 'york city', 'ice', 'ridge']

Most Important Sentences: ['At the start of the last ice age, 2.6 million years ago, a sheet of frozen water formed atop North America that kept expanding and thickening until it reached a maximum depth of roughly two miles.', 'The intermittent ridge runs from Puget Sound to the Missouri River to Montauk Point on Long Island, forming the prominence that supports its old lighthouse.', 'The ice over Manhattan would have buried even the tallest skyscraper and was so heavy that it depressed the underlying bedrock.']

Figure 6: Topic Model Output

General Categories, which are gathered from Wikipedia, are successful at finding the topic of the article. Glaciology, which is the scientific study of ice, is not present in the article, however the topic model was able to retrieve that information using the past knowledge, from an encyclopedia.

3.4. Emotion Model

To build an Emotion Model, it is significant to understand how we define emotions. How are they present in our written texts? In which ways we try to express our emotions in our languages?

For this analysis, it is essential to figure out the method children were taught to express their emotions. Without knowing how to express themselves, babies cry. Then, when they start talking, they try to represent their sadness or happiness with words. If the sentence "Someone stole your favorite toy." is said to a child, what would that child think? He takes the verb and the subject of the sentence, "Toy" and "Steal". Other words are unable to contain emotions. To have the same understanding in a machine, most important sentences of articles should be parsed grammatically. After parsing, subject, verb and adverbs need to be identified to figure out the overall emotion of the text.

The overall emotion gathered from the sentence can be compared with the category and keyword emotions to validate the accuracy of emotion recognition.

To parse the important sentences, I used Spacy, a Python library for advanced Natural Language Processing. By using Spacy, verbs, adverbs and nouns can be classified. Here is the verbs in the example provided above:

```
formed VERB VBN acl sheet

kept VERB VBD relcl sheet

expanding VERB VBG xcomp kept

reached VERB VBD advcl kept
```

Figure 7: Spacy Parser Verbs

After gathering verbs, adverbs and subject nouns that are present in important sentences, there are two possible ways to extract emotions: Training a dataset, Using a dataset which has emotion parameters for words.

To train a dataset, Emotion labeled articles are needed, therefore that is a possible future work in Narrative Modeling to increase the accuracy. On the other hand, National Research Council of Canada had already developed a Word-Emotion Association Lexicon that contains words with the eight primary emotions, as well as with two sentiments, which are positive and negative.[6] The dataset is used for Emotion Model. Each verb's sentiments are analyzed first to determine if the article has a positive or a negative sentiment, which is used also in Call to Action Model. The table below shows which emotion belongs to which sentiment in the analysis.

Positive	**Neutral**	**Negative**
Joy	Anticipation	Anger
Trust	Surprise	Fear
		Disgust
		Sadness

Table 2: Sentiments and Emotions

After getting the sentiment, emotions are predicted from the weights in the dataset. In the example given above, since verbs do not have strong emotional weights, the article is considered neutral with no significant emotion.

3.5. Call to Action Model

Call-to-Action model is an extension to topic and emotion models. A Call-to-Action argument is not necessary for each article. For instance, the article shown in examples, "How the Ice Age Shaped New York", does not have a call to action, it is an informative narrative. The articles which do not have a call to action usually belong to the "Neutral" sentiment in Emotion Model.

For articles which has a Call-to-Action argument, Emotion Model's sentiment analysis part needs to be adopted. Since the topic is gathered from Topic Model, if the sentiment is positive, we can state that the author is supporting the topic. However, if the sentiment is negative, the author is opposing the topic.

4. Narrative Model in Economics

Narrative Model can be used widely in the field of Economics. Forecasting FED decisions, exchange rates and numerous other economic indicators is possible with Narrative Modeling. With the analysis of the articles in the given field, the pattern how topics spread and the emotions contained can visually be identified.

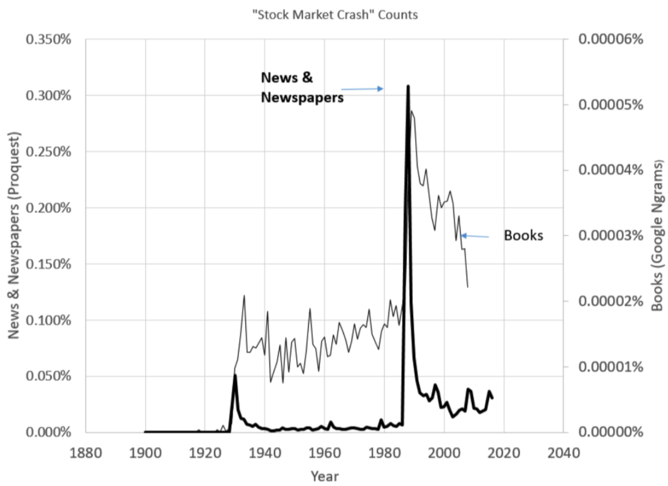

Figure 8: Frequency of appearance of phrase stock market crash by year

In the figure[1] above, frequency of the term "Stock Market Crash" in newspapers and articles after 1880 is plotted. As we can see, narratives spread right at the time of crisis. With Narrative Modeling, analyzing which topics are spreading and forecasting economic indicators and fluctuations are achievable.

5. Conclusion

In this paper, I proposed a Narrative Model by extracting topics and emotions of articles for further use in Narrative Economics. In order to fulfill computers with the knowledge humans have, I used Wikipedia to build a topic model and defined general categories of the articles. By this way, topic model was able to access the information of words' much easier than humans. With the topic model generated, it is straightforward to group articles and analyze how they spread across people. As proposed by Shiller, popular narratives spread among individuals rapidly and this can be modeled to forecast economic fluctuations.

An emotion model is proposed on the top of the topic model to interpret the sentiment of the individuals about the topic. With the sentiment and emotions, a more accurate forecast can be made. Considering the Stock Market articles, if they are spreading among individuals, sentiment is a significant measure to verify if our forecast will be a bull or a bear market. Using the emotion and the topic model, call to action can be generated using the proposed methods.

To sum up,by analyzing articles with the vast amount of computational power across the world, we can use the power of narratives, which are truly effecting the world we live in.

6. References

[1] R. J. Schiller, Narrative Economics, Cowles Foundation for Research in Economics (2017).

[2] R. Alghamdi, K. Alfalqi, A Survey of Topic Modeling in Text Mining, International Journal of Advanced Computer Science and Applications 6 (2015) 147–153.

[3] C. R. Chopade, Text-Based Emotion Recognition: A Survey, International Journal of Science and Research (2015).

[4] E. Gabrilovich, S. Markovitch, Wikipedia-based Semantic Interpretation for Natural Language Processing, Journal of Artificial Intelligence Research 34 (2009) 443–498.

[5] W. J. Broad, New York on Ice, New York Times (2018) D1.

[6] National Research Council Canada sentiment and emotion lexicons, `https://www.nrc-cnrc.gc.ca/eng/rd/ict/emotion_lexicons.html`

[7] I. Beaver, C. Freeman, Prioritization of Risky Chats for Intent Classifier Improvement, Flairs Conference (2016).

[8] J. Kaur, J. Saini, Emotion Detection and Sentiment Analysis in Text Corpus, International Journal of Computer Applications 101 (2014).

[9] K. Stevens, P. Kegelmeyer, D. Andrzejewski, D. Buttler, Exploring Topic Coherence over many Models and many Topics (2012).