



EXAMENSARBETE INOM TECHNOLOGY,  
GRUNDNIVÅ, 15 HP  
*STOCKHOLM, SVERIGE 2018*

# **Topic classification of Monetary Policy Minutes from the Swedish Central Bank**

**ANDREAS CEDERVALL, DANIEL JANSSON**

# Topic classification of Monetary Policy Minutes from the Swedish Central Bank

Andreas Cedervall<sup>1</sup> and Daniel Jansson<sup>2</sup>

*Abstract*—Over the last couple of years, Machine Learning has seen a very high increase in usage. Many previously manual tasks are becoming automated and it stands to reason that this development will continue in an incredible pace. This paper builds on the work in Topic Classification and attempts to provide a baseline on how to analyse the Swedish Central Bank Minutes and gather information using both Latent Dirichlet Allocation and a simple Neural Networks. Topic Classification is done on Monetary Policy Minutes from 2004 to 2018 to find how the distributions of topics change over time. The results are compared to empirical evidence that would confirm trends. Finally a business perspective of the work is analysed to reveal what the benefits of implementing this type of technique could be.

The results of these methods are compared and they differ. Specifically the Neural Network shows larger changes in topic distributions than the Latent Dirichlet Allocation. The neural network also proved to yield more trends that correlated with other observations such as the start of bond purchasing by the Swedish Central Bank. Thus, our results indicate that a Neural Network would perform better than the Latent Dirichlet Allocation when analyzing Swedish Monetary Policy Minutes.

*Sammanfattning*—Under de senaste åren har artificiell intelligens och maskininlärning fått mycket uppmärksamhet och växt otroligt. Tidigare manuella arbeten blir nu automatiserade och mycket tyder på att utvecklingen kommer att fortsätta i en hög takt. Detta arbete bygger vidare på arbeten inom topic modeling (ämnesklassifikation) och applicera detta i ett tidigare utforskat område, riksbanksprotokoll. Latent Dirichlet Allocation och Neural Network används för att undersöka huruvida fördelningen av diskussionspunkter (topics) förändras över tid. Slutligen presenteras en teoretisk diskussion av det potentiella affärsvärdet i att implementera en liknande metod.

Resultaten för de olika modellerna uppvisar stora skillnader över tid. Medan Latent Dirichlet Allocation inte finner några större trender i diskussionspunkter visar Neural Network på större förändringar över tid. De senare stämmer dessutom väl överens med andra observationer såsom påbörjandet av obligationsköp. Därav indikerar resultaten att Neural Network är en mer lämplig metod för analys av riksbankens mötesprotokoll.

*Index Terms*—Machine Learning, Latent Dirichlet Allocation, Neural Network, Central Bank, Riksbank, Topic Modeling, Monetary Policy Minutes

<sup>1</sup>D. Jansson is a student at KTH Royal Institute of Technology in Stockholm, majoring in Industrial Engineering and Management and specializing in Computer Science and Communication (e-mail: dajansso@kth.se).

<sup>2</sup>A. Cedervall is a student at KTH Royal Institute of Technology in Stockholm, majoring in Industrial Engineering and Management and specializing in Computer Science and Communication (e-mail: cederv@kth.se).

## I. INTRODUCTION

Every other to every third month, the Swedish central bank (will be referred to as Riksbanken, the central bank or the Swedish central bank) releases a new Monetary Policy Minute (MPM). The MPM is a report detailing the meeting of the executive board of the Swedish Central Bank. It mainly presents the members views on topics such as; housing market, bond purchases, GDP growth, growth in general, repo rate, inflation (and CPI inflation)[23]. Analysts study the MPM report looking for changes in how Riksbanken expresses their view on different topics compared to the previous report. Based on the changes (or lack thereof) the report presents, banks can decide on their trading strategy. Automating this task would reduce the risk of bias and make sure that the analysis is fast enough so that one is able to act on new information at the same time or earlier than competitors. Although the hard data, such as repo rate changes are currently being analysed autonomously, the soft data such as the percentage of the MPM that are allocated to discussions regarding inflation or the bond purchase program is still subject to manual analysis. Furthermore, it seems like the market reacts to the release of this soft data when looking at the currency market volatility. Figure 1 shows that there is a clear increase in the average change at the time of the release (9:30) on days when a minute is released. A first logical step in analysing the MPM would be to determine the extent to which different topics are discussed. One way of determining it would be using Topic Modelling algorithms, such as the Latent Dirichlet Allocation (an unsupervised model that can be used to find unobservable topics), or using a Feedforward-Neural Network (a supervised model). Both will be presented more thoroughly in the theory section.

### A. Goal and Scientific Questions

The ambition with this paper is to provide a baseline for applying machine learning on the Swedish Central Banks Monetary Policy Minutes. The scientific questions that will be answered is;

- How well can Latent Dirichlet Allocation (LDA) or a Feed-Forward Neural Network (NN) classify trends of what is discussed in the Monetary Policy Minutes?
- What would be the potential benefits and risks from a business perspective of applying models such as LDA and NN to the Monetary Policy Minutes?

### B. Scope and limitations

This paper addresses the problem of understanding how well a MPM can be interpreted using LDA or NN. However,

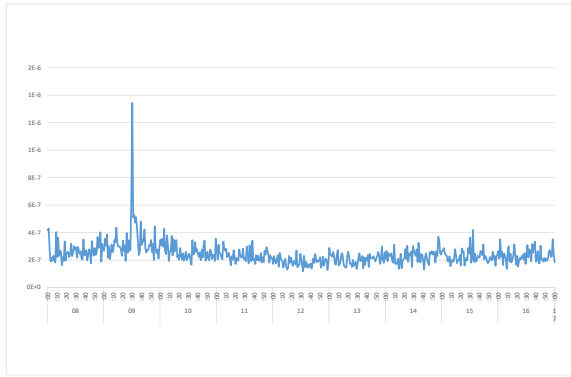


Fig. 1. Intraday average change per second EUR/SEK exchange rate on MPM release dates

there are some areas that will not be discussed. An example is parameter optimization for Latent Dirichlet Allocation and Neural Networks. Furthermore, this paper does not intend to give the reader a deep understanding in how the algorithms work but rather an overview of the algorithms. Instead, we propose articles and literature for further understanding of the methods. Some basic understanding of Natural Language Processing and machine learning is although required. Finally, implications of implementing a machine learning approach are discussed from a business perspective but we refrain from discussing ethical consequences as we do not deem this relevant to the work.

## II. BACKGROUND

### A. Previous Related Work in Topic Modeling

Because of the large amount of data that is becoming available (due to newspapers, social media, forums and other information services) a great amount of research has been conducted within the field of topic modelling. The reason is a need to structure data, an example is dividing recipes into different categories such as dinner recipes, lunch etc. Another topic that has been investigated is the possibility of analysing reviews on for example tech-products for retailers to be able to provide interesting information and increase sales. Some examples of the use of topic modelling are;

- Nils Everlings paper [4] on improving pricing models using topic modelling on companies' earnings calls to determine which industries are/seem relevant to the company. For this purpose, Everling uses non-negative matrix factorization, a topic modelling technique similar to Latent Dirichlet Allocation, which will be described in this report.
- Pang and Lee's [14] attempt to classify sentiment using Naive Bayes, Maximum Entropy and Support Vector Machines. Although it might seem unrelated, sentiment analysis is a topic modelling problem. One of their findings is that sentiment analysis is more difficult than

topic modelling due to more complex sentences. They achieve around 80% accuracy.

- Salinca [19] analyses the topics business reviews using convolutional neural networks and is able to, using this technique, perform very well with accuracies over 90% on larger texts. Salinca compares the results achieved using either pre-trained word vectors such as word2vec and other pre-trained representations of words. One finding is the large difference in performance between these.
- Kim Yoon [12] details to a great extent how he set up the convolutional neural network. Furthermore he is able to, using googles word2vec word representation get results which are very close to 100%. This is of course dependent on the dataset and whether or not the topics are easily recognizable or not.

The examples above give a brief background into what work has been conducted in the field of Topic Modelling. However, only Nils Everlings paper has similarities in terms of what type of document is analysed. Therefore, we will below provide a brief description of what research has been conducted on Central Bank Minutes.

### B. Previous Related Work on central bank minutes

To our knowledge, there has been no previous research on the use of natural language processing (NLP) on any of the Swedish central bank statements or reports. However, two reports have been written which analyse the US Federal Reserves minutes. One is: An AI approach to Fed Watching by Jeffrey N. Saret and Subhadeep Mitra. Saret and Mitra are examining the value of a machine interpretations of minutes from the Federal Open Market Committees (FOMC) [20]. FOMC is the US version of the Swedish central bank and are managing financial conditions such as inflation, repo rate and economic growth [2]. By using the Latent Dirichlet Allocation-algorithm (LDA), Saret and Mitra classifies the minutes in topics. Those topics are measured in order to see how much of the minute is about that topic. This enables them to compare how popular those topics have been through the years (1993-2016), e.g. they noted that after 2014 inflation have become a more frequent topic in the meetings. On the market, there are observers trying to get insight in how the government are going to act. Saret and Mitra argues that those market observers would benefit from a NLP-tool like the topic-modelling they created. This is due to the subjective view of the minute it produces compared to a more objective analysis made by professionals. Furthermore, Saret and Mitra does not present any methodology, the LDA algorithm is mentioned, but never how they implement it.

The other one is Deciphering FedSpeak: The Information Content of FOMC Meetings by Narasimhan Jegadeesh and Di Wu [7]. They like Saret and Mitra use the LDA-algorithm for their analysis, one of the reasons is that it does not require any classified data that could impose subjectivity on the results. However, they extend it to include a sentiment analysis on top of topic modelling. Jegadeesh and Wu divide the information the market get into two categories, soft and hard

information. Hard information is numbers e.g. the repo rate increases with X per cent. Soft information is the subliminal content of the minutes, information that can be acquired by reading the text. The hard information is released closed to the meeting (the source to the minute). The soft information is not released until around two weeks later, and then it is released as a minute of the meeting (MPM), which can be downloaded directly from Riksbankens homepage. By studying the volatility on the market, Jegadeesh and Wu find that the market reacts both when the hard information is released and when the soft information is released. They confirms that the market deems the soft information important as well. The study also proves that the information in the minutes does not matter in aspect of volatility. This is explained with the face that the market has expectations and at the release of a new minute those expectations is revised. Furthermore Jegadeesh and Wu concludes a sentiment analysis of the minute gives a higher predictive value than analysis made by professionals. They also find topics about financial market and dual mandate topics, such as inflation, has a higher informative value.

In order for our paper to complement the research by Saret and Mitra and Jegadeesh and Wu, we will perform a LDA method to analyse Riksbankens MPM. In addition, due to the promising results that Neural Networks have shown, we will provide a basic approach to using this sort of method to analyse the Minutes. This is also because the LDA might be more successful depending on text and which words are deemed relevant to a topic, this will be further elaborated upon subsequent sections of the paper.

### III. THEORY

For clarity this section is divided in pre-processing models, topic classification algorithms and lastly economic theory. Each model will briefly be presented and explained to some extent but not give a full account of all that is to know of each model. The pre-processing is defined as the process of converting raw-data into a more suitable data for the algorithms, e.g. non-significant words are removed, the text is normalized and classified, and so forth. The reason is that a lot of words that is used when writing or speaking have no meaning in the aspect of understanding the topic, the same applies to tense of a verb. More examples on this will follow as each of the pre-processing methods are described. Topic classification is the process of tagging a part of a text into subjects. In this paper the MPMs where split into sentences and all sentences got tagged with a subject (topic). Finally, a brief account of economic theory will be described to support the discussion.

#### A. Language Pre-processing Methods

- **Bag-of-words:** The bag of words method is based on reducing dimensionality by not taking into account the position of a word but only if it is present or not, creating a set of words [9]. Therefore one ends up with vectors with the count of each word in the vocabulary. A sentence such as 'I like cattle' might therefore be

represented with the vector [ 0 , 0 , 1 , ... , 0 , 1 , 0 , 1 ]. To be able to train a model to recognize patterns, this method will be used to convert sentences to vectors.

- **Stop Word Removal:** Stop word removal is a general technique to also reduce the vocabulary size [9], the idea being that words such as 'the', 'is' and 'in' have no effect on which class should be assigned. One such list is the Scikit-learn stop word list. An additional stop word list was created, to further remove non-significant words.
- **Lemmatization:** Lemmatization is the process of converting a word into its least inflected form, also known as its lemma [10]. The diminution in dimension with this method is elementary in terms of the many forms a word can take. For example all of the following words; 'having', 'had', 'have', 'has' and 'will' all map to have. Seeing as many classes have some words that map discriminatingly to their group in the chosen dataset, combining all the inflections into one feature will ensure that the model recognizes its importance.
- **Part-Of-Speech:** Part-of-speech tagging is used to identify which word class i.e. noun, verb etc. a word belongs to. There are several methods to do this. However, most fall into one of two types: Rule-based taggers and stochastic taggers. Rule-based taggers generally use a large database of rules for which types of combinations are possible [26]. Stochastic taggers on the other hand are based on calculating the likelihood of a given tag in a given context using a model trained on a corpus (large mass of text). Seeing as most of the words that are used to identify a topic in the Monetary Policy Minutes are nouns, there is a potential to filter out a lot of word using this type of technique.
- **Term Frequency - Inverse Document Frequency:** Term Frequency - Inverse Document Frequency (TF-IDF) is a weighting algorithm used to evaluate how important a term in a document is to a corpus or a set of documents. The method produces a score, and the higher the score is the more important the term is. The score gets higher every times a term occurs in a document, but is lowered for each document the word exists in [16]. However TF-IDF does not equalize fox with Fox or foxes, which Juan Ramos claims is disadvantageous in his report. Therefore, one can draw the conclusion that if TD-IDF was used on a lemmatized text, it would produce a better result. TD-IDF can be used as a word filtering method, where commonly used words like 'a' or 'the' occurs in a lot of document and therefore gets a lower score. The same goes for names or other words that does not occur that often in a document. This will tell what words that have value for the content of the sentence and what words does not. By knowing what words are important, the classification analysis can be made more efficient by removing the non-significant words.

## B. Topic modelling algorithms

- **Multilayer Feed-forward Neural Network** There exists a few different implementations of Neural Networks. For one example, there is the convolutional neural network which is more complex than the feed-forward neural network that is used in this paper.

A neural network consist of neurons, where every neuron connect to at least one other [26]. The neurons are classified into layers where the first layer is the input layer and the last one the output layer. The layers in between are hidden layers. A neural network with only an input layer and an output layer is a single layer network, and when hidden layers are added the network becomes a multilayer network. How many hidden layers the network should have is something that has to be tested. In some cases zero hidden layers give the best result and in some other several layers do. All neurons in a layer connects to all the neurons in the next, except for the output layer which does not have next layer. The connection between the nodes have weight coefficient attached (a real number). The weight determines the importance of the connection in the neural network and when training the algorithm the weight is revised. Furthermore, a Multi-Layer Feed-forward (MLF) neural network uses supervised training, i.e. the network is taught the desired output, and when trained the algorithm will adjust the weights to improve the output.

Training the network is done through backpropagation, starting from the output layer, updating the weights until the input layer is reached. Initially the weights are arbitrary numbers (they can be randomized). Through the backpropagation the weights are updated using gradient descent method (used for finding a local minimum). To avoid overfitting a dropout layer is used, i.e. to avoid getting stuck in a local minimum some nodes are ignored in the backpropagation. Which nodes are based on a dropout probability. The process of a training set (also called epoch) is running through the algorithm and then backpropagates. This process or epoch is repeated until the weights converge to an optimal set of values.

- **Latent Dirichlet Allocation**

The Latent Dirichlet Allocation (LDA) method for topic modelling was first presented by David Blei, Andrew Ng and Michael Jordan in 2003 [1]. The basic idea, as explained in their article is that each document is composed of several latent topics and that each topic is characterized by a distribution over words. Figure 2 is presented in most literature concerning LDA. The Figure illustrates the scope of each parameter using different plates representing, beginning with the innermost rectangle, the word/token, the document and lastly the corpus. Hence, the parameters  $\alpha$  and  $\beta$  are corpus-level parameters,  $\theta$  is a document level parameter and  $z$  and  $w$  are word/token level parameters. The parameters that can be easily understood are;  $w$  which is the

actual word,  $z$  which is the topic assigned to  $w$  and  $\theta$  which is the topic distribution for the document.  $\alpha$  and  $\beta$  are parameters of the dirichlet priors for the per document topic distributions and per topic word distributions respectively. Important to note is that  $w$  is the only observable variable. The following generative process is assumed by LDA;

- 1) For every document in the corpus
  - a) Choose  $\theta$  from a distribution of  $\text{Dir}(\alpha)$
  - b) For every word in a document
    - i) Choose a topic  $z$  from the distribution  $\text{Multinomial}(\theta)$
    - ii) Choose a word  $w$  from  $P(w|z, \beta)$ . This is a multinomial probability conditioned on  $z$ .

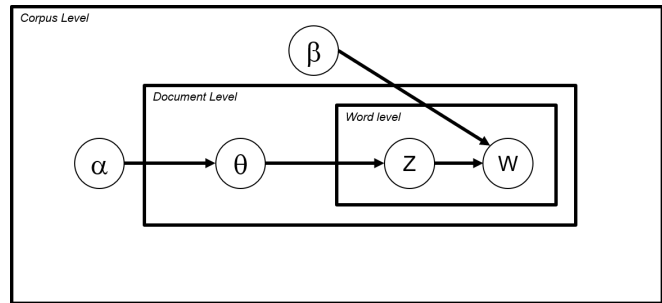


Fig. 2. Illustration of Latent Dirichlet Allocation methodology

The problem of using LDA becomes one of inference. As is explained in their (David Blei et al.) original article, there is a need to calculate the posterior distribution of the hidden variables given a document.

$$P(\theta, z|w, \alpha, \beta) = \frac{P(\theta, z, w|\alpha, \beta)}{P(w|\alpha, \beta)}$$

This however, is not possible to calculate in general. Instead, it is possible to use variational Bayes method to solve the problem. For further details see the article by David Blei, Andrew Ng, Michael Jordan. Furthermore, a learning method is required to be able to apply LDA to problems. As Matthew Hoffman, Francis Bach and David Blei show in their article 'Online learning for Latent Dirichlet Allocation', significant improvement concerning time consumption in training the model is accomplished when using a stochastic gradient learning algorithm instead of a batch gradient descent [6]. Hence, this is used in this project. One interesting property of LDA when comparing to other topic modelling methods is the possibility of a document being assigned more than one topic.

1) *Industrial Dynamics and Multi-Level Perspective*: To give a brief background, theories in the Industrial Dynamics field has its roots in Evolutionary Economics from early 20th century with one of the main characters in this revolution being Joseph Schumpeter. He, among others, criticized the work of previous researchers within Neoclassical Economics for the, very central, equilibrium aspect. It could be argued that this was the baseline for the development of Systems thinking as a research topic. One type of System model is Multi-Level Perspective, a model introduced by Frank Geels in 2002 [5]. One of the key properties that Multi-Level Perspective system highlights is that Industrial Dynamics needs to be analysed in a sociotechnical context. Therefore, it is divided into three levels namely landscape, regime and niche. What the discussion will focus on is especially to what extent landscape and regime changes affect the system. This will be connected to why an analysis like the one described in this paper can be highly useful.

2) *Efficient Market Hypothesis*: The Efficient Market Hypothesis (EMH) is a theory that claims that it is impossible to beat the market. This is due to that on an effective market all stock prices reflects all relevant information at that time. There is no speculative value, the value of a stock is solely based on available information. When new information is released the stock jumps to the new value instead. Furthermore, the EMH is divided in three different forms; the weak, semi-strong and strong. What differs between the forms is how much information that is available. The weak and semi-strong will not be discussed in this paper as the market for the central bank does not reflect any of their content. Why this is will be described in the discussion. In a strong EMH all relevant information is available to everyone, and possible inside information cannot give an investor advantages.

#### IV. METHOD

##### A. Data

The Monetary Policy Minutes used as input are in total 86 pdf files ranging from February 2004 to January 2018. This represents all monetary policy minutes in the time span with the exception of special minutes. The MPMs consist of four different sections describing;

- 1) Economic developments
- 2) Economic outlook abroad
- 3) Monetary policy discussion
- 4) Monetary policy decisions

However, only the third and fourth paragraphs are considered as relevant to analyse. Thus, the former sections of each minute are removed. The MPM were obtained using a custom built web-scraper that downloaded them from the Riksbanken Archive available at [24].

##### B. Modules

The main open-source python modules that were used to complete this project were;

- Keras - used to create the neural network model [11].
- Sklearn - used for TF-IDF vectorization and the creation of the LDA model [21].

- Spacy - used for part-of-speech tagging, lemmatization, stop word removal [22].
- Numpy - used for matrix and array creation [13].
- Py2PDF - used to read and parse PDF files [15].
- Requests - used to build the web-scraper [18].

##### C. Latent Dirichlet Allocation (LDA)

To obtain all monetary policy minutes, a web-scraper was built to download all minutes available on Riksbankens official webpage. Subsequently, a primitive PDF-reader was programmed using the open source PyPDF2 module for python. Here, any page with a string length of less than 200 characters is removed. Although sometimes the MPM has a cover page, which would suggest that the first page should always be removed, this is not always the case. Furthermore, a lot of information, such as numbers and signs, is removed using regular expressions. For parsing the text, the Spacy open source module is used. The Spacy module allows for sophisticated splitting of the string into sentences, identifying part-of-speech (POS) tag, extracting lemmas as well as removing any stop words. Using the objects returned by the PDF-reader, for every sentence;

- Stop words are removed
- Words with POS-tags proper noun, determinant, conjugation, space, number, coordinating conjugation, punctuation, apposition and 'could not be classified' are removed
- Only lemmas of the words remaining are included

The result of this operation is a list for each sentence containing only the lemmas of somewhat relevant words. Applied to this is a TF-IDF vectorizer (using the module sklearn). The TF-IDF vectorizer is set to remove any words with a document frequency higher than 50% or an absolute document frequency (i.e.. the times the word is seen in all documents) less than 10. Furthermore, a maximum number of features returned is set to 300. This returns a matrix of the size X (number of sentences in all documents) \* Y (number of features, which is 300). At this point, the pre-processing is complete and the LDA can be applied. To perform the LDA analysis, the Sklearn module is once again used with the Online Variational Bayes method for learning, a learning offset of 50, a seed for the random algorithm of 0 and a maximum number of iterations of 10. LDA is performed with the number of topics ranging from 2 to 10 to identify an optimal number of unobserved topics. Topics are analysed based on the top 10 words and put into one of the group categories ['House Market', 'Bond program', 'Inflation', 'Currency', 'Growth' and 'None'] where the 'None' category is intended for sentences that does not add any value or information or other topics such as 'repo rate'.

##### D. Neural Network

The same dataset that was obtained using a web-scraper was used for the neural network. However, since a neural network requires classified data, sentences from the reports were manually chosen arbitrarily and classified into six topics visible in Table I. The classes available are the same as

the ones described above, namely [House Market, Bond program, Inflation, Currency, Growth, None].

TABLE I  
NUMBER OF CLASSIFIED SENTENCES PER CLASS

Topic	# of Sentences
Currency	25
Housing	23
Inflation	34
Bond Program	38
Growth	35
None	38

The classified sentences are read and processed using Spacy. For every sentence, the same tokens that were removed as described for the LDA are removed for the neural network as well. Subsequently, the order of classified sentences is shuffled to avoid having the same training and validation set every iteration using the neural network. The size of the training and validation set is 80% and 20% respectively. The neural network model consists of (in order); Three Dense Layers, One Dropout Layer, One Dense Layer, One Dropout Layer and finally an Activation Layer with a softmax activation function. Note that this is one of the settings for the neural network. An attempt to remove all but one Dense layer was made, to determine the effect this had on the precision and recall of the model. The reason for using the softmax function is that it allows for the results of the neural network to be interpreted as a probability distribution over the classes. When training the set, the following applies;

- Training is done in batch sizes of 64 elements.
- Stochastic gradient descent is the optimizer used to backpropagate
- 1000 Epochs are performed to train the model.
- Dropout layers all have the same level of dropout.

The success of the neural network is determined based on precision, recall and F-score. Some constants, such as dropout level, whether or not specific words are removed and the number of layers are varied to analyse the degree to which the neural network is sufficiently optimized for the task of classifying sentences. To every iteration of the neural network, a confusion matrix is produced to calculate precision, recall and F-score. In order to show the gains of applying this model to entire Monetary Policy Minutes, one neural network model is applied to all the minutes and trends are analysed. This will be further elaborated upon in the results and discussion.

## V. RESULTS

The results that will be presented in this report are the following

- Latent Dirichlet Allocation Topics and interpretation
- Latent Dirichlet Allocation topic distribution over the entire dataset
- Neural Network scores on classified sentences
- Neural Network topic distribution over the entire dataset

### A. Latent Dirichlet Allocation Topics and Interpretation

Since the LDA model does not always produce coherent topics, work is required to analyse which number of unobservable topics is optimal. Even though words that mostly affect the classification are to some degree in the same predefined categories, the topics are not perfectly coherent, which can be seen in table II. Although more than 300 words are manually removed from the set such as *rate*, *increase*, *structural*, *Karolina* and *öberg*, many words that do not relate to any specific relatable remain. However, a low number of unobservable topics appear to result in less distinct topics. For example Topic 3 in Table II have words one would classify into several different categories. Furthermore, a higher number of unobservable topics appears to result in more distinct topics with fewer kinds of words, but with more irrelevant words included, an example is the word Euroarea in Topic 6 visible in Table III. Important to note is that the words presented in both Table II and Table III are sorted in order of how much they affect the classification to each topic i.e. the first word 'Growth' in topic 1 below is the word that has the highest weight for topic 1. The topics presented in Table III are put into five different categories: Growth, Inflation, Housing, Bond Program and Currency. Looking at table III the suggestion is that Topics 2,3,8 and 9 mainly deal with Inflation, Topics 5 and 6 should be considered as part of None, Topic 1 is mainly Bond Program, Topic 4 deals with Housing, Currency is the main subject in Topic 10 and lastly Topic 7 is considered a growth topic. All coherent topics are grouped together when used and presented. The topics used are the one in table III, and the distribution is presented in figure 3. The following is the interpretation of the data in table III; Topic 1 is classified as 'Bond', Topic 2, 3, 8 and 9 as 'Inflation', Topic 4 as 'Housing', Topic 7 as 'Growth', Topic 10 as 'Currency'. Rest of the topics i.e. Topic 5 and 6 is not classified to any of the topics, therefore, classified as 'None'.

TABLE II

TOP WORDS PER LDA TOPIC CLASSIFIED WITH THREE TOPICS

	Topic 1		Topic 2		Topic 3
growth	GROWTH	reparate	NONE	inflation	INFLATION
utilization	GROWTH	repo	NONE	interest	NONE
demand	GROWTH	unemployment	GROWTH	price	INFLATION
purchase	NONE	labour	GROWTH	household	HOUSING
bond	BONDPURCHASE	GDP	GROWTH	krona	CURRENCY
wage	GROWTH	recovery	GROWTH	house	HOUSING
employment	GROWTH	inflationary	INFLATION	euroarea	NONE
price	INFLATION	activity	GROWTH	CPI	INFLATION
upturn	GROWTH	share	NONE	debt	HOUSING
oil	NONE	event	NONE	indebtiness	HOUSING

### B. Neural Network scores on classified sentences

As was mentioned in the method section for the neural network, some different settings were changed and their affect determined. These settings are described below. The best results are, as is shown in Table IV, achieved when stop words are manually removed, Dropout is at 30% and the model consists of multiple dense layers. The results in Table IV are for three different settings, namely;

- Settings 1; stop words not removed, dropout 30% and one dense layer

TABLE III  
TOP WORDS PER LDA TOPIC WITH TEN TOPICS

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
bond	utilisation	inflation	interest	reparate
upturn	inflation	cpif	price	repo
reparate	cpu	energy	household	growth
share	credit	price	house	productivity
mortgage	reparate	suggest	gdp	develope
actual	summarise	little	debt	event
variable	experience	exclude	indebtedness	soon
possibility	enter	value	oil	past
group	relate	importance	stability	wait
refer	reservation	march	growth	maintain
Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
euroarea	unemployment	inflationary	stabilise	krona
purchase	labour	trade	inflation	exchangerate
recovery	wage	grow	supply	policyrate
activity	demand	associate	push	inflation
export	employment	upwards	develop	ecb
inflation	consumption	counteract	predict	interest
growth	slowdown	borrow	particular	fast
stable	inflation	claim	martin	know
investment	growth	build	lower	maturity
focus	near	price	prepare	pace

- Settings 2; stop words removed, dropout 30% and one dense layer
- Settings 3; stop words removed, dropout 30% and multiple dense layers

TABLE IV  
PRECISION, RECALL AND F-SCORE FOR DIFFERENT NEURAL NETWORK SETTINGS

	Precision	Recall	F1 — score
Settings 1	0,785	0,734	0,739
Settings 2	0,819	0,751	0,759
Settings 3	0,834	0,790	0,792

### C. Latent Dirichlet Allocation Topic Distribution over the entire dataset

After the LDA is trained on TF-IDF vectorized lemmas, each sentence in the data set is classified. This allows for an analysis of how the distribution over topics has changed over time. There are six different topics that are analysed; Currency, Bond Program, Housing, Inflation, Growth and None. None is not only the unclassifiable topics but also the topics that does not fit in under any of the other categories. This is due to our scope.

The results indicate that the topic distribution does not change in any significant way over time, shown in figure 3. The largest topic is by far Inflation with an average close to 37%. Inflation peaks at 42,4% in December 2007, and continues with a high trend until September 2008. During this time the topic average is close to 41%. Post 2008 the Inflation stabilize at an average of 36,4%. As the Inflation decreases 2008 the Bond program increases to its peak in October 2008 at 16,8%. Other than that the Bond program has a stable average at 9,2%. Regarding Housing, there are two distinct peaks. One 2004 at 18,1% and the other

in 2013 at 18,2%. The latter is a the peak of a longer period (December 2012 to July 2014) where Housing had an average of 16,1% compared to the average of the full duration (2004-2018) at 12,8%. The first peak is during a shorter period of time (between May 2004 and October 2004). Other than those two peaks, the Housing topic is stable close to 12%. As for Growth there are three peaks compared to the average at 10,3%, in April 2005 at 16,4%, in December 2006 to June 2007 with an average of 14,6% and lastly in February 2011 at 14,6%. Lastly, the topics Currency and None has no apparent peak. They vary little around their average at 9,4% (Currency) and 21,5% (None).

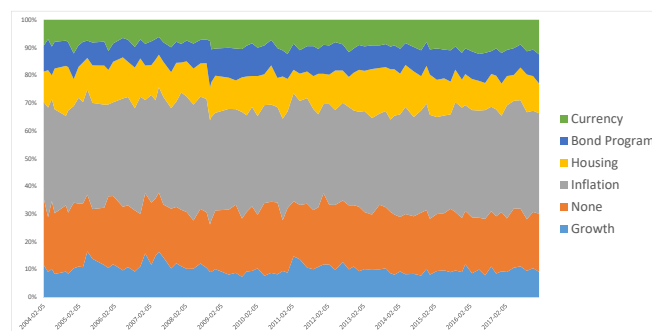


Fig. 3. Distribution of topics on Monetary Policy Minutes from 2004-02-05 to 2018-01-10 using LDA

### D. Neural Network Topic Distribution over entire dataset

Using the trained Neural Network, the MPMs is classified into the six topics determined as interesting, namely; Inflation, Housing, Bond Purchase, Growth, Currency and None. In Figure 4 there is a split of topic contributions. The results indicate that None is a large topic, peaking at 73% in 2009. This might seem odd but is largely due to sentences addressing the repo rate being classified as None. Furthermore, the topic Bond Program is almost not mentioned before 2015 (with an average of 0,9% in 2013 and 0,5% in 2014) and quickly rises to an average of 7,6% in 2015. Similarly, the inflation topic increases from an average of 17,3% in 2012 to an average of 28,5% in 2014. As for the topic Growth, there is a decreasing trend of -1,02% per year using a linear trendline from 2004 to 2018. Lastly, for the topics Currency and Housing there are no apparent long term trends. They vary between a maximum of 5,5% and 8,2% and a minimum of 0,6% and 1,5% respectively.

## VI. DISCUSSION

The two algorithms LDA and NN, are very different for several reasons, the biggest is that one is supervised (NN) the other unsupervised (LDA). Using both LDA and NN allows for two methods with different strengths and weaknesses to



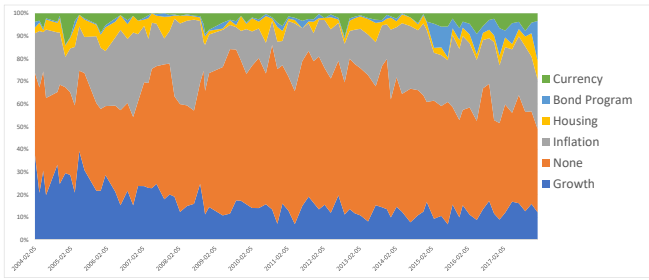


Fig. 4. Distribution of topics on Monetary Policy Minutes from 2004-02-05 to 2018-01-10 using Neural Network

be compared. One important feature that is compared is the inherent objectivity of LDA model which is lost in the Neural Network model when one has to manually classify training data. However, there is the question how objective the LDA-results are, when the results still have to be classified into topics that are fit for analysis e.g. a topic is considered to be about inflation. Another aspect is that the LDA results cannot be measured in accuracy precision etc.

We find that although the LDA appears to give quite objective results, they appear to show no significant changes in topic distribution which goes against our initial hypothesis. The reason for this could supposedly be that the analysis is made on a sentence level instead of paragraph level. The choice to analyse the minutes on a sentence level was chosen mainly with two things in mind. Firstly, it is a simpler task to split an entire PDF file into sentences due to them being separated by either a punctuation or a comma, than to split them by paragraphs. Secondly and just as relevant we believe that if one is to evaluate a sentiment or score different topics, this should be done on a sentence level as well. The reasoning behind this statement is that the more sentences one includes in each data point, the more topics will be represented thus making it more difficult to analyse which words affect sentiment on which topics. We still consider that it is an issue that a lot of sentences are composed of more than one topic.

Saret and Mitra, however, found a great variance over time, but as they did not include any methodology in their report. Thus, a comparison between our implementation and theirs is not possible [20]. For example it is not obvious whether stop word were removed, any lemmatization was performed and most importantly if the model was trained on all the data or a specific subset. The results they present show a significant level of volatility that is not present in our own LDA results. We speculate that a possible reason for this is that the FED Minutes are less controlled by an agenda than Riksbankens. However, we found no information that support this theory.

The LDA results in Figure 3 would suggest that there is a very even distribution of the topics.

One problem with the LDA is that the algorithm classifies the topics by itself and they are not necessarily easy to interpret. A topic about Housing could actually contain keywords from other categories, examples of this can be found in Table II. However, this is something a neural network solves, as the training is supervised (we manually tell the algorithm what sentences belongs to each one of the categories in the training data). The algorithm finds patterns in the training data and then look for those patterns in the text. Because the categories are predefined, less analysis is required to understand the results.

By analysing the results from the NN, shown in Figure 4, we can conclude that a majority of the MPMs is about the topic None. A reason for the large amount of None classified sentences in the MPMs is that the sentences about the repo rate are considered to belong to the None topic. This is due to the repo rate being Riksbankens main tool to control the economy. Thus, in every Minute, the repo rate will be extensively discussed. Furthermore, as is visible in Figure 4, the Bond Program is a topic that is more or less not discussed until 2015. This is in line with recent developments in the economy and that in order to affect the economy, Riksbanken has had to purchase bonds in addition to lowering the repo rate. Figure 5 shows Riksbank's holdings of government bonds from 2015-02-15 to 2018-03-15. We would argue that because of the increase to SEK 321 billion by 2018-03-15, it is obvious that more time has to be dedicated to discussions regarding the significant position Riksbank currently controls. Sweden has also had problems in recent years with reaching the inflation target of 2% annually which is in line with the inflation topic increasing in interest between 2012 and 2014.

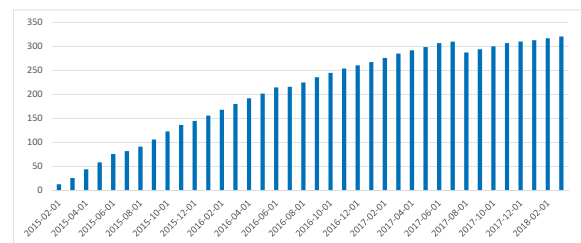


Fig. 5. The Riksbank's holdings of government bonds in nominal amounts from 2015-02-15 to 2018-03-15

Below we will present some of our analysis on the effects of this work, the potential shortcomings and improvement that can be made to the models and lastly some possible

additional work that can be conducted. We will begin by connecting the work to the field of Industrial Dynamics and what the value of similar work put in practice could be.

#### A. Industrial Dynamics

The basis of most Industrial Dynamics models is that the neoclassical view on economy is insufficient as equilibrium is, at best, a temporary state. One instead needs to consider the entire environment as a system. We argue that this type of work could be used by companies to quickly gain insight into some of the developments that are happening within the landscape and regime (defined in the Multi-Level Perspective model). Although this model typically considers the development of technological niches i.e. innovations, there is no reason however that we cannot use the same model to analyse the benefits of this type of project. Getting an overview of the distribution of topics or performing sentiment analyses on minutes or similar documents is a way to gather information on both regime and landscape changes. An example of this would be for a housing construction company in to keep up to date with what developments are happening with regards to new legislation and laws concerning housing or indebtedness. In Geels paper on Multi-Level Perspective [5], he presents the case of the technological transformation from sailing ships to steamships. By applying the Multi-Level Perspective, he finds that landscape changes, one of which was the increased emigration from Europe to America, to a great extent supported the market. This type of change might seem obvious in retrospect, but it is difficult to detect in the moment. Therefore, we believe that there exists an opportunity for market actors to, using techniques such as the one presented in this paper, detect changes in the regime and landscape level. The next section will further discuss the value of this type of analysis but with the Efficient Market Hypothesis in mind.

#### B. Efficient Market Hypothesis

The strong form of EMH is a version where all information is available to everyone on the market, there is no inside information, or the inside information does not affect the price of a stock. We would argue that Riksbanken is acting on a strong-form of EMH. This is due to the Swedish law The principle of public access to official documents, where the public, individuals or companies have access to information about government activities [17].

On page 375 in [3] we find the quote "Information is often said to be the most precious commodity on Wall Street, and the competition for it is insane."

As the quote mentions, information is extremely important in the aspect of an advantage. On a market where all the information is available to everyone at the same time, the only way to benefit from information is to understand new information first [3]. By implementing an algorithm like ours, an overview of the MPM can be gathered far more quickly than by any human. Thus, leading to an information advantage in the aspect discussed above.

#### C. Practical business value

In the following sections we will outline what we believe to be the main practical strengths, weaknesses, opportunities and threats of the usage of machine learning on MPMs.

- **Strengths**

By using any of the two algorithms presented, a quick overview of the report will be created. By going with the hypothesis that important topics are discussed more, a reader can even before reading the report get a hint of what the central bank deems as important subjects right now. Furthermore, in some aspect both of the algorithms creates a subjective view on the topic. Which could be seen as a strength as it just present raw-data without any interfering thoughts of an analyst.

- **Weaknesses**

A neural network is heavily based on the training data it has. Where poor training data gives poor results. To avoid this, a thorough work of setting up the training data is required. First of all there has to be enough data. Secondly, all the data has to be classified, which can be time consuming. The LDA however, creates its own topics. Those topics are not always a pure macroeconomic, but rather a mix of several different topics. This could make them hard to analyse and get something useful from. Besides this, analysing the data is based on hypotheses, where a relation between the data and the hypothesis could be a coincidence. Furthermore, the LDA has a drawback that someone has to classify the topics the algorithm finds, and regarding NN someone has to classify the training- data. Both of those activities take long time to complete and therefore uses up resources.

- **Opportunities**

As mentioned earlier, the key to get an advantage on a market where all the information is available to everyone, is to understand it first. Both algorithms we provide will create a quick overview of a report, thus possibly creating such an opportunity.

- **Threats**

As NLP-models find similarities, structure is the key. If the publisher, in our case the Swedish Central Bank, where to change the format of how those reports are conducted, it is possible the algorithm will fail or performed bad. Also the hypotheses that are used to analyse the data is based on similarities are weak to changes. The personnel at the central bank have their own opinions, which would indicate that a change of personnel could possibly result in a change of how much each of the topics is discussed. This without any other underlying reason than personal preference.

#### D. LDA Potential Improvements

The LDA is a model that requires specification of the number of unobservable topics, which makes a big difference in regards to which words are identified to each topics. The patterns that we are interested in and would prefer to identify

using LDA are ones that seem coherent to us, for example a topic that includes all sentences that refer to inflation, one with all sentences that refer to currency and so on. However, since LDA will attempt to identify topic patterns without bias, it might instead find all sentences with words such as higher which could refer to sentences such as inflation is higher than or indebtedness is higher in Sweden compared to. Although this is an accurate topic distribution using a statistical model, it is not what is required to analyse the minutes.

#### E. Neural Network Potential Improvements

The results from the neural network are, although a precision and recall over 80%, still not as good as they can be. Many studies have shown that using convolutional neural networks or recursive neural network, one can expect precision and recall levels to become even higher [23], [7]. This begs the question, what would be required to implement for example a convolutional neural network on the dataset that is being investigated. Although the Keras module supports the designing of convolutional layers some additional pre-processing would be required. Convolutional layers act as filters on matrix or vector inputs. Therefore, the filter would probably, with input in the form of bag-of-word vectors such as the ones used as input in this paper, not yield any interesting or coherent results. One way to go about solving this problem to be able to use convolutional layers could be to, instead of simply creating a bag-of-words vector with counts for each word, build a matrix using a word to vector implementation such as Googles word2vec. This would result in the model becoming more resilient to analysing words not previously seen. One drawback from such an implementation is that it in general require a lot more training data to ensure that all weights are updated.

One major drawback with the current method is the very few training examples. Although there is no specific limit to the least amount of training data required i.e. it depends a lot on how the data looks, what patterns are recurring and what kind of model one is training, generally training data are in the thousands. Therefore, to determine whether the accuracy achieved and presented in the method is accurate, more work should be conducted with more training data.

#### F. Possible Further Research

To end the discussion we will provide what we believe to be necessary future work on the subject. Because our results from the LDA analysis show no significant trends which we do believe should be present, more qualitative and quantitative analysis of these results has to be performed. One possible way of doing this is having experts in the field analyse several reports and classify to allow for comparisons. Furthermore, although we find that our neural network results show some trends that can be confirmed, a larger amount of training and verification data should be included. Since this work has mainly dealt with machine learning and a theoretical value of approaches such as this, subsequent work

should confirm whether the trends are actually present or simply random observations.

## VII. CONCLUSIONS

To conclude everything, how well does the LDA and the neural network perform when classifying trends in the MPM? Between the two methods the results vary greatly, where the neural network showed distributions which greatly varied over time and where trends are detectable (such as the increase of the bond program topic since 2015 which correlates well to observations). On the contrary, the LDA result gave very little variation, and no visible trends were observed. We argue that this might be because of either our interpretations, the classification being on a sentence level or the way pre-processing is done. Thus, further research is necessary to confirm either of the two.

Regarding what potential benefits and risks there is for a business implementing models such as LDA and neural network, there is a heavy argument that information is power. Through the use of methods like the ones that is presented, an overview is created almost instantly. If the information in that overview is correct this would be greatly beneficial. However, information is like a two edged sword. If the information in some way is wrong, may it be from bad training data or that the new MPM is disposed differently, there is a risk that decisions are based on false information. Thus, there is a possible benefit from having the latest information on the market first, but if acting upon this information and its wrong it could have negative financial consequences.

## VIII. ACKNOWLEDGEMENTS

We would like to thank Olov Engvall for his support and input as well as Mats Hyden at Nordea for the help he provided identifying the interesting subjects, classifying and interpreting the results.

## REFERENCES

- [1] D. M. Blei, A. Y. Ng, and M. I. Jordan, Latent dirichlet allocation, 2003, *J. Mach. Learn. Res.* 3 (March 2003), 993-1022.
- [2] Board of Governors of the Federal Reserve System, The Fed - Federal Open Market Committee, Available at: <https://www.federalreserve.gov/monetarypolicy/fomc.htm> [Accessed 10 May 2018].
- [3] Z. Bodie, A. Kane, A. J. Marcus. Investment and portfolio management: global edition, , 9th edition, McGraw-Hill Companies, 2011.
- [4] N. Everling, Extending the explanatory power of factor pricing models using topic modelling, Master thesis, Kungliga Tekniska Hgskolan, Sweden, 2017.
- [5] F. Geels, "Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study", *Research Policy*, vol. 31, no. 8-9, pp. 1257-1274, 2002.
- [6] M. D. Hoffman, D. M. Blei, and F. Bach, Online learning for Latent Dirichlet Allocation, 2010, In *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 1 (NIPS'10)*, J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta (Eds.), Vol. 1. Curran Associates Inc., USA, 856-864.
- [7] N. Jegadeesh, and D. Wu, Deciphering FedSpeak: The Information Content of FOMC Meetings, 2017, Available at: <https://ssrn.com/abstract=2939937> or <http://dx.doi.org/10.2139/ssrn.2939937>.
- [8] D. Jurafsky, and J. Martin, *Speech and language processing*. 2nd ed. Harlow: Pearson Education, 2014, pp.738.

- [9] D. Jurafsky, and J. Martin, *Speech and language processing*. 2nd ed. Harlow: Pearson Education, 2014, pp. 52-53.
- [10] D. Jurafsky, and J. Martin, *Speech and language processing*. 2nd ed. Harlow: Pearson Education, 2014, pp. 147-151.
- [11] Keras, *Keras: The Python Deep Learning library*, Available at: <https://keras.io/>, 2018 [Accessed 25 May 2018].
- [12] Y. Kim, Convolutional Neural Networks for Sentence Classification, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746-1751, October 25-29, Doha, Qatar, 2014.
- [13] Numpy, *Numpy*, Available at: <http://www.numpy.org/>, 2018 [Accessed 25 May 2018].
- [14] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment Classification using Machine Learning Techniques, Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10, 2002, pp. 79-86
- [15] Py2PDF, *PyPDF2 Documentation*, Available at: <https://pythonhosted.org/PyPDF2/>, 2018 [Accessed 25 May 2018].
- [16] J.Ramos, Using tf-idf to determine word relevance in document queries, 2003, In Proceedings of the first instructional conference on machine learning, 1331-42.
- [17] Regeringskansliet, *Offentlighetsprincipen* Available at: <http://www.regeringen.se/sa-styrs-sverige/det-demokratiska-systemet-i-sverige/offentlighetsprincipen/>, 2018, [Accessed 8 May 2018].
- [18] Requests, *Requests: HTTP for Humans*, Available at: <http://docs.python-requests.org/en/master/>, 2018 [Accessed 25 May 2018].
- [19] A. Salinca, 2017, Convolutional Neural Networks for Sentiment Classification on Business Reviews, 2017, Proceedings of IJCAI Workshop on Semantic Machine Learning (SML 2017): 35-39 <https://arxiv.org/abs/1710.05978>
- [20] J. O. Saret, and S. Mitra, An AI Approach to Fed Watching, Available at: <https://www.twosigma.com/wp-content/uploads/Ai-approach.pdf>, 2018 [Accessed 2018-05-10].
- [21] Scikit-learn, *scikit-learn Machine Learning in Python*, Available at: <http://scikit-learn.org/stable/index.html>, 2018 [Accessed 25 May 2018].
- [22] Spacy, *Industrial-Strength Natural Language Processing*, Available at: <https://spacy.io/>, 2018 [Accessed 25 May 2018].
- [23] Sveriges Riksbank, *Monetary Policy Report*, Available at: <https://www.riksbank.se/en-gb/monetary-policy/monetary-policy-report/>, 2018, [Accessed 10 May 2018].
- [24] Sveriges Riksbank, *Monetary Policy Report archive*. Available at: <http://archive.riksbank.se/en/Web-archiver/Published/Published-from-the-Riksbank/Monetary-policy/Monetary-Policy-Report/?all=1>, 2018 [Accessed 10 May 2018].
- [25] Sveriges Riksbank, *Numerical data, Monetary Policy Report April 2018*, Available at: <https://www.riksbank.se/en-gb/monetary-policy/monetary-policy-report/2018/monetary-policy-april-2018/> [Accessed 24 May 2018]
- [26] D. Svozil, V. Kvasnicka, and J. Pospichal, Introduction to multi-layer feed-forward neural networks, 1997, *Chemometrics and Intelligent Laboratory Systems*, 39(1), pp.43-62.

TRITA EECS-EX-2018:439