Concepts in Topics. Using Word Embeddings to Leverage the Outcomes of Topic Modeling for the Exploration of Digitized Archival Collections

Mathias Coeckelbergs(✉) and Seth Van Hooland

Université libre de Bruxelles, Brussels, Belgium
{mcoeckel,svhoolan}@ulb.ac.be
http://mastic.ulb.ac.be/

Abstract. Within the field of Digital Humanities, unsupervised machine learning techniques such as topic modeling have gained a lot of attention over the last years to explore vast volumes of non-structured textual data. Even if this technique is useful to capture recurring themes across document sets which have no metadata, the interpretation of topics has been consistently highlighted in the literature as problematic. This paper proposes a novel method based on Word Embeddings to facilitate the interpretation of terms which constituted a topic, allowing to discern different concepts automatically within a topic. In order to demonstrate this method, the paper uses the "Cabinet Papers" held and digitised by the The National Archives (TNA) of the United Kingdom (UK). After a discussion of our results, based on coherence measures, we provide details of how we can linguistically interpret these results.

Keywords: Topic modeling · Word embeddings · Document classification · Information retrieval

1 Introduction

The central bottleneck in the current topic modeling practice is how to interpret the developed models. As evidenced early by [7], it is difficult to compare different models, even if some methods such as topic intrusion have become well-known. Although these papers provide an important basis for the evaluation procedure, they remain stuck with the large amount of seemingly unavoidable subjectivity in the evaluation of topic models. Next to these interpretational tasks, a series of coherence measures have been developed to evaluate the intra-topic coherence of top terms. As we will discover later in this paper, these methods measure the overall coherence, but do not evaluate the local context of these words.

On the other hand, word embeddings have proven their worth to derive semantic information from a given corpus, for which tasks such as word analogies

R. Mugnaini (Ed.): DIONE 2020, LNICST 319, pp. 41–52, 2020.
https://doi.org/10.1007/978-3-030-50072-6_4

receive salient scores. Although word embeddings can also be used for document classification, the results focus on the discovery of concepts, understood as combinations of words which are strongly related on a semantic level. The usefulness of this feature has been amply discussed in the literature, but the question as to what extent concepts occur together, remains difficult to approach using only word embeddings.

This article present a novel method to combine word embeddings and topic models in order to compare their unique way of modeling a document collection. Both methods reside within their own research space, respectively neural network models for word embeddings, and information retrieval models for topic modeling. Whereas the former are more interested in modeling language in and of itself, and hence are more interested in the meaning of texts, the latter focus on the retrieval of the most relevant documents. In this article, we seek to answer two research questions. Firstly, we address the way in which Topic Modeling can be used in a specific archival context, in particular a large subset of digitized archival holdings of the National Archives of London (TNA)[1]. [10] demonstrated how to extract re-occurring themes via topic modeling (Latent Dirichlet Allocation). This is particularly useful to circumvent the problem of limited accessibility of digitized archival collections due to the minimal metadata which are available, hence severely limiting the possibilities in which historians and other interested people can interact with these documents.

Secondly, we wish to assess the viability of pre-trained word embeddings on a very large corpus (more than one billion tokens) to serve as a model of an entire (synchronic) language, which can help in semantic problems such as text summarization and query expansion. In this way we can use topic modeling to discover the topics important to the document collection, whereas word embeddings are used to discover concepts inherent in those topics. In other words, we compare the global context modeled by topic modeling with the local context described by word embeddings. By using these two techniques together, we are able to assess the concepts found throughout the document collection, as well as the way in which they co-occur. The remainder of this article will be structured in three parts. In the second sections, an overview of existing work on both topic models and word embeddings is presented, while putting an emphasis on the current open-ended questions. In the third section, then, we propose our novel methodology for combining both techniques, and describe the main results based on some examples. The last section discusses the ways in which we can expand our research.

2 Brief Overview of Topic Modeling and Word Embeddings

The use of vector space models in natural language processing (NLP) has proven very useful since its inception in the 1990s. Since [8] released their paper on

[1] https://www.nationalarchives.gov.uk/cabinetpapers/.

Latent Semantic Indexing (LSI), these methods have been applied in various domains of NLP, such as text summarisation, document classification and sentiment analysis. Their method is capable of finding (semantic) similarities between terms, starting from the word-document co-occurrence matrix, and then continue to use singular value decomposition to achieve dimensionality reduction. From a historical standpoint, this approach could be seen as topic modeling avant-la-lettre, because it results in the words of the original matrix to be clustered together in lower dimension. However, as we will see in the following subsection, the term topic modeling is currently understood to refer to the set of algorithms which use a probabilistic method to achieve this clustering. After this introduction to topic modeling, we will see how the use of vector spaces arises again within the context of word embeddings.

2.1 Topic Modeling

The use of the term 'topic model' has seen a semantic shift in its short but intense lifetime. At the moment, the term 'topic model' is near synonymous with the -by far- most widely used algorithm, namely Latent Dirichlet Allocation (LDA). This generative probabilistic model, published in the seminal 2003 paper by [3], clusters key words extracted from a document collection together in such a way, that they can serve as a source to generate the document collection. This model presumes that topics are hidden in the document collection, which can be rendered explicitly by the aforementioned clusters. Before this seminal paper, other methods of clustering key terms together were developed, which could also be marked as a type of topic model, such as the above mentioned LSI technique by [8]. After the seminal publication of [3], other research has vastly expanded on the conception of topic models. The most important works on the algorithmic evolution include hierarchical LDA [4]), which allows to hierarchically structure the extracted topics, and correlated LDA [5], which models the correlation of different topics. Next to these works which bear witness to the usefulness and wide applicative potential of these models, other voices have underlined some challenges. Two main ones arise within the literature, which both put an emphasis on the inherent subjective aspects of topic models. Firstly, the user needs to make choices concerning the amount of topics the algorithm needs to end within the document collection, as well as to assess whether the corpus under scrutiny is ready to be modeled. As has been described by other researchers, good decisions on both aspects can only be done after trial and error experimentation with the corpus and the algorithm(s), which then afterwards have to be assessed for their salience and applicability. This result brings us directly to the second problematic aspect of topic models, namely their interpretation. As [7] have indicated, it is difficult to present objective standards to monitor which interpretations of the topic model are valid and which not. Through their proposed methods of topic intrusion and word intrusion, they provide a measure of the stability of the topic through semantic relatedness.

Going further in this same direction, [10] built on this idea within a multilingual setting, although they still confess to the a priori interpretational difficulty

of topic models. This difficulty arises from the fact that it is psychologically attractive for humans to give a meaningful interpretation to a list of words they are presented. Even though given several clear cases -which often are cherry-picked-, we can see that a clear interpretation is sometimes allowed, but it is difficult to discern where the grey area of interpretation is located. This results from an interpretational difficulty inherent in topic models, namely that we would like to represent concepts hidden within the text. Although we know that the clusters of keywords are merely a representation of their occurrence within the document collection, we expect them to correspond to clear-cut concepts. This is due to the distributional hypothesis within the field of linguistic semantics, which states that the meaning of a word is determined by the company it keeps. Expressed differently, this hypothesis understands words which can occur in similar contexts to have a semantic relatedness. In practice we see that it can be the case that topics express concepts, for example the concept of DNA, biology, in the famous example from [3]. However, looking from a practical perspective, we see that topics in the grey zone of interpretation do not allow such easy identification with concepts, but rather as a combination of two or more concepts. This is because topic models derive from an information retrieval context, focussing on co-occurrence of terms, rather than a computational linguistic background which focusses on meaning. Of course, it is also attested that noise can be present, even in well-trained models, due to their inherent probabilistic nature. In the next section we explore how word embeddings are used to represent concepts in a local context.

2.2 Word Embeddings

In contrast to the topic models, which rely on probabilistic measures, word embeddings rely on the vector space model. The term word embeddings was first coined by [2] as a constitutive part of their neural language model. The term was made popular ten years later by the seminal paper of [11], in which they describe Word2Vec, an online, freely available toolkit to either train word embeddings on a corpus, or to use their pre-trained word vectors. Given the increasing attention for word embeddings in the NLP community, a year later GloVe was introduced by [13], a comparable toolkit to Word2Vec, which takes a slightly different approach and is the main competitor of Word2Vec. Since these two models, Word2Vec and GloVe are responsible for the current high output of NLP work based on word embeddings, we will limit our presentation to these two, since they diverge on a few key points, necessary for the remainder of this article.

The two models correspond to each other in some general aspects. They both learn a vectorial representation from the co-occurrence of words with the broader context in which they appear. In this way, both models represent geometrical encodings of the words from a corpus. The main difference between both is that Word2Vec is a predictive model, while GloVe is count-based. More specifically, this means that Word2Vec uses the vectorial representation to minimise a loss function predicting the target words from the context words. The lower the

function, the better the predictive power of the model. Word2Vec uses a feed-forward neural network for this task, while using Stochastic Gradient Descent for smoothing. In particular, every feed-forward neural network turns the input vocabulary into word embeddings, found in the weights of the first layer, which hence is also known as the embedding layer. [11] describe two architectures for learning these word embeddings, namely Continuous Bag-of-Words (CBOW) and skip-grams. These CBOW architecture, as the name already implies, uses continuous representations of the documents, and hence does not pay attention to the word order. The model starts from different input words, forming the context, and from this it aims to derive the probability for a specific word. The skip-gram architecture does the exact opposite, meaning it starts from a specific word and tries to derive probabilities for the context words. In contrast to this neural network approach of Word2Vec, GloVe starts of from the co-occurrence counts, represented through a word-context matrix. This is a high-dimensional matrix, both in rows and columns, which needs to be reduced in dimensionality to be able to derive semantic information. For GloVe, this is done by an objective function with weighted least squares with the differences between two words encoded as vector differences. This approach allows GloVe to make the relationships between words its inherent point of departure, whereas for Word2Vec it rather is a by-product.

Word embeddings are strongly related to distributional semantics models (DSM), and as a matter of fact outperform them. In fact, given that the distributional hypothesis states that the meaning of a word is determined by the company it keeps [9], we could infer that this is precisely what the above models also try to convey. The difference between DSMs and word embeddings is that the former keep track of contextual information by registering co-occurrences and counting them, whereas the latter use the same information, but use this mainly to predict surrounding words. As is shown by [1], the latter models nearly always outperform the former models. While DSMs are clearly count models, and Word2Vec clearly is a predict model, outperforming the former due to its neural architecture, the nature of the GloVe model is contested. Since it is factorising a co-occurrence matrix which links word to their context, it is based on the counts of words, being very close to traditional vector-based methods such as Latent Semantic Indexing. From this perspective, it should be classified as a count model, while [1] consider it a predict model as it also tries to minimise a loss function. Within the context of this paper, we seek to use the word embeddings as useful vectorial representations of the language as a whole, which then allows to estimate the semantic relatedness of terms found in the same topic.

2.3 Syntagmatic and Paradigmatic Axes from Structural Linguistics

As we have established in the previous sections, we obtain a global and local perspective on textual meaning by respectively using topic modeling or word embeddings. Since word embeddings trained on a large dataset can be interpreted as representative of the language as a whole, as in the GoogleNews dataset, they can improve the coherence found by the document-level top-ranked words of

topic modeling. On a linguistic level, the difference between these results have been interpreted to accord with Ferdinand De Saussure's seminal distinction between syntagmatic and paradigmatic contexts. The syntagmatic axis looks at the phrase-level at the direct context of a word under scrutiny. This accords well with the word embeddings approach, which takes a context window into account for every word. The paradigmatic axis on the other hand looks at the replaceability of each word in context. This accords well with topic modeling, since the same word can occur in different topic distributions if it can be injected in different contexts. This is a good test of polysemy. Explained differently, the syntagmatic axis for the word 'man' will register typical contextual words such as 'cried' and 'the', whereas the paradigmatic axis measures which words can be placed in the same contexts as 'man'. In this case, 'girl' is an example of a paradigmatical equivalent to 'man'. Figure 1 shows a simple visual example of the syntagmatic-paradigmatic axes.

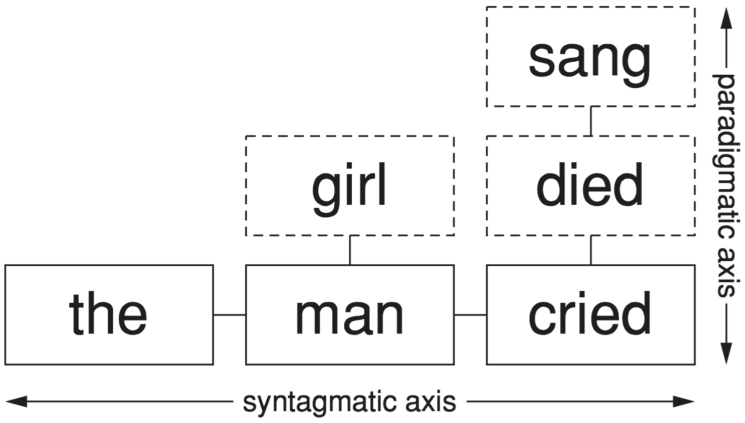

Fig. 1. Example of Syntagmatic-paradigmatic axes [6]

Interpreting the results of the previous section, this means that the keywords occurring within a topic have a high likelihood of being interchangeable, as the paradigmatic axis indicates. Although they have a high rate of interchangeability, this does not necessarily mean that the local contexts in which these words generally appear, would also be similar. This can be measured by word embeddings on a big dataset, to measure the relevance of local contexts within the English language. Reordering the topic modeling results based on the vectorial representation of word embeddings thus means that we evaluate the top paradigmatically similar terms based on their syntagmatic similarity. In our results section we will provide a concrete example of how to interpret the output of the topic modelling and word embeddings step along the syntagmatic-paradigmatic axes.

3 Mobilizing Word Embeddings to Discern Concepts Within Topics - A Novel Methodology

In this section we develop our methodology to use (pre-trained) Word2Vec word embeddings as vectorial representation to determine the semantic relatedness of keywords clustered together as a topic. This method should allow us to decrease the subjectivity of interpretation of topics, since it is easier to determine the underlying concepts. In this way, we could state that the topics are mainly used to discern what the documents are talking about -their global context-, whereas word embeddings are used to find semantic structure -the local context- in the probabilistic outcome of topic modeling. To be clear, we do not propose to use word embeddings as a rival classification method to topic modeling. As we have shown above, this viability has already been documented within the literature. As a connection between concrete documents and a specific vocabulary we use topic modeling, and propose word embeddings to help answer one of the inherently subjective questions of topic modeling techniques in general, namely how many topics to establish.

As we have learned from the literature overview, topic modeling helps to explore themes from a corpus under scrutiny, whereas (pre-trained) word embeddings can be seen as a general, vectorial representation of the language itself. Bringing together both sources of information allows the information gathered by topic modeling to be refined, allowing us to estimate the content of the extracted topics more concretely. Within this paper, we do not wish to evaluate the influence of different topic distributions, that is, to measure the influence of the choice of amount of topics to be retrieved on their internal coherence. Nevertheless, the methodology followed here can be re-used to other topic distributions, which then in turn can be compared. Different questions can be answered in this regard, depending on the research question leading to the conception of the topic modeling. The most prominent examples include the search for the topic distribution which most saliently models a certain concept. For this step, a reconciliation with a controlled vocabulary such as a thesaurus has to be made, but this step is out of the scope of the current article. Within the methodology developed, we divide the approach in separate sections on respectively topic modeling and word embeddings. By doing so, individual parts of the methodology can be used again, for example by exchanging the topic modeling approach by another clustering method, or by testing the coherence of the clusters by another method than word embeddings.

Although the methodological problems of producing salient topic models are cumbersome, we do not wish to go into the results achieved through different parameters entered in the algorithm. The main purpose of this article is to discuss how the semantic relatedness of words in a topic cluster can be evaluated, and how in turn different models can be assessed for their salience concerning a specific query. For this purpose, it suffices to outline the general methodology for modeling our corpus. For the purpose of creating the topic model, we used the machine learning package Gensim in Python, which also allows other machine learning applications such as document classification or information extraction.

We have to make certain decisions concerning stop word removal and the amount of topics, two aspects we will explain now. The stop word list contains a list of words which have no semantic contribution to the model, and hence can be deleted from consideration. Some words deserve a place within this list without hesitation, and can be easily accessed online, where several general-purpose stop word lists are readily available. These lists include such often attested, semantically vacuous words, such as interjections, conjunctions, and personal pronouns. Since the purpose of the current article is to propose a method to automatically identify the semantic coherence of the words in a topic, we will not go further than this standard list in appropriating it for our corpus. A further improvement could include a type of iterative selection, where firstly a topic model is created with only the standard stop words deleted, after which specific words, which either appear as outliers within the dataset, given the general scope of the topics under scrutiny, or words such a broad meaning that they do not have any delimiting power. For example, if our corpus is very strongly economic in nature, words such as 'economics', 'investment' and 'payment' are so widespread that they together constitute an important topic for each document. The number of topics which needs to be selected is a more difficult aspect than the selection of stop words. As we have mentioned in our state of the art on topic modeling, there is no specific standard to derive the amount of topics which should be extracted given a certain corpus size, language of content. The only general advice or rule which can be given is that the number should be such that it produces coherent topics, which in turn also is a concept which is difficult to define. It can be said in general that there is a middle zone in which the 'ideal' number of topics can be found, between having only topics with the most famous words, which do not indicate any semantic distinction between documents or topics when the number of topics selected is too low, or having an overproduction of nearly the same topics, with several noise terms when the number is too high. The reason why 'ideal' is placed between quotes is that in general practice, an approximation seems to be used, whereas only whole- number amounts of topics are tested, leaving the appropriateness of 100 topics rather than 97 or 102 unaccounted for. While admitting our subjective bias, we estimate that the amount of 50 topics gives a workable basis for the continuing questions of the article. This assessment is based on hyperparameter optimization, where the results are interpreted by the authors of this article.

Word embeddings deliver a vectorial representation for every word present the training corpus, and when trained on a representative corpus of sufficient size, these vectors can represent an entire natural language. This can be used subsequently for the representation of each keyword in the topic cluster. As was discussed in the state of the art, Word2Vec generally outperforms GloVe and was therefore chosen in the context for this paper. Also, preference has been given to use pre-trained word vectors, as we wish to underline their general-purpose use, which we apply here to topic modeling.

4 Results of the Co-occurrence of Terms

Word embeddings offer the possibility to select the best topic distribution, but this does not yet give us a clear vision of what we can find inside these topics precisely. From the LDA algorithm we derive that topics are formed according to the co-occurrence of certain words. Hence, words which are found inside the same topic have a high tendency to occur closely to each other in the document collection. From the general outline of topic modeling research, we find that a topic model should be able to identify what a text is about. In an ideal setting, users can identify the collection of keywords grouped by the algorithm to constitute a topic with a human concept, hence allowing for easy interpretation of the topics. For example, in the underlying example of the fifth topic, we see how the words *oil price commission energy coal company market supply demand opec increase percent production scheme index industry countries spot world stocks* have a high tendency of occurring together throughout the documents. Applying word embeddings to these words, and re-arranging them from the word most similar to the other words all the way to the word least similar to the other ones, we find the following: *market price demand supply industry company stocks production opec index percent increase oil energy coal country world commission scheme spot.* Throughout the experience of dealing with topic models, several researchers have pointed out that, although some clear-cut cases are available, the lion's share of clusters of key words lead to interpretational difficulties. Hence, we could state that it is a rare case to find that a topic corresponds to a human concept, and that in general we find multiple concepts represented in the same topic. In the above example we could state that 'market price demand supply industry company stocks' forms a concept of economy, whereas 'oil energy coal' forms a concept of energy. Taking a bigger window than the twenty most relevant terms for the topic might reveal other terms which word embeddings would group belonging to either of these two concepts. Of course it is difficult to objectively score whether the re-arrangement of topical keywords based on their vectorial representation. Showing the coherence according to three different metrics of three randomly chosen topics, before and after re-arrangement, shows that in general improvement is attested. This means that indexation of the documents will be more coherent and relevant if the top term after filtering by word embeddings information is performed than before. The three coherence measures are based on the Palmetto Toolbox, where CV is based on a sliding window, a one-set segmentation of the top words and an indirect confirmation measure that uses normalized pointwise mutual information (NPMI) and the cosinus similarity., CP on a sliding window, a one-preceding segmentation of the top words and the confirmation measure of Fitelson's coherence, and CUCI on a sliding window and the pointwise mutual information (PMI) of all word pairs of the given top words. CV and CP are based on [14], CUCI on [12].

Topic number	Measure	Before Re-arrangement	After Re-arrangement
3	CV	0.2608	0.2990
3	CP	−0.2748	−0.0776
3	CUCI	−0.5651	0.1757
18	CV	0.3872	0.4011
18	CP	0.4096	0.5233
18	CUCI	1.1214	1.4375
39	CV	0.3078	0.3479
39	CP	−0.3573	−0.2385
39	CUCI	1.0682	1.2685

These results show that re-arranging the top keywords from the topic modeling results using word embeddings gives an increase in the score across these three measures. We have selected three topics at random, in order to overload the table with numbers.

5 Linguistic Interpretation of the Results

We have shown how a local context window of word embeddings can improve the coherence of the global topic modeling results. As we have indicated in our overview section, these results can be interpreted using the linguistic distinction between syntagmatic and paradigmatic axes. Interpreting the results of the previous section, this means that the keywords occurring within a topic have a high likelihood of being interchangeable, as the paradigmatic axis indicates. Although they have a high rate of interchangeability, this does not necessarily mean that the local contexts in which these words generally appear, would also be similar. This can be measured by word embeddings on a big dataset, to measure the relevance of local contexts within the English language. Reordering the topic modeling results based on the vectorial representation of word embeddings thus means that we evaluate the top paradigmatically similar terms based on their syntagmatic similarity.

Going further, we can also compare rankings based on both models. Using word embeddings, the local context for every word can be compared by ranking the remainder of the vocabulary according to vector similarity to the word under scrutiny. For topic models, we need to find all occurrences of a given word among the different topic distributions. Each distribution will give a score for every word in the vocabulary, allowing us to find the mean value for every term with respect to a word under scrutiny. This ranking using the topic modeling results is more difficult, because the algorithm does not provide results on the word-level, only on the topic-level and document-level. When both rankings, the one of topic modeling and word embeddings, are similar, this means that paradigmatic and syntagmatic axis overlap. In other words, this means that the document-level word relationships are similar to the ones based on the smaller context window

of word embeddings. In such case, the word under scrutiny is used in an expected way.

Taking the word 'electricity' for example, we find that based on the paradigmatic axis, 'fuel' is the most similar word. This means that, given the corpus under investigation, these two words have highly similar contexts. As two of the most important energy resources, we can imagine that indeed on a global level many words have the same likelihood for both words. Based on the word embeddings, 'fuel' is only ranked on the eleventh place for similarity to 'electricity'. This means that in the average sentence, both of these words do not have a very high likelihood of occurring together. The most similar local level word for 'electricity' is 'megawatts', and for fuel 'gasoline'. These are expected results, since megawatts and gasoline are very likely to be in sentences which also have respectively 'electricity' and 'fuel' in them (Fig. 2).

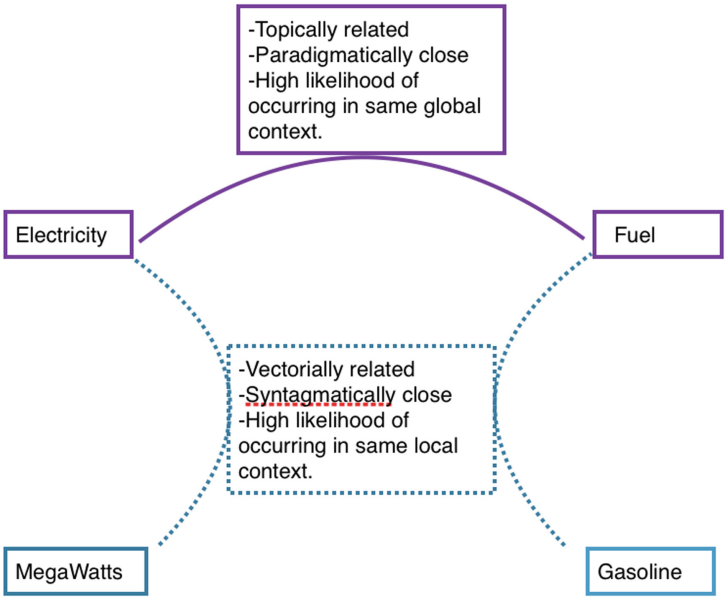

Fig. 2. Schematic Rendering of the relationship between Topically and Vectorially related Words

6 Conclusions and Future Work

In this article, we have sought to explore to what extent the topic modeling and word embedding models show comparable results. We have explained how the former models the global context of a document collection, whereas the latter models the local context. Using local information on global data improves the coherence of the top keywords from topic modeling. This leads to more salient words for indexing or query expansion models. Next to this fully automatic

pipeline to improve indexing using topic modeling, we have also shown how the results of both models can be interpreted linguistically, using the concept of syntagmatic and paradigmatic axes from Ferdinand de Saussure. This interpretation has allowed us to interpret the local and global differences for a concrete example from the corpus. Future work will consist of comparing different architectures within topic models and word embedding models. In this paper we have opted to only use the most well-known incarnation of both, respectively Latent Dirichlet Allocation and Word2Vec.

References

1. Baroni, M., Dinu, G., Kruszewski, G.: Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pp. 238–247 (2014)
2. Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C.: A neural probabilistic language model. J. Mach. Learn. Res. **3**, 1137–1155 (2003)
3. Blei, D.M., Ng, A., Jordan, M.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
4. Blei, D.M., Griffiths, T.L., Jordan, M., Tenenbaum, J.: Hierarchical topic models and the nested Chinese restaurant process. In: Advances in Neural Information Processing Systems 16 (2004)
5. Blei, D.M., Lafferty, J.D.: Correlated topic models. In: Weiss, Y., Schölkopf, B., Platt, J. (eds.) Advances in Neural Information Processing Systems 18. MIT Press, Cambridge (2006)
6. Chandler, D.: Semiotics: The Basics, 2nd edn. Routledge, London (2007)
7. Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., Blei, D.M.: Reading tea leaves: how humans interpret topic models. In: Proceedings of the 22nd International Conference on Neural Information Processing Systems, pp. 288–296 (2016)
8. Deerwester, S.C., Dumais, S.T., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. J. Am. Soc. Inf. Sci. **41**, 391–407 (1990)
9. Firth, J.R.: Papers in Linguistics 1934–1951. Oxford, London (1957)
10. Hengchen, S., Coeckelbergs, M., Van Hooland, S.: Exploring archives with probabilistic models: topic modeling for the valorization of digitised archives of the European Commission. In: IEEE International Conference on Big Data Workshop on Computational Archival Science, Washington D.C., pp. 3245–3249 (2016)
11. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th International Conference on Neural Information Processing Systems 2, pp. 3111–3119 (2013)
12. Newman, D., Lau, J.H., Grieser, K., Baldwin, T.: Automatic evaluation of topic coherence. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 100–108 (2010)
13. Pennington, J., Socher, R., Manning, C.D.: Glove: global vectors for word representation. In: Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
14. Röder, M., Both, A., Hinneburg, A.: Exploring the space of topic coherence measures. In: Proceedings of the Eighth International Conference on Web Search and Data Mining, pp. 399–408 (2015)