



HAL
open science

Patent Specialization for Deep Learning Information Retrieval Algorithms

Denis Cavallucci, Guillaume Guarino, Ahmed Samet

► **To cite this version:**

Denis Cavallucci, Guillaume Guarino, Ahmed Samet. Patent Specialization for Deep Learning Information Retrieval Algorithms. TRIZ Future conference, Sep 2021, Bolzano, Italy. pp.162-169, 10.1007/978-3-030-86614-3_13. hal-04067794

HAL Id: hal-04067794

<https://inria.hal.science/hal-04067794v1>

Submitted on 13 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Patent specialization for deep learning information retrieval algorithms

Guillaume Guarino¹ ✉, Ahmed Samet² and Denis Cavallucci¹

¹ ICUBE/CSIP Team (UMR CNRS 7357)- INSA Strasbourg, France
guillaume.guarino@insa-strasbourg.fr

² ICUBE/SDC Team (UMR CNRS 7357)- INSA Strasbourg, France

Abstract. Extracting information from patents using machine learning algorithms in the context of TRIZ still faces a major problem: the very limited amount of annotated data. Therefore, most approaches, whether for parameter, problem or solution extraction, are based on unsupervised learning algorithms such as Latent Dirichlet Analysis (LDA), or very small supervised learning algorithms such as Support Vector Machine (SVM) or Multi-Layer Perceptron (MLP), which are relatively inefficient. The use of Deep Learning is still relatively uncommon in the extraction of knowledge from patents despite the significant capabilities of these algorithms, particularly BERT. The objective is therefore to present a method for specializing Deep Learning's supervised algorithms on patents while using a very low volume of annotated data. The effectiveness of the method will be analyzed in a task of extracting contradictions from patents, which is a complex task that cannot be performed using unsupervised techniques.

Keywords: Deep Learning, TRIZ, Patent, Contradictions, Summarization.

1 Introduction

Patents contain a great deal of human knowledge, yet they are not widely used because of their large number and their legal nature which makes them difficult to read. Engineers are able to extract useful information from a patent but do not have the capacity to study thousands of patents simultaneously. This is why the use of machine learning algorithms seems to be a valuable option to quickly analyze large masses of data. However, supervised learning of neural networks requires a large dataset. Nevertheless, the analysis of patents for the elaboration of datasets is very time-consuming, which limits the number of annotated data. It therefore seems necessary to develop a method that would allow a high performance with a limited volume of starting data.

Transfer learning consists in training a model on a given dataset, usually of large size, and then specializing it on another nearby task with very little data. In the case of automatic language processing, the main issue is the quality of generated representations from words, sentences or even the document as a whole. These representations take into account the context and the links between words and sentences. This analysis of the links between words and sentences is common to all NLP algorithms, which

explains why these representations can be trained on data that are slightly different from those related to the target problem.

The extraction of contradictions from patents is a major research axis in the perspective of automating the search for a solution or at least assisting in the resolution of problems. Actually, being able to understand a problem is the first step in understanding the resolution process. In this article we will consider contradictions at the sentence level. This means that we will look for the sentences that contain the parameters of the contradictions without extracting these parameters. In fact, it is relatively easy to extract the parameters from a sentence. What is more challenging is to locate the contradictory sentences carrying these parameters. This selection of sentences will be considered as a summarization task in the rest of this paper. Labeling patents to extract contradictions is tedious and the dataset we have developed only contains a few hundred patents. The extraction of contradictions seems to offer a good ground for experimentation of transfer learning.

The main contributions of this paper are therefore:

- a transfer learning method to improve extractive summarization from abstract summarization learning.
- an application of transfer learning to contradiction extraction.

The paper consists of the following sections. Section 2 presents a brief state of art about information retrieval, contradictions and parameters extraction from patent documents. In Section 3 we introduce our method and in Section 4 we detail the model's architecture. We then show some extraction results in Section 5.

We finally conclude our work and show perspectives for future works.

2 Related work

2.1 TRIZ-based automatic analyses

The automatic extraction of contradictions from technical documents, and in particular from patents, is a known and central problem since it would allow the reconstruction of a new updated matrix. However, no prior art approach has succeeded in addressing this problem.

Automatic content analysis with the prism of TRIZ is a common theme in the community whether it is through the classification of inventive principles [1,2,3], the extraction of parameters [4,5], the reconstruction of TRIZ matrices for targeted domains [6].

However, these methods often use simplifications such as the reduction of the number of inventive principles [1,2], the use of keywords or key phrases [3,6] or assumptions on the structure of patents [5] which makes these approaches unusable in practice on "new" contents.

2.2 Automatic summarization

Extractive summarization generally consists in a binary classification of sentences from a source document. A positive label is assigned to a sentence if it contains important information in the context of the targeted summary. A computation of sentences' representations is therefore necessary. These representations should contain enough context information to be sufficient for the decision mechanism (classifier) to assign a label to the sentence.

Numerous automatic summarization approaches exist such as graph-based methods [7,8,9,10], Bayesian approaches [11,12], with Markov models [13] or CRF (Conditional Random Field). Neural networks have become popular since the development of recurrent models like LSTMs (Long Short-Term Memory) or GRU (Gated Recurrent Unit) because of their ability to model the context and dependencies between words and sentences [14,15].

The Transformers [16] allow to build richer representations than the recurrent networks because they are, by design, bidirectional. Nevertheless, the number of parameters is proportional to the squared input length which means that they can be used only for short texts. The main reason of their massive use is the ability to be pre-trained (BERT [17]), XLNet [18]). They will, therefore, be better for complex tasks without requiring large amounts of labelled data. These bidirectional encoders can be trained on all kinds of documents such as Wikipedia pages, articles or journals. These models learn word representations on different datasets before being applied to a specific domain. The huge amounts of data used for pre-training make these models very powerful [19].

3 When abstractive becomes extractive summarization

Abstract summarization consists in generating a word-by-word summary from a starting document. These words may or may not be part of the original text. Extractive summarization consists in selecting key sentences in a source document. A key sentence is a sentence that contains important information about the purpose of the summary. In the case of contradictions, the goal will be to select the sentences containing the parameters. In the case of a generative summary, the goal will also be that the generated sentences contain the parameters of the contradiction.

The main difference lies in the amount of data needed and available for training. The extractive summarization is more easily trainable since it is in fact only a matter of scoring each sentence with its probability of belonging to the summary. However, it is very difficult, if not impossible, to find data that can be used to learn extractive summarization. On the contrary, public data such as patents or press or scientific articles can be easily exploited for abstracting. Indeed, very often, abstracts are available. These abstracts contain the key elements of the documents without being exact copies, word for word, of sentences from the documents. A model can therefore be learned to generate the abstract or summary of an article from it. Nevertheless, the abstract summary will encounter the problem of language. Indeed, even if the data are very abundant (in millions for accessible patents, in hundreds of thousands for press articles or scientific

papers) the generation of realistic sentences remains the main problem. Abstract models can lack consistency and generate misunderstandings even if the keywords are actually present in the abstracts.

The goal of the method is to make the most of both extractive and abstractive techniques. The objective is to develop an extractive model to keep the exact statement of parameters or parameter descriptions in patents using a model that will have already been specialized on patent analysis via abstractive summary learning. This method will overcome the limited size of the extractive summary dataset.

A first abstractive model is trained to generate patent abstracts from their descriptions. This model consists of an encoder and a decoder. The encoder is in charge of extracting the salient information while the decoder exploits this information to generate the abstract. The encoder will therefore be able to build rich representations of the tokens of the input document. These representations will also be adapted to information selection which is exactly the goal of the final model. The encoder of this abstractive model will then be taken and integrated in an extractive model that will learn to select the sentences containing the parameters of the contradiction.

4 Architecture

The abstractive summarization model is presented in Fig. 1. The encoding is provided by BERT [17] which will generate a vector representation for each input token. BERT is a Transformer based on attention mechanisms to model bidirectional dependencies between tokens. A token will be influenced by the set of tokens that precedes it and will influence the set of tokens that follows it. This encoder has allowed results' improvements on various benchmarks in Natural Language Processing such as similarity, Question Answering or automatic summarization which will interest us here. The attention mechanisms allow for a precise selection of information in relation to the information sought and the context, and thus the representations generated are richer and allow for greater extraction precision. A decoder, also transformed, will then exploit the extracted information to generate a summary, in our case, the patent abstract. To generate the abstract, each output of the decoder will point to a word of the dictionary, which allows to reconstruct a whole text. At the beginning of the learning process, this text will be made of random words, and the more the algorithm learns, the more the choice of words and the construction of sentences will be refined.

This model cannot be used directly for extractive summarization (selection of sentences containing the parameters of the contradiction). The first adaptation is the addition of special tokens for classification ([CLS]) at the beginning of each sentence. The representations associated with these tokens will be the representations of the sentences, and it will be these representations that will be classified as the first or second part of the contradiction (depending on whether they contain the improved or degraded evaluation parameter). The goal is therefore to adapt the abstract summary model so that it can contain almost the entire extractive model and thus be able to learn the extractive model on the abstract data. This adapted model is presented in Fig. 2. The representations of the tokens composing the sentences are not used anymore. The representations

of the sentences then pass through additional layers of transformations to perfect the encoding before being decoded. This integration of the extractive model will definitely degrade the abstract summary since all the information is forced to be contained in the sentences' representations which are much less numerous. However, these representations will be richer and will allow an easier classification of the contradictions.

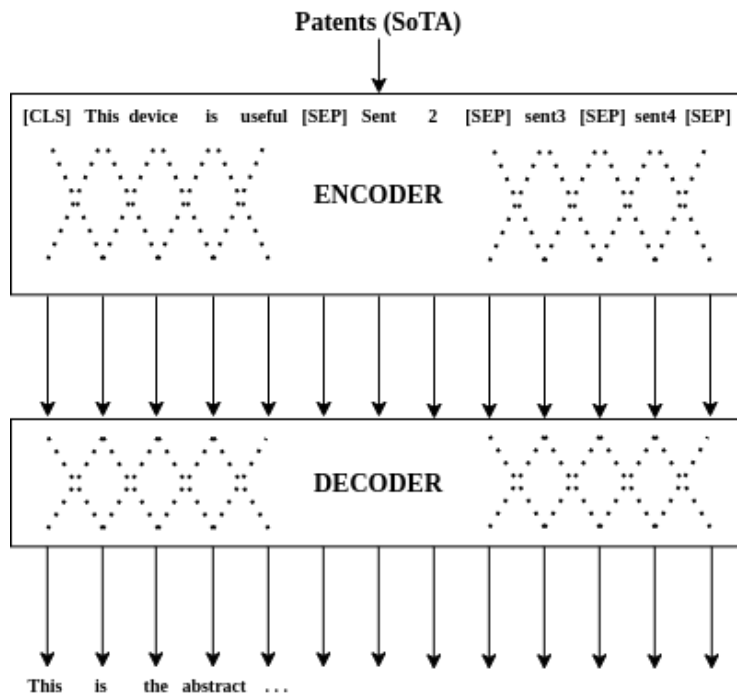


Fig. 1. Abstractive summarization model

5 Experiments

Some results are shown in this section. For each patent, the sentences chosen by the model for the first part and the second part of the contradiction are shown.

US10286848-20190514

Vehicles have been provided with various approaches to providing privacy, security, or additional storage in the vehicle.

While some clipless privacy, security, or storage implements have been developed, these clipless implements can also be difficult to install or remove.

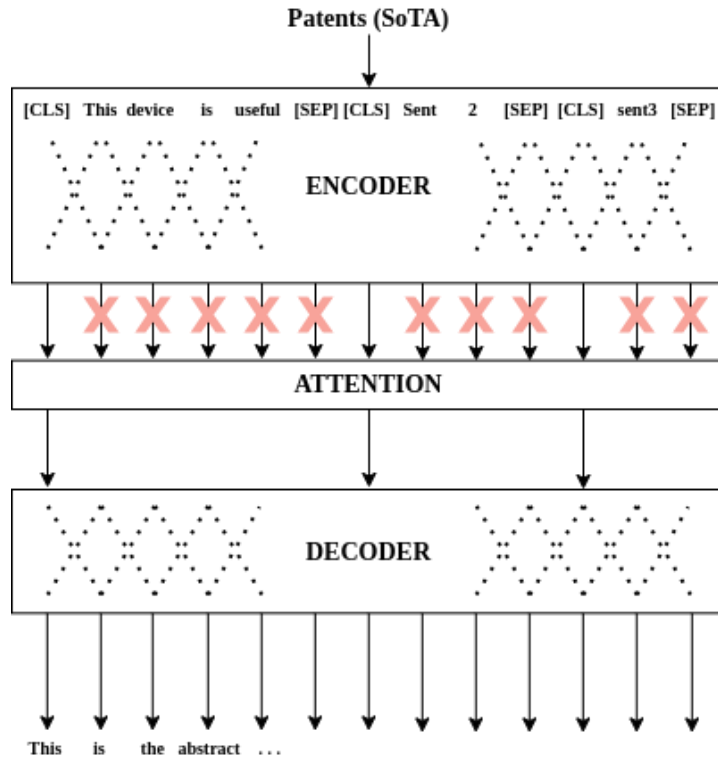


Fig. 2. Adapted model for extractive summarization

US09009869-20150421

When undertaking an activity causing sweating, a person can suffer from the effects of sweat dripping into his eyes.

*These other types of sweat headbands, to remain effective, must be remain in tight ap-
position to the forehead, which may require an uncomfortably tight fit of the headband
around the wearer's head.*

US09939370-20180410

*Thus, distorted band shapes result not only from the interaction between reflectance
and absorbance but from anomalous dispersion of the refractive index relating to other
physical parameters, such as the resonant and non-resonant signal components at play
in car there have been ongoing attempts to compensate for distorted band shapes as
arising from the superposition of reflective components onto the absorbance features
of infrared spectra.*

*This process also accounted for multiple interference spectra; however, the definition
of such interference spectra added subjectivity to the process.*

US07383707-20080610

Therefore, although the locking apparatus of the prior art achieves at least an advantage of capable of simply locking and unlocking the glove box, the operating space of putting the finger between the operating handle and the glove door needs to partition in view of a relationship of pivoting the operating handle from the lower side to the upper side and therefore, the glove door is integrally formed with the containing portion in the recessed shape, however, since the containing portion is significantly projected to an inner side of the glove box, there is a concern of constituting a hindrance in putting small articles into and from the glove box, or restricting an effective space of the glove box.

We can see that most of the time, the parameters are extracted in the sentences. Nevertheless, the extracted sentences may still be long which might cause noise when the parameters will be extracted. In patent US07383707-20080610, the same sentence is chosen as first and second part of the contradiction as it contains both parameters.

It is also important to note that these two sentences alone give a real understanding of the invention. For example, in the patent US10286848-20190514, we understand that the issue is around an additional storage for a car while guaranteeing a flexibility on the installation which is effectively shown in the abstract (without notion of contradiction this time):

A vehicle cargo area includes a retention structure having an upper portion, a vertical portion, an arcuate portion, and an angled portion. The vehicle cargo area further includes a storage tray having a coupling portion that removably couples to the retention structure by passing over the vertical portion and engages with the vertical portion such that the coupling portion is secured in the retention structure.

The notion of summary (in the sense of TRIZ) makes sense since in only two sentences one can extract the "meaning" of an invention, which can be particularly useful in a targeted search for information within the framework of a problem-solving process.

6 Conclusion

The amount of data required by deep learning algorithms makes it difficult to use these models in the context of TRIZ. A solution, called transfer learning, consists in pre-training models on other data before specializing them on their "TRIZ" task.

In this paper we presented a new transfer learning method for patent analysis that can be exploited for contradiction extraction. This method consists in integrating an extractive summary model in an abstractive summary model and training the extractive model on an abstractive task.

Contradiction extraction is a major issue for a massive and detailed understanding of the available inventions and thus of the invention process in order to automatically exploit them to find new inventive solution. It would be possible, for example, starting from an encountered contradiction, to propose patents that would allow this same contradiction to be solved.

The first results are encouraging on the extraction of contradictions even if the chosen sentences are not always the most relevant.

Future works include the extraction of parameters from the sentences chosen by this summarization model, the reconstruction of an updated version of the matrix and the generation of new solutions from patents.

References

1. Han Loh, Cong He, and Lixiang Shen. "Automatic classification of patent documents for TRIZ users". In: *World Patent Information* 28 (Mar. 2006), pp. 6–13. doi:10.1016/j.wpi.2005.07.007.
2. Cong He and Han Tong Loh. "Grouping of TRIZ Inventive Principles to facilitate automatic patent classification". In: *Expert Syst. Appl.* 34 (2008), pp. 788–795.
3. Y. Liang et al. "Computer-aided classification of patents oriented to TRIZ". In: 2009 IEEE International Conference on Industrial Engineering and Engineering Management. 2009, pp. 2389–2393.
4. Hsiang-Tang Chang, Chen-Yen Chang, and Wen-Kuei Wu. "Computerized innovation inspired by existing patents". In: *May 2017*, pp. 1134–1137. doi:10.1109/ICASI.2017.7988268.
5. Gaetano Cascini and Davide Russo. "Computer-aided analysis of patents and search for TRIZ contradictions". In: *International Journal of Product Development* 4 (Jan. 2007), pp. 52–67. doi:10.1504/IJPD.2007.011533.
6. Daria Berdyugina and Denis Cavallucci. "Setting Up Context-Sensitive Real-Time Contradiction Matrix of a Given Field Using Unstructured Texts of Patent Contents and Natural Language Processing". In: *System-atic Complex Problem Solving in the Age of Digitalization and Open Innovation*. Ed. by Denis Cavallucci, Stelian Brad, and Pavel Livotov. Cham: Springer International Publishing, 2020, pp. 30–39. isbn: 978-3-030-61295-5.
7. Larry Page et al. *The PageRank Citation Ranking: Bringing Order to the Web*. 1998.
8. Jon Kleinberg. "Authoritative Sources in a Hyperlinked Environment". In: *Journal of The ACM - JACM* 46 (Jan. 1999).
9. Rada Mihalcea. "Graph-based Ranking Algorithms for Sentence Extraction Applied to Text Summarization". In: *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, companion volume (ACL 2004)*. 2004. url: <http://www.cs.unt.edu/~rada/papers.html>.
10. Marina Litvak and Mark Last. "Graph-based keyword extraction for single-document summarization". In: *(Aug. 2008)*, pp. 17–24.
11. J. Kupiec, J. Pedersen, and F. Chen. "A trainable document summarizer". In: *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Seattle, Washington, United States: ACM Press, 1995, pp. 68–73. isbn: 0-89791-714-6.
12. Chinatsu Aone et al. "A Scalable Summarization System Using Robust NLP". In: *Intelligent Scalable Text Summarization*. 1997. url: <https://www.aclweb.org/anthology/W97-0711>.
13. John Conroy and Dianne O'leary. "Text summarization via hidden Markov models". In: *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. Jan. 2001, pp. 406–407. isbn: 1-58113-331-6. doi:10.1145/383952.384042.

14. Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. “SummaRuNNer: A Re-current Neural Network Based Sequence Model for Extractive Summarization of Documents”. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. AAAI’17. San Francisco, California, USA: AAAIPress, 2017, pp. 3075–3081.
15. Qingyu Zhou et al. “Neural Document Summarization by Jointly Learning to Score and Select Sentences”. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics, July 2018, pp. 654–663. doi:10.18653/v1/P18-1061. url: <https://www.aclweb.org/anthology/P18-1061>.
16. Ashish Vaswani et al. “Attention is All you Need”. In: Advances in Neural Information Processing Systems 30. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 5998–6008. url: <http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf>.
17. Jacob Devlin et al. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: NAACL-HLT. 2019.
18. Zhilin Yang et al. “XLNet: Generalized Autoregressive Pretraining for Language Understanding”. In: Advances in Neural Information Processing Systems 32. Ed. by H. Wallach et al. Curran Associates, Inc., 2019, pp. 5754–5764. url: <http://papers.nips.cc/paper/8812-xlnet-generalized-autoregressive-pretraining-for-language-understanding.pdf>.
19. Yang Liu and Mirella Lapata. “Text Summarization with Pretrained Encoders”. In: EMNLP/IJCNLP. 2019.