



HAL
open science

SummaTRIZ: Summarization Networks for Mining Patent Contradiction

Guillaume Guarino, Ahmed Samet, Amir Nafi, Denis Cavallucci

► **To cite this version:**

Guillaume Guarino, Ahmed Samet, Amir Nafi, Denis Cavallucci. SummaTRIZ: Summarization Networks for Mining Patent Contradiction. 19th IEEE International Conference on Machine Learning and Applications - 2020, 2020, Miami, United States. hal-03398232

HAL Id: hal-03398232

<https://hal.science/hal-03398232>

Submitted on 22 Oct 2021

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.

SummaTRIZ : Summarization Networks for Mining Patent Contradiction

Guillaume Guarino*
 ICUBE (UMR CNRS 7357)
 INSA Strasbourg, France

Ahmed Samet*
 ICUBE (UMR CNRS 7357)
 INSA Strasbourg, France

Amir Nafi†
 ICUBE (UMR CNRS 7357)
 ENGEES, France

Denis Cavallucci*
 ICUBE (UMR CNRS 7357)
 INSA Strasbourg, France

Abstract—TRIZ theory is considered as a kind of innovative theory, which mainly functions in solving contradiction and aims to improve engineers creativity. It may reason on different support such as web and scientific documents. Most of the current knowledge, especially in industrial domains lies in patents. They are, nevertheless, an underused resource because of their complexity, their length and the need of domain knowledge to make use of patent. In this paper, we introduce an innovative application of deep learning to mine knowledge from patents. We show that our summarization based approach, called SummaTRIZ¹, enables the extraction of these contradictions essential for TRIZ problem solving engine. A BERT based summarization approach is introduced to retain contradiction sentences. Our approach is experimentally evaluated on a real data set showing the performance of SummaTRIZ.

Index Terms—Summary, patent contradiction, BERT, Deep learning, TRIZ.

I. INTRODUCTION

TRIZ method was developed by an engineer from the former USSR Guenrich Altshüller from the 1950s onwards [1]. This engineer aimed to find general laws governing the innovation process of engineers. His study of innovative problems has shown that solutions and problems in all fields are very similar so as principles used to invent. From this analysis, he derived 40 inventive principles which are general principles describing the used solutions. Altshüller also worked on the formulation of the problems by defining 39 parameters allowing to form contradictions. The essence of his theory is based on the fact that each problem can be formulated as a contradiction between two parameters (for example, weight and speed). Thanks to the thousands of patents that he has had analysed by several hundred engineers, he was able to predict, for each type of contradiction between parameters, the solution, and therefore the inventive principle that is likely to be applied. Altshüller introduced a matrix, called the TRIZ matrix, which lists the inventive principles to be used according to the parameters forming the contradiction. Figure 1 pictures a part of this TRIZ matrix and the corresponding inventive principles. To illustrate the usage of the matrix, let us assume the following contradiction expressed in this patent: *“To improve the precision of the pipes soldering the operator is replaced by a robot. [...] Low temperatures are likely to cause robot’s failures”*.

Here in this citation, a contradiction between the precision and the reliability is expressed. Indeed, when the precision is increased through the introduction of a robot, the reliability is decreased. The engineer may then link the parameters in contradictions (precision and reliability) with the closest TRIZ parameters which are *Accuracy of manufacture* and *Reliability*. In the corresponding box in the TRIZ matrix, there are the numbers of the inventive principles to be prioritized in this case 1 (Segmentation).

Characteristics (39) Worsening Feature → Improving Feature ↓							Inventive Principles (40)
	1	2	3	4	5	6	
1: Weight of a Mobile Object		-	8,15, 29,34	-	29,17, 38,34	-	1: Segmentation
2: Weight of a stationary Object	-		-	10,1, 29,35	-	35,30, 13,2	2: Extraction
3: Length of a Mobile Object	8,15, 29,34	-		-	15,17, 4	-	3: Local Quality
4: Length of a Stationary Object	-	35,28 40,29	-		-	17,7, 10,40	4: Asymmetry
5: Area of a Mobile Object	2,17, 29,4	-	14,15 18,4	-		-	5: Consolidation
6: Area of a Stationary Object	-	30,2, 14,18	-	26,7, 9,39	-		6: Universality
7: Volume of a Mobile Object	2,25, 29,40	-	1,7, 4,35	-	1,7, 4,17	-	7: Nesting
8: Volume of a Stationary Object	-	35,10 19,14	19,14	35,8, 2,14	-	-	8: Counterweight
9: Speed	2,28, 13,38	-	13, 14,8	-	29,30, 34	-	9: Prior Counteraction

Fig. 1: A part of TRIZ matrix

Common TRIZ based inventive solution approach consists in formulating a problem in the form of a contradiction between “TRIZ” parameters, i.e. the parameters he proposed based on the analysis of the patents. A generic solution is proposed in the form of an inventive principle. Finally, the translation of this generic solution into the initial domain of the problem allows engineers to get a solution that has a high probability of being inventive. The relevance of inventiveness of the solution depends of finding precisely this contradiction.

Since Altshüller work, patents are an underused source of knowledge. The number of granted patents is continuously growing and increasing rapidly creating an immense database of knowledge and invention. Therefore, extracting the contradiction from patents is of paramount importance to automatize contradiction solving using TRIZ. To do so, natural language processing techniques are needed to find contradictions from

*firstname.lastname@insa-strasbourg.fr †firstname.lastname@engees.unistra.fr

¹SummaTRIZ code is made available for reviewers via the following link <https://github.com/nlpSummaTRIZ/SummaTRIZ>

patent corpus. Unfortunately, even with evolution of the NLP research field, results from patents remain noisy [2]–[5]. Results are usually so noisy due to the linguistic specificity of patents and unsuitability of current question answering training datasets.

In this paper, we tackle the problem of finding contradiction sentences inside patents using a BERT summarization technique [6]. We focus on using summarization networks because these methods are based on the meaning of the sentence while other approaches like Word2Vec [7] or SBERT-WK [8] only rely on tokens or words representations to generate sentences representation. The context is then not sufficiently taken into account. As contradictions in patents are expressed within two sentences (i.e., first and second part of the contradiction), we intend to locate them.

To summarize, this paper introduces three contributions: *(i)* the adaptation of a deep neural summarization network to retrieve sentences containing the main contradiction of a patent; *(ii)* an end to end automated framework called SummaTRIZ to extract contradictory sentences from patents web scrapped from USPTO database; *(iii)* finally, an experimental evaluation of the introduced contributions on a real dataset.

The paper consists of the following sections. Section II presents a brief state of art about information retrieval techniques related to patent documents. We also introduce existing summarization algorithms and explain the decision of using BERT. In Section III, we detail the structure of BERT and the adaptation of its architecture for summarization tasks. Section IV describes the structure of our SummaTRIZ model. Finally we show a extensive study on SummaTRIZ performance and some qualitative results of our approach in Section V.

II. RELATED WORK

A. Patent analysis

Several approaches attempted to address the problem of patents classification, information and parameters retrieval in the context of TRIZ. They rely on underperforming classification methods.

Linguistic tools [9] and keywords [10] were the first approaches to be tested to extract information from patents. The results of the extraction are either given to an expert or used to represent knowledge through graphs or discovery matrices. The discovery matrices associate keywords linked to technological systems to keywords linked to systems parameters.

TF-IDF (Term Frequency-Inverse Document Frequency) and LDA clustering are two well used approaches that were applied on patents to retrieve information like physical effects [11]. Physical effects regroup phenomenons like phase transitions, electromagnetic phenomena and thermal phenomena capable of solving a contradiction. For this reason, these effects are widely used in TRIZ literature.

Classifiers like Support Vector Machine (SVM) were also broadly used to extract targeted information. The inventive principles are the general methods to solve a problem described by Altshuller. Extracting the inventive principles used

in patents is a easier task than extracting a full contradiction. Nevertheless, classification of patents according to the inventive principles [2]–[4] shows poor results. The dataset used was small (200 patents) and Loh et al [5] only classified the patents within 6 inventive principles instead of the 40 original inventive principles. He et al. decided to build groups of inventive principles to make the classification simpler but despite these efforts, the results were not improved [2]–[4].

Recognizing patents which use the same principles to solve the problem using clustering methods [12] was one of the tested workaround to address the problem of contradictions extraction but no probing results are shown.

The NLP techniques used in aforementioned papers lead to noisy and often unusable results. Moreover, none of these approaches ever succeeded in contradiction extraction or tried to use the potential of deep learning and summarization method to retrieve information from patents.

B. Summarization networks

In this part, we focus on extractive summarization techniques which consists in selecting the best sentences regarding the summarization objective, in our case retrieving the sentences introducing contradictions. An extractive summarization enable retrieving unbiased information.

The first summarization algorithms were based on graphs [13]–[16]. Nodes (or vertices) of a graph representation are objects (or entities) and the interconnections between these objects are edges (or links). The nodes can be words or concepts and the edges can be for instance a meaning distance between words, coreferences for web-based applications. This approach has been used in PageRank [13] or HITS [14]. Nevertheless, the accuracy of such algorithm remains limited. Naive bayesian approaches [17], [18], Hidden Markov models [19] or Conditional Random Field-based models [20] were also proposed.

With the emergence of deep learning, most of neural based extractive summarization algorithms consist in recurrent networks. Recurrent networks are able to catch time dependencies which make them very suitable for NLP applications. LSTMs (Long Short Term Memory networks) or GRU (Gated Recurrent unit) were especially used. Recurrent networks are by nature unidirectional. They can catch the time dependencies along one direction. This problem was addressed with the introduction of stacked recurrent networks to catch dependencies in both directions (SummaRuNNer [21] and NeuSum [22]). Reinforcement learning was also studied [23] but these algorithms only show their full potential in abstractive summarization.

Currently the widely used performing extractive summarization networks are built with pretrained encoders like Bidirectional Encoder Representations from Transformers (BERT [24]) or XLNet [25]. BERT and XLNet are encoders designed to be trainable in a non-supervised way. They can be trained on all sort of documents like Wikipedia pages, articles, journals. These models learn therefore words representations on various dataset before being applied to a specific domain.

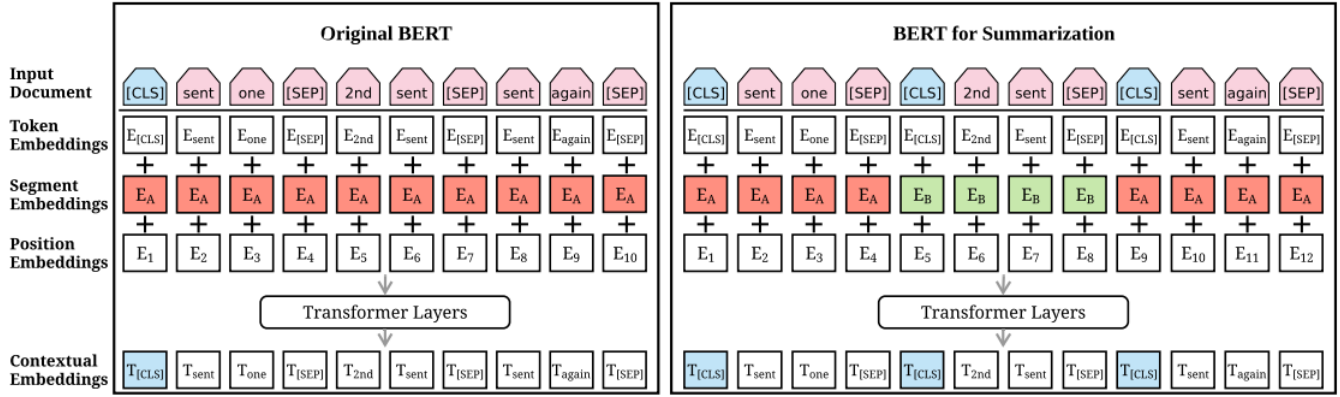


Fig. 2: Summarization network

The huge datasets used for the pretraining model make these models very performing. Moreover, transformer networks are interesting for language understanding and language generation because of their bidirectionality. These networks were developed to take into account right and left contexts in the same time. New research directions have been proposed to see the problem of summarization at another level than traditional sentence-level. Summarization can for instance be considered as a matching problem at the summary-level [26]. Nevertheless, as the number of sentences to extract from the patents is unknown, these approaches cannot be used.

The main criteria for choosing the summerization model are the initial performance of the model in classic summary tasks (newspaper summaries like CNN and Daily Mail for example) and the possibility to avoid creating a large dataset. Building a database in TRIZ domain is particularly difficult because this task require experts interventions and reading a patent is time consuming. We therefore focus on transfer learning and pretrained models. BERT has been singled out for his performance as it constitutes the state of the art for extractive summarization and its pretraining make him available for a fine-tuning on a target domain.

III. BERT FOR SUMMARY

BERT is a representation model pretrained with two different tasks. The first task is called "masked language model" and aims at encoding context bidirectionally for representing each token. It means that both tokens on the left and on the right of the studied token need to be taken into account. To train the model in such way, tokens are masked and BERT is trained to predict these masked tokens given the tokens next to it. To give BERT the notion of relationship between sentences, another pretraining task called "next sentence prediction" is performed. In this binary classification task, given two sentences, BERT is trained to predict if these two sentences are one after the other or is the two sentences were randomly sampled from the corpus. BERT is trained on a corpus of 3,3 billion words.

The original architecture of BERT is detailed in the left of Figure 2. The input of BERT is a sequence of tokens. Special tokens are first inserted in the input text. [CLS] token is inserted at the beginning of the input sequence. This token is used for the classification task. [SEP] tokens are inserted at the end of each sentence. Before being fed into the first transformer layer the token embedding is summed with two others embeddings: a segment embedding to indicate the different sequences/sentences and a positional embedding to indicate the position of each token in a sequence. Initially the maximum size of BERT's input is 512 tokens. The input embeddings are fed into several stacked transformer layers designed to select the richest information from a flow of data. Transformers layers are based on an attention mechanisms as follows,

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (1)$$

with Q , K , V the Queries (what information the layer needs), the Keys and Values (what information it actually has). d is a rescale factor linked to the dimension of the input. The attention is equivalent to a dot product to select the right data.

The output of BERT is a contextual embedding i.e. one vector representation of each input token. The token's contextual embeddings are made to integrate maximum information coming from other tokens. BERT can then be fine-tuned on various domains and applications.

BERT can be adapted for summarization tasks [6] (see on the right of Figure 2). Transformers layers and a binary classifier are added on top of the encoder. The binary classifier takes as input the representation of each sentence and decides on the pertinence of keeping the sentence as a part of the summary. The model uses therefore a sentence level extractor which is particularly interesting for patents summaries in which the length of the summaries are varying. The maximum length of input tokens is also increased by modifying the positional encoding. It can therefore deal with token sequences

longer than 512 tokens which is often the case for states of the art.

We use [6] as summarization model as it complies with all our constraints in term of input length and sentence-level analysis. The model is first fine-tuned on CNN/Daily mail dataset.

IV. GLOBAL APPROACH AND OBJECTIVES

SummaTRIZ aims at retrieving information (i.e. sentences containing contradictions and TRIZ parameters) with neural based summarization methods.

A. SummaTRIZ motivation

A patent consists of different parts: *abstract*, *summary*, *state of the art*, *description*, *description of the drawings* and *claims*. These different parts are extracted separately because, depending on the information to be searched, one of these parts could be pertinent to use. One difficulty is the difference in the names of the different parts according to the patents. It is therefore necessary to go through a recognition of the names of the parts in order to be able to assign standardised names to the parts. This recognition is done thanks to keywords.

The part we are interested in is the *state of the art*. Indeed, it is the part that details what is at stake in the invention. The state of the art presents, usually, the goal of the object of the invention and the difficulties encountered by the current solutions of the prior art. In order to extract the contradiction between parameters solved by the patent, several scenarios are then possible. A starting problem can be presented like this: *"When primary antioxidants, such as hindered phenols, are utilized, polymers can have a more yellow color than unstabilized polymers, therefore decreasing the commercial value of polymers"*. This sentence contains a parameter that worsens, in this case the commercial value. Solutions to this problem will then be presented but they have disadvantages such as: *"Applications employing phosphite additives can result in a reduced quality in the physical properties of polyolefins"*. In this second sentence, a second parameter is recovered which can be degraded by the application of the solution. Therefore, the solution mentioned in the state of the art allows the improvement of the *"commercial value"* parameter but will lead to the degradation of the *"physical properties"* parameter. Consequently, there is a contradiction between these two parameters. In the same way, if a parameter to be improved is presented in the state of the art: *"This necessitates proper positioning of the upwardly-extending portion of the below-ground structure"*, here the positioning. The disadvantages of the solutions of the prior art will form a contradiction with this parameter as for example the risks of injuries: *"Therefore, repetitively placing, removing and replacing such device in the process of determining how best to complete the upper end of the upwardly-projecting portion of the below-ground structure involves considerable physical strain and accompanying risks"*. Finding this contradiction makes it possible to formulate the problem in an original way and allows an in-depth description of the object of the

invention which, by hypothesis, responds to this (or these) contradiction(s). The sentences containing the parameters to be improved or the parameters involved in the initial problem will be associated with the **first part of the contradiction**. The disadvantages of the solutions of the prior art constitute the **second part of the contradiction**. This choice to separate the two parts of the contradiction is motivated by the fact that the meaning of the contradiction must be detected. Indeed, Altshüller's matrix presenting the inventive principles as a function of the contradictions between parameters is not symmetrical. Therefore, if a parameter A is in contradiction with a parameter B, the inventive principles that are statistically the most used and therefore the types of solutions will be different.

B. SummaTRIZ for contradiction retrieval

The overall objective of our SummaTRIZ approach is a classification of state of the art sentences. We distinguish between three main classes: 1st part of contradiction, 2nd part of contradiction and a reject class.

We develop a global approach with the final goal of extracting sentences that form contradictions (see Figure 3). Patents are first web scrapped from the official website of the United States Patents Trademark Office (USPTO) website and parsed into a local database as soon as they are published.

From all these stored patents, we recognize all sections. Only the state of the art part is retained since it is where the contradiction sentences are expressed. They are used to train and test our model to summarize and retain the contradiction. Since we use a supervised model, experts are requested to annotate the 1st and 2nd part of contradiction from all state of the art. The patents' states of the art are tokenized and constitute the input of the classification model. The classification model is based on a summarization approach using BERT summary. The sentences representations are computed as follows with Out^l the output of BERT's layer l , Out^0 BERT's input embeddings, $MHA(Out^{l-1})$ the Multi-Head Attention (MHA) of layer l computed using Attention (Equation 1), Norm a normalization layer and FF a feed forward layer:

$$Out^l = Norm(Out^{l-1} + MHA(Out^{l-1})) \quad (2)$$

$$Out^l = Norm(Out^l + FF(Out^l)) \quad (3)$$

Two instances of the model are used to classify the sentences belonging to the first part and the second part of the contradiction sentences (Equation 4). We assume that a sentence may contain a full contradiction and therefore be classified as first and second part of the contradiction. That's why we use only binary classifiers.

$$Prediction_{part\ c} = Cl_c(Out^n) \quad (4)$$

with Cl_c the classifier associated to the c^{th} part of the contradiction and Out^n the output of the last transformer layer.

The labels are built using the same scheme for both parts of the contradictions. From all the sentences tagged by the

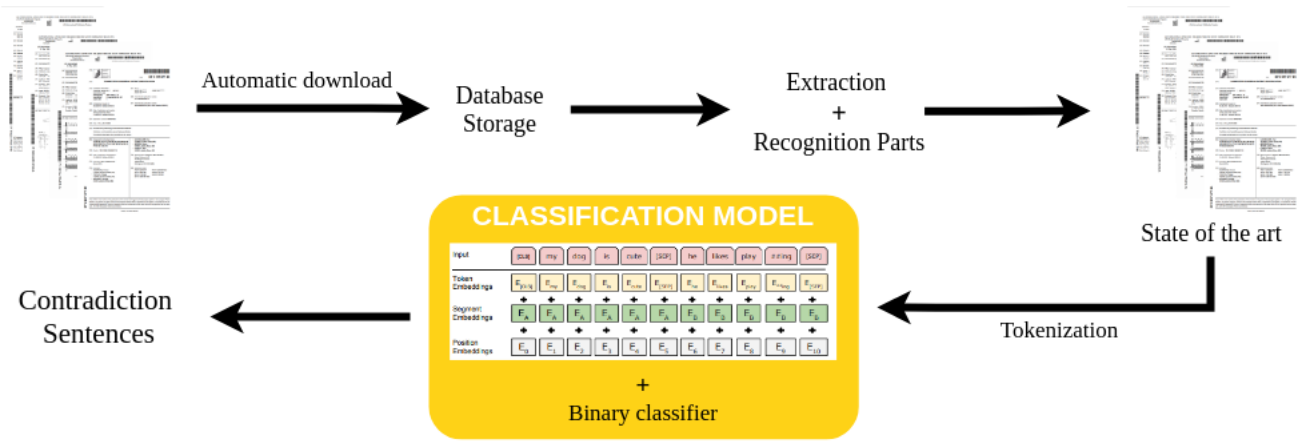


Fig. 3: SummaTRIZ : a global approach

experts, the sentences linked to the studied part are extracted and constitute a new "summary". All the sentences in this summary are associated to the label 1 and the other are considered as being part of the reject class 0. The classifiers are then trained with the sentences representations provided by BERT as inputs.

V. EXPERIMENTS

All the experiments and runtime analyses are performed using a four RTX 2080Ti and Intel Core i9-9820X 3.30 GHz machine. Pytorch framework and Scikit-learn are used for all experiments.

A. Dataset

Our dataset consists of 200 patents annotated by experts. The patents have been randomly selected from patents published in the United States and available on the official website of the United States Patents Trademark Office (USPTO). The years of publication are between 2008 and 2020. A precise annotation policy has been put in place. Patents are retained only if they present a total contradiction in their state of the art i.e. both sides of the contradiction exist. If several contradictions are expressed in the state of the art and if they have no parameters in common, only the most obvious one is tagged to avoid misinterpretations when these contradictions are used. If multiple contradictions are expressed and if the first part of the contradictions is the same, they are tagged. It implies that the models trained on this dataset will only look for this kind of linked contradictions and each sentence of the first part will form a contradiction with each sentence of the second part. We have also implemented a cross-validation by experts to decrease the inter-expert variance in the annotation. The dataset is splitted into a train/validation set of 150 patents and a test set of 50 patents. 5-fold cross-validation is used.

B. Sentences representations

To validate the choice of a summarization network, we show that the sentences representations generated by the

summarization bring unique meaning information. We compare the summarization network with three different sentence representation models: Word2Vec [7], Word2Vec+term frequency-inverse document frequency, SBERT-WK [8]. Word2Vec [7] is a well known static representations model. We add TF-IDF to Word2Vec model to improve the results. Word2Vec's output is a feature vector per word. To get sentences representations, the mean of the word's representations is computed. The mean is weighted with the TF-IDF coefficient for the second model. Word2Vec's representations are static, it means that no matter the current context of the word, the representation would not change. Moreover, a sentence representation is not easily computed from Word2Vec. One solution is a global average of all word's representations in the sentence but as the different words may not have the same importance it causes a loss of information. The use of TF-IDF, in order to take into account the relative importance of words, helps to limit this loss of information. SBERT-WK [8] aims at addressing these limitations by creating dynamic sentences representations i.e. context-aware representations. SBERT-WK is based on BERT and computes the sentence vector through pattern analysis in BERT's layers. SBERT-WK achieves the state of the art in several semantic similarity tasks.

These four models are tested using the same tokenizer and the same output classification algorithm (Logistic regression). Several indicators are used to compare sentence representations. The classification is evaluated using accuracy (Equation 5), precision (Equation 6), recall (Equation 7) and F1 score (Equation 8) expressed with TP the True Positives, FP the False Positives and FN the False Negatives:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>	<i>Extracted sents</i>	<i>Extracted sents(margin)</i>
Word2Vec [7]	0.94	0	0	0	9/64	23/64
Word2Vec + tf-idf	0.93	0	0	0	15/64	24/64
SBERT-WK [8]	0.87	0.12	0.16	0.13	11/64	22/64
SummaTRIZ	0.87	0.26	0.46	0.33	26/64	35/64

TABLE I: Comparative results in the 1st part of contradiction classification

	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 score</i>	<i>Extracted sents</i>	<i>Extracted sents(margin)</i>
Word2Vec [7]	0.74	0.14	0.42	0.19	28/97	46/97
Word2Vec + tf-idf	0.74	0.12	0.34	0.17	29/97	43/97
SBERT-WK [8]	0.84	0.33	0.60	0.42	50/97	66/97
SummaTRIZ	0.77	0.25	0.66	0.36	36/97	48/97

TABLE II: Comparative results in the 2nd part of contradiction classification

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (8)$$

To evaluate the quality of the summary i.e. the quality of sentence extraction, two others indicators are computed. The first and the second part of the contradictions are processed independently. For each test patent and for each part of the contradiction, the number of extracted sentences is set in accordance to the number n of extracted sentences by the expert. Based on the classifier’s output probabilities, the n sentences with the highest probability of being extracted are compared to the gold summary i.e. the sentences extracted by the expert. This indicator is called *Extracted sents*. To evaluate the quality of learning, a second sampling with this time the $n + 1$ highest probabilities sentences is performed. They are then also compared with the gold summary. This last indicator gives an insight on how close are the gold sentences to be actually extracted by the model.

Results are shown in Tables I and II. The limitation of Word2Vec approaches is observed. For the first part of the contradiction, none of the gold sentences are classified as “first part” (so with a probability of being part of contradiction greater than 0.5). That’s why the precision and recall and therefore F1 score are null. Nevertheless, the indicator *Extracted sents* is not null because some of the gold sentences belong highest output probabilities sentences. The precision is also very low for the second part of the contradiction. TF-IDF brings a significant improvement in the *Extracted sents* but it is not enough to compete with the other models. SummaTRIZ has the best results on the first part of the contradiction by a large margin and for all classification metrics and *Extracted sents*. Nevertheless, SBERT-WK is the most performing model on the second part of the contradiction. The explanation lies in the syntax of the sentences. Sentences linked to the second part are often based on keywords like “However”, “Nevertheless”, “decreased” while the syntax of the sentences in the first part is very variable. The presence of these keywords is favorable to models based on words like Word2Vec and SBERT-WK (based on BERT’s tokens vector). SummaTRIZ is trained to extract the sentences meaning by the summarization task. That’s why the model has a high accuracy and recall on the second part but especially on the first part where the meaning is crucial. To confirm these results the

number of full contradictions extraction (1st part and 2nd part) is computed (see Table III). SummaTRIZ is able to extract two time more full contradictions than all the other models without margin and is still the most accurate with margin.

<i>Model</i>	<i>Without margin</i>	<i>With margin</i>
Word2Vec [7]	1/50	13/50
Word2Vec + tf-idf	2/50	13/50
SBERT-WK [8]	5/50	15/50
SummaTRIZ	11/50	18/50

TABLE III: Full contradictions extraction

These experiments show that building sentences representations using summarization is useful to extract TRIZ contradictions as highlighted by the outperforming results of SummaTRIZ comparatively to other state of the art models.

C. Performance of the binary classifier

A classifier lies on top of the architecture of SummaTRIZ model. In this part, we study the impact of this classifier on the results.

The following classifiers are studied: logistic regression [27], SVM [28], decision tree [29], Naïve Bayes [30], Multi-Layer Perceptron [31]. Results for both parts of the contradictions are shown in Table IV. All the classification metrics are computed as well as quality of the extraction sentences. The decision tree is not able to select meaningful features as only 3 contradictions may be extracted. As Naïve Bayes show poor results, the features’ likelihood are therefore more complex than Gaussians distributions. Logistic regression and Multi-Layer Perceptron algorithms show the best results. The classification metrics are not sufficient to decide between logistic regression and MLP. The number of extracted gold sentences is also similar and the same number of full contradictions are extracted.

Logistic regression is a linear classifier, it means that some of the sentences representations are linearly separable but as the other linear classifiers don’t show such results, a non-linear classifier may be a better choice for future uses.

D. Evaluation of the data augmentation

Data augmentation may help to improve results without increasing the size of the dataset. Several data augmentation

	<i>Accuracy</i> (1 st – 2 nd)	<i>Precision</i> (1 – 2)	<i>Recall</i> (1 – 2)	<i>F1 score</i> (1 – 2)	1 st <i>part</i>	2 nd <i>part</i>	<i>Contradictions</i>
Decision Tree	0.91-0.85	0.26-0.23	0.20-0.24	0.23-0.24	9/64	24/97	3/50
Naïve Bayes	0.76-0.67	0.16-0.20	0.64-0.74	0.26-0.31	10/64	31/97	5/50
SVM	0.85-0.68	0.25-0.20	0.59- 0.75	0.35-0.32	24/64	32/97	9/50
MLP	0.93-0.90	0.47-0.40	0.13-0.06	0.20-0.11	27/64	35/97	11/50
Logistic Regression	0.87-0.77	0.26-0.25	0.46-0.66	0.33- 0.36	26/64	36/97	11/50

TABLE IV: Classifiers performance to retrieve the 1st and 2nd part of the patent contradiction

	1 st part Loss improve- ment (%)	2 nd part Loss improve- ment (%)	% Average loss im- provement
Pre-training	16.49	13.48	14.99
Transformation Synonyms	0.26	1.38	0.82
Transformation BERT	2.31	1.97	2.14
Transformation Word2vec	0.51	1.97	1.24
Translation korean	1.28	0.99	1.14
Translation chinese	2.31	0.99	1.65
Translation catalan	0.26	0.39	0.33
Translation chinese + BERT	2.56	1.18	1.87

TABLE V: Global influence of the pretraining and data augmentation methods

techniques are studied in this part. Data augmentation is applied during the training of SummaTRIZ summarization model on the patents dataset. The influence of the pretraining with the summarization task on CNN/Daily Mail is also studied.

Translation techniques consist in translating the patents (originally in english) in a target language and re-translate it in english. The difficulty lies in the high performance of the current translation tools. The target language must be far enough away from English for the wording of sentences to change but close enough for the meaning to remain the same. Double translations with korean, chinese and catalan languages are investigated.

Different transformations are also studied. Synonyms transformation consists in replacing randomly sampled words by synonyms using WordNet dictionary [32]. Transformation with BERT and Word2Vec consist in replacing randomly chosen words with words having a similar representation with BERT or Word2Vec model.

The impact of data augmentation algorithm is studied on the test loss. The pre-training on the summarization task clearly improves the ability of the model to build the sentences representations (15% loss improvement). The other algorithms have less influence. BERT transformation show the best results with a 2% test loss improvement for both parts of contradictions. The fusion of BERT transformation and double translation in chinese which both show the best results in their category does not improve the results compared to BERT transformation alone.

According to this study, BERT transformation will be used as data augmentation module.

E. Use cases

In this part, we show qualitative results of the summarization model with several patent’s state of the art as input.

US8950687B2 is a patent published in 2015 and dedicated to an improvement in HVAC controllers. The state of the art shows that a contradiction exist between the control of the comfort level and the device complexity. The expert

chose the following sentences as first and second part of the contradictions:

First part: "Heating, ventilation, and/or air conditioning (HVAC) systems are often used to control the comfort level within a building or other structure."

Secondpart: "Because of the increased complexity of such modern HVAC controllers, programming and/or operating such devices can be considered difficult and/or confusing, particularly for novice and/or non-technical users."

The sentences extracted by SummaTRIZ are the following:

First part: "Because of the increased complexity of such modern HVAC controllers, programming and/or operating such devices can be considered difficult and/or confusing, particularly for novice and/or non-technical users."

Second part: "There is a need for HVAC controllers that are more intuitive and user friendly to program and operate."

The sentence selected as the first part is in fact the second part for the expert. The sentence alone is usually not enough to predict if it’s the first part or the second part of the contradiction. The context is supposed to do so. Nevertheless it is not a problem if the second extracted sentence is related to expert’s first part. In this example it is not the case though. Only one part of the contradiction is therefore extracted. Not extracting exactly the experts sentences is not imperative as many sentences may be similar in the state of the art. In another patent, US7243212B1, the expert’s sentence "Notably, it may not be practical to provide an embedded coprocessor along with an embedded processor in an integrated circuit due to having to slow performance of the embedded processor to operate the coprocessor." is replaced by "This duality of operating at a same frequency is achievable by having dedicated circuitry for the processor and the coprocessor being implemented in the same technology.". In the context of the state of the art we understand that these two sentences are similar because there is a need to slow the frequency of the processor so that it can be synchronized with the coprocessor. The contradiction between the use of the coprocessor and the high frequencies is correctly extracted and the object of the

invention is understandable.

VI. CONCLUSION

In this paper, we presented SummaTRIZ, an end-to-end patent analysis algorithm to extract the contradictions solved by a patent. The model is based on BERT adapted for summarization. We have shown that sentences representations provided by summarization models improve the extraction of the targeted information. An extensive study on SummaTRIZ performance has been conducted to validate all design choices. Some qualitative results are also presented. SummaTRIZ shows really encouraging results as the extraction of full contradictions is a difficult task.

Future works include a fusion between sentences representations coming from token-based algorithms and summarization algorithms, the extension of the dataset and the training of a discriminative function capable of differentiate a correct contradiction between different contradiction candidates. The development of a pre-processing method to measure the a priori probability that the patent contain a contradiction in the state of the art is also a lead.

REFERENCES

- [1] G. Altshuller, L. Shulyak, and S. Rodman, *The Innovation Algorithm: TRIZ, Systematic Innovation and Technical Creativity*. Technical Innovation Center, 1999. [Online]. Available: <https://books.google.fr/books?id=TOKk0XBH5KgC>
- [2] H. Loh, C. He, and L. Shen, "Automatic classification of patent documents for triz users," *World Patent Information*, vol. 28, pp. 6–13, 03 2006.
- [3] C. He and H. T. Loh, "Grouping of triz inventive principles to facilitate automatic patent classification," *Expert Syst. Appl.*, vol. 34, pp. 788–795, 2008.
- [4] Y. Liang, R. Tan, C. Wang, and Z. Li, "Computer-aided classification of patents oriented to triz," in *2009 IEEE International Conference on Industrial Engineering and Engineering Management*, 2009, pp. 2389–2393.
- [5] Y. Li, K. Bontcheva, and H. Cunningham, "Svm based learning system for information extraction," in *Deterministic and Statistical Methods in Machine Learning*, J. Winkler, M. Niranjana, and N. Lawrence, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 319–339.
- [6] Y. Liu and M. Lapata, "Text summarization with pretrained encoders," in *EMNLP/IJCNLP*, 2019.
- [7] T. Mikolov, G. Corrado, K. Chen, and J. Dean, "Efficient estimation of word representations in vector space," in *Efficient Estimation of Word Representations in Vector Space*, 01 2013, pp. 1–12.
- [8] B. Wang and C. Kuo, "Sbert-wk: A sentence embedding method by dissecting bert-based word models," 02 2020.
- [9] A. Souili and D. Cavallucci, *Automated Extraction of Knowledge Useful to Populate Inventive Design Ontology from Patents*, 07 2017, pp. 43–62.
- [10] U. Y. Valverde, J.-P. Nadeau, and D. Scaravetti, "A new method for extracting knowledge from patents to inspire designers during the problem-solving phase," *Journal of Engineering Design*, vol. 28, no. 6, pp. 369–407, 2017.
- [11] D. M. Korobkin, S. A. Fomenkov, and A. G. Kravets, "Extraction of physical effects practical applications from patent database," in *2017 8th International Conference on Information, Intelligence, Systems Applications (IISA)*, Aug 2017, pp. 1–5.
- [12] Y. Liang and R. Tan, "A text-mining-based patent analysis in product innovative process," in *Trends in Computer Aided Innovation*, N. León-Rovira, Ed. Boston, MA: Springer US, 2007, pp. 89–96.
- [13] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," 1998.
- [14] J. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of The ACM - JACM*, vol. 46, 01 1999.
- [15] R. 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. [Online]. Available: <http://www.cs.unt.edu/~rada/papers.html>
- [16] M. Litvak and M. Last, "Graph-based keyword extraction for single-document summarization," pp. 17–24, 08 2008.
- [17] 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*. New York, NY, USA: ACM Press, 1995, pp. 68–73.
- [18] C. Aone, M. E. Okun, J. Gortalsky, and B. Larsen, "A scalable summarization system using robust NLP," in *Intelligent Scalable Text Summarization*, 1997. [Online]. Available: <https://www.aclweb.org/anthology/W97-0711>
- [19] J. Conroy and D. 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*, 01 2001, pp. 406–407.
- [20] D. Shen, J.-T. Sun, H. Li, Q. Yang, and Z. Chen, "Document summarization using conditional random fields," *IJCAI International Joint Conference on Artificial Intelligence*, pp. 2862–2867, 01 2007.
- [21] R. Nallapati, F. Zhai, and B. Zhou, "Summarunner: A recurrent neural network based sequence model for extractive summarization of documents," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, ser. AAAI'17. AAAI Press, 2017, p. 3075–3081.
- [22] Q. Zhou, N. Yang, F. Wei, S. Huang, M. Zhou, and T. Zhao, "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, Jul. 2018, pp. 654–663. [Online]. Available: <https://www.aclweb.org/anthology/P18-1061>
- [23] S. Narayan, S. B. Cohen, and M. Lapata, "Ranking sentences for extractive summarization with reinforcement learning," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, Jun. 2018, pp. 1747–1759. [Online]. Available: <https://www.aclweb.org/anthology/N18-1158>
- [24] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *NAACL-HLT*, 2019.
- [25] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," in *Advances in Neural Information Processing Systems* 32, H. Wallach, H. Larochelle, A. Beygelzimer, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 5754–5764. [Online]. Available: <http://papers.nips.cc/paper/8812-xlnet-generalized-autoregressive-pretraining-for-language-understanding.pdf>
- [26] M. Zhong, P. Liu, Y. Chen, D. Wang, X. Qiu, and X. Huang, "Extractive summarization as text matching," in *Proceedings of the 2018 Conference of the North*, 04 2020.
- [27] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.
- [28] M. A. Hearst, "Support vector machines," *IEEE Intelligent Systems*, vol. 13, no. 4, p. 18–28, Jul. 1998.
- [29] J. R. Quinlan, "Induction of decision trees," *MACH. LEARN*, vol. 1, pp. 81–106, 1986.
- [30] S. Xu, "Bayesian naïve bayes classifiers to text classification," *J. Inf. Sci.*, vol. 44, no. 1, p. 48–59, Feb. 2018.
- [31] M.-C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer perceptron and neural networks," *WSEAS Trans. Cir. and Sys.*, vol. 8, no. 7, p. 579–588, Jul. 2009.
- [32] C. Fellbaum, *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.