

A Hybrid Approach for Biomedical Relation Extraction Using Finite State Automata and Random Forest- Weighted Fusion

Thanassis Mavropoulos¹, Dimitris Liparas¹, Spyridon Symeonidis¹, Stefanos Vrochidis¹ and Ioannis Kompatsiaris¹

¹ Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece
{mavrathan, dliparas, spyridons, stefanos, ikom}@iti.gr

Abstract. The automatic extraction of relations between medical entities found in related texts is considered to be a very important task, due to the multitude of applications that it can support, from question answering systems to the development of medical ontologies. Many different methodologies have been presented and applied to this task over the years. Of particular interest are hybrid approaches, in which different techniques are combined in order to improve the individual performance of either one of them. In this study, we extend a previously established hybrid framework for medical relation extraction, which we modify by enhancing the pattern-based part of the framework and by applying a more sophisticated weighting method. Most notably, we replace the use of regular expressions with finite state automata for the pattern-building part, while the fusion part is replaced by a weighting strategy that is based on the operational capabilities of the Random Forests algorithm. The experimental results indicate the superiority of the proposed approach against the aforementioned well-established hybrid methodology and other state-of-the-art approaches.

Keywords: Natural Language Processing, Relation Extraction, Supervised Learning, Support Vector Machines, Random Forests, Weighted Fusion

1 Introduction

The onset of the digital era and notably the advent of the internet have not only changed the way people communicate and entertain themselves but have also altered fundamentally their working practices and needs. The medical domain has been on the forefront of these changes, as medical professionals have been exploiting the latest advancements of research and technology in order to improve their services since the very beginning. But this wealth of information is sometimes overwhelming and difficult to tackle manually. A certain level of automation in information extraction is imperative, especially when non-medical practitioners, like patients or their families, are involved. In most cases these people do not possess the ability to fully understand the language used by the professionals since there is a great knowledge gap between

the two groups. The rich in terminology patient history reports is one such area, especially when these are riddled with acronyms tailored to the medical domain. The same holds for online resources, like dedicated medical sites and forums, which users often consider when soliciting for information on drugs, diseases or treatments.

Medical concept relation extraction deals with the automatic extraction of relations that exist between entity types relevant to this domain, such as treatment, test or disease, among others. This task has been the focal point for a lot of researchers, due to many applications that it can support, such as the creation of medical ontologies and content representation that could serve as basis for medical content retrieval and question answering systems, as well as decision support services for doctors. According to [1], "*identifying relations between medical entities in clinical data can help in stratifying patients by disease susceptibility and response to therapy, reducing the size, duration, and cost of clinical trials, leading to the development of new treatments, diagnostics, and prevention therapies*".

Traditionally, studies on medical relation extraction have relied on rule/pattern-based linguistic approaches, machine learning ones and also on hybrid systems that combine linguistic templates and machine learning in order to improve their results. An example of a hybrid framework for medical relation extraction is the approach introduced in [2] and further evaluated in [3], which relied on two different methodologies: a) relation patterns defined by human experts via regular expressions and b) Support Vector Machine (SVM)-based classification based on three types of extracted features, namely lexical, morphosyntactic and semantic features. Fusion of the results from these two methodologies was achieved by means of a strategy, which relied on the training examples of a given dataset, giving more influence to the relation patterns when few training examples were available for a certain relation type and more influence to the machine learning approach when enough examples were provided.

In this paper, the focus is shifted towards the relation extraction task of the 2010 i2b2/VA challenge, which required the extraction of eight types of semantic relationships found between the medical concepts of the given dataset. The other parts of the contest involved the extraction of the medical concepts themselves and also the annotation of the assertions made about these concepts. We are inspired by the hybrid approach described above and we extend it with an innovative pattern-construction method, based on finite state automata, and a novel weighted fusion strategy. More specifically, we approach the creation of linguistic patterns not via the use of regular expressions, as in the case of [2], but by using node-based finite state automata, which can include information like the part of speech (POS) and the inflection of a lexical unit or even contain whole gazetteers of words inside a node.

As an additional novelty, we introduce the use of a Random Forests (RF) classification model, which provides the weighted fusion values for the pattern-based and machine learning modules of the relation extraction framework based on its operational performance on the training set, with the use of the out-of-bag (OOB) error estimate [4]. It should be noted that we keep the use of the SVM classifier for the machine learning module of our framework, due to its demonstrated superiority in many natural language processing (NLP)-related classification tasks. Our hybrid framework is applied to the currently available partial version of the 2010 i2b2/VA

challenge dataset [5] and the experimental results demonstrate its superior performance, compared to a number of considered approaches.

The rest of this paper is organised as follows: In Section 2, the theoretical background and an outline of the relevant literature are provided. In Section 3 the proposed hybrid relation extraction approach is described, while Section 4 provides the experimental framework of our study. In Section 5, the results of the experiments are presented and discussed. Finally, Section 6 concludes the paper.

2 Related work and theoretical background

In this section, since the biomedical domain constitutes the point of interest of the current study, we report previous work on relation extraction in this field. In addition, we provide information on the theoretical background, as well as the related work for the Random Forests (RF) and Support Vector Machines (SVMs) machine learning methods.

As already mentioned in Section 1, three main types of methodologies have been proposed over the years for concept relation extraction: the rule/pattern-based linguistic approaches, the statistical/machine learning approaches and the hybrid ones, which combine both approaches.

Pattern-based systems have been used in the biomedical domain since the early 2000s and have mainly approached the problem as a text classification one. [6] tried to extract and structure information related to molecular pathways with their GeneWays system. A year later, [7] attempted to extract similar relationships between genes, proteins, drugs and diseases.

However, the term “relation extraction” is only part of the problem called “relation classification”, which was first introduced in [8] and entails the extraction of the semantic roles and the recognition of the relationship that holds between them. It was a very influential study that explored five generative graphical models and a neural network to identify seven different relationships that can be found between “treatment” and “disease” entities. The corpus that was used in their work originates from “The BioText Project”, is known as the “MEDLINE 2001” corpus and has since been widely used in relation extraction tasks. In [9], a Conditional Random Fields (CRF) classifier was used because of the need to detect the medical entities and at the same time, the relations between them. The semantic relations between diseases and treatments, as well as between genes and treatments were targeted, which were classified into seven and five predefined types respectively. All experiments were conducted on the “MEDLINE 2001” corpus. Relation extraction between entities in literature text (Medline abstracts) was conducted by [10], via the use of kernel-based learning methods. The method involved a customization of the standard tree kernel function “by incorporating a trace kernel to capture richer contextual information” and resulted in outperforming word and sequence kernels.

The framework that currently claims the best results between treatments and diseases on the MEDLINE 2001 corpus is the one presented in [11], which uses a hybrid feature set for the classification of relations. The major differentiation is in the seman-

tic feature set, where verb phrases are ranked using the Unified Medical Language System (UMLS), while the relations are classified by SVM and Naïve Bayes models.

2010 was a year that marked a great insurgence of research in the medical concept extraction domain and this was due in no small part to the respective i2b2 Shared-Task and Workshop. The contest gave the research community the incentive by supplying a pre-annotated corpus with concepts, relations and assertions. Since then, the contest's best ranking systems are considered as the reference, against which all new ones are benchmarked.

The research, which is underway in the extraction of biomedical relationships, has also been receiving growing attention, “with numerous biological and clinical applications including those in pharmacogenomics, clinical trial screening and adverse drug reaction detection”, as [12] are outlining in great detail. In addition, there have been some recent approaches based solely on Convolutional Neural Network (CNN) models. For instance, in [13], a CNN-based model is implemented in order to extract the semantic relations found between medical concepts and with the goal “to learn features automatically and thus reduce the dependency on manual feature engineering”. The method is applied to the currently available partial version of the 2010 i2b2/VA challenge dataset with promising results.

Random Forests (RF) is a well-known machine learning method [4], used with great success in many applications. Its basic idea is the construction of a multitude of decision trees, which can be used for classification and regression purposes. There is randomness in the operational procedures of RF in two different ways: 1) Each decision tree is constructed on a different group of data, sampled randomly with replacement (bootstrap) from the training set, and 2) During the construction of each decision tree, the best split at each node is determined based on a randomly selected subset of the variable set. An estimation of the generalisation error of RF can be provided by means of an inherent method called out-of-bag (OOB) error. In a nutshell, only approximately 2/3 of the original data examples are used in a specific bootstrap sample during the construction of a decision tree. The rest of the original data examples (approximately 1/3), called OOB data, are used for testing the performance of the constructed decision tree. The OOB error is the averaged prediction error for each training case, using only the decision trees that do not have that training case in their bootstrap sample. As already mentioned, RF has been successfully applied to many disciplines. Specifically in the biomedical domain, there have been applications of RF for automated diagnosis of diseases [14], electromyography (EMG) signal classification [15], or in the context of brain-computer interfaces (BCI) [16], among others. Finally, the use of late fusion strategies based on RF's operational capabilities in the context of multimodal news articles classification has been investigated in [17].

Support Vector Machines (SVMs) [18] are supervised learning methods used for solving pattern recognition problems. Their basic notion lies in hyperplanes, which are used to separate sets of data points with different class memberships in multidimensional spaces. The effectiveness of SVMs in NLP classification tasks and more specifically, for relation extraction, can be highlighted by the fact that the highest performance for the relation extraction task in the 2010 i2b2/VA challenge was achieved by [19] with their supervised approach. This approach employed an SVM

classifier to identify relations, which was informed by several resources such as Wikipedia, WordNet, General Inquirer and a relation similarity metric. Furthermore, the only hybrid system participating in the challenge, employing an SVM classifier together with manually constructed linguistic patterns was developed by [20]. Finally, [1] used an SVM classifier with a combination of lexical, syntactic and semantic features, terms extracted from a vector-space model created using a random projection algorithm, as well as additional contextual information extracted at sentence-level to detect relations.

3 Hybrid relation extraction approach

In this section we present the proposed framework for the medical relation extraction problem, which is illustrated in Figure 1. It consists of two main modules for relation extraction (a pattern-based and a machine learning one) and a weighting module for the fusion of the results provided by each module.

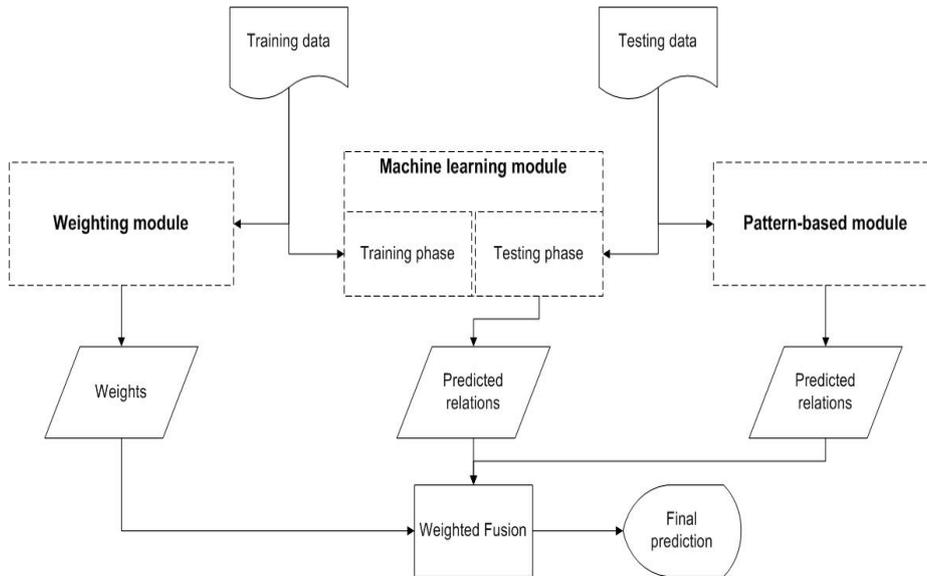


Fig. 1. Proposed relation extraction framework

Pattern-based module. While developing a pattern based method one has to consider the many forms that are often utilised in natural language to express the same thing. These variations need to be taken into consideration when devising the manually constructed rules and patterns, in order for the system to deliver the optimal results. This exact fact is also what makes pattern based methods complex and time consuming to develop. The procedure is not without limitations, with the most important being that the patterns need to be exhaustive enough to cover all possible language variations but manage it without overlapping each other. Therefore, the set of patterns

that is created must be finely tuned in order to avoid conflicts that can invalidate the extraction results. Another limitation is that the pattern creation is largely dependant on the corpus, which dictates a certain degree of flexibility and adaptation when there is the need to apply the method to a new corpus.

The method of choice revolves around finite state automata, which, while being the simplest level of grammar and well understood by users who write rules, is also a technique versatile enough to enable detailed description of complex linguistic phenomena as well as permit the generation of output files rich in linguistic information.

Thus, for the semantic relation extraction task, a set of patterns is constructed for each target relation after examining the structure of certain natural language expressions and detecting common forms in them. This is usually possible with the use of regular expressions and by exploiting keywords usually found in clinical texts, like *cure*, *treat*, *drug* and *side effect*. It is the most commonly used method and the one employed by [3] in their MEANS system. However, the current paper adopts an approach which is based on the exploitation of finite state automata (or graphs) via the use of the corpus processing suite *Unitex* [21], in order to overcome any limitations that are encountered when utilising regular expressions. The pattern-building procedure is done through a powerful interface that enables the manipulation of interconnecting nodes, in order for the user to achieve the most descriptive pattern possible. These nodes may contain a POS, a regular expression, a multitude of linguistic filters (e.g. the feminine plural forms of an adjective) or even whole graphs. A major differentiation compared to simple regular expressions, which ultimately plays a pivotal role in the effectiveness of a *Unitex*-made graph, is the ability to exploit the rich in linguistic information incorporated dictionaries. These have been manually created and contain the grammatical attributes, such as POS or inflection, for the whole of the English vocabulary. In addition to the default integrated dictionaries, *Unitex* also supports the creation of custom ones which can be populated with specialised entries such as disease or treatment terminology.

Each relation targeted by the pattern-based module is being represented by a number of dedicated, manually constructed patterns that locate medical entities/concepts, which appear in pairs in a sentence. A weighted label of specificity is allocated to each pattern in order to solve ambiguous matches, since different relations can be expressed in similar manners (for each pattern, the more detailed the representation of the lexical context, the more specific the weight that gets allocated). The pattern weights that correspond to the assigned labels take the values of 1 for the most specific relation type pattern, 0.75 for a fairly specific one and 0.50 for low specificity patterns (i.e. $R1=1$, $R2=0.75$, $R3=0.50$, with $R1$ being the most specific relation (R)). When the entity pair meets the criteria laid out by one of these patterns, the respective label is assigned. To be more precise via an example, the phrase “He had been noting night sweats, increasing fatigue, anorexia, and dyspnea, which were not particularly improved by increased transfusions or alterations of hydroxy urea.” can be represented with the automaton of Figure 2, while one of the possible output sentences is represented as ($E1=entity1$ and $E2=entity2$): *He had been noting night sweats, increasing fatigue, anorexia, and <E2>dyspnea</E2>which were not particularly*

<TrWP2>improved by</TrWP2><E1>increased transfusions </E1>or alterations of hydroxy urea.

All grey boxes invoke secondary graphs with similar formalism to this one, which contain relevant information to their title. The nodes “disease/signORsymptom” and “treat/cadec_drug/gene_unknown” enclose the relevant dictionaries, while the nodes “negation”, “possession”, “conjunction” describe the respective syntactic functions. Lastly, the white node, which is the only one not evoking another graph, is determining the output of the box, which in this case is the relation type *<TrWP2>* (Treatment Worsens Problem with level 2 specificity). In total, around 350 patterns were created, a number that also includes assistive graphs, like the ones used to handle lexical units of trivial importance found between or around the target entities (*test_{10}*/*test_{20}*).

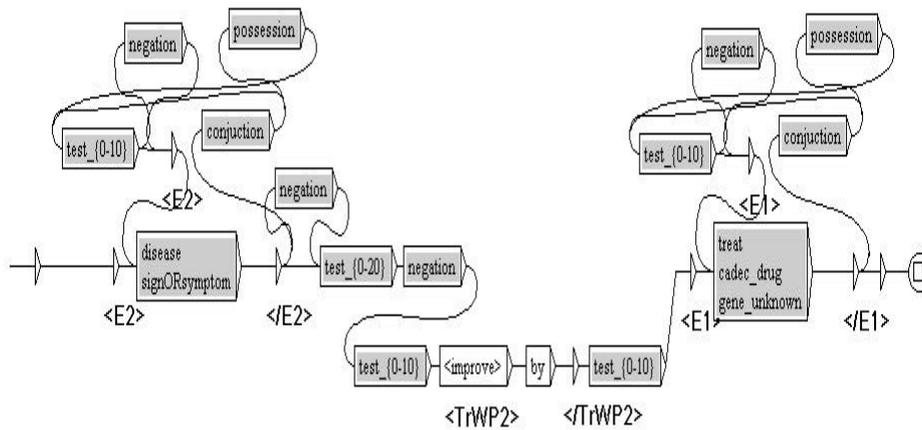


Fig. 2. Finite state automaton representing the “TrWP” relation type.

Machine learning module. In the training phase, a linear SVM classifier is trained on features extracted from a given dataset in order to describe each example. The extracted features fall into three types: lexical, morphosyntactic and semantic features.

The lexical features include the entities’ position in the phrase, the words that form each entity and their immediate context; the words before, after and between them. Also of importance are their lemmas. The morphosyntactic features include the POS (extracted by the Stanford CoreNLP suite [22]) of the lexical units in question, the number of words that form each entity, the verbs before, after and between the entity pairs. Finally, the semantic features refer to the concepts associated to the target entities, as well as those found in their close vicinity; before, after and between them. They are all derived from the online resource UMLS [23], which is a software suite

that encompasses various health-related vocabularies and standards to allow for interactions between computer systems. Another type of feature, which carries semantic information and is provided in the dataset, is the concept type of each entity. However, it was decided that, while such a feature is positively helpful and already available in the given dataset, it wouldn't be included in the feature set of the used classifier. The reason behind this decision lies in the non-existent availability of a reliable resource/procedure that can provide equivalent values in a real-life, non-laboratory scenario.

In the testing phase, for any instance where its relation type is considered to be unknown, the trained SVM model outputs a prediction of the relation type in the form of probability scores.

Weighting module. The probability scores from the pattern-based and machine learning modules are combined using weighted fusion. Different weights are assigned to each module and for each class (relation type). In order to output the final probability that a case is relevant to a class R , the predicted scores P_{pb} (from the pattern-based module) and P_{ml} (from the machine learning module) are first multiplied by their corresponding weights W_{pb} and W_{ml} and are then summed, as in equation (1). The relation type with the highest fused probability score is assigned to each test set instance.

$$P_{\text{fused}}(R) = (W_{pb}(R) * P_{pb}(R)) + (W_{ml}(R) * P_{ml}(R)) \quad (1)$$

In this study, we propose a weighting method, which relies on a different classifier than the one used in the machine learning module. Specifically, an RF model is trained on the training examples in order to leverage an operational capability exclusive to this algorithm. This capability is the out-of-bag (OOB) error, which provides an estimation of the generalisation error of RF. During the training of the RF model, a portion of the original data examples (approximately 1/3), called OOB data, are used for testing the performance of each constructed decision tree. The accuracy of the trained RF model on the OOB data is calculated for each class (relation type) separately and the corresponding scores are assigned as weight values to the machine learning module. The sum of the weights for the two modules must be strictly equal to 1. This means that the pattern-based weight for a relation R is the complement of the corresponding machine learning weight, $W_{pb}(R) + W_{ml}(R) = 1$.

4 Experimental framework

Dataset. The proposed approach was evaluated on the relation extraction task of the 2010 i2b2/VA challenge, which has been the reference for nearly every competing system working on medical relation extraction. The task's focus was on eight relation categories, as it can be seen in Table 1. The eight relationships can be further classified into three sub-groups of the treatment-problem (TrIP, TrWP, TrCP, TrAP, TrNAP), test-problem (TeRP, TeCP) and problem-problem (PIP) variety. The vast majority of training examples that can be found in the dataset belongs to the TrAP, PIP and TeRP relations, with 885, 755 and 992 examples respectively. This number

amounts to 84.39% of the dataset examples, which is a problem in itself as the remaining 15.61% that represents the five less populated classes is not enough to effectively feed the training procedure of the classifier in order to produce acceptable results. This fact alone renders the presence of a pattern-based module imperative, which not only rectifies the problem of the sub-populated classes, but also aids in the amelioration of the final results in their entirety.

The original dataset consisted of 394 training reports, 477 test reports, and 877 unannotated reports, while currently, the dataset is only partially available for research, due to IRB limitations, with 170 training and 256 test reports, respectively.

Experimental setup. The LibSVM [24] wrapper class contained in the Weka machine learning software was used to train the linear SVM models of the machine learning module. The main SVM parameters C and gamma, received values of 1 and 0, respectively. In the training procedure one binary classifier (mono-class) was trained for each relation type. For weight assignment, two different strategies were tested. In the first strategy (proposed in [2]), the weight values are directly analogous to the frequency of each relation type in the training set examples. The second strategy is the one we propose for our hybrid approach, based on the RF OOB error estimate. The RF classification model was trained using the scikit-learn python library. The number of trees in the trained RF model was set to 1000. Finally, for the evaluation of the performance of all configurations the micro-averaged values for the precision, recall and F-score measures were computed.

Table 1. Details of the dataset.

	Relation Type	Examples
TrIP	Treatment improves medical problem relations.	51
TrWP	Treatment worsens medical problem relations.	24
TrCP	Treatment causes medical problem relations.	184
TrAP	Treatment is administered for medical problem relations.	885
TrNAP	Treatment is not administered because of medical problem relations.	62
PIP	Medical problem indicates medical problem relations.	755
TeRP	Test reveals medical problem relations.	992
TeCP	Test conducted to investigate medical problem relations.	166

5 Experimental results

The test set results from the experiments conducted in this study are compared in Table 2 with state-of-the-art systems. Rows 2 and 3 of Table 2 contain the results from our system and from the one we use as a baseline approach. It should be noted that all experiments for these two hybrid systems were conducted with the use of our own patterns, as it is not possible to recreate the exact patterns used in [2]. We observe a 2.6% relative improvement (in terms of micro-averaged F-score) in the performance of our system, when compared to the baseline system. This improvement is satisfactory, considering that only the weighting strategy changes are taken into account. No reliable comparison can be made on a pattern level, until the two systems are compared on the same dataset. We would like to note that in general, the finite state automata of our pattern-based module were specific in their performance, with a high micro-averaged precision value (0.772) and a low micro-averaged recall value (0.292). In row 4 of Table 2, [13] trained a convolutional neural network on the exact same limited I2b2 dataset that we also used in our experiments. Rows 5-8 of Table 2 present the performance and type of the relation extraction systems that scored the highest in the I2b2/VA challenge (they used the full dataset, so the machine learning part was trained with more data). We notice that our proposed system outperforms all considered state-of-the-art approaches, to a lesser or greater extent. Most notably, there is an approximate 7% relative improvement in our system’s performance, compared to the best I2b2 hybrid system [20].

Furthermore, Table 3 presents the added value that the integration of the pattern-based module brings to our hybrid system, compared to the use of the machine learning module only. We notice an overall improvement in the F-score values for the different relation types of the dataset. The biggest gains are observed in the TrNAP and TeCP relation types, with a relative improvement of 320.6% and 133.6%, respectively. If we specifically observe the results for the relation types, for which there are fewer training examples available in the dataset (compared to the TrAP, PIP and TeRP relations), namely TrIP, TrWP, TrCP, TrNAP and TeCP, it is evident that there are noticeable improvements from the integration of the pattern-based module, not only in relative terms, but also in terms of absolute values. For instance, for the TrIP relation we have a 3.9% (absolute) improvement, for the TrWP relation the F-score of the hybrid system is computed only on the performance of the pattern-based module and for the TrCP relation, there is a 6% (absolute) improvement. Hence, it can be said that the aforementioned performance improvements warrant the manual effort needed for the construction of our hybrid system’s pattern-based module. Finally, we used the nonparametric Wilcoxon signed ranks test for comparing the 8 F-score values in Table 3 from the supervised (machine learning) module with the corresponding values from the hybrid system. The p-value is 0.008, showing that there is evidence for statistically significant difference between the performance of the machine learning module and that of the hybrid system.

Table 2. Performance evaluation of the proposed hybrid system vs. the baseline system and state-of-the-art approaches.

System	Approach	F-score
Our method	Hybrid	0.758
Abacha & Zweigenbaum [2]	Hybrid	0.739
Sahu et al. [13]	Semi-supervised	0.712
Roberts et al. [25]	Supervised	0.737
DeBruijn et al. [26]	Semi-supervised	0.731
Grouin et al. [20]	Hybrid	0.709
Patrick et al. [27]	Supervised	0.702

Table 3. Performance difference (in terms of F-score) from the integration of the pattern-based module into the proposed system.

Relation type	Supervised	Hybrid	Relative difference
TrIP	0.240	0.279	+16.2%
TrWP	0.0	0.275	N/A
TrCP	0.456	0.516	+13.2%
TrAP	0.749	0.782	+4.4%
TrNAP	0.068	0.286	+320.6%
PIP	0.792	0.823	+3.9%
TeRP	0.817	0.829	+1.5%
TeCP	0.125	0.292	+133.6%

6 Conclusions and future work

In this study, we have proposed a novel medical concept relation extraction framework by extending [2] with the use of a more sophisticated pattern-constructing method and a weighting strategy, which leverages an inherent operational feature of the RF algorithm. Based on experiments conducted on a well-known dataset for relation extraction, we have demonstrated that our methodology outperforms a number of state-of-the-art approaches. It should be noted that in [2] the evaluation is conducted on the MEDLINE 2001 corpus and the patterns of the corresponding module are constructed in a different way. In the future, we plan to fully compare our approach with the latter on the MEDLINE 2001 corpus, as well as investigate the use of alternative weighting strategies for our framework.

Acknowledgments. This work was supported by the project KRISTINA (H2020-645012), funded by the European Commission. Deidentified clinical records used in this research were provided by the i2b2 National Center for Biomedical Computing funded by U54LM008748 and were originally prepared for the Shared Tasks for Challenges in NLP for Clinical Data organized by Dr. Ozlem Uzuner, i2b2 and SUNY.

References

1. Frunza, O., & Inkpen, D.: Extracting relations between diseases, treatments, and tests from clinical data. In Canadian Conference on Artificial Intelligence, pp. 140-145, Springer Berlin Heidelberg (2011, May).
2. Ben Abacha, A., & Zweigenbaum, P.: A hybrid approach for the extraction of semantic relations from medline abstracts. In International Conference on Intelligent Text Processing and Computational Linguistics, pages 139–150, Springer (2011).
3. Ben Abacha, A., & Zweigenbaum, P.: Means: A medical question-answering system combining nlp techniques and semantic web technologies. *Information Processing & Management*, 51(5):570–594 (2015).
4. Breiman, L.: Random forests. *Machine learning*, 45(1):5–32 (2001).
5. Uzuner, Ö., South, B.R., Shen, S., & DuVall, S.L.: 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 18(5):552–556 (2011).
6. Friedman C., Kra, P., Yu, H., Krauthammer, M., & Rzhetsky, A.: Genies: a natural-language processing system for the extraction of molecular pathways from journal articles. *Bioinformatics*, 17(suppl 1):S74–S82 (2001).
7. Feldman, R., Regev, Y., Finkelstein-Landau M., Hurvitz, E., & Kogan, B.: Mining biomedical literature using information extraction. *Current Drug Discovery*, 2(10):19–23 (2002).
8. Rosario, B., & Hearst, M.A.: Classifying semantic relations in bioscience texts. In Proceedings of the 42nd annual meeting on association for computational linguistics (p. 430). Association for Computational Linguistics (2004, July).
9. Bundschuh, M., Dejori, M., Stetter, M., Tresp, V., & Kriegel, H.P.: Extraction of semantic biomedical relations from text using conditional random fields. *BMC bioinformatics*, 9(1), p.1 (2008).
10. Li, J., Zhang, Z., Li, X., & Chen, H.: Kernel-based learning for biomedical relation extraction. *Journal of the American Society for Information Science and Technology*, 59(5), pp.756-769 (2008).
11. Muzaffar, A.W., Azam, F., & Qamar, U.: A Relation Extraction Framework for Biomedical Text Using Hybrid Feature Set. *Computational and mathematical methods in medicine* (2015).
12. Luo, Y., Uzuner, Ö., & Szolovits, P.: Bridging semantics and syntax with graph algorithms state-of-the-art of extracting biomedical relations. *Briefings in bioinformatics* (2016).
13. Sahu, S.K., Anand, A., Oruganty, K., & Gattu, M.: Relation extraction from clinical texts using domain invariant convolutional neural network. *arXiv preprint arXiv:1606.09370* (2016).
14. Tripoliti, E.E., Fotiadis, D.I., & Manis, G.: Automated diagnosis of diseases based on classification: dynamic determination of the number of trees in random forests algorithm. *IEEE transactions on information technology in biomedicine*, 16(4), pp.615-622 (2012).
15. Gokgoz, E., & Subasi, A.: Comparison of decision tree algorithms for EMG signal classification using DWT. *Biomedical Signal Processing and Control*, 18, pp.138-144 (2015).
16. Steyrl, D., Scherer, R., Faller, J., & Müller-Putz, G.R.: Random forests in non-invasive sensorimotor rhythm brain-computer interfaces: a practical and convenient non-linear classifier. *Biomedical Engineering/Biomedizinische Technik*, 61(1), pp.77-86 (2016).
17. Liparas, D., HaCohen-Kerner, Y., Mourtzidou, A., Vrochidis, S., & Kompatsiaris, I.: November. News articles classification using Random Forests and weighted multimodal features. In *Information Retrieval Facility Conference* (pp. 63-75). Springer (2014).

18. Vapnik, V.N.: *The Nature of Statistical Learning Theory* (1995).
19. Rink, B., Harabagiu, S., & Roberts, K.: Automatic extraction of relations between medical concepts in clinical texts. *Journal of the American Medical Informatics Association*, 18(5), pp.594-600 (2011).
20. Grouin, C., Abacha, A.B., Bernhard, D., Cartoni, B., Deleger, L., Grau, B., Ligozat, A.L., Minard, A.L., Rosset, S., & Zweigenbaum, P.: CARAMBA: concept, assertion, and relation annotation using machine-learning based approaches. In *i2b2 Medication Extraction Challenge Workshop* (2010, November).
21. Paumier, S., & Nagel, J.S.: *UNITEX 3.1BETA. User Manual* (2013).
22. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., & McClosky, D.: The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*, pp. 55-60 (2014, June).
23. Lindberg, D.A., Humphreys, B.L., & McCray, A.T.: The unified medical language system. *IMIA Yearbook*, pp.41-51 (1993).
24. Chang, C.C., & Lin, C.J.: LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), p.27 (2011).
25. Roberts, K., Rink, B., & Harabagiu, S.: Extraction of medical concepts, assertions, and relations from discharge summaries for the fourth i2b2/VA shared task. In *Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data*. Boston, MA, USA: i2b2 (2010).
26. de Bruijn, B., Cherry, C., Kiritchenko, S., Martin, J., & Zhu, X.: NRC at i2b2: one challenge, three practical tasks, nine statistical systems, hundreds of clinical records, millions of useful features. In *Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data*. Boston, MA, USA: i2b2 (2010).
27. Patrick, J.D., Nguyen, D.H.M., & Wang, Y.: I2b2 challenges in clinical natural language processing 2010. In *Proceedings of the 2010 i2b2/VA Workshop on Challenges in Natural Language Processing for Clinical Data*. Boston, MA, USA: i2b2 (2010).