Relationship Extraction based on Category of Medical Concepts from Lexical Contexts

Anupam Mondal **Dipankar Das** **Sivaji Bandyopadhyay**
Department of Computer Science and Engineering
Jadavpur University, Kolkata, India
link.anupam@gmail.com, dipankar.dipnil2005@gmail.com,
sivaji_cse_ju@yahoo.com

Abstract

Medical information extraction is an emerging task in healthcare services aim to acquire crucial information of the concepts like *diseases, symptoms, and drugs* and also to identify their *relations* from corpora. In the present article, we have proposed a relationship extraction system based on such categories of medical concepts. We have employed rule-based as well as Support Vector Machine (SVM) based feature-oriented approach along with a domain-specific lexicon viz WordNet of Medical Event (WME 2.0). The lexicon assists in recognizing medical concepts and their related features like Parts-Of-Speech (POS), categories, and Similar Sentiment Words (SSW). We have opted only four fundamental categories *diseases, drugs, symptoms, and human_anatomy* of medical concepts as provided in WME lexicon. Such categories play a crucial role in identifying eight types of different semantic relations viz. *drug-drug, disease-drug, and human_anatomy-symptom* from the medical context. Thereafter, we have validated both rules and features-oriented approaches and offers an average F-Measures of 0.79 and 0.86 individually.

1 Introduction

The availability of medical documents such as reports, discharge summaries, and prescriptions and their related information are growing quickly. In order to extract critical and crucial information, the researchers have applied various statistical and ontology-based approaches with well-known machine learning classifiers (Mondal et al., 2016b; Uzuner et al., 2011). The extracted informations are medical concepts (terms), categories (classes), and their relations, which assist the experts such as doctors and other medical practitioners as well as the non-experts as patients in understanding the problems (e.g. *diseases, symptoms*) and their related remedies (e.g. *drugs*).

The medical concepts are presented by the key terms like words or phrases of the corpus whereas the category refers to the fundamental classes of medical concepts such as diseases and symptoms. The assigned categories of medical concepts and their in-between relations help to build a medical annotation system. Besides, each sentence of the corpus is presented as a medical context in this paper. For an example, "***abdominal_pain***" denotes the medical concept and its category is denoted by "*symptom*" in the following medical context.

"***Abdominal_pain*** *is a* ***sign*** *of early* ***pregnancy***.".

In order to design our category based relationship extraction system, we observed the following major challenges:

A. The first challenge was how to identify the medical concepts and their textual spans from unstructured or semi-structured medical corpora. To address this challenge, we have used WordNet of Medical Events (WME), a domain-specific lexicon (Mondal et al., 2016a, 2015). The lexicon provides a good coverage while extracting medical concepts from our experimental dataset viz. SemEval-2015 Task-6[1] and MedicineNet[2].

B. The second challenge was how to decide the set of categories for the medical concepts and assign them. To overcome the first sub-challenge, we adopted the help of a group of medical practitioners and to cope-up with the second sub-

[1] http://alt.qcri.org/semeval2015/task6/

[2] http://www.medicinenet.com/script/main/hp.asp

S Bandyopadhyay, D S Sharma and R Sangal. Proc. of the 14th Intl. Conference on Natural Language Processing, pages 212–219,
Kolkata, India. December 2017.

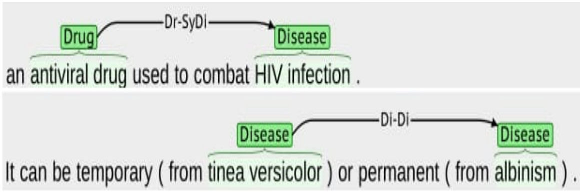

Figure 1: An example of extracted relations using proposed system.

challenge, we used the categorization system of medical concepts (Mondal et al., 2017, 2016a). The categorization system assists in assigning one of the four medical categories such as *diseases, symptoms, drugs, and human_anatomy* to the concepts.

C. The third and final challenge of the present work was how to identify the relations between a pair of medical concepts in a context and evaluate the relations. To address this issue, we have built a rule-based and a feature-based relationship extraction systems, which help to predict the type of relations between a pair of medical concepts. Table 1 shows the proposed eight types of relations with illustrative examples. Besides, we have manually prepared a labeled dataset, which contains 2000 medical contexts and their tagged medical concepts, categories, and relations as shown in Figure 1.

In the present research, our primary motivation was to build an annotation system which helps to assign all four types of medical categories such as *diseases, symptoms, drugs, and human_anatomy* and their related relations in a context. According to the best of our knowledge, we are unable to find any medical corpora which contain all the mentioned categories at a time. Afterwards, we have discussed the contribution of the paper as follows,

I. A labeled dataset preparation by a group of medical practitioners. The dataset has been labeled with medical concepts and their categories and proposed eight types of category-based relations in a context. We have acquired the dataset from SemEval-2015 Task-6 and MedicineNet resources which contain around 2000 number of medical contexts. The dataset helps to design and validate the relationship extraction system.

II. Relationship extraction plays a key role in identifying the semantic information from the corpus. To extract these relations, we have proposed a linguistic rule-based (Abacha and Zweigenbaum, 2011a; Hearst, 1992) and a feature-oriented machine learning (Rink et al., 2011; Zhu et al., 2009) approach. The rule-based patterns help to identify the specific relations from the dataset, whereas machine learning approach assists in extracting generalize relations with promising accuracy. For an example, the following medical context is able to extract *disease - symptom* (illustration, inflammation symptom for the adnexitis disease) and *symptom - human_anatomy* (illustration, inflammation affect the uterus) relations.

*"The **adnexitis**_disease characterizes **inflammation**_symptom of attachments of the **uterus**_human_anatomy."*

The proposed relation extraction system assists in understanding the subjective information of the corpus. Besides, these systems guide to build various applications namely, annotation and recommendation system in healthcare services.

The overall structure of the paper is as follows. Section 2 presents the related work carried out in this field. Section 3 describes the dataset preparation and brat representation technique. Section 4 and Section 5 present the proposed relation extraction system and its evaluation approach. Finally, Section 6 presents the concluding remarks related to our study.

2 Related Work

2.1 Medical Ontologies and Lexicons

Biomedical information extraction research is challenging due to the availability of a large number of daily produced unstructured and semi-structured medical corpus. To represent the structured corpus and extracting the subjective and conceptual information from the corpus, a domain-specific lexicon is essential (Borthwick et al., 1998). To this end, the standard vocabularies and ontologies such as UMLS (Unified Medical Language System) and SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms), and lexicons like MEN (Medical WordNet) and WME (WordNet of Medical Event) have used by the researchers (Smith and Fellbaum, 2004; Kilgarriff and Fellbaum, 2000; Mondal et al., 2016a).

2.2 Medical Category and Relation Extraction

These ontologies and lexicons assist in extracting the relevant information from the corpus such as medical concept categories and relations between medical concepts.

Relation	Explanation and Example
Dr-SyDi	A drug how helps to improve or cure or side effects the diseases or symptoms.
	Warfarin *is also used to reduce the risk of* ***clots*** *causing* ***strokes*** *or* ***heart_attacks***.
Ha-SyDi	A disease or symptom which effects a part of the body.
	A ***painful_inflammation*** *of the* ***big_toe*** *and* ***foot***.
Di-Sy	The symptoms which reflect a disease.
	Anal_fissures *typically cause* ***pain*** *and* ***bleeding*** *with* ***bowel_movements***.
Dr-Dr	How the drugs are related each other.
	An ***oral_lipid-lowering_medicine*** *(* ***trade_name_Zocor*** *) administered to reduce blood cholesterol levels; recommended after heart attacks.*
Sy-Sy	How the symptoms are related each other.
	A ***rhythmic_tightening*** *in* ***labor_of_the_upper_uterine_musculature*** *that contracts the size of the uterus and pushes the fetus toward the birth canal.*
Di-Di	How the diseases are related each other.
	Kaposi's_sarcoma *is a form of* ***skin_cancer*** *that can involve internal organs.*
MMT-SyDi	The medical terms such as process and chemical components etc. how helps to refer the diseases or symptoms.
	When you swallow or inhale these highly ***toxic_products****, you can experience* ***life-threatening_symptoms***.
Ha-Ha	How various body parts are related or effected each other under a situation.
	Nodding movement of the ***head*** *or* ***body***.
Note: Di− > Disease, Dr− > Drug, Sy− > Symptom, Ha− > Human_anatomy, and MMT− > Miscellaneous Medical Term	

Table 1: An illustration of the proposed eight types of relations for the medical context.

Eklund (Eklund, 2011) developed an annotation system to extract the relations as *diseases* for *treatments* from the scientific medical corpus. Yao, et al. (Yao et al., 2010) extracted category relations as *cures, prevents, and side effects*, which describe the distinctive nature for the biomedical text (medical papers) (Abacha and Zweigenbaum, 2011b; Frunza and Inkpen, 2010). Franzen et al. (Franzén et al., 2002) have annotated Yapex corpus with 200 medical abstracts to extract the category as *proteins*. These ontologies are fundamentally looking for extracting *protein-protein* interaction and *disease-treatment* relations from corpora under a BioText project (Rosario and Hearst, 2005).

Khoo et al. (Khoo et al., 2000) developed a causal relations extraction system from abstracts of biomedical articles by aligning manually constructed graph patterns with syntactic dependency trees. Lee et al. (Lee et al., 2003) applied UMLS to identify semantic relations between medical entities. Their first method was able to extract 68% of the semantic relations in their test corpus but if many relations were possible between the relation arguments no disambiguation performed. Their second method (Lee et al., 2004) targeted the precise extraction of "treatment" relations between drugs and diseases. Manually written linguistic patterns were constructed from medical abstracts talking about cancer. Their system reached 84% recall but an overall 48.14% precision. Embarek and Ferret (Embarek and Ferret, 2008) developed an approach to extract four kinds of relations (*Detect, Treat, Sign, and Cure*) between five kinds of medical entities. The employed patterns were constructed automatically using an alignment algorithm which maps sentence parts using an edit distance (defined between two sentences) and different word-level clues.

3 Dataset Preparation

The present section describes how we employed a domain-specific lexicon namely WME 2.0 and prepared an annotated dataset for relation extraction system. Also, discussed the brat environment to visualize the tagged concepts, categories and relations of our medical corpora.

Evaluation Data: We initially acquired corpora from SemEval-2015 Task-6 and MedicineNet resources. Thereafter, the corpora have been converted into medical contexts in the form of single sentences according to the presence of medical concepts for our experiment. Table 2 shows the distributions of medical concepts and contexts from SemEval, MedicineNet as well as WME 2.0 resources.

We randomly collected 2000 medical contexts from the acquired corpora and manually labeled

	SemEval-2015 Task-6	MedicineNet	WME 2.0
Medical concepts	9786	9834	10186
Contexts (medical + non-medical)	10985	9076	-
Medical contexts	6774	7042	-

Table 2: A statistical distribution of unique number of the medical concepts and contexts from various resources.

by a group of medical practitioners in the brat environment. Table 3 shows the distributions of manually labeled category-based relations. In order to label the corpus, the medical practitioners have used WME 2.0 lexicon, which assists in understanding medical concepts and their categories based on their glosses and semantics.

Relation	Manually labeled
All relation	2071
Dr-SyDi	52
Ha-SyDi	198
Di-Sy	312
Dr-Dr	15
Sy-Sy	132
Di-Di	282
MMT-SyDi	927
Ha-Ha	153

Table 3: A statistics of manually labeled various relations

WME Lexicon: In healthcare, a lexicon from the medical domain is demanding to identify the conceptual information such as category or sentiment from the corpus (Cambria, 2016). To this end, we borrow the knowledge from WordNet of Medical Event (WME 2.0), a domain-specific lexicon (Mondal et al., 2016a, 2015).

However, the current version of WME namely WME 2.0 was enhanced with more sentiment and semantic features for 10186 number of medical concepts (Mondal et al., 2017, 2016a). WME 2.0 was added with affinity score, gravity score, Similar Sentiment Words (SSW), and category feature along with the existing features of WME 1.0, e.g., gloss (descriptive explanation), Parts-Of-Speech (POS), polarity score, and sentiment.

The conventional WordNet [3], a preprocessed medical dictionary, and SenticNet [4] were applied to prepare our present resource for extracting semantic relations of concepts. Affinity score indicates what extend two concepts are close to each other by measuring the number of common sentiment words (SSW) appeared for a pair of concepts within the range of 0 to 1. On the other hand, gravity score identifies sentiment-oriented relevance between medical concepts and their various glosses (descriptive explanations) and ranges from -1 to 1. While -1 suggests no relation and 1 indicates strong relations either positive or negative, which helps to identify a proper gloss for concepts. Besides, the assigned categories such as *diseases, drugs, treatments, human_anatomy,* and *MMT* assist in extracting the subjective information of the concepts.

For example, WME 2.0 lexicon presents the properties of a concept say ***amnesia*** as of category (*disease*), POS (*noun*), gloss (*"loss of memory sometimes including the memory of personal identity due to brain injury, shock, fatigue, repression, or illness or sometimes induced by anesthesia."*), SSW (*memory_loss, blackout, fugue, stupor*), polarity score (*-0.375*), affinity score (*0.429*), gravity score (*0.170*), and sentiment (*negative*).

Brat Annotation: We have used an annotation tool namely brat to manually label the relations between a pair of medical concepts within their contexts. The tool helps to easily label the contexts, which generate an annotation (.ann) file for each input text (.txt) file. The .ann file has labeled medical concepts along with their IDs (Ti), categories (Disease, Drugs etc.), textual spans and concepts and relations with IDs, categories and arguments. Thereafter, we have written a python script to convert the manually tagged annotation file into the format according to our proposed system output. The script assists in evaluating the proposed extracted relations.

For example, the following medical context denotes an annotated output as

"T1 Disease 1 39 Giant cell interstitial pneumonia (GIP)"

"T2 Disease 59 77 pulmonary_fibrosis"

"R1 Di-Di Arg1:T1 Arg2:T2".

"***Giant_cell_interstitial_pneumonia_(GIP)****_disease is a rare form of* **pulmonary_fibrosis***_disease.*"

4 Relationship Extraction

Biomedical texts are primarily rich with subjective information such as *problems and treatments* and they are represented as medical concepts, category and their relations in case of ours. Recognition of important relations between medical concepts is a challenging task due to lack of involvement of domain-experts. Thus, to overcome the challenge, we have proposed a relation extraction system by utilizing the categories of medical concepts. To

[3] https://wordnet.princeton.edu/
[4] http://sentic.net/

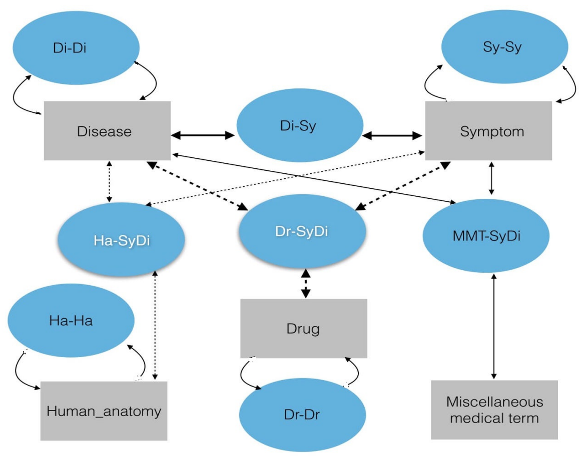

Figure 2: A flow diagram of different types of relations in medical context.

develop the category-based relation extraction system, we have considered the following hypothesis and proposed methodology viz. rule-based and feature-oriented approach.

4.1 Hypothesis:

In the present research, we have considered four primary categories of medical concepts *diseases, drugs, human_anatomy, and symptoms* and a combined category *MMT* that refers to unrecognized categories of medical concepts. To identify the relationship between these categories, we have adopted eight types of relations after close observations done by experts. Figure 2 shows the overall presentation i.e how we proposed eight relations based on various concepts and their categories in a context. Finally, we have classified these relations into two major groups as combined and direct. The combined relations are *Ha-SyDi, Dr-SyDi*, and *MMT-SyDi* and rest of the five relations are presented as direct relation as mentioned in Table 1.

We have also observed that two of the categories (e.g., Disease and Symptom) are very close to each other in their context level appearances and therefore we merged them to make a single category (e.g., SyDi) instead of making the individual pair of relation with other categories.

4.2 Proposed Methodology:

Rule-based Approach In case of relation extraction, rule-based approach adopts various linguistic textual patterns between the pair of medical concepts. To identify these patterns, we have collected the promising pairs of the medical category that are semantically or subjectively related each other as shown in Figure 2. The consecutive appearance of concepts has not been taken into account in case of identifying rules because such concepts are ambiguous in nature and conflicting in their medical senses.

These identified patterns converted to Static Surface Patterns (SSP) using various regular expressions and are labeled by our proposed relations. An example, the linguistic textual pattern ”*<Drug> used to combat <Disease>*” is converted to static surface pattern as ”*<Drug>(.*?)<Disease>*” with *Dr-SyDi* relation label.

The linguistic patterns help to reduce the manual effort and enhance the list of patterns with new relations where SSP assists in designing an automated relationship extraction system. Table 4 presents the number of extracted SSPs for eight relations with the specific example.

Therefore, the following algorithm has been applied to extract the category-based relations between the pair of medical concepts in a context. The output of the proposed system is shown in Figure 3.

Algorithm:

1. Identify the category of annotated medical concepts in a context and present them as C = {MC1,MC2,...,MCn}, where MCn is nth identified medical concept with category in a context.

2. Compare the Static Surface Patterns (SSP) with the pair of medical concepts from C.

2.1. If SSP matches with the pair of concepts in C, then assigned the corresponding relation.

2.2. Else look for the next pair of concepts in C as mentioned in Step 2.

3. If not found any relation in C then label ”*No relation*” and move to next context (C).

4. Else combine the assigned relations with labeled concepts for the context (C) and move to next context (C).

Feature-oriented Approach Relation extraction is presented as a multi-label classification problem due to the presence of various types of semantic relations between medical concepts. To address this problem, we have considered feature-oriented machine learning approach over rule-based approach. The concerned features are category, intermediate word sequence, POS, and SSW of the pair of medical concepts in a context for our experiment.

Relation	Patterns	Example
Dr-SyDi	12	*.... <Warfarin/Drug> is also used to reduce the risk of <clots/Symptom> ....*
Ha-SyDi	14	*.... <coronary artery/Human_anatomy> by a <thrombus/Disease> ....*
Di-Sy	24	*.... <halitosis/Disease> , is an <unpleasant odor/Symptom>....*
Dr-Dr	6	*.... <oral lipid-lowering medicine/Drug> trade name <Zocor/Drug> ....*
Sy-Sy	10	*.... <Frozen shoulder/Symptom> , also known as <adhesive capsulitis/Symptom> ....*
Di-Di	18	*.... <Kaposi's sarcoma/Disease> is a form of <skin cancer/Disease> ....*
MMT-SyDi	27	*.... <toxic products/MMT> , you can experience <life-threatening symptoms/Symptom> ....*
Ha-Ha	6	*.... <vagus nerve/Human_anatomy> included , emerge from or enter the <skull/Human_anatomy> ....*

Table 4: A statistics and examples of relation patterns in the contexts.

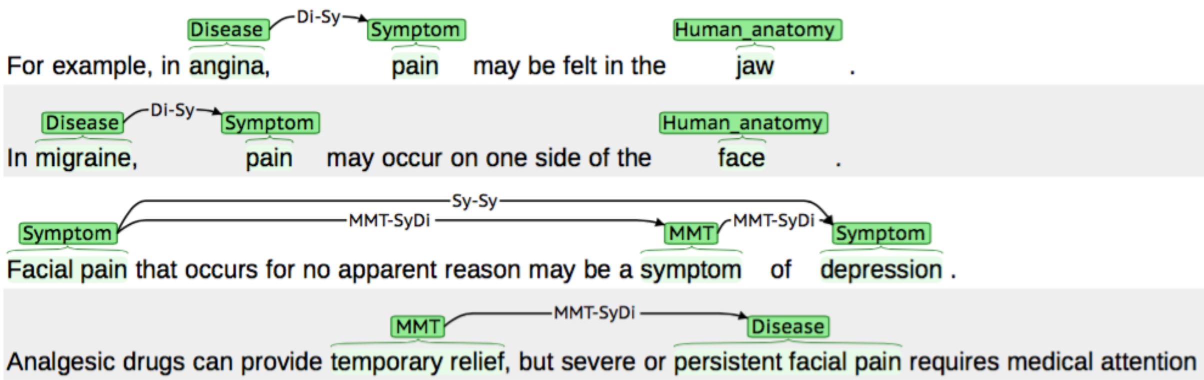

Figure 3: Output of the extracted relation using rule-based approach.

To identify the mentioned features, we have also employed WordNet and WME 2.0 lexicons. These lexicons help to assign the category, POS, and Similar Sentiment Words (SSW) for the medical concepts. Besides, we have written a python script to recognize the intermediate word sequence between the pair of the concepts.

For example, the following medical context identifies the features say 1. annotated medical concepts (”***degenerative_brain_disorder***” and ”***dementia***”) 2. POS labels (noun and verb) 3. intermediate word sequence (”(.*) that leads to (.*)”) 4. categories of medical concepts (*disease* and *disease*), and 5. SSW (”***Alzheimer's disease***, ***Huntington's disease***, *and* ***Parkinson's disease***” and ”***mental_illness***, ***madness***, *and* ***insanity***”), respectively.

”***Degenerative_brain_disorder*** *that leads to* ***dementia***.”

Figure 4 illustrates the steps to extract the relations between a pair of medical concepts using proposed feature-oriented approach along with machine learning classifier. Besides, we have extracted these features from the evaluation dataset and processed through the linear Support Vector Machine (SVM) classifier to predict the relations. The following section discusses the evaluation process for both of the proposed approaches.

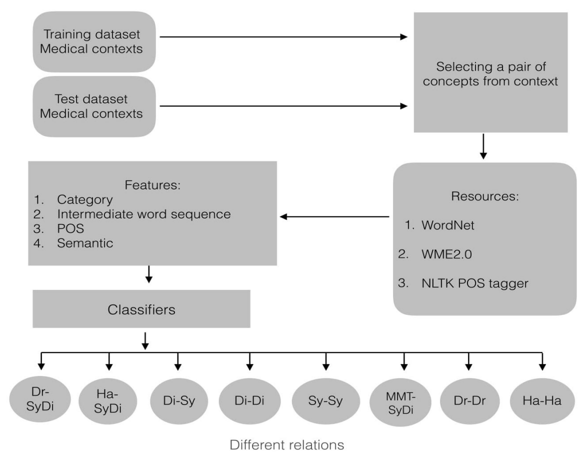

Figure 4: A flow diagram of the feature-oriented relationship extraction system.

5 Evaluation

We have used the state-of-the-art evaluation metrics such as precision, recall, and F-Measure [5] to validate the proposed relation extraction approaches.

Rule-based Relation Extraction In order to reduce the ambiguity of the extracted relations,

[5]https://en.wikipedia.org/wiki/Precision_and_recall

we have employed the same annotated medical concepts and their categories in both rule-based as well as feature-oriented approaches. Table 5 presents the distribution of manual and system tagged category-based relations and their related precision, recall, and F-measure.

Relation	Manually labeled	Extracted/Correct/Incorrect	Precision	Recall	F-Measure
All relations	2071	2681 / 1881 / 780	0.70	0.91	0.79
Dr-SyDi	52	102 / 46 / 56	0.45	0.88	0.59
Ha-SyDi	198	233 / 178 / 35	0.76	0.90	0.82
Di-Sy	312	386 / 282 / 104	0.73	0.90	0.81
Dr-Dr	15	22 / 13 / 9	0.59	0.87	0.70
Sy-Sy	132	193 / 115 / 78	0.60	0.87	0.71
Di-Di	282	341 / 254 / 87	0.74	0.90	0.81
MMT-SyDi	927	1227 / 871 / 356	0.71	0.94	0.81
Ha-Ha	153	177 / 122 / 55	0.69	0.80	0.74

Table 5: A statistics of various relation identification between the pair of medical concepts in context using rule-based approach.

Feature-oriented Relationship Extraction Besides, to validate the feature-oriented system, we additionally prepared a test dataset along with our built-in evaluation dataset. The test dataset contains rest of 11816 medical contexts as referred in Table 2.

Thereafter, the features have been extracted from the evaluation dataset and processed through linear Support Vector Machine (SVM) classifier to learn the proposed relation extraction model. Hence, the test dataset has been applied to the learned model for predicting and validating the extracted relations. Table 6 summarizes the result as precision, recall, and F-Measure.

Relation	Precision	Recall	F-Measure
All relations	0.92	0.81	0.86
Dr-SyDi	0.90	0.72	0.80
Ha-SyDi	0.91	0.80	0.85
Di-Sy	0.90	0.79	0.84
Dr-Dr	0.88	0.76	0.82
Sy-Sy	0.89	0.78	0.83
Di-Di	0.91	0.80	0.85
MMT-SyDi	0.93	0.82	0.87
Ha-Ha	0.88	0.72	0.79

Table 6: A statistics of various relation identification between the pair of medical concepts in context using feature-oriented approach.

Finally, we have observed that the feature-oriented approach provides a better F-Measure (0.86) over the rule-based approach (0.79) for extracting relations. So, we conclude that both approaches are important to extract the mentioned eight relations from the unstructured corpus.

6 Conclusion and Future Scope

In this article, we have focused on extracting eight types of category-based relations of medical concepts from the context. The relation extraction system facilitates to design various domain-specific applications. We have employed two well-known approaches such as rule-based and feature-oriented. Also, we have manually prepared an evaluation dataset to design and validate the relation extraction system. The rule-based approach helps to understand the semantic knowledge where linguistic features assist in improving the accuracy of the system. Finally, the evaluation section shows the effectiveness of the proposed relation extraction approaches by offering the average F-Measures of 0.79 and 0.86 for the rules and features-oriented techniques, respectively in healthcare. In future, we will try to introduce new relations for the medical concepts to build a relation database, which helps to design a medical recommendation system.

References

Asma Ben Abacha and Pierre Zweigenbaum. 2011a. Automatic extraction of semantic relations between medical entities: a rule based approach. *Journal of biomedical semantics* 2(5):1.

Asma Ben Abacha and Pierre Zweigenbaum. 2011b. A hybrid approach for the extraction of semantic relations from medline abstracts. In *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, pages 139–150.

Andrew Borthwick, John Sterling, Eugene Agichtein, and Ralph Grishman. 1998. Exploiting diverse knowledge sources via maximum entropy in named entity recognition. In *Proc. of the Sixth Workshop on Very Large Corpora*. volume 182.

Erik Cambria. 2016. Affective computing and sentiment analysis. *IEEE Intelligent Systems* 31(2):102–107.

Ann-Marie Eklund. 2011. Relational annotation of scientific medical corpora. In *LOUHI 2011 Third International Workshop on Health Document Text Mining and Information Analysis*. page 27.

Mehdi Embarek and Olivier Ferret. 2008. Learning patterns for building resources about semantic relations in the medical domain. In *LREC*.

Kristofer Franzén, Gunnar Eriksson, Fredrik Olsson, Lars Asker, Per Lidén, and Joakim Cöster. 2002. Protein names and how to find them. *International journal of medical informatics* 67(1):49–61.

Oana Frunza and Diana Inkpen. 2010. Extraction of disease-treatment semantic relations from biomedical sentences. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*. Association for Computational Linguistics, pages 91–98.

Marti A Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2*. Association for Computational Linguistics, pages 539–545.

Christopher SG Khoo, Syin Chan, and Yun Niu. 2000. Extracting causal knowledge from a medical database using graphical patterns. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, pages 336–343.

Adam Kilgarriff and Christiane Fellbaum. 2000. Wordnet: An electronic lexical database.

Chew-Hung Lee, Christopher Khoo, and Jin-Cheon Na. 2004. Automatic identification of treatment relations for medical ontology learning: An exploratory study. *ADVANCES IN KNOWLEDGE ORGANIZATION* 9:245–250.

Chew-Hung Lee, Jin-Cheon Na, and Christopher Khoo. 2003. Ontology learning for medical digital libraries. In *International Conference on Asian Digital Libraries*. Springer, pages 302–305.

Anupam Mondal, Erik Cambria, Dipankar Das, and Sivaji Bandyopadhyay. 2017. Auto-categorization of medical concepts and contexts. *Research in Computing Science* .

Anupam Mondal, Iti Chaturvedi, Dipankar Das, Rajiv Bajpai, and Sivaji Bandyopadhyay. 2015. Lexical resource for medical events: A polarity based approach. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE, pages 1302–1309.

Anupam Mondal, Dipankar Das, Erik Cambria, and Sivaji Bandyopadhyay. 2016a. Wme: Sense, polarity and affinity based concept resource for medical events. *Proceedings of the Eighth Global WordNet Conference* pages 242–246.

Anupam Mondal, Ranjan Satapathy, Dipankar Das, and Sivaji Bandyopadhyay. 2016b. A hybrid approach based sentiment extraction from medical context. In *4th Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2016), IJCAI 2016 Workshop, July 10, Hilton, New York City, USA*.

Bryan Rink, Sanda Harabagiu, and Kirk Roberts. 2011. Automatic extraction of relations between medical concepts in clinical texts. *Journal of the American Medical Informatics Association* 18(5):594–600.

Barbara Rosario and Marti A Hearst. 2005. Multi-way relation classification: application to protein-protein interactions. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pages 732–739.

Barry Smith and Christiane Fellbaum. 2004. Medical wordnet: a new methodology for the construction and validation of information resources for consumer health. In *Proceedings of the 20th international conference on Computational Linguistics*. Association for Computational Linguistics, page 371.

Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall. 2011. 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association* 18(5):552–556.

Lin Yao, Cheng-Jie Sun, Xiao-Long Wang, and Xuan Wang. 2010. Relationship extraction from biomedical literature using maximum entropy based on rich features. In *2010 International Conference on Machine Learning and Cybernetics*. IEEE, volume 6, pages 3358–3361.

Jun Zhu, Zaiqing Nie, Xiaojiang Liu, Bo Zhang, and Ji-Rong Wen. 2009. Statsnowball: a statistical approach to extracting entity relationships. In *Proceedings of the 18th international conference on World wide web*. ACM, pages 101–110.