

Exploiting the Scope of Negations and Heterogeneous Features for Relation Extraction: A Case Study for Drug-Drug Interaction Extraction

Md. Faisal Mahbub Chowdhury^{†‡} and Alberto Lavelli[‡]

[‡]Fondazione Bruno Kessler (FBK-irst), Italy

[†]University of Trento, Italy

fmchowdhury@gmail.com, lavelli@fbk.eu

Abstract

This paper presents an approach that exploits the scope of negation cues for relation extraction (RE) without the need of using any specifically annotated dataset for building a separate negation scope detection classifier. New features are proposed which are used in two different stages. These also include non-target entity specific features. The proposed RE approach outperforms the previous state of the art for drug-drug interaction (DDI) extraction.

1 Introduction

Negation is a linguistic phenomenon where a *negation cue* (e.g. *not*) can alter the meaning of a particular text segment or of a fact. This text segment (or fact) is said to be inside the *scope of that negation (cue)*. In the context of RE, there is not much work that aims to exploit the scope of negations.¹ The only work on RE that we are aware of is Sanchez-Graillet and Poesio (2007) where they used various heuristics to extract negative protein interaction.

Despite the recent interest on automatically detecting the scope of negation² till now there seems to be no empirical evidence supporting its exploitation for the purpose of RE. Even if we could manage to obtain highly accurate automatically detected

¹In the context of event extraction (a closely related task of RE), there have been efforts in BioNLP shared tasks of 2009 and 2011 for (non-mandatory sub-task of) event negation detection (3 participants in 2009; 2 in 2011) (Kim et al., 2009; Kim et al., 2011). The participants approached the sub-task using either pre-defined patterns or some heuristics.

²This task is popularized by various recently held shared tasks (Farkas et al., 2010; Morante and Blanco, 2012).

negation scopes, it is not clear how to feed this information inside the RE approach. Simply considering whether a pair of candidate mentions falls under the scope of a negation cue might not be helpful.

In this paper, we propose that the scope of negations can be exploited at two different levels. Firstly, the system would check whether all the target entity³ mentions inside a sentence along with possible relation clues (or trigger words), if any, fall (directly or indirectly) under the scope of a negation cue. If such a sentence is found, then it should be discarded (i.e. candidate mention pairs⁴ inside that sentence would not be considered). Secondly, for each of the remaining pairs of candidate mentions, the system should exploit features related to the scope of negation (rather than simply adding a feature for negation cue, approach adopted in various RE systems) that can provide indication (if any such evidence exists) that the corresponding relation of interest actually does not hold in that particular context.

In the subsequent sections, we describe our approach. The RE task considered is drug-drug interaction (DDI) extraction. The task has significant importance for public health safety.⁵ We used

³The target entities, for example, for *DDI* extraction and for *EMP-ORG* relation extraction would be $\{DRUG\}$ and $\{PER, GPE, ORG\}$ respectively. Any entity other than the target entities (w.r.t. the particular RE task) belongs to non-target entities.

⁴Candidate mention pairs for RE are taken from target entity mentions.

⁵After the death of pop star Michael Jackson, allegedly due to DDI, it was reported that about 2.2 million people in USA, age 57 to 85, were taking potentially dangerous combinations of drugs (Landau, 2009). An earlier report mentioned that deaths from accidental drug interactions rose 68 percent between 1999 and 2004 (Payne, 2007).

the DDIExtraction-2011 challenge corpus (Segura-Bedmar et al., 2011). The official training and test data of the corpus contain 4,267 and 1,539 sentences, and 2,402 and 755 DDI annotations respectively.

2 Proposed Approach

2.1 Stage 1: Exploiting scope of negation to filter out sentences

We propose a two stage RE approach. In the first stage, our goal is to exploit the scope of negations to reduce the number of candidate mention pairs by discarding sentences. For this purpose, we propose the following features to train a binary classifier:

- *has2TM*: If the sentence has exactly 2 target entity mentions (i.e. drug mentions for DDI extraction).
- *has3OrMoreTM*: Whether the sentence has more than 2 target entity mentions.
- *allTMonRight*: Whether all target entity mentions inside the sentence appear after the negation cue.
- *neitherAllTMonLeftOrRight*: Whether some but not all target entity mentions appear after the negation cue.
- *negCue*: The negation cue itself.
- *immediateGovernor*: The word on which the cue is directly syntactically dependent.
- *nearestVerbGovernor*: The nearest verb in the dependency graph on which the cue is syntactically dependent.
- *isVerbGovernorRoot*: Whether the *nearestVerbGovernor* is root of the dependency graph of the sentence.
- *allTMdependentOnNVG*: Whether all target entity mentions are syntactically dependent (directly/indirectly) on the *nearestVerbGovernor*.
- *allButOneTMdependentOnNVG*: Whether all but one target entity mentions are syntactically dependent on the *nearestVerbGovernor*.
- *although*PrecedeCue*: Whether the syntactic clause containing the negation cue begins with “although / though / despite / in spite”.
- *commaBeforeNextTM*: Whether there is a comma in the text between the negation cue and the next target entity mention after the cue.
- *commaAfterPrevTM*: Whether there is a comma in the text between the previous target entity mention before the negation cue and the cue itself.

- *sentHasBut*: Whether the sentence contains the word “but”.

The objective of the classifier is to decide whether all of the target entity mentions (i.e. drugs) as well as any possible evidence of the relation of interest (for which we assume the immediate and the nearest verb governors of the negation cue would be good candidates) inside the corresponding sentence fall under the scope of a negation cue in such a way that the sentence is unlikely to contain a DDI.

At present, we limit our focus only on the first occurrence of the following negation cues: “no”, “n’t” or “not”.⁶ In the Stage 1, any sentence that contains at least one DDI is considered by the classifier as a positive (training/test) instance. Other sentences are considered as negative instances. We rule out any sentence (i.e. we do not consider as training/test instance for the classifier that filters less informative sentences) during both training and testing if any of the following conditions holds:

- The sentence contains less than two target entity mentions (such sentence would not contain the relation of interest anyway).
- It has any of the following phrases – “not recommended”, “should not be” or “must not be”.⁷
- There is no “no”, “n’t” or “not” in the sentence.
- No target entity mention appears in the sentence after “no”, “n’t” or “not”.

To assess the effectiveness of the proposed Stage 1 classifier, we defined a **baseline** classifier that filters any sentence that contains “no”, “n’t” or “not”.

2.2 Stage 2

Once the sentences which are likely to have no DDI are identified and removed, the next step is to apply a state-of-the-art RE approach on the remaining sentences. In this section, we propose a new hybrid kernel, K_{Hybrid} , for this purpose. It is defined as follows:

$$K_{Hybrid}(R_1, R_2) = K_{HF}(R_1, R_2) + K_{SL}(R_1, R_2) + w * K_{PET}(R_1, R_2)$$

⁶These cues usually occur more frequently and generally have larger negation scope than other negation cues.

⁷These expressions often provide clues that one of the bio-entity mentions negatively influences the level of activity of the other.

Here, K_{HF} stands for a new feature based kernel (proposed in this paper) that uses a heterogeneous set of features. K_{SL} stands for the Shallow Linguistic (SL) kernel proposed by Giuliano et al. (2006). K_{PET} stands for the Path-enclosed Tree (PET) kernel (Moschitti, 2004). w is a multiplicative constant used for the PET kernel. It allows the hybrid kernel to assign more (or less) weight to the information obtained using tree structures depending on the corpus.

The proposed kernel composition is valid according to the closure properties of kernels. We exploit the SVM-Light-TK toolkit (Moschitti, 2006; Joachims, 1999) for kernel computation. In Stage 2, each candidate drug mention pair represents an instance.

2.2.1 Proposed K_{HF} kernel

As mentioned earlier, this proposed kernel uses heterogeneous features. The first version of the heterogeneous feature set (henceforth, $HF v1$) combines features proposed by two previous RE works. The former is Zhou et al. (2005), which uses 51 different features. We select the following 27 of their features for our feature set:

*WBNUL*L, *WBFL*, *WBF*, *WBL*, *WBO*,
BM1F, *BM1L*, *AM2F*, *AM2L*, *#MB*, *#WB*,
*CPHBNUL*L, *CPHBFL*, *CPHBF*, *CPHBL*,
CPHBO, *CPHBM1F*, *CPHBM1L*, *CPHAM2F*,
CPHAM2L, *CPP*, *CPPH*, *ET12SameNP*,
ET12SamePP, *ET12SameVP*, *PTP*, *PTPH*

The latter is the TPWF kernel (Chowdhury and Lavelli, 2012a) from which we use following features:

HasTriggerWord, *Trigger-X*, *DepPattern-i*, *e-walk*, *v-walk*

The TPWF kernel extracts the *HasTriggerWord*, *Trigger-X* and *DepPattern-i* features from a sub-graph called *reduced graph*. We also follow this approach with one minor difference. Unlike Chowdhury and Lavelli (2012a), we look for trigger words in the whole reduced graph instead of using only the root of the sub-graph.

Due to space limitation we refer the readers to the corresponding papers for the description of the above mentioned features and the definition of *reduced graph*.

In addition, $HF v1$ also includes surrounding tokens within the window of $\{-2,+2\}$ for each candidate mention. We are unaware of any available list of trigger words for drug-drug interaction. So, we created such a list.⁸

We extend the heterogeneous feature set by adding features related to the scope of negation (henceforth, $HF v2$). We use a list of 13 negation cues⁹ to search inside the reduced graph of a candidate pair. If the reduced graph contains any of the negation cues or their morphological variants then we add the following features:

- *negCue*: The corresponding negation cue.
- *immediateNegatedWord*: If the word following the negation cue is neither a preposition nor a “be verb”, then that word, otherwise the word after the next word.¹⁰

Furthermore, if the corresponding matched negation cue is either “no”, “n’t” or “not”, then we add additional features related to negation scope:

- *bothEntDependOnImmediateGovernor*: Whether the immediate governor (if any) of the negation cue is also governor of a dependency sub-tree (of the dependency graph of the corresponding sentence) that includes both of the candidate mentions.
- *immediateGovernorIsVerbGovernor*: Whether the immediate governor of the negation cue is a verb.
- *nearestVerbGovernor*: The closest verb governor (i.e. parent or grandparent inside the dependency graph), if any, of the negation cue.

We further extend the heterogeneous feature set by adding features related to relevant non-target entities (with respect to the relation of interest; henceforth, $HF v3$). For the purpose of DDI extraction, we deem the presence of *DISEASE* mentions (which might result as a consequence of a DDI) can provide some clues. So, we use a publicly available state-of-the-art disease NER system called BioEnEx (Chowdhury and Lavelli, 2010) to annotate the DDIExtraction-2011 challenge corpus. For

⁸The RE system developed for this work and the created list of trigger words for DDI can be downloaded from <https://github.com/fmchowdhury/HyREX>.

⁹No, not, neither, without, lack, fail, unable, abrogate, absence, prevent, unlikely, unchanged, rarely.

¹⁰For example, “interested” from “... not interested ...”, and “confused” from “... not to be confused ...”.

each candidate (drug) mention pair, we add the following features in *HF v3*:

- *NTEInsideSentence*: Whether the corresponding sentence contains important non-target entity mention(s) (e.g. disease for DDI).
- *immediateGovernorIsVerbGovernorOfNTEM*: The immediate governor (if any) of the non-target entity mention, only if such governor is also governing a dependency sub-tree that includes both of the target candidate entity mentions.
- *nearestVerbGovernorOfNTEM*: The closest verb governor (if any) of the non-target entity mention, only if it also governs the candidate entity mentions.
- *immediateGovernorIsVerbGovernorOfNTEM*: Whether the immediate governor is a verb.

3 Results and Discussion

We train a linear SVM classifier in Stage 1 and tune the hyper-parameters (by doing 5-fold cross-validation) for obtaining maximum possible recall. In this way we minimize the number of false negatives (i.e. sentences that contain DDIs but are wrongly identified as not having any).

During the cross-validation experiments on the training data, 334 sentences (7.83% of the total sentences) containing at least 2 drug mentions were identified by our proposed classifier (in Section 2.1) as unlikely to have any DDI and hence are candidates for discarding. Only 19 of these sentences were incorrectly identified. When we trained on the training data and tested on the official test data of DDIExtraction-2011 challenge corpus, 121 sentences (7.86% of the total test sentences) were identified by the classifier as candidates for discarding. Only 5 of them were incorrectly identified.

Unlike Stage 1, in Stage 2 where we train the hybrid kernel based RE classifier and use it for RE (i.e. DDI extraction) from the test data, sentences are not the RE training/test instances. Instead, a RE instance corresponds to a candidate mention pair.

All the DDIs (i.e. positive RE instances) of the incorrectly identified sentences in Stage 1 (i.e. the sentences which are incorrectly labelled as not having any DDI and filtered) are automatically considered as false negatives during the calculation of DDI extraction results in Stage 2.

To verify whether our proposed hybrid kernel achieves state-of-the-art results without taking benefits of the output of Stage 1, we did some experiments without discarding any sentence. These experiments are done using Zhou et al. (2005), TPWF kernel, SL kernel, different versions of proposed K_{HF} kernel and K_{Hybrid} kernel. Table 1 shows the results of 5-fold cross-validation experiments (hyper-parameters are tuned for obtaining maximum F-score). As the results show, there is a gain +0.9 points in F-score (mainly due to the boost in recall) after the addition of features related to negation scope. There is also some minor improvement due to the proposed non-target entity specific features.

We also performed (5-fold cross validation) experiments by combining the Stage 1 classifier with each of the Zhou et al. (2005), TPWF kernel, SL kernel, PET kernel, K_{HF} kernel and K_{Hybrid} kernel separately (only the results of K_{Hybrid} are reported in Table 1 due to space limitation). In each case, there were improvements in precision, recall and F-score. The gain in F-score ranged from 1.0 to 1.4 points.

	P / R / F-score
Using SL kernel (Giuliano et al., 2006)	51.3 / 64.7 / 57.3
Using (Zhou et al., 2005)	58.7 / 37.1 / 45.5
Using PET kernel (Moschitti, 2004)	46.8 / 60.2 / 52.7
TPWF (Chowdhury and Lavelli, 2012a)	43.7 / 60.7 / 50.8
Proposed approaches	
Proposed $K_{HF v1}$	53.4 / 51.5 / 52.4
$K_{HF v2}$ (i.e. + neg scope feat.)	53.9 / 52.6 / 53.3 (+0.9)
$K_{HF v3}$ (i.e. + non-target entity feat.)	53.6 / 53.5 / 53.6 (+0.3)
Proposed K_{Hybrid}	56.3 / 68.5 / 61.8
Proposed K_{Hybrid} with Stage 1	57.3 / 69.4 / 62.8 (+1.0)

Table 1: 5-fold cross-validation results on training data.

Table 2 reports the results of the previously published studies that used the same corpus. Our proposed K_{Hybrid} kernel obtains an F-score that is higher than that of the previous state of the art.

When the Stage 1 classifier (based on negation scope features) is exploited before using the K_{Hybrid} kernel, the F-score reaches up to 67.4. This is +1.0 points higher than without exploiting the Stage 1 classifier and +1.7 higher than previous state of

the art. We did separate experiments (also reported in Table 2) to assess the performance improvement when the output of Stage 1 is used to filter sentences from either training or test data only. The results remain the same when only training sentences are filtered; while there are some improvements when only test sentences are filtered. Filtering both training and test sentences provides the larger gain which is statistically significant.

Usually, the number of negative instances in a corpus is much higher than that of the positive instances. In a recent work, Chowdhury and Lavelli (2012b) showed that by removing less informative (negative) instances (henceforth, *LIIs*), not only the skewness in instance distribution could be reduced but it also leads to a better result. The proposed Stage 1 classifier, presented in this work, also reduces skewness in instance distribution. This is because we are only removing those sentences that are unlikely to contain any positive instance. So, in principle, the Stage 1 classifier is focused on removing only negative instances (although the classifier mistakenly discards few positive instances, too).

We wanted to study how the Stage 1 classifier would contribute if we use it on top of the techniques that were proposed in Chowdhury and Lavelli (2012b) to remove LIIs. As Table 2 shows, by using the Stage 1 classifier along with LLI filtering, we could further improve the results (+3.2 points difference in F-score with the previous state of the art).

4 Conclusion

A major flexibility in the proposed approach is that it does not require a separate dataset (which needs to match the genre of the text to be used for RE) annotated with negation scopes. Instead, the proposed Stage 1 classifier uses the RE training data (which do not have negation scope annotations) to self-supervise itself. Various new features have been exploited (both in stages 1 and 2) that can provide strong indications of the scope of negation cues with respect to the relation to be extracted. The only thing needed is the list of possible negation cues (Morante (2010) includes such a comprehensive list).

Our proposed kernel, which has a component that exploits a heterogeneous set of features including negation scope and presence of non-target entities, already obtains better results than previous studies.

	P	R	F-score
(Thomas et al., 2011)	60.5	71.9	65.7
(Chowdhury et al., 2011)	58.6	70.5	64.0
(Chowdhury and Lavelli, 2011)	58.4	70.1	63.7
(Bjorne et al., 2011)	58.0	68.9	63.0
Proposed K_{Hybrid}	60.0	74.3	66.4
K_{Hybrid} + Stage 1 baseline	61.8	68.9	65.1
K_{Hybrid} + proposed Stage 1 (only training sentences are filtered)	60.0	74.2	66.4
K_{Hybrid} + proposed Stage 1 (only test sentences are filtered)	61.4	73.8	67.0
K_{Hybrid} + proposed Stage 1 (both training and test sentences are filtered)	62.1	73.8	67.4 stat. sig.
Proposed K_{Hybrid} + LII filtering	61.1	75.1	67.4 stat. sig.
Proposed K_{Hybrid} + LII filtering + proposed Stage 1	63.5	75.2	68.9 stat. sig.

Table 2: Results obtained on the official test set of the 2011 DDI Extraction challenge. LII filtering refers to the techniques proposed in Chowdhury and Lavelli (2012b) for reducing skewness in RE data distribution. **stat. sig.** indicates that the improvement of F-score, due to usage of Stage 1 classifier, is statistically significant (verified using *Approximate Randomization Procedure* (Noreen, 1989); number of iterations = 1,000, confidence level = 0.01).

The results considerably improve when possible irrelevant sentences from both training and test data are filtered by exploiting features related to the scope of negations.

In future, we would like to exploit the scope of more negation cues, apart from the three cues that are used in this study. We believe our approach would help to improve RE in other genres of text (such as newspaper) as well.

Acknowledgement

This work was carried out in the context of the project “eOnco - Pervasive knowledge and data management in cancer care”.

References

- J Bjorne, A Airola, T Pahikkala, and T Salakoski. 2011. Drug-drug interaction extraction with RLS and SVM classifiers. In *Proceedings of the 1st Challenge task on Drug-Drug Interaction Extraction (DDIExtraction 2011)*, pages 35–42, Huelva, Spain, September.

- MFM Chowdhury and A Lavelli. 2010. Disease mention recognition with specific features. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, pages 83–90, Uppsala, Sweden, July. Association for Computational Linguistics.
- MFM Chowdhury and A Lavelli. 2011. Drug-drug interaction extraction using composite kernels. In *Proceedings of the 1st Challenge task on Drug-Drug Interaction Extraction (DDIExtraction 2011)*, pages 27–33, Huelva, Spain, September.
- MFM Chowdhury and A Lavelli. 2012a. Combining tree structures, flat features and patterns for biomedical relation extraction. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012)*, pages 420–429, Avignon, France, April. Association for Computational Linguistics.
- MFM Chowdhury and A Lavelli. 2012b. Impact of Less Skewed Distributions on Efficiency and Effectiveness of Biomedical Relation Extraction. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012) : Posters*, pages 205–216, Mumbai, India, December.
- MFM Chowdhury, AB Abacha, A Lavelli, and P Zweigenbaum. 2011. Two different machine learning techniques for drug-drug interaction extraction. In *Proceedings of the 1st Challenge task on Drug-Drug Interaction Extraction (DDIExtraction 2011)*, pages 19–26, Huelva, Spain, September.
- R Farkas, V Vincze, G Móra, J Csirik, and G Szarvas. 2010. The CoNLL-2010 shared task: Learning to detect hedges and their scope in natural language text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 1–12, Uppsala, Sweden, July. Association for Computational Linguistics.
- C Giuliano, A Lavelli, and L Romano. 2006. Exploiting shallow linguistic information for relation extraction from biomedical literature. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006)*, pages 401–408.
- T Joachims. 1999. Making large-scale support vector machine learning practical. In *Advances in kernel methods: support vector learning*, pages 169–184. MIT Press, Cambridge, MA, USA.
- JD Kim, T Ohta, S Pyysalo, Y Kano, and J Tsujii. 2009. Overview of BioNLP’09 shared task on event extraction. In *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task*, pages 1–9, Boulder, Colorado, June. Association for Computational Linguistics.
- JD Kim, Y Wang, T Takagi, and A Yonezawa. 2011. Overview of Genia event task in BioNLP shared task 2011. In *Proceedings of BioNLP Shared Task 2011 Workshop*, pages 7–15, Portland, Oregon, USA, June. Association for Computational Linguistics.
- E Landau. 2009. Jackson’s death raises questions about drug interactions [Published in CNN; June 26, 2009]. http://articles.cnn.com/2009-06-26/health/jackson.drug.interaction.caution_1_drug-interactions-heart-rhythms-antidepressants?s=PM:HEALTH.
- R Morante and E Blanco. 2012. *SEM 2012 shared task: Resolving the scope and focus of negation. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 265–274, Montréal, Canada, 7-8 June. Association for Computational Linguistics.
- R Morante. 2010. Descriptive Analysis of Negation Cue in Biomedical Texts. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, Malta.
- A Moschitti. 2004. A study on convolution kernels for shallow semantic parsing. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004)*, Barcelona, Spain.
- A Moschitti. 2006. Making Tree Kernels Practical for Natural Language Learning. In *Proceedings of 11th Conference of the European Chapter of the Association for computational Linguistics (EACL 2006)*, Trento, Italy.
- EW Noreen. 1989. *Computer-Intensive Methods for Testing Hypotheses : An Introduction*. Wiley-Interscience, April.
- JW Payne. 2007. A Dangerous Mix [Published in The Washington Post; February 27, 2007]. <http://www.washingtonpost.com/wp-dyn/content/article/2007/02/23/AR2007022301780.html>.
- O Sanchez-Graillet and M Poesio. 2007. Negation of protein-protein interactions: analysis and extraction. *Bioinformatics*, 23(13):i424–i432.
- I Segura-Bedmar, P Martínez, and CD Pablo-Sánchez. 2011. The 1st DDIExtraction-2011 challenge task: Extraction of Drug-Drug Interactions from biomedical texts. In *Proceedings of the 1st Challenge task on Drug-Drug Interaction Extraction (DDIExtraction 2011)*, pages 1–9, Huelva, Spain, September.
- P Thomas, M Neves, I Solt, D Tikik, and U Leser. 2011. Relation extraction for drug-drug interactions using ensemble learning. In *Proceedings of the 1st Challenge task on Drug-Drug Interaction Extraction*

(*DDIExtraction 2011*), pages 11–18, Huelva, Spain, September.

GD Zhou, J Su, J Zhang, and M Zhang. 2005. Exploring various knowledge in relation extraction. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL 2005)*, pages 427–434, Ann Arbor, Michigan, USA.