

Learning to Identify Treatment Relations in Clinical Text

Cosmin A. Bejan, PhD¹ and Joshua C. Denny, MD, MS^{1,2}

¹Department of Biomedical Informatics, Vanderbilt University, Nashville, TN;

²Department of Medicine, Vanderbilt University, Nashville, TN

Abstract

In clinical notes, physicians commonly describe reasons why certain treatments are given. However, this information is not typically available in a computable form. We describe a supervised learning system that is able to predict whether or not a treatment relation exists between any two medical concepts mentioned in clinical notes. To train our prediction model, we manually annotated 958 treatment relations in sentences selected from 6,864 discharge summaries. The features used to indicate the existence of a treatment relation between two medical concepts consisted of lexical and semantic information associated with the two concepts as well as information derived from the MEDication Indication (MEDI) resource and SemRep. The best F1-measure results of our supervised learning system (84.90) were significantly better than the F1-measure results achieved by SemRep (72.34).

Introduction

Discovering treatment relations from clinical text is a fundamental task in clinical information extraction and has various applications in medical research. For instance, the availability of treatment relations could enable a more comprehensive understanding of a patient's treatment course,[1,2] improve adverse reaction detection,[3] and assess healthcare quality.[4,5] However, such relations are not stored in a structured format in most electronic medical records (EMRs), but rather they are encoded in narrative patient reports. Therefore, natural language processing technologies need to be employed to facilitate the automatic extraction of treatment relations from clinical text.

A treatment relation is a relation holding between two medical concepts in which one of the concepts (e.g., a medication or procedure) is a treatment for the other concept (e.g., a disease). The treatment relation between the concepts highlighted in (1), for instance, is defined by a medication(*levofloxacin*) as a treatment administered for a disease(*pneumonia*). The relation in (2), holding between a procedure(*surgery*) and a disease(*lumbar stenosis*), represents another common category of treatment relation. On the other hand, the relations between the emphasized concepts in (3) and (4) do not constitute treatment relations. As observed in (3), in spite of a medication being prescribed for a specific medical problem, the treatment did not cure or improve the medical condition of the corresponding patient. Similarly in (4), although *lorazepam* can be indicated for *nausea*, the context in which these concepts are described invalidate the existence of a treatment relation. Due to this observation, the only treatment relations in (4) are between *morphine PCA* and *pain*, *IV lorazepam* and *anxiety*, and *IV reglan* and *nausea*. The remaining pair combinations of the concepts from this example constitute non-treatment relations.

- (1) *She switched to [levofloxacin] for treatment of [pneumonia] as evidenced on CT.*
- (2) *He had a previous [surgery] for [lumbar stenosis].*
- (3) *She was treated with [morphine], which did not control her [pain].*
- (4) *She will be discharged with her morphine PCA for pain, scheduled [IV lorazepam] for anxiety, and scheduled IV reglan for [nausea].*

Our approach for identifying treatment relations in clinical text is based on (a) the idea of exploring the contextual information in which medical concepts are described and (b) the idea of using predefined medication-indication pairs. For this purpose, we used MEDication Indication (MEDI),[6] a large database of medication-indication pairs. Another resource we employed in the feature extraction phase of our machine learning framework is SemRep,[7] which uses linguistic rules to identify treatment relations in text. For example, a rule of the form *X for treatment of Y* can be applied to find the treatment relation in (1).

The goal of our study was not only to build a system for accurately extracting treatment relations from clinical notes, but also to use this system for expanding MEDI with new medication-indication pairs.

Related Work

SemRep is a publicly available biomedical information extraction tool which was developed at the U.S. National Library of Medicine (<http://semrep.nlm.nih.gov>) and has been used by a number of investigators. Given a text document, this system analyzes each sentence from the document and identifies multiple types of semantic relations between the concepts described in the sentence. Examples of relations that SemRep is able to extract are DIAGNOSES, CAUSES, LOCATION_OF, ISA, TREATS, PREVENTS, etc. The system relies on MetaMap[8] to extract the Unified Medical Language System (UMLS) concepts from text and on linguistic and semantic rules specific to each relation. Its output for each sentence consists of a list of all the UMLS concepts mentioned in the sentence followed by the semantic relations that exist between these concepts. SemRep has been used in a wide range of applications in biomedical informatics including automatic summarization and literature based discovery.[9,10] While SemRep was primarily designed for processing documents from the biomedical research literature, only a few studies involving this system have been performed on clinical documents. In one of these studies, drug-disorder co-occurrences were computed from a large collection of clinical notes to improve the SemRep performance on extracting treatment relations from Medline citations.[11] Another study focused on how the semantic relations extracted by SemRep from Medline abstracts can guide the process of labeling concept associations from clinical text.[12]

One of the first machine learning systems developed to extract treatment relations from clinical text is described in Roberts et al.[13] In this work, the evaluation was performed on a small set of 77 oncology narratives, which was manually annotated with 7 categories of semantic relations. The feature set comprises various lexical, syntactic, and semantic features designed to capture different aspects of the relation arguments. Using a classification framework based on support vector machines (SVMs), the system achieved an average F1-measure of 72 over the 7 relation categories. An SVM-based framework was also developed by Uzuner et al.[14] to identify treatment relations defined for a more specific scope. To represent treatment relations, the authors of this study utilized semantic categories of concepts and the assertion values associated with these concepts. Examples of relation categories consist of present disease-treatment, possible disease-treatment, and possible symptom-treatment. Using a rich set of features, the SVM-based relation classifier recognized 84% and 72% of the relations annotated in two different corpora. Furthermore, due to its importance, the task of treatment relation extraction was part of the 2010 Informatics for Integrating Biology and the Bedside (i2b2)/Veteran's Affairs (VA) challenge.[15] Examples of treatment relations devised for this competition include relations in which the corresponding treatment (a) has cured or improved a medical problem, (b) has worsened a medical problem, (c) has caused a medical problem, (d) has been administered for a medical problem, and (e) has not been administered because of a medical problem. The concept pairs that occurred in the same sentence and did not fit this criteria were not assigned a relationship. The majority of the systems solving the 2010 i2b2/VA task on relation extraction relied on supervised machine learning approaches.[16–19]

In our preliminary studies,[20] we have implemented a simple algorithm using MEDI and have shown that it is a reliable method on assessing the validity of treatment relations identified in clinical notes. Like SemRep, MEDI is also publicly available (<http://knowledgemap.mc.vanderbilt.edu/research/>). It was developed by aggregating RxNorm, Side Effect Resource (SIDER) 2,[21] MedlinePlus, and Wikipedia and was designed to capture both on-label and off-label (e.g., absent in the Food and Drug Administration's approved drug labels) uses of medications. While RxNorm and SIDER 2 store the medication and indication information in a structured format, MedlinePlus and Wikipedia encode this information in narrative text. Therefore, further processing of the documents from MedlinePlus and Wikipedia was performed including the use of KnowledgeMap Concept Indexer[22,23] and custom-developed section rules to identify the text expressions describing indications and to map them into the UMLS database. The current version of MEDI contains 3,112 medications and 63,343 medication-indication pairs.

Method

We implemented a supervised learning framework that is able to predict whether or not two concepts co-occurring within the same sentence are in a treatment relation. To capture the relationship between the two concepts, we extracted various features based on the lexical information surrounding the concepts, on the semantic properties associated with the two concepts, and on the information derived from both MEDI and SemRep.

MEDI-based treatment relation extraction

A simple algorithm for treatment relation extraction is based on the assumption that any two concepts that co-occur within the same sentence and match a medication-indication pair in MEDI are likely to be in a treatment relation.

Table 1 The set of features for predicting whether two UMLS concepts are in a treatment relation.

| Feature | Description |
|------------------------|--|
| $f_1:semrep$ | Boolean feature that is true whether SemRep indicates a treatment relation between the two concepts. |
| $f_2:medi$ | Boolean feature that is true whether there is a match between the two concepts and a medication-indication pair in MEDI. |
| $f_3:semrep\ or\ medi$ | Boolean feature that is true whether f_1 is true or f_2 is true. |
| $f_4:unigram$ | The lowercased word unigrams surrounding the two concepts. |
| $f_5:bigram$ | The lowercased word bigrams surrounding the two concepts. |
| $f_6:expression$ | The lowercased word expression describing each of the two concepts. |
| $f_7:cui$ | The concept unique identifiers (CUIs) of the two concepts. |
| $f_8:sem\ type$ | The semantic types associated with the two concepts. |

Despite the fact that this assumption does not take into account the context in which the two concepts are mentioned, our review of clinical documents before this study revealed that it holds true for the majority of the cases.

To increase the coverage of MEDI, we expanded the initial set of medication-indication pairs by using ontology relationships from the RxNorm database. For instance, because the medications in MEDI are mapped to generic ingredients,[6] the resource contains pairs involving RxCUI#1000082 (*alcaftadine*), but it does not include pairs with medications containing alcaftadine ingredients such as RxCUI#1000083 (*alcaftadine 2.5 MG/ML*) or brand medication names of alcaftadine as, e.g., RxCUI#1000086 (*lastacraft*). The relations we used to perform this expansion are *has_ingredient* and *tradename_of* from MRREL.

Features for predicting treatment relations

To learn a prediction model that is able to differentiate between pairs of UMLS concepts in treatment relations and the ones not belonging in such relations, we extracted the set of features described in Table 1. In this table, each feature was designed to capture a specific property associated with a pair of concepts. For instance, the lowercased word bigrams surrounding the emphasized concepts in (1) are “switched to”, “for treatment”, “treatment of”, and “as evidenced”. Also in (1), the semantic types extracted by MetaMap for *levofloxacin* are ‘antibiotic’ and ‘organic chemical’, and the semantic type for *pneumonia* is ‘disease or syndrome’. As observed, since a concept can be associated with multiple semantic types in the UMLS Metathesaurus, the *sem type* feature can have multiple values for each concept. Of note, before the feature extraction phase, we assumed that the medical concepts were already identified in text and mapped to concepts in the UMLS Metathesaurus using SemRep.

In addition to the features listed in Table 1, we investigated the contribution of several other features including more word n-gram features and the version of f_4 , f_5 , and f_6 without lowercasing their corresponding textual expressions. We also extracted the concept preferred names and the semantic group(s) in which the semantic type(s) of each concept belongs to. None of these features were able to improve the overall performance of our prediction system.

Evaluation

To train our supervised learning system, we first constructed a dataset annotated with treatment and non-treatment relations. Based on this dataset, we then evaluated the performance results of our system and compared them against the results achieved by SemRep.

Dataset

For creating the dataset with annotated treatment relations, we randomly selected 6,864 discharge summaries from the Vanderbilt Synthetic Derivative, a de-identified version of the electronic medical record. In the data processing phase, we first split the content of each report into sentences using the OpenNLP sentence detector (<http://opennlp.apache.org/>) and removed the duplicate sentences. The output generated by this process consisted of 290,911 sentences. We then parsed these sentences with the current version of SemRep, v1.5, which identified 943,306 UMLS concepts (~3.2 concepts/sentence), and 3,386 treatment relations in 2,841 sentences. Next, we ran the MEDI algorithm over the same concepts extracted by SemRep in the previous step. Since SemRep is designed to identify treatment relations at sentence level, we constrained the algorithm based on MEDI to match any pair of concepts mentioned within the same sentence. As a result, 3,716 MEDI relations were obtained.

Table 2 Results for extracting treatment relations from text.

| System | TP | FP | FN | TN | P | R | F |
|------------|-----|-----|-----|------|--------|--------|--------|
| SemRep | 625 | 145 | 333 | 9483 | 81.17 | 65.24 | 72.34 |
| Our system | 790 | 113 | 168 | 9515 | 87.49* | 82.46* | 84.90* |

* $p < 0.001$; statistically significant differences in performance between our system and SemRep.

F, F1-measure; FN, false negatives; FP, false positives; P, precision; R, recall, TN, true negatives, TP, true positives.

In the manual annotation phase, two reviewers examined 620 sentences in which the MEDI algorithm and SemRep identified at least one relation. We decided on this set of sentences to cover as many SemRep and MEDI relations as possible and, at the same time, to minimize the annotation effort. However, since all these sentences contain relations identified by both SemRep and the MEDI algorithm, one limitation of this selection is that the evaluation of the two systems may result in overestimating their corresponding recall values. Blinded from the results extracted by SemRep and the MEDI algorithm, the annotation process consisted of manually linking pairs of concepts that represent treatment relations inside every sentence. Once a sentence was annotated, the remaining combinations of concept pairs in the sentence were automatically marked as non-treatment relations. During this process, the reviewers performed a double annotation on 25% of the data obtaining a percentage agreement of 97.9, with the Cohen’s Kappa value of 0.86. In the manual annotated dataset, the disagreements were adjudicated by an experienced clinical expert. For annotation, the BRAT annotation tool[24] was employed resulting in 958 and 9,628 treatment and non-treatment relations, respectively.

Results

We evaluated our machine learning framework on the manual annotated dataset using a 5-fold cross validation scheme. For performing the classification of treatment and non-treatment relations, we employed LIBSVM (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>), an implementation of the SVM algorithm. In all experiments, we used the radial basis function as the kernel function of choice. Since our interest was in detecting the treatment relations as accurately as possible, we measured the performance results of our system in terms of precision, recall, and F1-measure.

In Table 2, we report the results obtained from comparing the manual annotated relations with the relations extracted by SemRep and our system. Our system results were obtained by aggregating the results over the test folds during cross validation. As observed, our system managed to achieve significantly higher results than the SemRep results. To measure the statistically significant differences in performance between the two systems, we employed a randomization test based on stratified shuffling.[25]

To get a better insight of the features extracted for identifying treatment relations, we performed several feature ablation studies. The findings of these studies are listed in Tables 3(a) and 3(b). The experiments in Table 3(a) show the contribution of each feature to the overall performance of our system. As indicated, the largest drop in performance is caused by the *sem type* feature. This behavior is expected since there is a clear pattern in the semantic types corresponding to most of the concepts participating in treatment relations. For instance, the medication-disease and medication-medication relation types represent strong indicators for treatment and non-treatment relations, respectively. However, to properly estimate the contribution of SemRep and the MEDI algorithm to the overall performance of our system, we also performed experiments using the All – {*semrep*, *semrep or medi*} and All – {*medi*, *semrep or medi*} feature configurations. This is because the *semrep or medi* feature is highly correlated with the *semrep* and *medi* features. For instance, if *medi* is true for a given concept pair, the *semrep or medi* feature is also true for the same concept pair. Using the All – {*semrep*, *semrep or medi*} configuration, our system achieved a recall of 78.39 and an F1-measure of 82.03. When All – {*medi*, *semrep or medi*} was employed, the recall and F1-measure values dropped to 73.49 and 79.82, respectively.

The experiments from Table 3(b) show how our system performed when only one feature from Table 1 was selected to differentiate between treatment and non-treatment relations. The best performing experiments in this table are the ones employing the results of the MEDI algorithm (i.e., *medi* and *semrep or medi*). The results of these experiments are also statistically significant from the results of the *semrep* experiment ($p < 0.001$). Also interestingly, the machine learning framework using only the *sem type* feature was able to obtain better results than SemRep (74.08 vs. 72.34). Nevertheless, the difference in performance was not statistically significant. Of note, the machine learning framework using only the *semrep* feature was able to find the same separation of the relations as the SemRep

Table 3 Feature ablation studies for treatment relation extraction.

| (a) | | | | (b) | | | |
|-----------------------------|-------|-------|-------|-----------------------|-------|-------|-------|
| Features | P | R | F | Features | P | R | F |
| All – <i>semrep</i> | 85.12 | 82.99 | 84.04 | <i>semrep</i> | 81.17 | 65.24 | 72.34 |
| All – <i>medi</i> | 84.95 | 83.09 | 84.01 | <i>medi</i> | 80.48 | 76.62 | 78.50 |
| All – <i>semrep or medi</i> | 87.88 | 81.00 | 84.30 | <i>semrep or medi</i> | 73.61 | 85.91 | 79.29 |
| All – <i>unigram</i> | 86.75 | 82.05 | 84.33 | <i>unigram</i> | 63.56 | 39.87 | 49.01 |
| All – <i>bigram</i> | 86.74 | 81.94 | 84.27 | <i>bigram</i> | 77.89 | 16.18 | 26.79 |
| All – <i>expression</i> | 87.58 | 81.73 | 84.56 | <i>expression</i> | 74.27 | 39.77 | 51.80 |
| All – <i>sem type</i> | 85.94 | 81.00 | 83.40 | <i>sem type</i> | 68.78 | 80.27 | 74.08 |
| All – <i>cui</i> | 87.00 | 81.00 | 83.89 | <i>cui</i> | 68.80 | 49.48 | 57.56 |

All, the entire set of features from Table 1; F, F1-measure; P, precision; R, recall.

system. As observed, the SemRep results in Table 2 are identical with the results of the *semrep* experiment in Table 3(b). Similarly, the MEDI algorithm achieved the same performance results as the results of the *medi* experiment.

Error Analysis

The cases when the MEDI algorithm was not able to find an exact match for a given concept represented some of the most frequent false negative examples of our system. For instance, the relation *sucralfate*→*heartburn* does not have a corresponding medication-symptom pair in MEDI despite the fact that related concepts such as *esophagitis*, *burn of esophagus*, and *esophageal reflux* are included in the list of indications for *sucralfate* in this resource. Likewise, the relation *unasyn*→*pneumonia* cannot be matched by the algorithm although a medication-indication pair between *unasyn* and a more general concept of *pneumonia*, *communicable diseases*, exists in MEDI. Furthermore, from the false positive examples we analyzed, many of them occurred in complex sentences in which the context is critical in determining the relationship between concepts. The emphasized concepts in (4), e.g., represent a false positive instance because they correspond to a medication-indication pair in MEDI.

Discussion

Our experiments indicate that a machine learning framework is a successful approach for capturing treatment relations in clinical text. By incorporating various sources of lexical and semantic information associated with relation concepts as well as information extracted from a knowledge base of medication-indication pairs, our system managed to improve the performance results over SemRep, a widely used rule-based system in information extraction applications. The major improvements in recall over SemRep due to the information derived from MEDI confirmed our assumption that a valid medication-indication pair expressed in a sentence corresponds to a treatment relation in the majority of cases. Furthermore, as observed from the results of our experiments, the most significant decrease in recall and F1-measure is achieved when discarding all MEDI related features (i.e., using the All – {*medi*, *semrep or medi*} feature configuration). Despite the fact that SemRep was not particularly implemented for the clinical domain, our experiments from Tables 2 and 3(b) showed that this system is able to extract treatment relations with high precision. This study also adds to the relatively few evaluations of SemRep on clinical text, demonstrating that it does work in this domain also.

It is worth mentioning that both the MEDI algorithm and SemRep are not able to identify all treatment relation types. For instance, since the first argument of the treatment relation in (1) is a procedure, the concept pair from this example will not have a corresponding match in MEDI. Similarly, in SemRep, the types of the relations extracted from text are constrained to match the types of their corresponding relations in the UMLS Semantic Network.[7,26,27] As illustrated in Figure 1, only 9% of relations are identified by the two algorithms from the total number of relations extracted over the entire collection of 6,864 discharge summaries. The top 3 most frequent types of these relations are listed in Table 4. In this table, each relation type is described using the UMLS semantic types associated with the relation arguments. Not surprisingly, the most frequent relation type identified by both algorithms is the one abbreviated as *orch,phsu*→*sosy*, which represents the generic type of medication *treats* disease. On the other hand, the next two most frequent SemRep relation types (i.e., *topp*→*dsyn* and *topp*→*podg*) have procedure as first argument and therefore, they were not found among the MEDI relation types. From the types identified by the MEDI algorithm in Table 4, *clnd*→*sosy* was not found in the relation types extracted by SemRep.

Figure 1 The connection between the relations identified by the MEDI algorithm and SemRep.

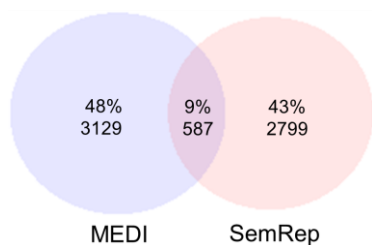


Table 4 Top 3 most frequent relation types identified by the MEDI algorithm and SemRep. The semantic types of the relation arguments are abbreviated as returned by MetaMap.

| SemRep relation types | Freq. | MEDI relation types | Freq. |
|-----------------------|-------|---------------------|-------|
| orch,phsu→soty | 425 | orch,phsu→soty | 1582 |
| topp→dsyn | 324 | orch,phsu→dsyn | 616 |
| topp→podg | 264 | clnd→soty | 270 |

clnd, clinical drug; dsyn, disease or syndrome; orch, organic chemical; phsu, pharmacologic substance; podg, patient or disabled group; soty, sign or symptom; topp, therapeutic or preventive procedure.

In future research, we plan to improve our machine learning framework by implementing additional features that are able to better differentiate treatment relations from non-treatment relations. For instance, features using the structural and syntactic information derived from constituent and dependency trees could better capture the contextual properties between two medical concepts.[13,14,28] For this task, assertion classification[29] can also be investigated to better detect the relations whose corresponding treatment did not improve or cure a medical condition. Other technologies that may improve treatment relation extraction are statistical feature selection[30] and learning methods for imbalanced data.[31] Moreover, we intend to run our system over a large collection of clinical notes that could enable the discovery of new medication-indication pairs. Examples of pairs not in MEDI that our machine learning system was able to extract included valid relations such as *ethambutol*→*infection*, *lidocaine*→*stump pain*, *famvir*→*oral ulcers*, *levaquin*→*pyuria*, and *vesicare*→*bladder spasm*.

Conclusion

In this paper, we described a supervised learning system for identifying treatment relations in clinical notes. Our system successfully integrated various types of information which lead to achieving significantly better performance results than SemRep. One relevant source of information which had a major impact in boosting our system's recall is MEDI. As we empirically proved, MEDI is a broad and reliable resource on assessing the validity of treatment relations. We believe that future information extraction systems in the clinical domain should rely on a knowledge base of medication-indication pairs to accurately identify treatment relations in text. We plan to further improve this task and to assess its usability in various clinical applications.

Acknowledgements

This work was supported by NLM/NIH grants 5 T15 LM007450-12 and 1 R01 LM010685. The dataset used for the analyses described were obtained from Vanderbilt University Medical Center's Synthetic Derivative which is supported by institutional funding and by the Vanderbilt CTSA grant ULTR000445 from NCATS/NIH.

References

- 1 Cebul RD, Love TE, Jain AK, *et al.* Electronic health records and quality of diabetes care. *N Engl J Med* 2011;**365**:825–33.
- 2 Ghitza UE, Sparenborg S, Tai B. Improving drug abuse treatment delivery through adoption of harmonized electronic health record systems. *Subst Abuse Rehabil* 2011;**2011**:125–31.
- 3 Liu M, Wu Y, Chen Y, *et al.* Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J Am Med Inform Assoc* 2012;**19**:28–35.
- 4 Roth CP, Lim YW, Pevnick JM, *et al.* The challenge of measuring quality of care from the electronic health record. *Am J Med Qual* 2009;**24**:385–94.
- 5 Roth MT, Weinberger M, Campbell WH. Measuring the quality of medication use in older adults. *J Am Geriatr Soc* 2009;**57**:1096–102.
- 6 Wei WQ, Cronin RM, Xu H, *et al.* Development and evaluation of an ensemble resource linking medications to their indications. *J Am Med Inform Assoc* 2013;**20**:954–61.

- 7 Rindflesch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *J Biomed Inform* 2003;**36**:462–77.
- 8 Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. In: *Proc AMIA Symp*. 2001. 17–21.
- 9 Fiszman M, Rindflesch TC, Kilicoglu H. Abstraction summarization for managing the biomedical research literature. In: *HLT-NAACL Workshop on Computational Lexical Semantics*. 2004. 76–83.
- 10 Hristovski D, Friedman C, Rindflesch TC, et al. Exploiting Semantic Relations for Literature-Based Discovery. In: *Proc AMIA Symp*. 2006. 349–53.
- 11 Rindflesch TC, Pakhomov SV, Fiszman M, et al. Medical Facts to Support Inferencing in Natural Language Processing. In: *Proc AMIA Symp*. 2005. 634–8.
- 12 Liu Y, Bill R, Fiszman M, et al. Using SemRep to label semantic relations extracted from clinical text. In: *Proc AMIA Symp*. 2012. 587–95.
- 13 Roberts A, Gaizauskas R, Hepple M, et al. Mining clinical relationships from patient narratives. *BMC Bioinformatics* 2008;**9** (Suppl. 11).
- 14 Uzuner Ö, Mailoa J, Ryan R, et al. Semantic relations for problem-oriented medical records. *Artif Intell Med* 2010;**50**:63–73.
- 15 Uzuner Ö, South BR, Shen S, et al. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J Am Med Inform Assoc* 2011;**18**:552–6.
- 16 Rink B, Harabagiu S, Roberts K. Automatic extraction of relations between medical concepts in clinical texts. *J Am Med Inform Assoc* 2011;**18**:594–600.
- 17 De Bruijn B, Cherry C, Kiritchenko S, et al. Machine-learned solutions for three stages of clinical information extraction: the state of the art at i2b2 2010. *J Am Med Inform Assoc* 2011;**18**:557–62.
- 18 Minard AL, Ligozat AL, Ben Abacha A, et al. Hybrid methods for improving information access in clinical documents: concept, assertion, and relation identification. *J Am Med Inform Assoc* 2011;**18**:588–93.
- 19 Patrick JD, Nguyen DHM, Wang Y, et al. A knowledge discovery and reuse pipeline for information extraction in clinical notes. *J Am Med Inform Assoc* 2011;**18**:574–9.
- 20 Bejan CA, Wei WQ, Denny JC. Using SemRep and a medication indication resource to extract treatment relations from clinical notes. In: *AMIA Jt Summits Transl Sci Proc*. 2014.
- 21 Kuhn M, Campillos M, Letunic I, et al. A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* 2010;**6**.
- 22 Denny JC, Smithers JD, Miller RA, et al. ‘Understanding’ medical school curriculum content using KnowledgeMap. *J Am Med Inform Assoc* 2003;**10**:351–62.
- 23 Denny JC, Spickard A 3rd, Miller RA, et al. Identifying UMLS concepts from ECG Impressions using KnowledgeMap. In: *Proc AMIA Symp*. 2005. 196–200.
- 24 Stenetorp P, Pyysalo S, Topić G, et al. brat: a Web-based Tool for NLP-Assisted Text Annotation. In: *Demonstrations Session at EACL*. 2012. 102–7.
- 25 Noreen EW. *Computer-Intensive Methods for Testing Hypotheses: An Introduction*. 1st ed. New York: : John Wiley & Sons 1989.
- 26 Rindflesch TC, Fiszman M, Libbus B. Semantic interpretation for the biomedical research literature. In: Chen H, Fuller SS, Friedman C, et al., eds. *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*. 2005. 399–422.
- 27 Ahlers CB, Fiszman M, Demner-Fushman D, et al. Extracting semantic predications from Medline citations for pharmacogenomics. In: *Pac Symp Biocomput*. 2007. 209–20.
- 28 GuoDong Z, Jian S, Jie Z, et al. Exploring various knowledge in relation extraction. In: *Annual Meeting on Association for Computational Linguistics*. 2005. 427–34.
- 29 Bejan CA, Vanderwende L, Xia F, et al. Assertion modeling and its role in clinical phenotype identification. *J Biomed Inform* 2013;**46**:68–74.
- 30 Bejan CA, Xia F, Vanderwende L, et al. Pneumonia identification using statistical feature selection. *J Am Med Inform Assoc* 2012;**19**:817–23.
- 31 He H, Garcia EA. Learning from Imbalanced Data. *IEEE Trans Knowl Data Eng* 2009;**21**:1263–84.