MediConceptNet: An Affinity Score Based Medical Concept Network

Anupam Mondal, Erik Cambria
School of Computer Science and Engineering
Nanyang Technological University
{manupam,cambria}@ntu.edu.sg

Dipankar Das, Sivaji Bandyopadhyay
Department of Computer Science and Engineering
Jadavpur University
{ddas,sbandyopadhyay}@cse.jdvu.ac.in

Abstract

In healthcare, information extraction is essential in building automatic domain-specific applications. Medical concepts and their semantic identification take an important role to develop a network for visualizing medical concepts and their relations. The challenge appears while available medical corpora are only in either unstructured or semi-structured forms. In the present paper, to overcome the challenge and consequently to construct a structured corpus, we apply a domain-specific lexicon, namely WordNet of Medical Event. Medical concepts assigned by this lexicon and their affinity score, polarity score, sense, and semantic features assist in identifying conceptual and sentiment relations from the corpus. The lexicon and all these features provide an essential support to analyze an unstructured corpus and represent it in a structured corpus which we term MediConceptNet: the medical concepts are connected with each other through the concerned features. A previously suggested network for the same purpose, e.g., SemNet, is only based on the semantic and affinity features. The semantic relations of the concepts can be successfully determined in three distinct ranges, e.g., 0 for no relation, 0-1 for partial relations, and 1 corresponding a full relation. To evaluate the data of MediConceptNet, we apply an agreement analysis provided by the Cohen's kappa coefficient and achieve 0.66 agreement score, evaluating the comparative statistics of two medical practitioners working as manual annotators.

Introduction

In Biomedical Natural Language Processing (BioNLP) domain, medical concepts and their sentiment relation identification are introduced as contributory tasks to build user compatible applications. The tasks face difficulties due to a large number of unstructured corpora produced daily and lack of sufficient number of domain experts such as doctors and medical practitioners. On the other hand, a representation of a structured corpus from an unstructured corpus, delivered by a digital web as articles, prescriptions, reports, and web-blogs, is essential for building domain applications such as BioNLP.

To this end, various information identification approaches namely linguistic (e.g., rule based), tabulation (e.g., unigram, bigram, negation word count), and machine learning (e.g., supervised and unsupervised classifiers) have been already suggested to represent potentially available structured corpora (Cambria, Wang, and White 2014; Poria et al. 2015). Besides, the number of sentiment lexicons namely SenticNet, SentiWordNet, Bing Liu Subjective list, and Taboada adjective list have been used for extracting concepts and their sentiment (Cambria et al. 2016; Liu 2012; Taboada et al. 2011; Esuli and Sebastiani 2006). Unfortunately, these lexicons are not provided an output with enough accuracy in BioNLP because of the absence of domain relevant concepts and their related information (Muhammad et al. 2013).

In the present paper, we use a domain-specific lexicon, namely WordNet of Medical Event (WME), to identify medical concepts and their conceptual features (Mondal et al. 2016). WME lexicon refers two different versions, WME 1.0 (the first version of WME) and WME 2.0 (the current version of WME). WME 1.0 has been built by the extracted medical concepts (terms) of training and test datasets of SemEval-2015 Task-6. The conventional WordNet[1] and preprocessed English Medical dictionary[2] used for identifying the linguistic features of the concepts, which are gloss (descriptive explanation) and parts-of-speech (POS) (Rajagopal et al. 2013; Cambria 2013).

Moreover, SenticNet, SentiWordNet, Bing Liu subjective list, and Taboada adjective list sentiment lexicons have been applied to extract polarity score, and sense (sentiment) features for the medical concepts (Mondal et al. 2015). The current version of WME (WME 2.0) provide 10186 number of medical concepts in total and additional features such as affinity score, gravity score, and semantic (similar sentiment concepts) compared to the previous version of WME 1.0. The assigned semantic and affinity score features of medical concepts help to identify the conceptual and sentiment relevance between concepts (Cambria et al. 2009; Hsu and Chen 2006; Lenat et al. 1990; McCarthy 1960; Poria et al. 2013). Therefore, for the first time, we here introduce the conceptual and sentiment linking base seman-

[1] https://wordnet.princeton.edu

[2] http://alexabe.pbworks.com/f/Dictionary+of+Medical+Terms+4th+Ed.-+(Malestome).pdf

tic relations as no-relation, partial-relation, and full-relation. These relations are represented by the affinity score between the concepts as 0, 0-1, and 1, indicating no-relation, partial-relation, and full-relation. The semantic feature and affinity score based relations are conventionally applied to build a semantic network (SemNet), assigning the sentiment relevance between medical concepts. For example, the semantics *abdominal_breathing*, *hypopnea* and *respiration* are connected each other by partial-relations and presented in SemNet. To understand the conceptual relevance and visualize the linking, we add all semantics of medical concepts on the top of SemNet, which can assist in representing the three types of medical concept networks based on the linking strength described above. Then, the resultant medical concept networks MediConceptNet can be differentiated by the affinity score based semantic relations. SemNet identifies the sentiment relations between similar concepts, whereas MediConceptNet assigns the conceptual linking between medical concepts. Such networks are useful for the experts and non-experts to visualize similar concepts and their semantic relations, help to retrieve hidden relations between different concepts, and bring a more complete picture for a better understanding of the medical concepts.

To evaluate the consistency of the structured data produced by MediConceptNet, we will consider an agreement analysis approach by the Cohen's kappa coefficient[3]. The statistics provided by the manual annotators is used, where the annotators are medical practitioners. Moreover, the (network) visualization of MediConceptNet will be illustrated, aiming to help the domain researchers to develop the structured corpus, assist in building annotation, categorization, and clustering system in the medical domain.

The structure of the paper is as follows: I. Related work, II. Affinity score identification, III. Semantic network, IV. Medical concept network, V. Evaluation, and VI. Conclusion and future scope.

Related Work

Biomedical information extraction research is challenging due to lack of complete structured corpus on the contrary to a huge amount of semi-structured and unstructured medical corpora. The researchers have introduced the domain-specific lexicons with preserving the features such as polarity score, semantics, and sentiment (sense) for medical concepts to build information extraction systems considering annotation and relation identification from unstructured medical corpora (Asgarian and Kahani 2014; Abacha and Zweigenbaum 2011; Uzzaman and Allen 2010; Embarek and Ferret 2008). To this end, the standard tool as GENIA tagger (Kim et al. 2003) and the lexicons MEN (Medical WordNet) and WME (WordNet of Medical Event) have been invented (Tanabe et al. 2005; Kilgarriff and Fellbaum 2000; Mondal et al. 2016).

MEN lexicon has been built with two sub-networks, namely Medical FactNet (MFN) and Medical BeliefNet (MBN), to evaluate consumer health reports (Kilgarriff and Fellbaum 2000). The formal architecture of the Princeton WordNet has been used for MEN (Kilgarriff and Fellbaum 2000; Smith and Fellbaum 2004). In addition, while MFN aims to serve non-expert groups to extract and present a better understanding of basic medical information, MBN identifies the fraction of beliefs on medical phenomena. Their primary motivation was to evolve a visualization system for retrieving the medical information from corpora. (Kang et al. 2012) developed a medical concept recognition system from unstructured clinical records using two dictionaries following the ABNER, Lingpipe, MetaMap, OpenNLP, JNET, Peregrine, and StanfordNER approaches. They applied a simple voting schema to evaluate the output to decide the acceptance of the annotated concepts based on the predefined threshold value.

Moreover, Open Mind Common Sense, ConceptNet5, KASO, and Concept Extractor tools have been developed for identifying the relations between the concepts (Singh et al. 2002). Open Mind Common Sense (OMCS) resource provides a support to extract the fact-base relations (e.g., IsA, MadeOf, UsedFor, LoctedNear, PartOf, DefinedAs) for the concepts (Singh et al. 2002). (Speer and Havasi 2012) applied OMCS with ConceptNet4, DBPedia[4], ReVerb[5], English wiktionary[6], and "games with a purpose" resources[7] to build the ConceptNet5. ConceptNet5 resource is produced a large multidimensional graph of the concepts with the WordNet ontology. (Wang, Völker, and Haase 2006) developed KASO system to reduce the workload for both experts and non-experts using a hybrid approach, which is the combination of manual and automatic extraction process. Concept Extractor tool is introduced to compare the performance between different sentiment extraction methods of the concepts (Dinh and Tamine 2011). These lexicons and tools facilitate concepts and their relation identification, helping to establish structural corpora. Unfortunately, the mentioned lexicons and also tools haven't provided a semantic relation extraction system with an enough accuracy for the medical concepts as well as corpora, yet. To develop a semantic relation extraction system for the medical concepts, in this paper, we introduce WordNet of Medical Events (WME 2.0), a domain-specific lexicon (Mondal et al. 2016). The lexicon with affinity score, polarity score, semantic and sense features of the concepts helps to construct a medical concept network with visualization, which is able to identify sentiment and conceptual relevance between concepts.

Affinity score identification

Affinity represents the sentiment linking between pairs of medical concepts by determining their common semantics, which sets a degree. Affinity score calculates this degree of semantic relations of each concept pair and can bring concept clusters. The clusters are extremely important to build a concept network ensuring both organizing the semantic relations between the concepts, e.g., how the concept pairs associate with each other semantically. In addi-

[3]https://en.wikipedia.org/wiki/Cohen's kappa

[4]http//dbpedia.org/About

[5]http//reverb.cs.washington.edu/

[6]http//en.wiktionary.org/wiki/Wiktionary

[7]http//www.gwap.com/gwap/gamesPreview/verbosity/

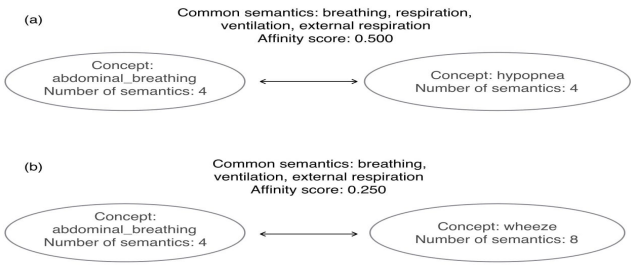

Figure 1: Affinity score assignment process between medical concept pairs. (a) 4 common semantics *breathing, respiration, ventilation,* and *external_respiration* are determined for the concept pair of *abdominal_breathing* and *hypopnea* out of 8 total semantics, resulting in affinity score *Affinity Score*$_c$ = 0.500 (b) 3 common semantics *breathing, ventilation,* and *external_respiration* out of 12 total semantics of the pair *abdominal_breathing* and *wheeze* provides *Affinity Score*$_c$ = 0.250, indicating finite but weaker relations of the pair in (a).

tion, the combined network can provide a visualization for a better understanding and reduce a communication gap between computer data and medical practitioners as well as patients (Cambria, Hussain, and Eckl 2011). Affinity score is obtained by a probabilistic counting of similar semantics and explained below.

First, we need to define the overlapping semantics of each concept pair as

$$\textit{Affinity}_c = MC_1 \cap MC_2, \tag{1}$$

where MC_1 and MC_2 represent semantic sets of two different medical concepts. Thus, *Affinity*$_c$ is simply the number of common semantics of MC_1 and MC_2. Then, *Affinity*$_c$ assists in identifying the final affinity score *Affinity Score*$_c$ as

$$\textit{Affinity Score}_c = \frac{\textit{Affinity}_c}{MC_1 + MC_2}, \tag{2}$$

where the sum of MC_1 and MC_2 represent the total number of all semantics in each set of the concepts.

Figure 1 summarizes the procedure to achieve *Affinity Score*$_c$ of example concept pairs. First, we determine *Affinity*$_c$ of each concept pair, which is 4 by the common semantics of *breathing, respiration, ventilation,* and *external_respiration* of the pair *abdominal_breathing* and *hypopnea* in Figure 1(a). Consequently, *Affinity Score*$_c$ results in the value of 0.500 dividing its *Affinity*$_c$ by the total semantics of 8, 4 semantics from each concept. Similarly, for the concept pair *abdominal_breathing* and *wheeze* in Figure 1(b), *Affinity Score*$_c$ is equal to 0.250 with *Affinity*$_c$ = 3 out of 12 total semantics, 4 from the concept *abdominal_breathing* and 8 from the concept *wheeze*. Therefore, we can conclude that *abdominal_breathing* is conceptually related with *hypopnea* more than *wheeze*.

Determining *Affinity Score*$_c$ for all concept pairs brings *Affinity Score*$_c$ for every semantic of the associated concepts.

<Concept>abdominal_breathing</Concept>
<Semantic (Affinity score)>
breathing (0.400)
respiration (0.250)
ventilation (0.220)
external respiration (0.200)
</Semantic (Affinity score)>

Figure 2: Affinity scores *Affinity Score*$_c$ of the medical concept *abdominal_breathing* with their semantics *breathing, respiration, ventilation,* and *external_respiration*. While *breathing* indicating the strongest relation (0.400), both *ventilation* and *external_respiration* presents the weakest relation (0.200) with the concept.

Here, the considered concept pairs are the concept and its semantics evaluated one by one (Simply, a concept appears as a semantic under another concept and vice versa). Figure 2 provides a complete *Affinity Score*$_c$ list for the semantics of an example concept *abdominal_breathing*.

Semantic network

The affinity score based semantic relations are used to build the semantic network (SemNet) for the medical concepts. The SemNet helps to understand the sentiment relevance of the concepts to represent the structured corpus using their semantic features. The current version of WME (WME 2.0) with assigned medical concepts and their conceptual features like polarity score, semantic, and sense are applied to develop SemNet, in the absence of domain experts (e.g., doctors, medical practitioners). In this paper, we introduce affinity score identification process for the medical concepts of WME 2.0 as mentioned in the previous section. On the other hand, polarity score and sense of the medical concepts are both taken from SenticNet, SentiWordNet, and Bing Liu sentiment lexicons, whereas semantics (similar sentiment based concepts) are extracted from conventional WordNet, preprocessed English Medical Dictionary[8] and SenticNet resources.

The affinity scores 0, 0 to 1, and 1 refer the semantic relations as no-relation, partial-relation, and full-relation, respectively, which indicates the sentiment relevance between the pair of concepts under SemNet. For example, the medical concept *suffer* is weakly related to the semantics *mantle, ailment,* and *winery* with the corresponding affinity scores 0.056, 0.067, and 0.091, where the semantics *shear* and *enoblement* are both strongly related with the affinity score of 0.700, respectively. Figure 3 shows the SemNet representation for the medical concept *suffer* in a visualization.

The SemNet is able to recognize the similar sentiment based medical concepts and their relations, which indicate medical concept clusters, and the concept cluster based SemNet helps to develop three types of medical concept networks (MediConceptNet) according to the strength of the semantic relations of the concepts.

[8] http://alexabe.pbworks.com/f/Dictionary+of+Medical+Terms+4th+Ed.-+(Malestrom).pdf

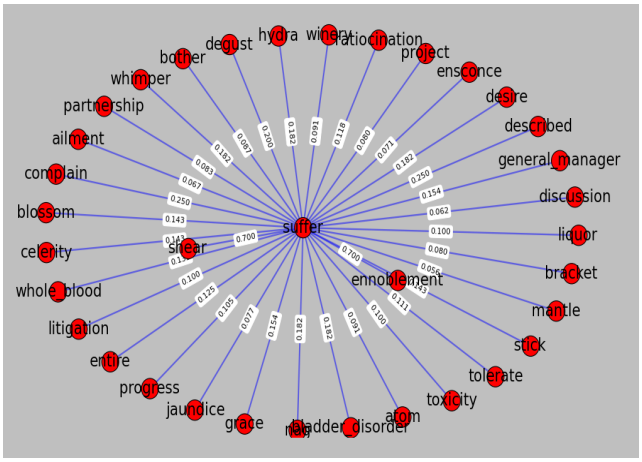

Figure 3: A semantic network for the medical concept *suffer*. The network represents sentiment linking between the medical concept *suffer* and its semantics such as *ennoblement* and *disorder* with the affinity scores 0.700 and 0.186, respectively.

Medical concept network

Medical concept network (MediConceptNet) indicates the conceptual and sentiment relevance between concepts, which helps to extract the hidden relations of the medical concepts with a visualization. In the present paper, we develop three types of MediConceptNet such as no-relation, partial-relation, and full-relation according to SemNet and their semantic relations of the concepts. No-relation based MediConceptNet refers the absence of the conceptual linking between medical concepts with 0 affinity score. In contrast, partial-relation based MediConceptNet indicates the weak or average conceptual relevance between the concepts of ranges from 0 to 1 in the corresponding affinity score. Furthermore, full-relation based MediConceptNet shows a strong conceptual and sentiment relevance between concepts with the affinity score of 1.

As an illustration of the described visualization, for instance, the medical concepts *plague* and *clot* present a no-relation based medical concept network with 0 affinity score, on the contrary, Figure 4 shows a partial-relation based medical concept network between medical concepts *brain* and *clog* with 0.130 affinity score.

Figure 5 indicates a full-relation based medical concept network between medical concepts *plague* and *attack* with 1 affinity score. The medical concept networks are able to identify the hidden sentiment and conceptual relations of the concepts, assisting in extracting common-sense knowledge from medical corpora with good performance.

Thereafter, we develop a Graphical User Interface (GUI) to represent all proposed networks as SemNet and MediConceptNet. Figure 6 shows an example GUI providing a single hand support with a visualization effect to identify these semantic relations of the medical concepts. Moreover, the network base interface can apply as a part of a concept-based search engine to assist the researchers as well as the

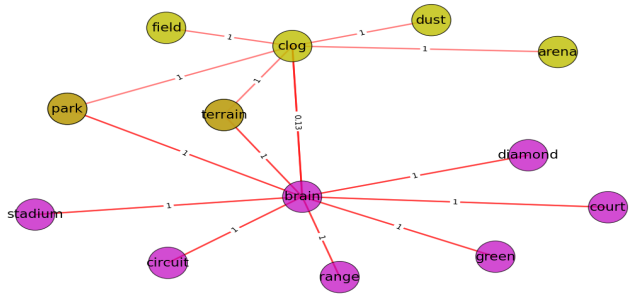

Figure 4: A partial-relation based medical concept network. The medical concepts *brain* and *clog* are weakly related to each other with 0.130 affinity score.

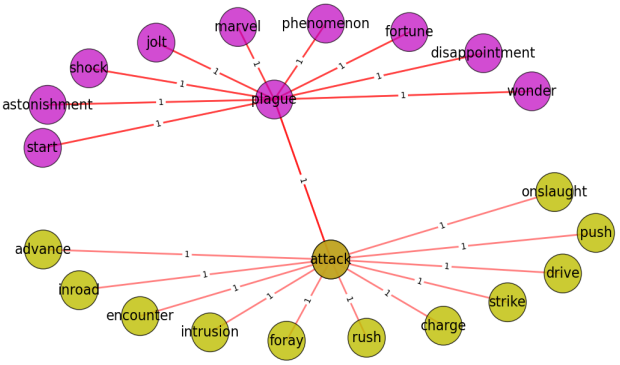

Figure 5: A full-relation based medical concept network. The medical concepts *attack* and *plague* are strongly related to each other with 1 affinity score.

experts (e.g., medical practitioners) in the biomedical domain and so to recognize similar medical concepts and their relations (Speer and Havasi 2012).

Evaluation

To build a concept network, a domain-specific lexicon is essential with conceptual features of the concepts under Natural Language Processing (NLP). This task is even more evident when the concern is to build a structured corpus in a specific domain such as BioNLP. Our aim is to develop an intelligent cognitive system in medical field using conceptual features like affinity and polarity scores and semantic with a visualization. The system helps to cluster similar sentiment base concepts and identify semantic relations between medical concepts.

To validate the proposed medical concept network and its data under WME 2.0 lexicon, we adopt an agreement analysis due to lack of annotated data. The agreement analysis involves the annotated statistics provided by medical practitioners. The annotators provide validated medical concepts and their semantic relations based on their practical experience and WME 2.0 lexicon. The reason for considering WME 2.0 lexicon as a baseline system due to the maximum presence of the medical concepts over other sentiment lex-

Figure 6: A GUI representation of the system to build a semantic network and medical concept network.

```
<Concept>amnesia<\Concept>
<Properties>
<POS>noun<\POS>
<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.
<\Gloss>
<Semantic>memory_loss,blackout,fugue,stupor<\Semantic>
<Polarity_score>-0.375<\Polarity_score>
<Affinity_score>0.429<\Affinity_score>
<Gravity_score>0.170<\Gravity_score>
<Sense>negative<\Sense>
<\Properties>
```

Figure 7: WME 2.0 lexicon representation for a concept *amnesia*. Each concept is presented with all features such as affinity score, gloss, gravity score, POS, polarity score, semantic as similar sentiment concepts, and sense as sentiment.

icons. Conventionally well-known sentiment lexicons such as SenticNet and SentiWordNet only consider 26% and 40% coverage of the medical concepts presented in WME 2.0, which are not effective to get the accuracy of medical concepts due to the shortage of medical words (concepts) in these resources. On the other hand, WME 2.0 lexicon satisfies 10186 number of medical concepts in total and their affinity score, gloss, gravity score, Parts Of Speech (POS), polarity score, semantics, and sense features as shown in Figure 7.

The agreement analysis is conducted by the Cohen's kappa coefficient and the ingredients are processed by two manual annotators to validate the medical concepts and the semantic relations of the network (Viera, Garrett, and others 2005). Both the concepts and the relations are first generated by our system. Then, the manual annotators label the concepts and the semantic relations of concepts as an agreement with the system output by Yes, otherwise, a disagreement by No. Table 1 shows counts of Yes and No from the two independent annotators and they are presented as Annotator-1 and Annotator-2.

Total number of identified relations 5862		Annotator-2	
		Yes	No
Annotator-1	Yes	4206	447
	No	245	964

Table 1: Validation of the medical concept network by annotators. Agree (Yes) and disagree (No) on the semantic relations of concepts by Annotator-1 and Annotator-2.

The Cohen's Kappa coefficient κ is defined

$$\kappa = \frac{Pr_a - Pr_e}{1 - Pr_e}, \tag{3}$$

where Pr_a is the observed proportion of full agreement between two annotators as well as the agreement of the system output with the labeling performed by the annotators. In addition, Pr_e is the proportion expected by a chance and so indicates a kind of random agreement between the annotators.

Consequently, we have the Cohen's Kappa κ = 0.66 for the identified relations of the medical concept network. The κ score proves a satisfactory agreement of the identified semantic relations between the medical concepts from MediConceptNet.

Conclusion and future scope

Semantic relations of the concepts are extremely important for extracting contextual information from unstructured corpora to represent structured corpora. The contextual information helps to identify domain knowledge under Biomedical Natural Language Processing (BioNLP) for the experts and non-experts. This paper introduces a medical concept network to identify the conceptual and sentiment linking between medical concepts with visualization, which helps to represent structured corpora from unstructured corpora. A domain-specific lexicon, namely WordNet of Medical Event (WME 2.0), its assigned medical concepts, and their statistical and sentiment features are applied to build the concept networks. The conceptual features refer as affinity score, which measures a sentiment linking between medical concepts, whereas the semantic feature uses to identify the similar sense based concepts. The proposed concept networks presented as the semantic network (SemNet) and the medical concept network (MediConceptNet). The SemNet shows the sentiment linking between similar medical concepts, whereas MediConceptNet indicates the conceptual and sentiment relevance of medical concepts identifying the hidden relations between the concepts.

To validate the extracted relations of the medical concept network, we employ agreement analysis. The agreement analysis is satisfied by the Cohens kappa coefficient of 0.66, verifying the semantic relations in a good performance, as a system output by manual confirmations of medical practitioners.

In future, such concept networks can be used as a part of a concept-based search engine to assist researchers and experts such as doctors and medical practitioners to retrieve similar medical concepts and their hidden relations to their applications in the domain of BioNLP.

References

Abacha, A. B., and Zweigenbaum, P. 2011. A hybrid approach for the extraction of semantic relations from medline abstracts. In *International Conference on Intelligent Text Processing and Computational Linguistics*, 139–150.

Asgarian, E., and Kahani, M. 2014. Designing an integrated semantic framework for structured opinion summarization. In *European Semantic Web Conference*, 885–894. Springer.

Cambria, E.; Hussain, A.; Havasi, C.; and Eckl, C. 2009. Common sense computing: from the society of mind to digital intuition and beyond. In *European Workshop on Biometrics and Identity Management*, 252–259. Springer.

Cambria, E.; Poria, S.; Bajpai, R.; and Schuller, B. 2016. SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives. In *the 26th International Conference on Computational Linguistics (COLING), Osaka*, 2666–2677.

Cambria, E.; Hussain, A.; and Eckl, C. 2011. Bridging the gap between structured and unstructured healthcare data through semantics and sentics. *ACM WebSci'11* 1–4.

Cambria, E.; Wang, H.; and White, B. 2014. Guest editorial: Big social data analysis. *Knowledge-Based Systems* 69:1–2.

Cambria, E. 2013. An introduction to concept-level sentiment analysis. In Castro, F.; Gelbukh, A.; and González, M., eds., *Advances in Soft Computing and Its Applications*, volume 8266 of *Lecture Notes in Computer Science*, 478–483. Berlin: Springer-Verlag.

Dinh, D., and Tamine, L. 2011. Biomedical concept extraction based on combining the content-based and word order similarities. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, 1159–1163. ACM.

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

Esuli, A., and Sebastiani, F. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, 417–422. Citeseer.

Hsu, M.-H., and Chen, H.-H. 2006. Information retrieval with commonsense knowledge. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, 651–652. ACM.

Kang, N.; Afzal, Z.; Singh, B.; Van Mulligen, E. M.; and Kors, J. A. 2012. Using an ensemble system to improve concept extraction from clinical records. *Journal of biomedical informatics* 45(3):423–428.

Kilgarriff, A., and Fellbaum, C. 2000. Wordnet: An electronic lexical database.

Kim, J.-D.; Ohta, T.; Tateisi, Y.; and Tsujii, J. 2003. Genia corpus—a semantically annotated corpus for bio-textmining. *Bioinformatics* 19(suppl 1):i180–i182.

Lenat, D. B.; Guha, R. V.; Pittman, K.; Pratt, D.; and Shepherd, M. 1990. Cyc: toward programs with common sense. *Communications of the ACM* 33(8):30–49.

Liu, B. 2012. *Sentiment Analysis and Opinion Mining*. Morgan and Claypool Publishers.

McCarthy, J. 1960. *Programs with common sense*. RLE and MIT Computation Center.

Mondal, A.; Chaturvedi, I.; Das, D.; Bajpai, R.; and Bandyopadhyay, S. 2015. Lexical resource for medical events: A polarity based approach. In *ICDM*, 1302–1309.

Mondal, A.; Das, D.; Cambria, E.; and Bandyopadhyay, S. 2016. Wme: Sense, polarity and affinity based concept resource for medical events. *Proceedings of the Eighth Global WordNet Conference* 242–246.

Muhammad, A.; Wiratunga, N.; Lothian, R.; and Glassey, R. 2013. Domain-based lexicon enhancement for sentiment analysis. In *In SMA@ BCS-SGAI*, 7–18.

Poria, S.; Gelbukh, A.; Agarwal, B.; Cambria, E.; and Howard, N. 2013. *Common sense knowledge based personality recognition from text*. Advances in Soft Computing and Its Applications. Springer. 484–496.

Poria, S.; Cambria, E.; Gelbukh, A.; Bisio, F.; and Hussain, A. 2015. Sentiment data flow analysis by means of dynamic linguistic patterns. *IEEE Computational Intelligence Magazine* 10(4):26–36.

Rajagopal, D.; Cambria, E.; Olsher, D.; and Kwok, K. 2013. A graph-based approach to commonsense concept extraction and semantic similarity detection. In *WWW*, 565–570.

Singh, P.; Lin, T.; Mueller, E. T.; Lim, G.; Perkins, T.; and Zhu, W. L. 2002. Open mind common sense: Knowledge acquisition from the general public. In *On the Move to Meaningful Internet Systems*, 1223–1237.

Smith, B., and Fellbaum, C. 2004. Medical wordnet: a new methodology for the construction and validation of information resources for consumer health. In *ACL*, 371.

Speer, R., and Havasi, C. 2012. Representing general relational knowledge in conceptnet 5. In *LREC*, 3679–3686.

Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K.; and Stede, M. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics* 37(2):267–307.

Tanabe, L.; Xie, N.; Thom, L. H.; Matten, W.; and Wilbur, W. J. 2005. Genetag: a tagged corpus for gene/protein named entity recognition. *BMC bioinformatics* 6(1):1.

Uzzaman, N., and Allen, J. F. 2010. Trips and trios system for tempeval-2: Extracting temporal information from text. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, 276–283.

Viera, A. J.; Garrett, J. M.; et al. 2005. Understanding interobserver agreement: the kappa statistic. *Fam Med* 37(5):360–363.

Wang, Y.; Völker, J.; and Haase, P. 2006. Towards semi-automatic ontology building supported by large-scale knowledge acquisition. In *AAAI Fall Symposium On Semantic Web for Collaborative Knowledge Acquisition*, volume 6, 06.