The Generation of a Corpus for Clinical Sentiment Analysis

Yihan Deng[1], Thierry Declerck[2], Piroska Lendvai[2], and Kerstin Denecke[3](✉)

[1] Innovation Center Computer Assisted Surgery,
University of Leipzig, Leipzig, Germany
yihan.deng@medizin.uni-leipzig.de
[2] Department of Computational Linguistics,
Saarland University, Saarbrücken, Germany
declerck@dfki.de, piroska.r@gmail.com
[3] Department of Medical Informatics,
Bern University of Applied Sciences, Biel, Switzerland
kerstin.denecke@bfh.ch

Abstract. Clinical care providers express their judgments and observations towards the patient status in clinical narratives. In contrast to sentiment expressions in general domains targeted by language technology, clinical sentiments are influenced by related medical events such as clinical precondition or outcome of a treatment. We argue that patient status in terms of positive, negative and neutral judgements can only suboptimally be judged with generic approaches, and requires specific resources in term of a lexicon and training corpus targeting clinical sentiment. To address this challenge, we manually developed a corpus based on 300 ICU nurse letters derived from a clinical database, and an annotation scheme for clinical sentiment. The paper discusses influence patterns between clinical context and clinical sentiments as well as a semi-automatic method to generate a larger annotated corpus.

1 Introduction

Attention towards opinion and sentiment analysis has been growing over the past 10 years, in which most existing methods and corpora are developed to process and analyze general domain sentiment, most prominently in Web 2.0 texts. Besides, stance detection has evolved as an extension of opinion analysis towards a given topic or object [1,2]. Sentiment analysis and stance detection represent two levels of observation on polarity: sentiment analysis deals with the detection of polarized terms at token level, whereas the detection of stance, i.e., pro, contra, and neutral attitudes towards a target address the discourse level. Information originating from these two levels necessarily influence one another, and deliver important knowledge for several applied tasks such as information retrieval, text summarization and textual entailment.

Neither sentiment analysis nor stance detection have been thoroughly scrutinized in the clinical domain yet – one of the first studies on sentiment in clinical

H. Sack et al. (Eds.): ESWC 2016 Satellite Events, LNCS 9989, pp. 311–324, 2016.
DOI: 10.1007/978-3-319-47602-5_46

documents that we are aware of is from Denecke and Deng [3] –, whereas their utility in automated analysis is undeniable. One particular challenge that clinical sentiment analysis could support is the collaborative decision making process in clinical practice. According to this, a successful diagnosis or treatment is achieved based on the collaboration and knowledge sharing of care providers from different specialities. Since the communication dialogue directly influences the quality of the treatment, misunderstandings or missing out on evidence can increase the risk of operational failure and thus threaten patient safety. Our development scenario takes place exactly in this context, motivated by the importance of automated methods for sentiment analysis in the clinical domain in order to reconstruct patient status based on consolidated judgments from care providers.

In contrast to the definition of sentiment analysis in non-clinical domains, we consider **clinical sentiment** as an event that reflects the patient status, in which the care provider additionally expresses stance towards clinical and social situations (cf. [3,4]). In this paper, we characterize the patient status event and its corresponding context objects in detail, formulating annotation entities that can accommodate sentiment values.

To our knowledge, there is no annotated corpus or sentiment dictionary available for the clinical domain, and we argue that such resources form crucial prerequisites to develop and evaluate methods for processing clinical narratives. The main contributions of this work are:

- introduction of an annotation schema for clinical sentiment,
- creation of an annotated clinical sentiment dataset,
- assessment of the annotation scheme, collecting suggestions for automated corpus generation.

The paper is structured as follows: Sect. 2 summarizes related work in corpus generation and annotation. The annotation scheme is introduced in Sect. 3. In Sect. 4, we describe our method of corpus generation procedure. The results of the annotation study are presented in Sect. 5. At the end of the paper, we discuss methods for larger corpus generation and pointers for future work.

2 Sentiment Resources and Annotation Level for Clinical Sentiments

To process texts of social media and online news, general domain sentiment lexicons (SentiWordNet [5], WordNetAffect [6], etc.) are utilized in the field of language technology, which some researchers have adapted to the medical domain. Goeuriot et al. [7] extended general domain sentiment lexicons with domain terminology from drug reviews to constitute a sentiment lexicon for pharmaceutical evaluation. In more detail, the MPQA lexicon [8] and SentiWordNet were employed as basic sources for getting opinionated terms. However, a more comprehensive study using these two lexicons based on six different data sets of health and clinical texts [3] have already proved the insufficiency of the merged sentiment lexicons with respect to clinical sentiment analysis, noting that clinical

practice has a completely different usage of sentiment terminology and emphasis in the expression. Experts are more likely to use patient reaction and clinical outcome to express polarity instead of sentiment terms seen in generic domains (adjective, adverb and part of the noun). Deng presented a merged event/entity level sentiment corpus [9].

Most of the work in sentiment analysis focused on English texts. Recently, some German lexicons [10] and annotated corpora based on a Twitter data set for sentiment analysis have been developed [11]. Besides, a German aspect-oriented sentiment corpus of smartphone app store reviews were developed by Saenger et al. [12]. Their study builds on data-driven mechanisms and conditional random field prediction to build the baseline. In addition, a newly emerged annotation schema for verb-centered sentiment inference has been reported by Klenner et al. [13], discussing sentiment features extracted from or implied by verbs.

Importantly, conventional sentiment corpora typically assign labels to spans of tokens, which does not allow for representing the influence of context on the target event (in our case, patient status). Our annotation effort aims to overcome this limitation and allows to gather influence patterns from context objects towards patient status events.

3 Annotation Scheme

Existing corpora in generic domains feature annotations at document level [14], sentence or phrase level [15] to event level [9]. Our targeted sentiment objects relate to fine-grained sentiment aspects of the patient status and its healthcare context. Given the limitations we discussed above, our work focuses on designing an annotation scheme targeting clinical texts, involving patient status as well as clinical context objects. Our intention is that the scheme should simulate the judgment process between clinical experts, and thus should be able to represent multiple perspectives pertaining to a single patient status.

We regard clinical sentiment as an event that reflects the patient bodily status, coupled with clinical, pharmaceutical, and social objects (i.e., procedures, events, states, phenomena) that influence this status:

$$Patient\,status\,event + set\,of\ \ (CI|PI|SC)$$

with CI = clinical intervention, PI = pharmaceutical intervention, SC = social connection. The main target of clinical sentiment is the event that represents the change of the patient status (see Fig. 1).

The patient status event can be further categorized into different aspects of the status. The 'concrete', i.e., physical aspect refers to descriptions of body parts, movement, input and output of patients as well as the vital signs. Next to the patient status, another type of annotation target are the context objects, which are divided into three types: clinical intervention, pharmaceutical intervention and social connections. The context objects annotation class provides important context information for the expression of the patient status. The patient event and its context objects are defined below in detail.

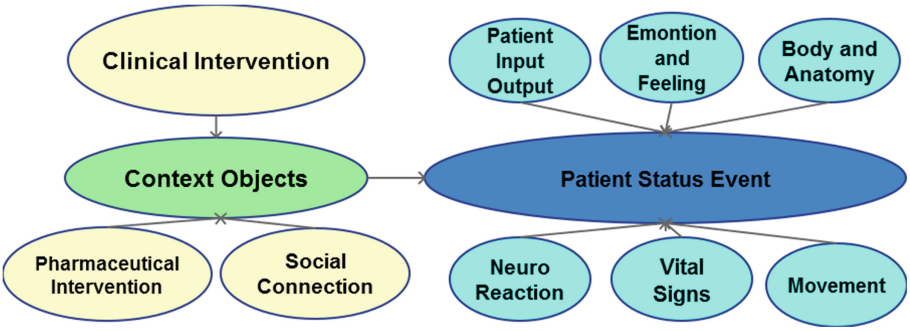

Fig. 1. Annotation schema for patient status event and context objects

Patient status event: Patient reactions, physical condition, emotional state, input and output, movement.

Context objects: (1) Clinical interventions (physical therapy, ventilator, paralyzed, balloon placement), laboratory values. (2) Pharmaceutical intervention (treatment, anesthesia, end of life care). (3) Patient's social connection (relatives' visiting, religious support).

Sentiment polarity is associated to the patient status, while the context objects additionally influence the patient status and the entire polarity outcome. Three sentiment values (positive, negative, neutral) are annotated with respect to the patient status, while the polarity of context objects is only limited to two categories: positive and negative (see Fig. 2). The neutral polarity of context objects is excluded, as it is assumed not to alter the patient status. The polarity of the context object and patient status can overlap or interfere with each other, forming an aggregated sentiment value.

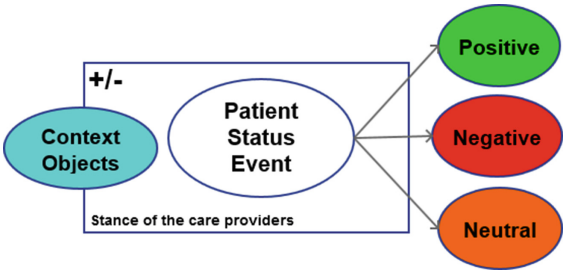

Fig. 2. Patient status event and its context objects

4 Corpus Generation

4.1 Raw Data

Our raw data set consists of 300 nurse letters from MIMIC II database. MIMIC (Multi-parameter Intelligent Monitoring in Intensive Care[1]) is an openly available dataset managed by the MIT Lab for Computational Physiology, comprising anonymized health data associated with more than 40,000 critical care patients. More specifically, we collected nurse letters from this database. A nurse letter is part of a patient record written by nurses on duty while monitoring the patients, containing information such as patient health status and response of the patient to treatments. These documents are written in a relatively free text manner. Acronyms and typos appear very often when describing e.g. vital signs, laboratory tests, medications. In order to ensure that the annotated sentiment will only depend on the treating care provider and the reflected patient status, we have identified a patient cohort with a limited set of diagnoses and treatments, based on the disease classification code (International Classification of Diseases) and procedure code.

4.2 Annotation Guidelines and Examples

The annotation guidelines are derived from the annotation scheme, and pertain to the *patient status event* and the three context objects (*clinical, pharmaceutical, social*). The patient status event can be reflected by up to six types of descriptions, as illustrated in Fig. 1.

As indicated earlier, we regard the polarity of the patient status event as positive, negative and neutral, while the polarity of context object is defined as stance of the care providers, where only two values can be assigned: for and against. Concrete examples of classes, annotation guidelines, and color coding are presented in the following section.

In general, blue marks words referring to the patient status. Red (*negative*), orange (*neutral*) and green(*positive*) reflect the polarity, the yellow color highlights the positive clinical context object whereas brown indicates the negative clinical objects. Silver represents the social connection with positive effect. Purple indicates words referring to social connection with negative effect. Magenta shows positive pharmaceutical context objects, while pink represents the negative pharmaceutical context object.

Patient Status. As illustrated in Fig. 1, the patient status (e.g. movement or emotion) is the basic type of status signs. The patient status concerns the patient's body part, reaction or status changes. Typically, the author of a clinical text directly expresses polarity to show their attitude towards the patient

[1] https://physionet.org/mimic2/.

status. The patient status is sometimes self-explained with polarity. Only minimal medical knowledge is required to conduct this kind of annotation. Given the explicit expression of polarity, existing conventional methods are able to detect and analyze these explicit sentiment expressions in clinical texts.

1. Drowsy but easily arousable seems irritated at times nods head Y/N appropriately follows simple commands perla.
   The patient status: Drowsy, arousable, irritated, nods head, follows commands
   Polarity: Easily, at times, appropriately –> positive
2. belly distended firm, abbd wound incision oozing moderate amnt serous secretions
   Patient status: Belly, abbd, secretions
   Polarity: Distended firm –> negative, wound incision in the area of abdomen –>negative, oozing moderate amnt serous –> negative
3. Multiple blisters on the trunk/chest, oozing from the swollen scrotum
   Patient status: blisters, trunk, chest, scrotum
   Polarity: multiple, oozing, swollen –> negative

Clinical Interventions. The objects for clinical interventions, readmission events and pharmaceutical interventions and their relations to the patient status need to be determined with sufficient medical knowledge and experiences. Since the patient status is not presented or only partly presented in the explicit polarity, conventional methods can not extract the implicit sentiments expressed in terms of clinical context objects. The annotators were therefore asked to pick out the clinical context objects and judge their polarity and relation towards the patient status, so that their impact could be determined. Positive and negative clinical events are determined by the annotator separately. Afterwards, the patient status under the influence of the context object has been selected.

1. Pt remains in AFIB(h/o chronic afib) rate initially 100–106 presently down to 90's AFIB with rare PVC noted.
   Patient status: Pt remains
   Clinical object: AFIB 100–106, 90's, premature ventricular contraction (pvc), negated
   Polarity: although it shows "down to" trend, the value 100–106 and 90 –> still negative, rare PVC –> positive
2. Echo done, no ASD or VSD. Found severe biventricular enlargement and biventricular systolic dysfunction, 4+ MR, torn mitral chordae.
   Patient status: Echo, biventricular, mitral chordae
   Clinical object: Atrial septal defect(ASD), Ventricular Septal Defect(VSD), enlargement, systolic dysfunction
   Polarity: Negated ASD and VSD –> positive effect –> positive in this sentence, systolic dysfunction and severe enlargement –> negative effect –> negative patient status

3. pt with ecchymolic area over trunk and arms bilateral
   Patient status: Pt, trunk, arms bilateral
   Clinical object: Ecchymolic
   Polarity: Ecchymolic area renders negative effect towards pt's trunk and arms –> negative
4. Pt remains on ventilator and IHO. Pt received esopheageal balloon placement and optimal peep was established
   Patient status: Pt
   Clinical object: Ventilator and IHO, esopheageal balloon placement, peep was established
   Polarity: On ventilator and IHO –> negative effect on patient status, balloon placement successful be conducted and optimal peeps–> positive effect –> positive

Pharmaceutical Interventions. Besides clinical intervention, the pharmaceutical therapy is another important measure to support the entire treatment process in clinical care. There are three main types of usage for the context object in this category. Firstly, the intervention can treat the patient with a certain disease, where recovery is the goal in the long run. Secondly, it can be a medication used in order to support a clinical intervention, such as stabilizing the blood pressure or regulating of the clotting time of blood during the surgery. Thirdly, it may refer to end of life care, where the aim of drug application is to relieve pain. The first two pharmaceutical interventions assist the conduction of therapy. A positive outcome and a potential recovery of patient through usage of medication can be expected, while the latter goal of intervention is only to relieve the pain whereas the negative outcome has already been determined when the medication was applied. As a result, the former two types of pharmaceutical interventions stand for a positive outcome whereas the latter one implies the negative outcome of the patient status.

1. Pt initially paralyzed on cisactricurium-atvian and mso4-, pt 4 twitches out of 4 with med on- cisactricurium to off- cont ativan and mso4- both dec by 1 mg each, pt eyes blinking to stimuli, pupils 4 cm- sluggish rx to light, not moving extremities at this time.
   Patient status: Pt, pt eyes blinking, pupils, rx, extremities
   Pharmaceutical object: Cisactricurium, atvian, mso4-
   Clinical object: paralyzed, stimuli
   Polarity: Cisatracurium-atvian has been applied for relaxation of muscle and assisting breath. The medication is used to support the treatment (positive effect). The status of the patient is as expected. Hence the entire status –> positive
2. continues to be febrile despite Tylenol, but the temp has come down from 39.2 to 38.7.
   Patient status: Implicit pt, continue, temp

Pharmaceutical object: Tylenol for fever treatment
Polarity: Although the "come down" trend from 39.2 to 38.7 has been confirmed, the body temperature is still higher than normal temperature, Tylenol has a positive effect but the patient status is still –> negative.

3. Fentanyl increased to 300 mcg/hr at 0830 for comfort. CVVHD discontinued. ECMO withdrawn by perfusionist and pt expired.
Patient status: Pt
Pharmaceutical object: Fentanyl
Clinical object: Continuous veno-venous hemodialysis (CVVHD), extracorporeal membrane oxygenation (ECMO)
Polarity: Patient expired after the application of Fentanyl for comfort (negative effect on outcome) CVVHD and ECMO withdrawn –> negative

Social Connections. The third type of objects indicate the patient social relations and relatives' visiting, which shows the social status of the patient. A patient's social connection directly influences the patient status. The activity and implicit sentiments for the appearance of certain social connections entail positive and negative effect towards the patient status.

1. Wife and brother in to visit, discussed with doctor and comforted pt.
Patient status: Pt
Social connections: Brother and wife
Polarity: Come to support and comfort –>Positive
2. Family called in middle of night and in to see pt. Priest called in and last rights given.
Patient status: Pt
Social connection: Family, priest
Polarity: Called in middle of night, last right –> negative
3. wife and daughter into visit and will return this evening daughter upset when leaving
Social connection: Wife, daughter
Polarity: Upset –> negative

5 Annotation Task and Results

5.1 Annotators and Annotation Task

To generate a high-quality corpus, clinical experts were asked to supply the annotations, with whom the annotation schema was jointly improved after initial discussions.

In a pre-annotation phase, all five annotators marked the same 25 nurse letters individually in an iterative process: In a first step, they were asked to label four entities (patient status and three context objects). Then, they had to identify and highlight the polarity terms and the polarity of the context objects (*+ effect* or *- effect*). At last, they were requested to connect the polarity term and

Table 1. Background of the annotators

ID	Knowledge background	Experience	Specialities
A1	Medicine	5 years	Cardiology, surgery
A2	Medicine	6 years	ENT, cardiology
A3	Pharma engineering	6 years	Pathology, targeted drug delivery
A4	Medical informatics	4 years	Text mining, medical ontology
A5	Medical informatics	3.5 years	Knowledge modeling

context objects with the patient status entity to show their relationship. Ehost[2] was the tool used in this annotation task. Annotation was conducted independently by each annotator without discussions. However, annotators were allowed to use the Internet to look up unfamiliar terms. The annotator's background is presented in Table 1: two students of medicine, two medical informatics students and one pharmacy student took part in the annotation. During the annotation, annotators received the annotation guideline and text corpus, which comprises the same texts to be annotated by the five annotators. One introduction seminar was held to explain the annotation guidelines and two rounds of pre-annotation with adjudications were conducted to train the annotators.

After this pre-stage, the annotation guidelines were verified and extended. Subsequently, the remaining corpus was annotated by two annotators who were selected based on good inter-annotator agreement and accordance with the guidelines (annotators A1 and A2 (medical annotators)).

5.2 Annotation Assessment Methods

The objective of annotation assessment was (i) to determine which background is necessary to perform the annotation, and (ii) to assess the recognition rate. The following methods have been applied: the average annotation agreements in the two pre-annotations based on the subset of 25 nurse letters were compared in pairs of annotators. The inter annotator agreement (IAA) was calculated as Kappa statistic value. IAA scores were calculated under exact match criteria (i.e., the annotations should match completely). The recognition of patient status events and context objects from the five annotators were compared with the adjudicated event list (gold standard). The event list was generated through majority voting and discussion among the five annotators. A correct recognition should agree on both the span and polarity (*+/− effect*) to this span.

During the adjudication, a subset of 25 documents was annotated by all five annotators. The recognition of patient status events and context objects from five annotators are compared with the adjudicated event list (gold standards) (see Table 2). A correct recognition requires the correct annotation of both the span and the polarity (+/− effect) to this span. The recognition rate is $recognized\,events + polarity/adjudicated\,events\,in\,list$.

[2] https://code.google.com/p/ehost/.

The final corpus consists of the consensus of the annotations. Altogether, the 300 documents were annotated with 7,080 patient status events, 2,040 clinical objects, 1,380 pharmaceutical events and 535 social context objects.

5.3 Recognition Rate and Inter-annotator Agreement

The medical annotators (A1 and A2) have achieved a high average recognition rate in all three categories. The pharma-annotator has recognized the highest percentage of pharmaceutical interventions. Generally, the social context objects are well detected by all annotator groups.

Table 2. Statistics of event annotation in two adjudications

Class statistics (total number)	A1	A2	A3	A4	A5
Patient status event (118)	109	106	83	118	79
Clinical context object (34)	33	30	9	17	7
Pharmaceutical context object (23)	20	21	22	15	10
Social context object (9)	9	8	9	9	8

As can be seen in Table 3, the average agreement between medical annotators (A1, A2) have reached 79.8 %. Annotations of A5 received the lowest agreement with others, only 22–34.2 %. The other pairs of annotators had an overlap of around 50 %. Further, the medical annotators (A1 and A2) have achieved a high average recognition rate in all three categories. The pharmacy-annotator A3 has recognized the highest percentage of pharmaceutical interventions. Generally, the social context objects are well detected by all annotator groups.

Table 3. Two-way average inter annotator agreement (Kappa) in two adjudications (all selected classes)

Annotators	A5	A4	A3	A2
A1	0.342	0.494	0.493	0.798
A2	0.291	0.511	0.532	
A3	0.22	0.36		
A4	0.24			

In summary, the statistics show that annotators with medical education background reached a high agreement and recognized more clinical objects than other annotator groups. Moreover, medical annotators can interpret the polarity and its influences on the patient status in a more professional way. For annotators A4 and A5, the other annotations were difficult given the missing medical background.

6 Discussions

6.1 Annotation Errors

Most of the errors that occurred during the annotations are missed classes. This type of error occurred at all annotators irrespective of knowledge backgrounds. We hypothesize that it is mainly due to lack of concentration or understanding of the annotation guidelines.

The second error group pertains to the determination of boundaries of entities Since our goal is to use this corpus for testing sentiment analysis methods, assigning the correct boundary for each class is vitally important, whereby only the target events should be labeled. With respect to the latter, we recognized inconsistencies, e.g. some patient status events such as (pt, abd) appeared several times in one document, but the annotators have only labeled the first occurrence. With respect to the former, some annotators simply marked all the tokens around the target span, often including stop words, conjunctions or other terms without meaning.

Moreover, acronyms with different meanings have proven to be problematic for our annotators. For example, the acronym *BS* can either refer to *breath sounds* or to *blood sugar* in different contexts. Our annotators have only recognized one of these meanings, which caused mistakes in polarity determination. After two rounds of pre-annotation and reversioning, these kinds of disagreements have been reduced.

6.2 Feedback from Annotators

We collected feedback from the annotators during the annotation process for future consideration. Disagreement among annotators was mainly caused by different event recognitions, or idiosynchratic interpretation of the annotation guidelines. For medical annotators, if they have identified the clinical objects, they were typically able to identify the correct polarity and relations between the patient status and context objects, whereas this has proven to be the most difficult part for the annotators with other backgrounds. A non-medical annotator could easily identify the span for context objects, but could not determine the polarity towards the patient status. The low recognition rates for context objects of non-medical annotators presented in Table 2 are mainly due to this.

Another special situation observed by annotators are complex influence patterns that cannot be described by the current annotation schema. For the current study, it was assumed that the context object is going to influence the polarity of patient status, whereby the annotation scope is limited to the sentence level. However, the influences accumulated from multiple objects may each have their impact on the patient status, especially for patients with multiple diseases. Sentiment may therefore need to be determined by the evidences from several annotation levels.

Further difficulties are caused by using unstandardized values or units for medical measurements and unclear precondition of the value pairs in text. For

example, *ABG 7.38/49/71/30* is the value for arterial blood gas, but the writing convention of the numeric value is unclear to the annotators. Since the ranges of the value are quite similar, it is impossible to differentiate it without additional hints. Besides, some values are noted with different units such as *mmol/L* or *mg/dl*. These kinds of ambiguities have made the judgment of polarity difficult. A more concrete interpretation should be provided to reach a better understanding about the medical concepts.

Sometimes, the annotation of trends has also rendered a high disagreement between annotators. For example, the phrase *MAP 55 -70, MAP increased* should be labeled as positive outcome for the patient status, if the status is judged according to the trend of value. The value of MAP 55 on its own needs to be considered negative. However, some of the annotators have annotated these values separately, while other just labeled the negative boundary of these both values. Again, this resulted in inconsistent annotations.

6.3 Towards Larger Sentiment Corpora

The annotated text does not contain complete sentences, but phrases with abbreviations (e.g. "Pt" refers to "Patient"). They are rather enumerations of patient status events, their polarity and context objects. These characteristics hamper the automatic processing or annotation as it could be realized using machine learning. The effectiveness of machine learning based methods relies on the size of the training corpus. The manual method can however only reach a limited corpus size. There are mainly two types of methods that can be put to use for corpus extension: data-driven and rule-based [16]. Rule-based approaches are mainly based on a lexicon, ontology or a dictionary, while the data-driven methods are exploiting unsupervised machine learning methods applied to an unannotated text corpus. The text snippets with similiar syntactic and semantic structure will be extracted and grouped together. A hybrid method combining both, rule-based and data-driven methods, could even achieve more reliable results. More specifically, after extending our corpus of 300 documents with manually generated annotations, it could work as a bootstrapping data set for unsupervised learning. In this way, a larger text corpus can be developed. For instance, the corpus can be enlarged gradually through the distributional similarity comparison between manually annotated corpus and unannotated text with the same cohort distribution.

7 Conclusions

In this paper, we introduced an annotation scheme for clinical sentiment and reported an annotation study that exploited the scheme for annotating clinical documents. We argued that existing resources have been developed for generic texts and are not suitable for representing the influence of special clinical context objects on our targeted event: patient status sentiment.

To fill this gap, we generated an annotated corpus of clinical documents that enabled analysis of relations in terms of influence patterns between patient status and context objects. Several terminology lists and acronym lists with corresponding polarities were established, constituting the first step towards a sentiment lexicon for the clinical domain.

Furthermore, the annotation and validation of 300 nurse letters have been prepared[3]. Our annotation scheme and annotated corpus have unveiled important characteristics of clinical sentiment.

Annotation mistakes and inconsistencies have been identified based on annotator feedback and analysing the major disagreements between annotators leading us to a revised annotation scheme and guidelines. As next step, the corpus will be used to generate a clinical sentiment lexicon semi-automatically and it can also be employed to develop and to evaluate methods for sentiment analysis considering the clinical context objects.

References

1. Sridhar, D., Getoor, L., Walker, M.: Collective stance classification of posts in online debate forums. In: ACL 2014, p. 109 (2014)
2. Somasundaran, S., Wiebe, J.: Recognizing stances in online debates. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL, the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1, ACL 2009, Stroudsburg, PA, USA, pp. 226–234. Association for Computational Linguistics (2011)
3. Denecke, K., Deng, Y.: Sentiment analysis in medical settings. Artif. Intell. Med. **64**(1), 17–27 (2015)
4. Deng, Y., Stoehr, M., Denecke, K.: Retrieving attitudes: sentiment analysis from clinical narratives. In: Proceedings of the Medical Information Retrieval Workshop at SIGIR Co-located with the 37th Annual International ACM SIGIR Conference (ACM SIGIR 2014), Gold Coast, Australia, 11 July 2014, pp. 12–15 (2014)
5. Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: LREC (2010)
6. Strapparava, C., Valitutti, A.: Wordnet affect: an affective extension of wordnet. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, Lisbon, Portugal, 26–28 May 2004 (2004)
7. Goeuriot, L., Na, J.-C., Kyaing, W.Y.M., Khoo, C., Chang, Y.-K., Theng, Y.-L., Kim, J.-J.: Sentiment lexicons for health-related opinion mining. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, IHI 2012, pp. 219–226, New York, NY, USA. ACM (2012)
8. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT 2005, pp. 347–354, Stroudsburg, PA, USA. Association for Computational Linguistics (2005)

[3] The corpus is available on request.

9. Deng, L., Wiebe, J.: MPQA 3.0: an entity/event-level sentiment corpus. In: NAACL HLT 2015, The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, Colorado, USA, 31 May–5 June 2015, pp. 1323–1328 (2015)
10. Emerson, G., Declerck, T.: SentiMerge: combining sentiment lexicons in a Bayesian framework. In: Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing. The COLING 2014 Organizing Committe, August 2014
11. Sidarenka, U.: PotTS: the potsdam twitter sentiment corpus. In: Calzolari, N., et al. (eds.) Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2016), Portoro++, Slovenia, European Language Resources Association (ELRA), May 2016
12. Snger, M., Leser, U., Kemmerer, S., Adolphs, P., Klinger, R.: SCARE - the sentiment corpus of app. reviews with fine-grained annotations in German. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portoro v z, Slovenia, May 2016. European Language Resources Association (ELRA) (2016)
13. Klenner, M., Amsler, M.: Sentiframes: a resource for verb-centered German sentiment inference. In: Calzolari, N., (Conference Chair), Choukri, K., Declerck, T., Grobelnik, M., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (eds.) Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Paris, France. European Language Resources Association (ELRA), May 2016
14. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. CoRR, cs.CL/0205070 (2002)
15. McDonald, R.T., Hannan, K., Neylon, T., Wells, M., Reynar, J.C.: Structured models for fine-to-coarse sentiment analysis. In: ACL 2007, Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic, 23–30 June 2007 (2007)
16. Manishina, E., Jabaian, B., Huet, S., Lefevre, F.: Automatic corpus extension for data-driven natural language generation. In: Calzolari, N. (Conference Chair), Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S. (eds.) Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Paris, France. European Language Resources Association (ELRA), May 2016