Don't Let Notes Be Misunderstood: A Negation Detection Method for Assessing Risk of Suicide in Mental Health Records

George Gkotsis[1], **Sumithra Velupillai**[1,2], **Anika Oellrich**[1],
Harry Dean[1], **Maria Liakata**[3] **and Rina Dutta**[1]
[1] IoPPN, King's College London
`firstname.lastname@kcl.ac.uk`
[2] School of Computer Science and Communication, KTH, Stockholm
[3] Department of Computer Science, University of Warwick
`m.liakata@warwick.ac.uk`

Abstract

Mental Health Records (MHRs) contain free-text documentation about patients' suicide and suicidality. In this paper, we address the problem of determining whether grammatic variants (inflections) of the word "suicide" are affirmed or negated. To achieve this, we populate and annotate a dataset with over 6,000 sentences originating from a large repository of MHRs. The resulting dataset has high Inter-Annotator Agreement (κ 0.93). Furthermore, we develop and propose a negation detection method that leverages syntactic features of text[1]. Using parse trees, we build a set of basic rules that rely on minimum domain knowledge and render the problem as binary classification (affirmed vs. negated). Since the overall goal is to identify patients who are expected to be at high risk of suicide, we focus on the evaluation of positive (affirmed) cases as determined by our classifier. Our negation detection approach yields a recall (sensitivity) value of 94.6% for the positive cases and an overall accuracy value of 91.9%. We believe that our approach can be integrated with other clinical Natural Language Processing tools in order to further advance information extraction capabilities.

1 Introduction

Suicide is a leading cause of death globally. Approximately 10% of people report having suicidal thoughts at some point in their lives (Nock et al., 2013) and each year 0.3% of the general population

make a suicide attempt (Borges et al., 2010). Mental disorders (particularly depression, substance abuse, schizophrenia and other psychoses) are associated with approximately 90% of all suicides (Arsenault-Lapierre et al., 2004). Assessment of suicide risk is therefore routine practice for clinicians in mental health services, but it is notoriously inaccurate as well as time-consuming (Ryan et al., 2010). Although individual risk factors associated with suicide have been reported in depth (e.g. Steeg et al., 2016), integrating them into an algorithm to analyse signatures of suicidality has been beset with difficulties.

Clinicians document the progress of mental health patients in Mental Health Records (MHRs), predominantly using free text, with sparse structured information. This poses new and interesting challenges for clinical Natural Language Processing (NLP) tool development that could assist in identifying which patients are most at risk of suicide, and when (Haerian et al., 2012; Westgate et al., 2015). Developing a classifier to identify times of greatest risk for suicide at an individual patient level (Kessler et al., 2015; Niculescu et al., 2015) would assist in targeting suicide prevention strategies to those patients who are most vulnerable (Mann et al., 2005). Negation can be used to denote absence or inversion of concepts. As a linguistic feature it can play a prominent role in monitoring both symptom context and risk in psychological conditions (Chu et al., 2006). For instance, one study found that almost 50% of the clinical concepts in narrative reports were negated (Chapman et al., 2001).

In this paper, we address the long-term goal of de-

---

[1] `https://github.com/gkotsis/negation-detection`

veloping improved information retrieval systems for clinicians and researchers, with a specific focus on suicide risk assessment. To achieve this, we focus on the problem of determining negation concerning mentions of suicide. Clinical concepts are most often defined as nouns ("suicide") or noun phrases ("suicide ideation"), and a negation detection algorithm needs to model the surrounding context to correctly ascertain whether the concept is negated or not ("patient has never expressed any suicidal ideation" vs. "patient expressing suicidal ideation").

Modelling the surrounding context of words can be done in different ways. Our work is motivated by the advances in Probabilistic Context Free Grammar Parsers (PCFGs), which allow us express widely generalisable negation patterns in terms of restrictions on constituents. A solution to negation detection that uses different aspects of linguistic structure can provide richer and more informative features. As a next step we want to extend our work and extract other important features from MHRs, such as the statements in the form of subject-predicate-object, temporal characteristics or degree of suicidality.

We propose an automated method for determining negation in relation to documented suicidality in MHRs. Our negation detection algorithm relies on syntactic information and is applied and evaluated on a manually annotated corpus of sentences containing mentions of *suicide*, or inflections thereof, from a repository of mental health notes. Our paper makes the following contributions:

- we create an annotated dataset containing over 6,000 sentences with mentions of suicide (affirmed or negated),

- we propose a new method for incorporating syntactic information for automatically determining whether a mention of interest is affirmed or negated.

To our knowledge, no previous research has addressed the problem of negation detection in the domain of MHRs and suicidality.

2 Related Work

Negation detection has long been recognized as a crucial part of improved information extraction sys-

tem development in the biomedical and clinical NLP research community (Morante and Sporleder, 2010), as negated concepts alter the meaning of what is extracted. In general, successful approaches have relied on terminological resources defining negation keywords, concepts, and other rules for determining negation values (lexicon- and rule-based), or based on machine learning methods, where models have been built based on large sets of manually annotated training examples (Meystre et al., 2008). In both cases, manually annotated corpora are needed for training, developing and evaluation. Systems are usually evaluated by calculating precision (positive predictive value) and recall (sensitivity).

One of the earliest, and still widely used, negation detection algorithms is NegEx, which determines whether or not a specified target concept is negated by searching for predefined lexical negation (e.g. *not*), pseudonegation (*no increase*) and conjunction (*except*) cues surrounding the concept (6 words before and after). On a test set of 1000 discharge summary sentences (1235 concepts), NegEx resulted in 84.5% precision and 77.8% recall (Chapman et al., 2001). The NegEx algorithm has also been extended to handle further semantic modifiers (e.g. uncertainty, experiencer) and a wider surrounding context with improved results (overall average precision 94%, recall 92%) when evaluated on 120 reports of six different types (Harkema et al., 2009), and to also perform document-level classifications including semantic modifiers (Chapman et al., 2011).

Lexical approaches relying on surface features are limited in that the *linguistic* relation between the target term and the negation is not captured. NegFinder (Mutalik et al., 2001) is a system that, in addition to defining lexical cues, uses a context free grammar parser for situations where the distance between a target term and negation is far. This approach resulted in 95.7% recall and 91.8% precision when evaluated on 1869 concepts from 10 documents. Syntactic parsers can provide a richer representation of the relationship between words in a sentence, which has been utilised also for negation detection solutions. For instance, DepNeg (Sohn et al., 2012) rely on the syntactic context of a target concept and negation cue, which improved negation detection performance, in particular for reducing the number of false positives (Type I errors) on a test set

of 160 Mayo clinical notes (96.6% precision,73.9% recall). Similarly, DEEPEN (Mehrabi et al., 2015) adds a step after applying NegEx on clinical notes. Syntactic information from a dependency parse tree is then used in a number of rules to determine the final negation value, resulting in precision of 89.2-96.6% and recall of 73.8-96.3% on two different clinical datasets and three different types of clinical concepts.

Machine learning approaches have also been applied to the negation detection problem with success. These approaches rely on the access to training data, which has been provided within the framework of shared tasks such as the 2010 i2b2 challenge (Uzuner et al., 2011) for clinical text, the BioScope (Vincze et al., 2008) corpus in the CoNLL-2010 shared task for biomedical research articles as well as clinical text, the ShARe corpus (Pradhan, Sameer and Elhadad, Noémie and South, Brett R and Martinez, David and Christensen, Lee and Vogel, Amy and Suominen, Hanna and Chapman, Wendy W and Savova, Guergana, 2015) in the ShARe/CLEF eHealth and SemEval challenges, and the GENIA corpus (Kim et al., 2003) in the BioNLP'09 shared task.

A comprehensive study on current state-of-the-art negation detection algorithms and their performance on different corpora is presented by Wu et al (2014). As is concluded in this study, none of the existing state-of-the-art systems are guaranteed to work well on a new domain or corpus, and there are still open issues when it comes to creating a generalizable negation detection solution.

3 Proposed framework

Two main stages were employed in this study: 1) data collection and creation of a MHR corpus with annotations of concepts marked as negated or affirmed, and 2) the development of our proposed methodology to detect negations for the purpose of assessing risk of suicide from MHRs[2]. Figure 1 provides an overview of the workflow we employed in this study. We discuss these stages in detail below.

3.1 Dataset and annotation

Pseudonymised and de-identified mental health records of all patients (both in and outpatients) from the Clinical Record Interactive Search (CRIS) database were used (Perera et al., 2016). CRIS has records from the South London and Maudsley NHS Foundation Trust (SLaM), one of the largest mental health providers in Europe. SLaM covers the Lambeth, Southwark, Lewisham and Croydon boroughs in South London. CRIS has full ethical approval as a database for secondary analysis (Oxford REC C, reference 08/H0606/71+5) under a robust, patient-led and externally approved governance structure. Currently, CRIS contains mental health records for around 226K patients, and approximately 18.6 million documents with free text. Out of these documents, 783K contain at least one mention of "suicid*" (111K patients). Monitoring suicide risk is an important task for mental health teams, and therefore use of the term "suicid*" was expected to be common.

The annotation task was defined on a concept-level: each target concept ("suicid*") in a sentence was to be marked as either *negative* (negated mention, e.g."denies suicidal thoughts") or *positive* (affirmed mention, e.g."patient with suicidal thoughts")[3]. In clinical narratives, there are cases where this distinction is not necessarily straightforward. For instance, in a sentence like "low risk of suicide based on current mental state", a clinician may be inclined to interpret this as *negated* (this is not a patient at risk of suicide), while a linguistic interpretation would be that this is *not negated* (there is no linguistic negation marker in this example). In this study, the annotators were asked to focus on linguistic negation markers, and disregard clinical interpretations, in order to create a well-defined and unambiguously annotated corpus. They were also instructed to annotate mentions of suicide regardless of whether comments concerned the patient, their family member or a friend.

A collection of 5000 randomly selected MHRs was extracted, divided (segmented) into individual sentences, keeping only sentences containing the target concept. This resulted in a corpus of *6066*

---

[2]The source code for our tool is available at `https://github.com/gkotsis/negation-detection`.

[3]Annotators were also allowed to assign the value "Irrelevant" for uncertain or otherwise problematic cases.
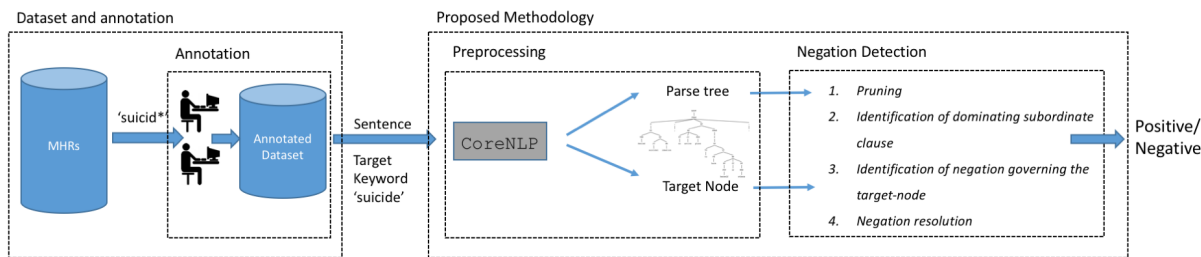
Figure 1: Workflow illustration. Two main stages were employed: 1) data extraction and annotation for corpus creation, and 2) development of the proposed methodology for negation resolution.

sentence-instances.

One annotator (domain expert) annotated the entire corpus. To assess the feasibility and estimate the upper performance levels that could be expected from an automated system, we employed a double-annotation procedure on a portion of the corpus. We calculated the Inter-Annotator Agreement (IAA) in order to examine if the task is well-defined. A randomly selected subset (1244 sentences, >20% of the corpus) was given to a second annotator (NLP researcher) to calculate IAA.

The IAA analysis showed that our annotators agreed on 97.9% of the instances (Cohen's κ 0.93, agreement over 1218 sentences). From this result, we concluded that: 1) the annotation task was indeed defined in an unambiguous way and was well-understood by humans, and 2) there are still some cases that are inherently difficult to assess, due to a degree of ambiguity, which is to be expected in real-world settings. The final corpus contains 2941 sentences annotated as positive (affirmation of suicide) and 3125 annotated as negative (i.e. suicide negated, 48.5% - 51.5% positive to negative ratio).

3.2 Proposed method for negation detection

Our proposed methodology consisted of two steps: 1) preprocessing and formatting the data, and 2) execution of the negation resolution task.

3.2.1 Preprocessing

Each sentence was preprocessed in order to prepare the input for the negation resolution algorithm in a suitable format: a syntactic representation (parse tree) and the target token ("suicide").

Our proposed methodology makes use of *constituency-based parse trees*. A constituency tree is a tree that categorises nodes as grammatical

constituents (e.g. NP, VP) using the Penn Treebank tagset (Marcus et al., 1993). Nodes are classified either as leaf nodes with terminal categories (such as noun, verb, adjective etc.) or interior nodes with non-terminal categories (e.g. verb phrase, sentence etc.). Therefore, constituency trees are quite expressive and provide us with rich information concerning the roles of elements and chunks of elements found in written natural language. In this study, we used the Probabilistic Context Free Grammar (PCFG) parser that is built into the Stanford Core NLP toolkit (Klein and Manning, 2003), a variant on the probabilistic CKY algorithm, which produces a constituency parse tree representation for each sentence. As will become clear in the sections below, we found constituency parse trees particularly useful in modelling global grammatical constraints on the scope of negation and in the context of surface mentions of the word "suicide". Such constraints would have been harder to express using dependency parsers, although we do plan to incorporate dependency triples in future analysis.

In addition, the target token was also searched for in the sentence tree, in a reduced form (by applying stemming) in order to identify all possible inflections of the word "suicide".

3.2.2 Negation resolution algorithm

Similar to other approaches, we reduced the problem of negation resolution to the problem of identifying the *scope of negation*. The basic premise of scope-based negation resolution algorithms is that a list of negation words (or phrases) is provided. In this study, we defined a list of 15 negation cues[4] based on an initial manual analysis of the data. Once

---

[4]See supplement for the complete list.

a negation cue is found in the syntactic tree, a scope-based algorithm attempts to mark the concept that is affected by this negation word.

Our approach starts from the target-node ("suicide") and *traverses* the tree moving upwards and visiting nodes of the tree accordingly. The function and role of each node-element in relation to negation resolution during this traversal is then considered through a set of operations:

- *Pruning* refers to the removal of interior nodes that are not expected to have an impact on the final output. Figure 2 shows an example of tree pruning. Node pruning occurs when two conditions are met: a) a node is tagged with subordinate conjunctions or clause-related Treebank categories, and b) the node and none of its children contain the target node. After pruning, the remainder of the tree is further processed.

- *Identification of the dominating subordinate clause* is an action that also results in the removal of selected nodes, but with an important difference: it leads to the generation of a new *subtree*. During pruning, once a node is considered irrelevant, *all* of its children are removed. Here, the aim is to isolate the target node from higher level nodes that do not propagate the negation to the lower levels of the tree, hence leading to a new *subtree*. For a node to be considered a root candidate in the new tree, it has to be classified as a "subordinate clause" (SBAR) and the subtree must contain the target node. Figure 3 illustrates an example of this operation, where the highlighted segment shows the nodes that are *not* participating in the formation of the new tree. The new tree is kept and used for further processing in the subsequent steps.

- *Identification of negation governing the target-node* aims to deal with tree structures, such as conjunctions, where negations can be propagated to the target node. Intuitively, the traversal continues upwards as long as the initial context remains the same. If a sentence ("S") is found, a stopping condition is met and only the node-child of the stopping node is examined. In this context, the algorithm will flag the target node as negated regardless of the negation-

words counted (at least one negation stopword must be present, see final step below for counting negations). If a negation word is found, its *relative* position with regards to the target node is considered. When the negation word is to the *left* of the target node, the target is considered negated. This approach allows us to capture cases of potential ambiguity. Figure 4 contains an example where the negation word is contained in a sibling noun phrase (NP), to the left of the target NP.

- *Negation resolution* is the last operation that is applied on the final version of the tree, after the previous operations have been executed. This step simply counts the number of negation words in the tree. If the number is odd, the algorithm predicts a "negative" value, else it returns "positive". This counting step allows us to take into account cases where multiple negations are propagated to the target node and are cancelling each other.
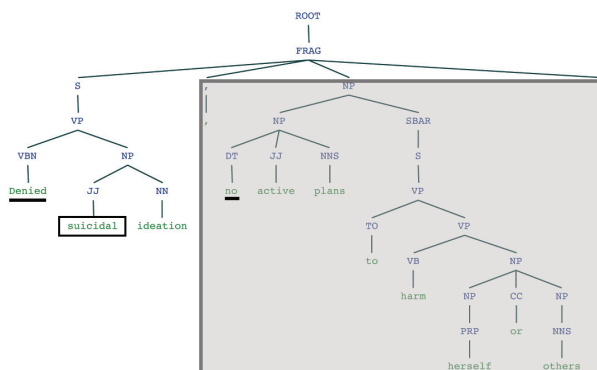

Figure 2: An example of pruning. The highlighted fragment is considered to be out of the scope of the target-node ("suicidal") and is therefore removed.

3.2.3 Evaluation metrics

We evaluate results with precision (positive predictive value), recall (sensitivity), F-measure (harmonic mean of precision and recall) and accuracy (correct classifications over all classifications). We also compare our algorithm against two other, openly available, lexical negation resolution approaches: pyConTextNLP (Chapman et al., 2011)[5]

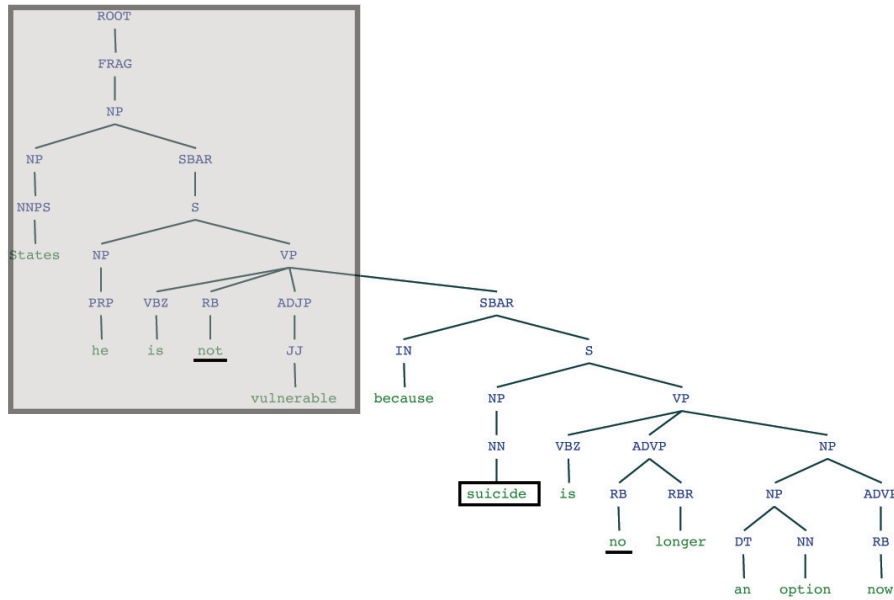
[5] available by pip install.

Figure 3: An example of a dominating subordinate clause. The highlighted fragment shows the nodes that are not included in the formation of the new tree.
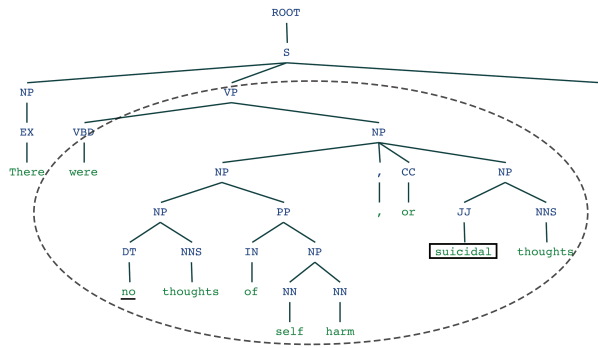

Figure 4: An example of a governing node. The highlighted verb phrase governs the target-node ("suicidal") and also exhibits ellipsis.

and the NegEx (Chapman et al., 2001) 2009 python implementation[6]. Since these approaches depend on lists of negation and termination cues, we compare results with three configurations: 1) NegEx as obtained from the online code repository, 2) pyConTextNLP with the negation and termination cues from configuration 1 (pyConTextNLP-N), and 3) pyConTextNLP with the negation and termination cues created for our proposed approach (pyConTextNLP-O).

Furthermore, we provide a more detailed performance analysis with regards to the length (in words) of a sentence, since the syntactic parses are more error-prone for longer sentences (lower accuracy and time-out requests).

4 Negation detection results

Our study focusses on assessing the risk of suicide based on information contained in mental health records. Since the overall goal is to identify patients who are expected to be at high risk of suicide, we focus on the evaluation of *positive* (affirmed) cases as determined by our classifier, i.e. cases without negation or where the negation does

not govern the target keyword ("suicide"). These affirmed cases are where, according to the clinician, patients have entered into a heightened state of risk (risk assessment), they must be re-assessed and have their suicide risk updated frequently like a time-dependent "weather forecast" (Bryan and Rudd, 2006). Short-term risk assessments, like weather forecasts, are much more accurate than longer-term assessments (Simon, 1992).

Table 1 presents the confusion matrix for our classifier when compared with the manual annotations. In addition, the numbers as obtained from pyConTextNLP, when installed and used with the NegEx lexicon (pyConTextNLP-N), are shown in brackets. The table shows that both classifiers produce few Type I (false positive) and II (false negative) errors. Our proposed approach manages to correctly identify more positive/affirmed cases (2782 vs. 2733), albeit at a higher cost compared to pyConTextNLP-N (more false positives, 331 vs. 172). On the other hand, our proposed solution identifies fewer negative cases (i.e. 159 instances wrongly identified as negative vs. 208). In summary, pyConTextNLP-N has a higher bias towards negative instances, which results in lower recall for the positive instances, but higher accuracy overall.

		Class	
		Positive	Negative
Prediction	Positive	**2782** (2733)	331 (172)
	Negative	**159** (208)	2794 (2953)
	Total	2941	3125

Table 1: Confusion matrix: Manually annotated (Class) vs. predicted (Prediction) instances from our proposed algorithm. Numbers in brackets report on pyContextNLP-N (negation lexicon from NegEx).

Table 2 reports on the precision, recall, F-Measure and accuracy for the positive (affirmed) cases when using the four different negation resolution systems and configurations[7]. Results are overall very similar, and very high, except perhaps for pyContextNLP-O (83.2% accuracy) which demonstrates how important the lexical resources and definitions are for improved performance. This also means that the high results for our proposed approach is promising, as

---

[7]Note that in our evaluation we have not selected a specific tool as a baseline.

there is less need for manual creation and curation of lexical resources. Furthermore, this result also reflects characteristics of this data: mentions of suicide in mental health records are negated in a fairly consistent and unambiguous way.

	P	R	FM	A
NegEx	93.4	92.1	92.8	93
pyConTextNLP-N	**94.1**	92.9	**93.5**	**93.7**
pyConTextNLP-O	80.7	86	83.2	83.2
Proposed	89.4	**94.6**	91.9	91.9

Table 2: Results for negation resolution using different tools: NegEx, PyConTextNLP with negation and termination cues from the original NegEx code (pyConTextNLP-N), PyConTextNLP with negation and termination cues from our proposed approach (pyConTextNLP-O), and our proposed approach (Proposed). Precision (P), Recall (R), F-Measure (FM) and Accuracy (A) report on the case of positive/affirmed instances.

Although the overall results are high, there are some aspects that could be studied further, for instance the effect of preprocessing. There are a few instances where the sentence chunking failed, which poses a severe challenge for the syntactic analysis. Figure 5 presents the cumulative word count of sentence instances. The vast majority of the instances contain less than 50 words, but there are a few instances where a "sentence" contains more than 300 words. These long sentences turned out to be complete documents. Clinical text is known for being noisy and hard to correctly tokenise in many cases, and instead of removing these cases, we decided to keep them so as to have a closer to real-world assessment of the efficiency of our methodology.
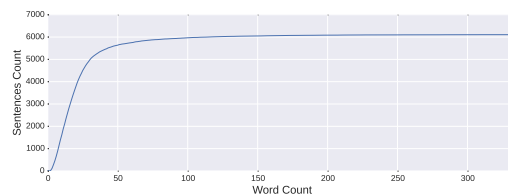

Figure 5: Cumulative word count of sentences in our dataset.

Furthermore, to understand the effect of keeping incorrectly tokenised sentences, we studied the performance of our proposed tool based on sentence length (as defined by word count). Figure 6 presents the mean cumulative accuracy of our algorithm with

regards to the word count of the sentences. The figure shows that the system performance is significantly higher for shorter sentences. This performance slowly declines as lengthier sentences are included[8].

5   Discussion and Future Work

The corpus that we have created for this study is, to our knowledge, the first of its kind, and also of a considerable size[9]. At the same time, for the annotation process, decisions were made that introduce some limitations in our study design (e.g. linguistic focus, target concepts). Hence, the results presented in this work are generalisable but are, to a certain degree, overestimating clinical reality, real world applicability and generalisability. Despite these limitations, we believe that our analysis of the dataset sheds light in the broad area of suicide risk assessment from MHRs.

Furthermore, our proposed negation resolution approach is competitive when compared to state-of-the-art tools. In particular, it performs slightly better for correctly classifying positive/affirmed mentions as opposed to negated mentions. This is a welcomed outcome, since in our use case we aim to focus on patients at risk for suicidal behaviour. Additionally, an early observation concerning its performance is that our tool is better for cases of short, simple and properly punctuated text, which is something that could be addressed by better writing support in MHR systems, and by the authors of health record notes. Small, incremental changes in the documentation creation process can increase the quality of the clinical NLP tools' output considerably.

Comparing our results to previous research is not straightforward, since we are using a new corpus and we study negation resolution on a new domain. However, in general, our results are very promising and in line with, or above, previously reported results on negation detection. For instance, NegEx, when applied on a variety of corpora and use cases, has resulted in precision ranging from 84.5% – 94% (Chapman et al., 2001; Harkema et al.,

2009). When compared to approaches that also incorporate syntactic information in the negation resolution algorithm, both DepNeg (Sohn et al., 2012) and DEEPEN (Mehrabi et al., 2015) report high overall results when evaluated on different types of clinical corpora, in particular for reducing false positives (i.e. overgenerating predictions of negation). However, DEEPEN is biased to the performance of NegEx, whereas our proposed approach is completely standalone. Furthermore, previous research studies report results with an emphasis on performance on negation detection, not on detecting affirmed instances, which is a crucial issue in our case.

There are several areas in which we plan to extend this work. As already discussed, negation detection tools can exhibit a drop in performance when applied on different corpora (Wu et al., 2014). In our approach, the dictionary of negation keywords is much smaller compared to other approaches. We believe that this feature is a sign that our method is robust and can be generalisable. We intend to evaluate the approach on other datasets – clinical as well as other text types, e.g. biomedical articles and abstracts, to assess the generalisability of our proposed system. Moreover, our approach to use parse trees allows us to extend our work and extract further semantic and syntactic layers of information. In particular, we plan to focus on the extraction of statements (e.g. in the form of subject-predicate-object), the identification of temporal characteristics as well as the extraction of the degree of suicidality. Most importantly, we also plan to use this algorithm for suicide risk modelling. We already have a cohort study in progress, where this system will be central to the model.

6   Conclusions

Free text found in Mental Health Records (MHRs) is a rich source of information for clinicians. In this paper, we focus on the problem of suicide risk assessment by studying mentions of suicide in MHRs. To that end, we 1) produced and presented a new corpus of MHRs annotated for negation or affirmation of mentions of suicidality, with high Inter-Annotator Agreement, and 2) developed an algorithm for negation resolution relying on constituency parse tree information. The results of our study confirm the

---

[8]In Figure 6, notice that we have clipped the X-axis to show sentences of up to 80 words. The plotted line converges to the overall mean accuracy of 91.9, as reported in Table 2.

[9]Access to this material is, however, restricted by IRB approval and data access protocols.
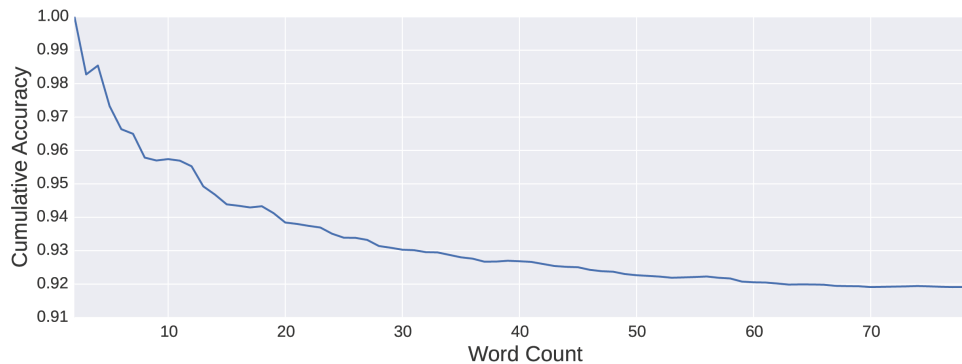
Figure 6: Mean cumulative accuracy on sentences containing up to N words using our proposed negation resolution algorithm.

prominence of negation in MHRs and justify the need for developing a negation detection mechanism. Our approach is competitive when compared to lexical negation resolution algorithms, and performs better for correctly classifying affirmed mentions. Finally, our negation detection algorithm can be applied on different datasets, and can be extended in order to extract more semantics.

Acknowledgments

References

Geneviève Arsenault-Lapierre, Caroline Kim, and Gustavo Turecki. 2004. Psychiatric diagnoses in 3275 suicides: a meta-analysis. *BMC Psychiatry*, 4:37.

Guilherme Borges, Matthew K Nock, Josep M Haro Abad, Irving Hwang, Nancy A Sampson, Jordi Alonso, Laura Helena Andrade, Matthias C Angermeyer, Annette Beautrais, Evelyn Bromet, Ronny Bruffaerts, Giovanni de Girolamo, Silvia Florescu, Oye Gureje, Chiyi Hu, Elie G Karam, Viviane Kovess-Masfety, Sing Lee, Daphna Levinson, Maria Elena Medina-Mora, Johan Ormel, Jose Posada-Villa, Rajesh Sagar, Toma Tomov, Hidenori Uda, David R Williams, and Ronald C Kessler. 2010. Twelve-month prevalence of and risk factors for suicide attempts in the world health organization world mental health surveys. *Journal of Clinical Psychology*, 71(12):1617–28, Dec.

Craig J Bryan and M David Rudd. 2006. Advances in the assessment of suicide risk. *Journal of Clinical Psychology*, 62(2):185–200.

Wendy W. Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. 2001. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, 34(5):301 – 310.

B.E. Chapman, S. Lee, H.P. Kang, and W.W. Chapman. 2011. Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm. *Journal of Biomedical Informatics*, 44(5):728–737.

David Chu, John N Dowling, and Wendy W Chapman. 2006. Evaluating the effectiveness of four contextual features in classifying annotated clinical conditions in emergency department reports. In *AMIA Annual Symposium Proceedings*, pages 141–145. American Medical Informatics Association.

Krystl Haerian, Hojjat Salmasian, and Carol Friedman. 2012. Methods for identifying suicide or suicidal ideation in EHRs. In *AMIA Annual Symposium Proceedings*, pages 1244–1253. American Medical Informatics Association.

Henk Harkema, John N. Dowling, Tyler Thornblade, and Wendy W. Chapman. 2009. ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of Biomedical Informatics*, 42(5):839 – 851. Biomedical Natural Language Processing.

Ronald C Kessler, Christopher H Warner, Christopher Ivany, Maria V Petukhova, Sherri Rose, Evelyn J Bromet, Millard Brown, Tianxi Cai, Lisa J Colpe, Kenneth L Cox, et al. 2015. Predicting suicides after psychiatric hospitalization in US Army soldiers: the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *JAMA Psychiatry*, 72(1):49–57.

J-D Kim, Tomoko Ohta, Yuka Tateisi, and Jun'ichi Tsujii. 2003. Genia corpus—a semantically annotated corpus for bio-textmining. *Bioinformatics*, 19(suppl 1):i180–i182.

Dan Klein and Christopher D Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pages 423–430. Association for Computational Linguistics.

J John Mann, Alan Apter, Jose Bertolote, Annette Beautrais, Dianne Currier, Ann Haas, Ulrich Hegerl, Jouko Lonnqvist, Kevin Malone, Andrej Marusic, et al. 2005. Suicide prevention strategies: A systematic review. *JAMA*, 294(16):2064–2074.

Mitchell P Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational linguistics*, 19(2):313–330.

Saeed Mehrabi, Anand Krishnan, Sunghwan Sohn, Alexandra M. Roch, Heidi Schmidt, Joe Kesterson, Chris Beesley, Paul Dexter, C. Max Schmidt, Hongfang Liu, and Mathew Palakal. 2015. DEEPEN: A negation detection system for clinical text incorporating dependency relation into NegEx. *Journal of Biomedical Informatics*, 54:213–219.

S M Meystre, G K Savova, K C Kipper-Schuler, and J F Hurdle. 2008. Extracting information from textual documents in the electronic health record: a review of recent research. *IMIA Yearbook*, (1):128–144.

Roser Morante and Caroline Sporleder, editors. 2010. *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*. University of Antwerp, Uppsala, Sweden, July.

P.G. Mutalik, A. Deshpande, and P.M. Nadkarni. 2001. Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the UMLS. *JAMIA*, 8(6):598–609.

AB Niculescu, DF Levey, PL Phalen, H Le-Niculescu, HD Dainton, N Jain, E Belanger, A James, S George, H Weber, et al. 2015. Understanding and predicting suicidality using a combined genomic and clinical risk assessment approach. *Molecular Psychiatry*, 20(11):1266–1285.

Matthew K Nock, Jennifer Greif Green, Irving Hwang, Katie A McLaughlin, Nancy A Sampson, Alan M Zaslavsky, and Ronald C Kessler. 2013. Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: Results from the national comorbidity survey replication adolescent supplement. *JAMA Psychiatry*, 70(3):300–310.

Gayan Perera, Matthew Broadbent, Felicity Callard, Chin-Kuo Chang, Johnny Downs, Rina Dutta, Andrea Fernandes, Richard D Hayes, Max Henderson, Richard Jackson, et al. 2016. Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource. *BMJ Open*, 6(3):e008721.

Pradhan, Sameer and Elhadad, Noémie and South, Brett R and Martinez, David and Christensen, Lee and Vogel, Amy and Suominen, Hanna and Chapman, Wendy W and Savova, Guergana. 2015. Evaluating the state of the art in disorder recognition and normalization of the clinical narrativet. *JAMIA*, 22(1):143–154.

Christopher Ryan, Olav Nielssen, Michael Paton, and Matthew Large. 2010. Clinical decisions in psychiatry should not be based on risk assessment. *Australasian Psychiatry*, 18(5):398–403.

RI Simon. 1992. Clinical risk management of the suicidal patient. *Clinical psychiatry and the Law. 2nd ed. American Psychiatric Press*, pages 259–96.

Sunghwan Sohn, Stephen Wu, and Christopher G Chute. 2012. Dependency parser-based negation detection in clinical narratives. *AMIA Summits on Translational Science*, 2012:1–8.

Sarah Steeg, Matthew Haigh, Roger T Webb, Nav Kapur, Yvonne Awenat, Patricia Gooding, Daniel Pratt, and Jayne Cooper. 2016. The exacerbating influence of hopelessness on other known risk factors for repeat self-harm and suicide. *Journal of affective disorders*, 190:522–528.

Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall. 2011. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *JAMIA*, 18(5):552–556.

V. Vincze, G. Szarvas, R. Farkas, G. Móra, and J. Csirik. 2008. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(11):1–9.

Christine Leonard Westgate, Brian Shiner, Paul Thompson, and Bradley V Watts. 2015. Evaluation of Veter-

ans' Suicide Risk With the Use of Linguistic Detection Methods. *Psychiatric services*, 66(10):1051.

Stephen Wu, Timothy Miller, James Masanz, Matt Coarr, Scott Halgrim, David Carrell, and Cheryl Clark. 2014. Negation's Not Solved: Generalizability Versus Optimizability in Clinical Natural Language Processing. *PLoS ONE*, 9(11):e112774.