Modeling human comprehension of Swedish medical records for intelligent access and summarization systems - Future vision, a physician's perspective

Maria Kvist[ab], Maria Skeppstedt[a], Sumithra Velupillai[a], Hercules Dalianis[a]

[a]*Dept. of Computer and Systems Sciences (DSV), Stockholm University, Kista, Sweden*
[b]*Dept. of clinical immunology and transfusion medicine, Karolinska University Hospital, Stockholm, Sweden*

Abstract

Physicians are daily demanded to read, understand and reach a summarized comprehension of earlier documentation for the patient at hand. This documentation includes medical procedures and clinical findings such as symptoms, observations, and diagnoses but also reasoning and speculation by previous physicians and nurses. The information is sometimes hidden in free-text in such a way that it requires an experienced medical background to decipher. Medical records are typically written with incomplete sentences, abbreviations, and medical jargon.
A wanted tool when reading electronic health records is computer generated text summaries with the possibility to pose questions to an intelligent search tool. To realize this it is necessary to build models of how physicians read and understand unstructured free-text in medical records.

Keywords: *Electronic Health Records, Natural Language Processing, Information Storage and Retrieval, Data Mining, Abstracting and Indexing as Topic*

Introduction

In many clinical situations, physicians, nurses and other caregivers are required to, in a short timeframe, reach an understanding of their patient and his or her disorders. Apart from speaking with and examining the patient, we rely on information in electronic health records (EHR) for medical history. It is essential to select the right information. One of the problems in this process concerns over documentation. Information was earlier limited, but in the digitalized world of today we drown in data about the patient, and much is unstructured. Computer generated text summarization can be of aid and have been developed also for Swedish, but only for structured texts such as news text [1].

Natural language processing (NLP) tools can interpret and generate natural language text. These include search engines, automatic text summarizers, information extraction and text mining tools, grammar- and spell checking tools, and also important subcomponents such as negation detectors, decompounders, lemmatizers and stemmers. For a review of NLP-tools in the clinical domain see Meystre et al. [2]. The complex task of creating NLP-tools for unstructured clinical free-text is addressed in, e.g., the yearly challenge i2b2[1]. NLP-tools for clinical free-text must be developed for each language. For example, a spelling suggester [3] and a patient chronicle-generator for patient event overview [4] has been constructed for clinical English, information retrieval for German [5], medical event recognizer for Japanese [6], topic segmentation and labeling for Finnish [7] and entity identification for Swedish clinical text [8].

In this paper we describe some steps to be taken towards an envisioned intelligent support tool in Swedish, and point at some obstacles for a small language. Examples of models from our research group show that this can be a realistic prospect.

Language in Swedish medical records

Medical records are rich in medical terminology, necessary for preciseness and for medically safe documentation. Entries in the free-text section of EHR are written under time pressure and the writer counts on that the reader is familiar with medical terminology, medical jargon and local abbreviations. Semantic economy drives towards an unstructured and ungrammatical text [9, 2] pregnant with abbreviations and acronyms, incomplete sentences that often lack subjects, verbs in passive form, and reasoning that skip intermediate clues and present conclusions directly. The records are written in an informal manner but are also spiced with words of Latin or Greek origin, as well as words from the English vocabulary of medical journals, often misspelled. Moreover, the Swedish language is highly inflectional and littered with compounding words.

Models of understanding

Models of how physicians read medical records and extract

[1] Informatics for Integrating Biology & the Bedside, https://www.i2b2.org.

knowledge about the patient and his or her diseases must mimic the physicians' way of determining which earlier symptoms, clinical findings, examinations, and diagnoses that are reliable and relevant for the question at hand. It is crucial to find the correct pieces of information with high precision and recall, but also to determine the certainty of that knowledge. The reference to a diagnosis can be connected to a negation or speculation (factuality), or historical events (temporality). Subject identification is also crucial since hereditary disorders frequently are included in EHR.

Human interpretation of free-text in EHR

Speculations often include probability expressions that are known to be perceived and interpreted differently by diverse individuals. Verbal expressions of uncertainty have been studied thoroughly [10], and modeled for instance in numerical scales. Also, humans can see deeper meanings in narratives, see patterns and new perspectives in a way computers do not.

Machine interpretation of free-text in EHR

Information about clinical findings is hard to reach through computer tools, as natural language expresses many subtleties. Negations and speculative language in English have been automatically determined by contextual cues and hedge phrases [11]. The traditional way of mapping clinical findings such as disorders, in a text, would be to match the text entries in EHR to terminologies for diagnoses (e.g. ICD-10[2], SNOMED-CT[3]). As diseases are rarely expressed in such precise wordings in free-text, an extensive preprocessing of EHR is needed to find many of the terms in the terminology [12]. EHR also contain implicit information that will not be revealed if simple string matching is applied to terminologies of disorders. Some information lodge in compound words where it is not directly accessible. A compound word can be split in its parts [13] but the contextual meaning of the word may be lost in that process. When normalizing word inflections to base form, using a stemmer or a lemmatizer, consideration must be taken to domain specific vocabulary.

Many of these problems are possible to overcome. A harder nut to crack is how to construct a machine that can read between the lines, in the way humans do.

Methods and some results - Addressing the issue of intelligent information access

We have performed some studies that confirm that, bit by bit, it may be possible to construct a system that can extract and assess data from unstructured clinical free-text in Swedish.

Human understanding of a text is modeled regarding the ability to identify clinically interesting words and expressions, determine their affirmation or negation, identify speculation, and assign knowledge certainty. These models of human comprehension are used to form annotation classes for annotation of text entries in order to create resources that subsequently are used for machine learning. Building automated systems using these resources and the results thereof is not the scope of this article and are published elsewhere [2, 14].

Selected medical records and annotation

Free-text entries were extracted from patient records in the Stockholm EPR Corpus [15][4]. We wanted a diversity of speculative words and expressions. For that, we looked for a variety of diagnoses, which would call for different types of speculation. The notes in the assessment field from an emergency ward fulfilled those criteria, representing different kinds of narratives with both speculative and non-speculative approaches to reasoning around diagnoses. Emergency ward EHR were used for similar research in English [16].

Clinical expressions in EHR entries were marked manually by two senior physicians according to guidelines drawn up for each task. In some tasks, diagnoses were pretagged and factuality levels or temporality was to be given. We used the Knowtator plugin in the Protégé tool [17] for all annotation work. A general language automatic lemmatizer for Swedish (*CST lemmatizer* [18]) was used for capturing inflections. Intra- and interannotator agreement was measured.

Linguistic properties

The Stockholm EPR Corpus was earlier described [15] but further linguistic exploration is required, e.g. regarding medical jargon. It was evident that many entries were written rapidly, resulting in spelling errors, incomplete sentences and referring errors. We found a multitude of expressions for the same clinical finding. Apart from the usual inflections and misspellings, abbreviations were used in a creative way. For example, the clinical finding that a patient has a normal blood level of troponin T (analysis of heart muscle tissue damage) was found to be expressed in 30 different ways by varying expressions, word orders and abbreviations, and still none of these were misspellings (43 annotations in 614 random free-text entries).

Words originating from Greek or Latin often give rise to misspellings to the extent that it must be foreseen in NLP. For example, four alternate spellings for *takykardi* (*tachycardi, tachykardi, takycardi* and *takykardi*) were found, together with the original Greek *tachycardia,* meaning "rapid heart".

Model for identification of clinical findings in EHR

The trained physician can rapidly spot clinically interesting facts such as symptoms, clinical observations, medical procedures and disorders. This process must be replicated by the computer. We have annotated findings, body structures and disorders in Swedish EHR, for subsequent machine learning.

[2]International Classification of Diseases (ICD), http://www.who.int/classifications/icd/en/.
[3]SNOMED Clinical Terms User Guide, July 2008 International Release. http://www.ihtsdo.org/

[4]This research has been carried out after approval from the Regional Ethical Review Board in Stockholm (Etikprövningsnämnden i Stockholm), permission number 2009/1742-31/5

This work is in progress, with the aim to train a machine-learning system to automatically detect words belonging to these classes through contextual or other markers. Clinical entities have been identified in English clinical text and machine-learning systems have successfully been trained to automatically detect them [19].

Our results showed that the same medical expression can be a finding in one instance and a disorder in another, according to context. However, only a few of them mapped to both categories in the Swedish version of SNOMED-CT, which is still under expansion. Tachycardia is a symptom or a clinical observation of several disorders, and a normal condition (running), but is also found in ICD-10 as a diagnosis. Hence, we need to construct a machine that can learn to differentiate which functions words have in situation specific contexts.

Model for identification of disorders implicit in text

Physicians sometimes express disorders in terms of a test result, given drug or other treatment. For example, ”pat. regulates insulin according to measured blood glucose” will be translated in the reading physicians mind to “patient has diabetes”.

By semantics, it should be possible to detect some disorders implicated in the text, also without medical knowledge. A disorder can be implicated by a circumscription such as ”pneumonic infiltrates in the lungs” (*pneumoniska infiltrat i lungorna*) and the reader is supposed to understand that the patient has pneumonia. With the proper settings, an NLP-tool could recognize that disorder as well. In order to identify such implicit disorder expressions, and to broaden the resource used for machine learning, we also annotated adverbs and adjectives related to the patients' condition, e.g. talks manically (*talar maniskt*) or asthmatic patient (*astmatiker*).

Model for knowledge certainty

Automated systems must be able to distinguish factuality levels of clinical findings; otherwise uncertain and negated diagnoses will be identified as factual diagnoses.

In one study, factuality levels for diagnoses were modeled in polarity: *Positive* or *Negative* and the knowledge was graded: *Certain, Probable* or *Possible*. A diagnosis expression from the assessment field of EHR entries was annotated as belonging to polarity and gradation resulting in six annotation classes. Annotation was performed creating a resource for subsequent machine learning [14].

Studying a subset of the annotations (1297 assessment fields), we saw a pattern in how some diagnosis expressions were used differently in negating, affirming or speculating sentences. Some diagnoses appeared to group solely to affirmed classes. One explanation can be that certain words are negated by the use of autonyms. Hypertension was almost never negated, as this state instead is expressed as “normal blood pressure”. Other clinical findings were found mostly in the negating classes, e.g. skeletal injury (*skelettskada*) and ischemia (*ischemi*), with related diagnoses in affirmed classes in a complementary fashion (figure 1). Diagnoses that rely on machine detection, in the way atrial fibrillation depend on ECG, showed little presence in classes for uncertainty.

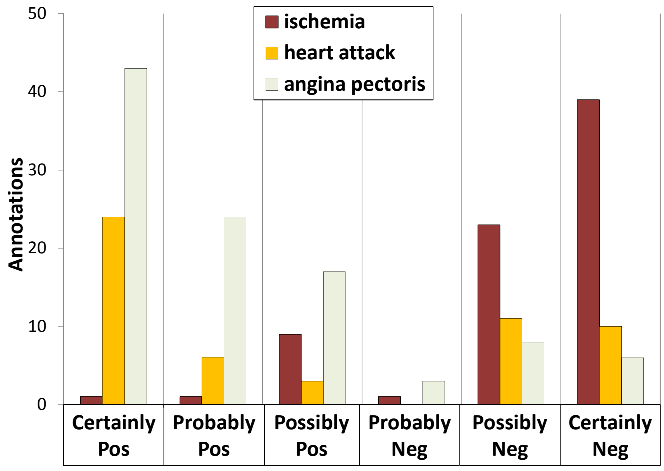

Figure 1- Complementary vocabulary. Ischemic diagnoses are negated as “ischemia” but named as disorders when affirmed.

A linguistically interesting disclosure was that *lunginflammation*, the Swedish word for pneumonia, was not used in the same pattern as *pneumoni*, a synonym with Greek origin. *Lunginflammation* was found mostly in the annotation class Probably positive, but rarely in Certainly positive. As patients were examined and the hypothesis of pneumonia arose, both *lunginflammation* and *pneumoni* was used, but when the patient returned from radiological examination with a conclusive answer, *pneumoni* was almost always the term affirmed or negated. Additional patterns are described in [20].

Discussion

Computer generated text summaries with the possibility to pose questions to an intelligent search tool is still just the wet dream of a physician at the end of a night shift. For researchers of common diseases, a huge amount of knowledge and experience is concealed in the free-text of EHR, too valuable to be left unexploited, but could be reachable with knowledge extraction methods such as text mining [2]. Reasoning processes and decisions documented in free-text is information waiting to be exploited for developing decision support. For surveillance of medical errors or near miss, automatic detection of adverse events could be implemented for at least one third of the GTT-triggers[5] that today are interpreted manually.

Clinical narratives interpreted by machines

We cannot construct machines that are more skilled than humans, but machines are tireless, consistent and can digest a huge amount of data.

In the clinical reality, the presence of a diagnosis is a

[5]Global Trigger Tool Kit, 2005, Version 6. Institute for Healthcare Improvement, Cambridge, MA, USA. For (In Swedish 2007).

continuum of varying degrees of certainty. Writers using verbal uncertainty expressions mean a certain level of factuality, which is then interpreted differently by diverse readers [10]. A computer will interpret phrases such as negation cues consistently, irrespective of the actual factuality intended. Research on modeling speculations in free-text has been performed using machine learning techniques [11, 16]. Negation detection in radiology reports was executed with good results [21]. The certainty of clinical findings can be found on several levels; in context, hedge phrases and cue lines, but as we have shown also inherent in medical expressions and choice of synonyms. This calls for context- and situation sensitive NLP-systems.

Domain specific abbreviations and misspellings

A large number of words in clinical text are unknown; studies typically show 30% in English speaking countries [3]. These words must be recognized and identified as medical terminology, jargon, abbreviations, named entities, or misspellings, and so on. Up to 10 % misspellings was seen in a study of French clinical text [22]. Spelling errors have not been measured quantitatively for the Stockholm EPR corpus, but were frequent and often of the keyboard slipping type indicating rapid typing. Considering the frequency of medical jargon and neologisms, traditional spell checkers are ineffective. The single example of four alternate spellings for a medical term in Swedish EHR (above) illustrates the need for domain specific spell checkers as part of information access systems. Such spelling suggesters have been constructed for the English vocabulary of medical records [3].

Abbreviations in Swedish clinical text have been cataloged [23], likewise vital for NLP of clinical narratives. The use of abbreviations is often local, ambiguous, and situation specific, e.g. “pat.” meaning “patient” or “pathological” depending on context. The medical substance noradrenalin was abbreviated in 60 different ways by nurses at the Intensive care unit at Karolinska University Hospital [9], and we found 30 variants of ”normal troponin T” in a small subset of entries by physicians at the emergency ward. This illustrates the need of a continuously expanding thesaurus.

Reading between the lines

When reading EHR, an experienced physician often scans the lists of medications and laboratory results, and checks which other physicians or departments the patient frequents, in order to get a quick overview of the patients’ problems. Some of these tricks could be mimicked by the machine. For instance, the fact that a patient is on insulin, can be machine translated to speculations on the disorder diabetes mellitus by connecting it to an ontology such as SNOMED-CT.

Physicians do not always write down what they think, as they assume that the reader will think along the same paths and make the same conclusions. Accordingly, a lot of diagnoses are implicit in EHR. Reasoning and speculation is often documented with reference to the origin of a disorder, e.g. “chest pain probably not of cardiac origin” (*bröstsmärtan troligen inte av kardiell genes*) or “can be something gynecological” (*kan vara något gynekologiskt*). This kind of disorder reference may need medical knowledge incorporated in the system, for deduction of diagnostic reasoning. Other communications, relying on subtleties of words, such as “Repeatedly denies abuse of alcohol” implies that the physician is of another opinion. Artificial intelligence of that level is yet to be seen.

Reading between the lines may seem unattainable, but great progress has been made in such an intriguing part of human communication as humor. Computers can identify jokes in free-text, and have long been known to produce puns [24].

Access to EHR for the research community and patients

Health records are rarely made available for legal and integrity reasons. Tools for de-identification can widen the research community working on information access solutions for clinical text. We have access to a unique collection of EHR for 2 million patients from Karolinska University Hospital 2006-2010. A subset of these has been described as the Stockholm EPR Corpus [15]. This resource is used for machine learning after modeling the human process of understanding text, and for linguistic studies of free-text entries in medical records.

Patients of today, and in the future, demand access to their EHR. Web access is under progress in Sweden. True access is accomplished only if the text is understandable and possible to overview, also for those who are not familiar with medical terminology and jargon. The demand for automatic text processing may increase. For example, it would be convenient to receive your radiology report translated for a layman.

Conclusions

The development of text summarization systems for Swedish EHR is hampered by the unstructured nature of the free-text, the lack of NLP-tools adjusted to the clinical domain and the fact that Swedish is a small language with a limited market.

It is possible to identify clinical entities, determine their factuality, and model reasoning and speculation also in unstructured and incomplete text. For this, domain-specific NLP-tools, with combinations of rule based and machine learning systems, are under progress.

In order to create information access systems for Swedish EHR we need a better understanding of the Swedish language in clinical narrative text, including semantic characterization of words and concepts with special concern to medical jargon. We also need models of how physicians decipher and summarize the information in EHR in order to mimic the process in a machine. It is crucial that the users, physicians and other medical personnel, are involved when systems for EHR and tools for information access are constructed.

For a small language such as Swedish, there is presently a limited number of NLP-tools for clinical text. We are collaborating within the Scandinavian research community to achieve better results, e.g. the research network HEXAnord[6].

[6]http://dsv.su.se/hexanord

Acknowledgments

The authors thank Dr. Martin Hassel for valuable discussions.

References

[1] Dalianis, H. SweSum - A Text Summarizer for Swedish. Technical report TRITA-NA-P0015, IPLab-174, NADA, KTH, October 2000

[2] Meystre SM, Savova GK, Kipper-Schuler KC and Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform 2008: 3; pp. 128-144.

[3] Patrick J, Sabbagh M, Jain S, and Zheng H. Spelling correction in clinical notes with emphasis on first suggestion accuracy. 2nd workshop on Building and Evaluating Rescources for Biomedical Text Mining 2010, pp. 2-8.

[4] Harkema H, Setzer A, Gaizaukas R, and Heppel M. Mining and modelling temporal clinical data. Proc. of the UK e-Service All Hands Meeting 2005, pp 507-514.

[5] Spat S, Cadonna B, Radovac I, Gütl C, Leiter H, Stark G, and Beck P. Enhanced information retrieval from narrative German-language clinical text documents using automated document classification. Proc of MIE2008. In: Studies in health technology and informatics. 2008: 136: 473-478.

[6] Aramaki E, Miura Y, Tonoike M, Ohkuma T, Mashuichi H, and Ohe K. TEXT2TABLE: Medical Text Summarization system based on named entity recognition and modal identification. In: Proc. Workshop on BioNLP 2009, pp.185-192.

[7] Ginter, F, Suominen, H, Pyysalo, S and Salakoski T. Combining hidden Markov models and latent semantic analysis for topic segmentation and labeling: Method and clinical application. J of Medical Informatics 2009: 78 (12): pp. e1-e6.

[8] Kokkinakis D, and Thurin A. Identification of Entity References in Hospital Discharge Letters. In: Proc. 16th Nordic Conf. of Computational Linguistics NODALIDA-2007. University of Tartu, Tartu, 2007, pp. 329–332.

[9] Allvin H, Carlsson E, Dalianis H, Danielsson-Ojala R, Daudaravicius V, Hassel M, Kokkinakis D, Lundgren-Laine H, Nilsson G,. Nytrø Ø, Salanterä S, Skeppstedt M, Suominen H, and Velupillai S. Characteristics of Finnish and Swedish intensive care nursing narratives: a comparative analysis to support the development of clinical language technologies. J of Biomedical Semantics, 2011, 2 (Supp 3): S1 (In press).

[10]Clark DA. Verbal Uncertainty Expressions: A critical review of two decades of research. Current Psychology: Research & Reviews, 1990: 9: 203-235.

[11]Farkas R, Vincze V, Móra G, Csirik J, and Szarvas G. The conll-2010 shared task: Learning to detect hedges and their scope in natural language text. In: Proc. of the Fourteenth Conference on Computational Natural Language Learning, Uppsala, Sweden. Assoc for Computational Linguistics. 2010; pp 1–12.

[12]Aronson AR. Effective mapping of biomedical text to the UMLS metathesaurus: The Metamap program. Proc AMIA Symp 2001, pp. 17–21.

[13]Sjöbergh, J and Kann, V. Finding the correct interpretation of Swedish compounds, a statistical approach. Proc of LREC 2004, pp 899-902.

[14]Velupillai S. Classifying factuality levels in diagnoses - an annotated corpus of Swedish clinical assessments. 2010. Submitted.

[15]Dalianis H, Hassel M, and Velupillai S. The Stockholm EPR Corpus - Characteristics and some initial findings. Proc. 14th ISHIMIR, Kalmar, Sweden, 2009, pp. 243-249.

[16]Harkema H, Dowling JN, Thornblade T, and Chapman W. ConText: An algorithm for determining negation, experience, and temporal status from clinical reports. J of Medical Informatics 2009:2: pp 839-851.

[17]Ogren P. Knowtator: a Protégé plugin for annotated corpus construction. Proc. HLT-NAACL 2006, Morristown, NJ, USA, ACL, 2006; pp. 273–275.

[18]Jongejan, B and Haltrup, D. The CST Lemmatiser. Center for Sprogteknologi, University of Copenhagen, 2005.

[19]Wang Y and Patrick J. Cascading classifiers for named entities recognition in clinical notes. In: Workshop Biomedical Information Extraction. 2009; pp. 42-49.

[20]Velupillai S, Dalianis H, and Kvist M. Factuality levels of diagnoses in Swedish medical text. MIE 2011, IOS Press, in print.

[21]Huang Y and Lowe HJ. A novel hybrid approach to automated negation in clinical radiology reports. J Am Med Inform Assoc 2007, pp 304-311.

[22]Ruch P, Baud R, and Geissbühler A. Using lexical disambiguation and named-entity recognition to improve spelling correction in the electronic patient record. In: Artificial intelligence in medicine. 2003; 29(1-2): pp. 169-84.

[23]Cederblom S. *Medicinska förkortningar och akronymer.* (In Swedish) Studentlitteratur 2005.

[24]Binsted K, Bergen B, Coulson S, Nijholt A, Oliviero Stock O, Carlo Strapparava C, Ritchie G, Manurung R, Pain H, Waller A, and O'Mara D. Computational humor. IEEE Intelligent Systems. 2006: 21(2): 59–69.

Address for correspondence

Maria Kvist, Dept. of Computer and Systems Sciences (DSV), Stockholm University, Forum 100, SE-164 40 Kista, Sweden. maria.kvist@karolinska.se