International Journal of Research in Signal Processing, Computing & Communication System Design
4 (2), 2018, 06-08
http://www.publishingindia.com/ijrsp

Review on Health Care Claim Processing Using Text Mining and Natural Language Processing

Tushar Gonawala[1*], Hima Khimani[2], Ruchi Patel[3] and Vatsal Shah[4]
[1]Birla Vishvakarma Mahavidyalaya Engineering College, Vidyanagar, Anand, Gujarat, India.
Email: tushargonawala96@gmail.com
[2]Birla Vishvakarma Mahavidyalaya Engineering College, Vidyanagar, Anand, Gujarat, India.
Email: himakhimani@gmail.com
[3]Birla Vishvakarma Mahavidyalaya Engineering College, Vidyanagar, Anand, Gujarat, India.
Email: ruchi14797@gmail.com
[4]Birla Vishvakarma Mahavidyalaya Engineering College, Vidyanagar, Anand, Gujarat, India.
Email: vatsal.shah@bvmengineering.ac.in
*Corresponding Author

Abstract: The processing of health care claims includes a combination of structured and unstructured data collected from various sources of information that are directly or indirectly related to the medical insurance claim. Such processing takes help of Natural Language processing along with some concept specific language. NLP Techniques along with Text Mining helps in finding dependencies between different entities which further generate scores for individual claims. These scores are considered in making decisions involving determination of fraud or genuine claims by the client.

Keywords: Categorization, Information retrieval, Medical claims, Natural Language Processing (NLP), Pattern matching, Text mining.

I. Introduction

As Data mining finds pattern in the database similarly, text mining takes care of detection of pattern in the linguistic texts. Information processing algorithms such as text mining leverages its potential to access both the structured data in the database as well as the unstructured data stored in form of documents. For more detailed linguistic analysis of the text, there are various techniques for analyzing natural language text [1].

Document Categorization is the most popular application of Text Mining [2]. In order to process the health care claims, this paper would be focusing on analysis of textual information and categorization in the context of Health care related issues. The text used in making the claims may be highly fragmented stating that it may include many acronyms and abbreviation in standards of treatment and diagnostics taxonomy [5], making it difficult to fetch the relevant information from it. As a result of this, it is necessary to merge the information from medical field and information retrieved from the insurance claims to provide effective identification of claims involving fraud and abuse or third party liability or subrogation.

The process of Automated Medical Claim Auditing is visualized in the Fig. 1, where the concept analyzer of the Natural Language Processing along with the domain specific information in form of Concept Taxonomies generates an output that is subjected to human analysis.

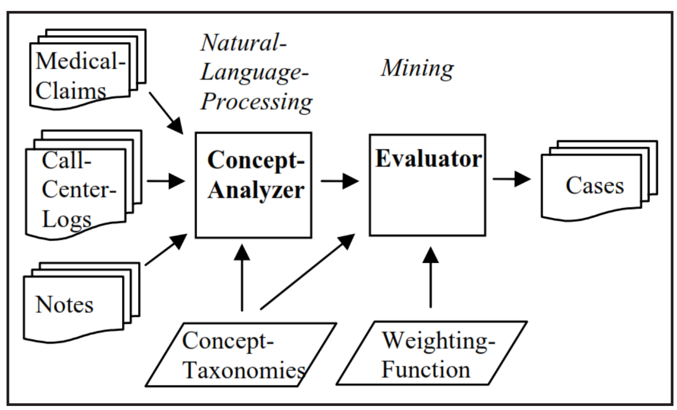

Fig. 1: Automated Health Care Claim Auditor [3]

II. Natural Language Processing

A. Overview

Natural Language Processing (NLP) deals with the analysis and automatic processing of the textual information. There are mainly two approaches that can be taken. One is to provide huge amount of training material which exemplifies the desirable and/or undesirable relationships and dependencies. A minor modification in the system would require subsequently retraining the whole system with probably a new training dataset. The second is a rule based technique that is probably more efficient but required detailed knowledge of the domain and the rules formed should not be ambiguous. The major advantage is that it can leverage the existing dictionaries, classification systems and other taxonomical works. The final decision of the usage of technology is made keeping in mind the availability of training datasets along with external resources and the resultant desired output of the application.

B. Concept Specific Language (CSL)

Concept Specific Language (CSL) is the core technology that takes care of the "Concept" mapping. The CLS specified is very rich in incorporating linguistic patterns and predicates [4]. The very first stage of analysis consist of abbreviation expansion and spelling correction which is later on sent for tagging and partial parsing. Specific information can be extracted from the text using the rules and concepts of CLS. The specification if interrelationship among concepts in the form of multiple operators, such as OR, NOT, Precedes, Immediately Precedes, Is related, or Causes [6].

In order to illustrate CSL, we can take an example of Accidents and Trauma wherein there are multiple scenarios in real world that can be helpful in the medical claims processing.

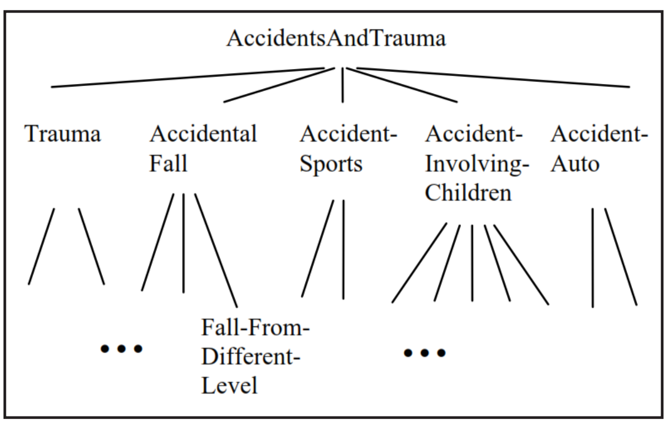

Fig. 2: An Example of Taxonomy [3]

In a single definition of concept there may be sub concepts depending upon the situation. For example a concept of Accidental Fall may include sub concept of Slipped or Fell which itself may include a sub concept Fall From Different Level. It is not mandatory that these categories have to be mutually exclusive. A low level CSL for such a concept is demonstrated in Fig. 2. Note is to be taken that concept containing individual words are linguistically linked to a word or a phrase that matches with concept of higher hierarchy. In our case individual words like "off", ""from", "to" and "feet" are linked with SlippedOrFell Concept [3].

```
Concept FallFromDifferentLevel(
        Related(%SlippedOrFell, (off | from | to | feet))
(%SlippedOrFell& down & /NOUN))
```

III. Creating Concepts

To create a very accurate and detailed specification for CSL is a tedious and time consuming task. Also, for such a task, it is required that both the linguistic experts and the domain experts have to work combine. The knowledge base of Natural Language Processing system can be leveraged along with the linguistic rules of the domain specific knowledge.

The text based concept algorithm is as follows:

1. Input of Text Fragment: The user is allowed to enter one or more text fragments which are considered for the input to the next step of this algorithm.
2. Fragment Split into Words: The fragments are split into words using the Concept Analyzer.
3. Selection of Relevant Words: The user selects the relevant words out of the whole fragment. Also the system can automatically select the default relevant words using the linguist rule concepts.
4. Optional Operation on Relevant Words: For a given word the user has the freedom to select any synonym, hypernym or hyponym available in the Wordnet.
5. Concept Matching: The predefined sets are used for the fragments to find all the matches and return the resulting match that is known as "Concept Match".
6. Removal of Concept Matches: There are two major criteria for removing concept matches.
   a) The words that are marked "relevant".
   b) The interpretation that the user marks relevant.
7. Building of Concept Chain (Tiling): The process of concept matching based on the matches of previous steps is known as tiling where a chain of sequence is formed. For performing such a task certain criteria should be kept in mind.
   a) No two matches of the chain should overlap.
   b) Chain's maximum length criteria should not be violated.
8. Chain Writing as CSL Concept: These chains should be written and passed through the CSL.

Algorithm Step Number	User Input	Example
1	Text fragments	*Mary was adored by John since high school*
2	Split into words	*Mary, be, adore, by, John, since, high, school*
3	Relevant words	*John, Mary, adore*
4	Synonyms (for *adore*)	*love intensely*
5		*Subj_Passive_Verb_Obj(john, adore, mary)* *Noun_Noun(john, mary)* *Noun_Verb(john,adore)* ...
6		*Subj_Passive_Verb_Obj(john, adore, mary)* *Noun_Noun(john, mary)*
7		*Subj_Passive_Verb_Obj(john, adore, mary)*

8	Adoration	Concept Adoration {
		Subj_Verb_Obj(john, @adore, mary)
		}

Fig. 3: Example of CSL from Text [3]

IV. Conclusion

In order to develop an application for medical claim auditor, the gap between text mining and Natural Language Processing is to be bridged. Using this technique we can enrich text with semantical error correction, abbreviations and acronym. Moreover, the efficiency to normalize the unstructured textual data into standard tags.

In the future, when more indicator-enhanced claim data becomes available, it will be possible to apply additional data-mining techniques [6] to detect previously unknown patterns. Of particular interest will be the use of association rules for fraud and abuse detection.

References

[1] D. Jurafsky, and J. Martin, *Speech and Language Processing*, Prentice Hall, Upper Sale River, NJ, 2000.

[2] D. Mladenic, and M. Grobelnik, "Text and Web Mining," in D. Mladenic, N. Lavrac, M. Bohanec, and S. Moyle, (eds.), *Data Mining and Decision Support Integration and Collaboration*, Kluwer, Dordrecht, 2003.

[3] F. Popowich, "Using text mining and natural language processing for health care claims processing," *SIGKDD Explorations*, vol. 7, no. 1, pp. 59-66, ACM, June 2005. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.5598&rep=rep1&type=pdf

[4] RecordOne, "Is natural language processing the key to data-driven health care?," 2015. Available: http://www.recordsone.com/wp-content/uploads/2015/05/R1_NLP_WP_web.pdf

[5] F. Popowich, "Use of text analytics and taxonomies for fraud and abuse detection in medical insurance claims," *Proceedings of Semantic Web Symposium of I2LOR-04 Towards the Educational Semantic Web (Université de Quebec à Montréal, Montréal)*, 19 November 2004. Available: http://www.cscsi.org/home/CSCSI/Members/swig/swig04papers/popowich-swig.pdf

[6] S. Abney, "Part-of-speech tagging and partial parsing," in S. Young, and G. Bloothooft, (eds.), *Corpus-Based Methods in Language and Speech Processing*, Kluwer, Dordrecht, pp. 118-136, 1997. J. Han, and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, San Francisco, CA, 2000.