Identifying Temporal Information and Tracking Sentiment in Cancer Patients’ Interviews

Braja Gopal Patra¹, Nilabjya Ghosh², Dipankar Das¹, and Sivaji Bandyopadhyay¹

¹Dept. of Computer Science & Engineering, Jadavpur University, Kolkata, India
²Yahoo, Bangalore, India
{brajagopal.cse,dipankar.dipnil2005}@gmail.com, nilabjyaghosh@hotmail.com, sivaji_cse_ju@yahoo.com

Abstract. Time is an essential component for the analysis of medical data, and the sentiment beneath the temporal information is intrinsically connected with the medical reasoning tasks. The present paper introduces the problem of identifying temporal information as well as tracking of the sentiments/emotions according to the temporal situations from the interviews of cancer patients. A supervised method has been used to identify the medical events using a list of temporal words along with various syntactic and semantic features. We also analyzed the sentiments of the patients with respect to the time-bins with the help of dependency based sentiment analysis techniques and several Sentiment lexicons. We have achieved the maximum accuracy of 75.38% and 65.06% in identifying the temporal and sentiment information, respectively.

Keywords: temporal information, sentiment analysis, support vector machines, cancer patients’ interviews.

1 Introduction

The extraction of temporal information and identification of the sentiment(s) associated with it from the clinical text is an interesting and popular research area. Several related work can be found on clinical temporal information, based on different admission and discharge summary reports of the patients [5, 13]. Various types of research have been carried out related to sentiment analysis from the textual data [4]. To the best of our knowledge, the extraction and analysis of clinical temporal data along with sentiment analysis has not been explored much till date.

Cancer is one of the most dangerous diseases in the world is associated with less chances of post-diagnosis survival as compared to other diseases. For all types of cancer, the latest available records are found during 2000 to 2001¹. One of such reports from the Cancer Research in UK shows that 43% of men and 56% of women were alive for more than 5 years and might have been cured. It is also observed that

so many cancer survivors face physical and mental challenges as resulted out of their disease and treatment\(^2\) even after a decade of being cured. Another research in [3] shows that around 22\%–29\% newly diagnosed cancer patients suffered from major depressive disorder (MDD). It is observed that if the diagnosis is carried out in prior to the critical temporal stages, somehow, we could improve the possibilities of recovery by incorporating proper treatment. Thus, the proposed labeled corpora and a prototype system could be helpful and therefore explored, in the context of cancer diagnosis, treatment and recovery.

In the present paper, we have labeled the temporal information in each of the sentences of the cancer patients’ interviews. We considered four time-bins to capture a medical event, such as before, during, after and not-informative where the not-informative sentences do not contain any temporal information. We also annotated the underlying sentiments (positive, negative and neutral) of the patients associated with the specific situations or temporal bins. However, the present research aims to help psychiatrists in order to resolve the psychological problems of cancer patients aroused for different emotions at different timestamps. Later on, we employed different machine learning algorithms for automatic identification of temporal and sentiment information. We used different syntactic and semantic features to extract the temporal and sentiment information from the clinical texts.

The rest of the paper is organized in the following manner. Section 2 provides related work in details. Section 3 provides an elaborate description of the data used in the task. Similarly, the features used in these experiments are described in Section 4. The detailed setup of experimentation and analysis of the results are described in Section 5. Finally, conclusions and future directions are presented in Section 6.

## 2 Related Work

Recently, several tasks have been carried out on identifying the temporal bins from different types of medical reports. Jindal and Roth [5] presented a joint inference model for the task of concept recognition in clinical domain. The major contribution of this paper was to identify the boundaries of medical concepts and to assign types to such concepts. Each concept was categorized to have 3 possible types e.g., Test, Treatment, and Problem. They used the datasets provided by i2b2/VA team as part of 2010 i2b2/VA shared task\(^3\). They achieved an overall accuracy of 86.1\% in identifying the concept types.

On the other hand, Raghavan et al., [13] modeled the problem of clinical temporal extraction as a sequence tagging task using Conditional Random Fields (CRFs). They extracted a combination of lexical, section-based and temporal features from medical events in each of the clinical narratives. Learning temporal relations, for fine-grained temporal ordering of medical events in clinical text is challenging: the temporal cues typically found in clinical text may not always be sufficient for these kinds of

\(^3\) https://www.i2b2.org/NLP/Relations/