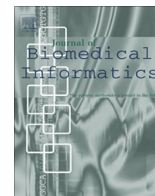




Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Identifying scientific artefacts in biomedical literature: The Evidence Based Medicine use case

Hamed Hassanzadeh, Tudor Groza*, Jane Hunter

School of ITEE, The University of Queensland, Australia

ARTICLE INFO

Article history:

Received 30 September 2013

Accepted 1 February 2014

Available online xxx

Keywords:

Evidence Based Medicine

Text classification

PICO

PIBOSO

Machine Learning

ABSTRACT

Evidence Based Medicine (EBM) provides a framework that makes use of the current best evidence in the domain to support clinicians in the decision making process. In most cases, the underlying foundational knowledge is captured in scientific publications that detail specific clinical studies or randomised controlled trials. Over the course of the last two decades, research has been performed on modelling key aspects described within publications (e.g., aims, methods, results), to enable the successful realisation of the goals of EBM. A significant outcome of this research has been the PICO (Population/Problem–Intervention–Comparison–Outcome) structure, and its refined version PIBOSO (Population–Intervention–Background–Outcome–Study Design–Other), both of which provide a formalisation of these scientific artefacts. Subsequently, using these schemes, diverse automatic extraction techniques have been proposed to streamline the knowledge discovery and exploration process in EBM. In this paper, we present a Machine Learning approach that aims to classify sentences according to the PIBOSO scheme. We use a discriminative set of features that do not rely on any external resources to achieve results comparable to the state of the art. A corpus of 1000 structured and unstructured abstracts – i.e., the NICTA-PIBOSO corpus – is used for training and testing. Our best CRF classifier achieves a micro-average *F*-score of 90.74% and 87.21%, respectively, over structured and unstructured abstracts, which represents an increase of 25.48 percentage points and 26.6 percentage points in *F*-score when compared to the best existing approaches.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Evidence Based Medicine (EBM) represents a framework that encompasses decision making in the healthcare domain based on the best existing evidence, with the goal of providing treatment options for individual patients. In order to provide patients with judicious decisions, practitioners must access the current best evidence in relevant published medical research, such as Randomised Control Trials (RCTs). The rhetorical structure of these publications generally follows the Population/Problem–Intervention–Comparison–Outcome (PICO) scheme [1]. For instance, based on the NICTA-PIBOSO corpus [2], the sentence “*The authors describe the case of a 13-year-old girl who was admitted with a history of back pain and acute-onset lower-extremity weakness.*” is an instance of Population class, since it gives some information about the individuals involved in the study. Similarly, the sentence “*One year postoperatively, the residual cyst had gradually shrunk and had*

almost disappeared.” discusses the result of the study, and hence can be considered an Outcome.

In the EBM process, clinicians search for PICO elements as evidence when making their judgments. Although most of today’s domain-specific text mining approaches [3,4] are able to identify and recognise concepts in the content of scientific publications (e.g., genes, proteins, chemical elements), they are still unable to capture and retrieve scientific artefacts. These vary in scope and granularity and can be framed within a particular area or domain, such as those comprised by the PICO structure, or may have more generic roles, like Aim, Method or Results. For example, some of the existing rhetorical schemes provide a fine-grained perspective of the narrative (e.g., [5]), which leads to mixed classification results but to an increased potential to realise fine-grained linking across rhetorical types (e.g., relate a Motivation statement to a Goal, and the Goal to a Observation). Other approaches model this knowledge at a more coarse grained level (e.g., [6]), which leads to better classification results (since the types follow a rather uniform distribution), but do not enable linking – except at a very high level (e.g., relate a Scientific statement to a Methodology).

* Corresponding author.

E-mail addresses: h.hassanzadeh@uq.edu.au (H. Hassanzadeh), tudor.groza@uq.edu.au (T. Groza), jane@itee.uq.edu.au (J. Hunter).

Independently of the underlying scheme, scientific artefacts articulate the essential knowledge emerging from the described research. The granularity of the scheme becomes, however, important when trying to observe and analyse evolving patterns and trends, for example from one *Goal* to a set of *Observations*, and in particular when such relations span multiple publications. Moreover, in order to gain a deeper understanding of this knowledge, the types of scientific artefacts represented by the scheme should map, as close as possible, to domain-specific aspects. Hence, while one could use generic models to represent and extract artefacts in the EBM domain, by using PICO one addresses directly the specific application and domain requirements. Finally, our focus on the PICO scheme is also motivated by the lack of well-established annotated corpora. To our knowledge the NICTA-PIBOSO corpus is currently the only such resource available for EBM, the other existing corpora being targeted towards other domains (e.g., biochemistry in the case of the corpus described in [5]) or being completely generic – for example, the Wilbur corpus [6].

Recognising scientific artefacts and their complex relationships within a publication and across multiple publications is extremely difficult. Furthermore, in order to capture their rhetorical nature and semantics, one needs to bridge the gap between unstructured text and some structured formalism (e.g., [1,7,5]). Finally, once formalised, they require integration, consolidation and linking, in order to create a comprehensive and interlinked overview of a domain. Our ultimate goal is to provide a holistic solution to the lifecycle of EBM scientific artefacts, from unstructured text to an enriched, consolidated and linked network. Achieving this goal would enable clinicians to gain a deeper understanding in the role of different PICO elements in similar studies and to discover new relations and trends within them. For instance, this would allow us to analyse better the impact of the variation of the number of participants (i.e., *Population*) on the *Outcomes* of two similar studies with the same *Intervention*, e.g., using the following two statements from two different publications: “A total of 23 patients with mandibular gingival cancer were treated with docetaxel by intra-arterial infusion and systemic chemoradiotherapy with cisplatin.” and “In all, 34 patients (21 men and 13 women) with squamous cell carcinoma of the gingiva underwent radiation therapy with concurrent intra-arterial infusion chemotherapy with cisplatin and docetaxel.”

In this paper, we focus on the first step of the above-described goal, i.e., automatic extraction of rhetorical artefacts (or PICO elements) from abstracts in the EBM domain. Research on automatic recognition of scientific artefacts in biomedical publications has been reported from the early 2000s, but has only recently become more prominent, with most approaches employing Machine Learning techniques guided by specific schemes. As shown in [5], three major research directions can be distinguished: (i) sentence or zone classification according to a predefined annotation scheme [8,9,5]; (ii) detection and analysis of speculative language and hedging [10,11]; and (iii) sentence classification according to a multi-dimensional scheme [6,7,12].

In the context of the EBM domain, all existing approaches that aim at extracting rhetorical artefacts are designed to work on publication abstracts [13–15,2]. Furthermore, some of them have been developed around an extended, fine-grained PICO scheme – PIBOSO [2]. PIBOSO refines PICO by categorising scientific artefacts in six categories, rather than four: (i) *Population* – the group of individuals participating in a study; (ii) *Intervention* – the act of interfering with a condition to modify it or with a process to change its course; (iii) *Background* – material that places the current study in perspective, e.g. work that preceded the current study; information about disease prevalence; etc.; (iv) *Outcome* – a summarisation of the consequences of an intervention; (v)

Study Design – the type of study that is being described; and (vi) *Other* – other information provided in the publication.

From a technical perspective, the underlying task is represented by sentence classification according to the PICO/PIBOSO scheme. Hence, an associated challenge is the encoding of the characteristics of a sentence in a format able to preserve both token and sentence features together. This would enable us to take advantage of all the information that can be inferred from the sentence tokens in a setting that deals with a sentence as the meaningful unit. The Machine Learning features used so far in this context vary greatly, yet they can be grouped into two major categories: structural and sequential features [14,2,16]. Structural features capture the position of the sentence in the context of the given abstract, while sequential features leverage token-based information, such as bag of words or nearest neighbours using sliding windows.

Our approach follows the same design principles as the existing solutions and employs Machine Learning techniques to perform sentence classification. In terms of features, we combine token-level and sentence-level features that capture both the positional (e.g., placement in the abstract), as well as the sequential (e.g., predicted classes of adjacent neighbours) aspects of the target classes. In addition, we introduce a new category of features – statistical features – that joins the two levels of feature granularity by computing sentence-wide token statistics. The intuition behind these new features is that each target class may be characterised by a specific statistical distribution of the types of tokens composing it – e.g., based on the verb or on other low-level linguistic information (exact details are provided in Section 3). Also, inferred sequential features derived from the co-occurrence of similar types of sentences are proposed. The overall combination of these features results in a representative feature vector that enables the training of an accurate classifier.

We have performed an extensive set of experiments using the NICTA-PIBOSO corpus [2] – described in detail in Section 3.1. Our results show that classifiers trained with a mixture of the above-presented features are able to achieve an accuracy comparable to the state of the art. Four different classification methods have been investigated: Conditional Random Fields (CRF) [17], Support Vector Machines (SVM) [18], Naive Bayes [19], and Multinomial Logistic Regression [20]. Among these, CRF has achieved a micro-average *F*-score of 90.74% and 87.21%, respectively, over structured and unstructured abstracts, which represents an increase of 25.48 percentage points and 26.6 percentage points in *F*-score when compared to best existing approach. In summary, the main contributions of this manuscript are: (i) a selection of features that enables an increased accuracy of sentence classification based on the PIBOSO scheme and (ii) a comprehensive experimental setup that provides a good overview of the behaviour of different classification mechanisms using diverse feature configurations.

The remainder of the paper is structured as follows: Section 2 describes relevant existing research. In Section 3 we detail the classification features and the data used within our experiments, as well as the experimental setup. Section 4 presents the evaluation results of achieving based on different feature configurations – (i.e. the presence or absence of particular feature sets in the experiments), and Section 5 discusses the experimental results and provides a thorough comparative analysis of the state of the art. Finally, we conclude in Section 6.

2. Related work

In order to reach an efficient and effective judgement in EBM, Richardson et al. [1] concluded that clinicians require an addition fundamental skill, i.e., to ask well-built clinical questions. Consequently, they propose the PICO criteria for formulating these

Download English Version:

<https://daneshyari.com/en/article/6928424>

Download Persian Version:

<https://daneshyari.com/article/6928424>

[Daneshyari.com](https://daneshyari.com)