Journal of Biomedical Informatics xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Automatic endpoint detection to support the systematic review process

Catherine Blake [a,*], Ana Lucic [b]

[a] Graduate School of Library and Information Science and Medical Information Science, Center for Informatics Research in Science and Scholarship (CIRSS), University of Illinois, Urbana Champaign, 501 E. Daniel Street, MC-493, Champaign, IL 61820-6211, USA
[b] Graduate School of Library and Information Science, University of Illinois, Urbana Champaign, 501 E. Daniel Street, MC-493, Champaign, IL 61820-6211, USA

ARTICLE INFO

Article history:
Received 20 November 2014
Revised 5 May 2015
Accepted 6 May 2015
Available online xxxx

Keywords:
Systematic review
Information extraction
Evidence-based medicine
Text mining

ABSTRACT

Preparing a systematic review can take hundreds of hours to complete, but the process of reconciling different results from multiple studies is the bedrock of evidence-based medicine. We introduce a two-step approach to automatically extract three facets – two entities (the agent and object) and the way in which the entities are compared (the endpoint) – from direct comparative sentences in full-text articles. The system does not require a user to predefine entities in advance and thus can be used in domains where entity recognition is difficult or unavailable. As with a systematic review, the tabular summary produced using the automatically extracted facets shows how experimental results differ between studies. Experiments were conducted using a collection of more than 2 million sentences from three journals *Diabetes*, *Carcinogenesis* and *Endocrinology* and two machine learning algorithms, support vector machines (SVM) and a general linear model (GLM). F_1 and accuracy measures for the SVM and GLM differed by only 0.01 across all three comparison facets in a randomly selected set of test sentences. The system achieved the best performance of 92% for objects, whereas the accuracy for both agent and endpoints was 73%. F_1 scores were higher for objects (0.77) than for endpoints (0.51) or agents (0.47). A situated evaluation of Metformin, a drug to treat diabetes, showed system accuracy of 95%, 83% and 79% for the object, endpoint and agent respectively. The situated evaluation had higher F_1 scores of 0.88, 0.64 and 0.62 for object, endpoint, and agent respectively. On average, only 5.31% of the sentences in a full-text article are direct comparisons, but the tabular summaries suggest that these sentences provide a rich source of currently underutilized information that can be used to accelerate the systematic review process and identify gaps where future research should be focused.

1. Introduction

Systematic reviews are the bedrock of evidence-based medicine (EBM), however the time required to conduct a review can be extensive (five or six people more than 1000 h to complete [1]), which can cause delays between when an experimental result is published and when those results are integrated into clinical practice. The systematic review process is fundamentally an information organization activity comprising of information retrieval, extraction, and analysis [2]. The increased availability of electronic abstracts and full text articles have led to several automated methods to accelerate the process, where most systems focus on the information retrieval stage [3–8] and to some extent, the information extraction stages [2,9]. In addition to strategies that operate over published literature, manual efforts have been proposed to capture data required in a systematic review [10,11]. The approach presented in this paper augments these efforts by leveraging a currently under-utilized resource that occurs in scientific articles– the direct comparison sentence.

The idiom "you should not compare apples to oranges" underscores the common practice of only comparing entities that are of the same type. This practice in common language usage also appears in scientific articles, such as in sentence 1, where we learn that the endpoint *uterine weights* was used to compare the two entities the *TAM* group (TAM in this article refers to Tamoxifen) and *controls*. The article from which this sentence was drawn also provides information about the dosage and duration of the TAM treatment and the type of animals used in the control group that may be necessary to contextualize the experimental results, but this single short sentence alone provides a succinct summary of one of the experimental findings.

* Corresponding author. Tel.: +1 (217) 333 0115.
E-mail addresses: clblake@illinois.edu (C. Blake), alucic2@illinois.edu (A. Lucic).

http://dx.doi.org/10.1016/j.jbi.2015.05.004
1532-0464/

(1) In the present study, uterine weights[endpoint] of intact animals treated with TAM[agent] was decreased as compared with controls[object], although not significantly. 12189200

Although comparative sentences provide a wealth of information, from a linguistic perspective these sentences have earned a reputation for being "notorious for its syntactic complexity" [12] and "very difficult" to process automatically [13]. Despite these challenges, the densely packed information contained in a comparison sentence have been used to identify aspects about products that customers prefer or dislike [14,15]. Similarly, there has been some work on comparatives in biomedical literature. For example one study focused on comparative sentences from MEDLINE abstracts that contained two drugs [16]; however, such a strategy would not consider sentence 1 because only one drug is mentioned. Comparing a drug to a placebo or control group is common in clinical trials, a practice that lead the comparative effective research (CER) community to call for more head-to-head trials [17]. Although adding a placebo entity type to the set of required entities in the earlier approach would alleviate this immediate issue, this fix does not mitigate against the need to decide the entities a priori. Our results show that authors use a range terms when describing the entities being compared, which would be very difficult to specify in advance.

In this paper, we present an automated process that identifies two entities (the agent and object) that are compared with respect to a given endpoint. The approach uses sentence structure, which removes the need for a user to provide a set of entities in advance, and minimizes the impact of errors during entity recognition. Moreover, this approach enables the results from the system to be used in domains where the entities of interest are unknown. With respect to the clinical domain, removing the need to define entities a priori means that the method presented here will work for both drug-placebo and head-to-head studies.

In addition to focusing on sentence structure, the automated approach presented here extends earlier work in biomedicine by providing system predictions at the noun phrase rather than sentence level. Consider an earlier study where semantic and syntactic features were used with three classifiers (Naïve Bayes, Support Vector Machines and a Bayesian network) to differentiate comparison from a non-comparison sentences [18]. The system described in this paper identifies the specific noun phrases that fill the agent, object and endpoint roles. Thus the system would identify *TAM* from sentence 1 as the agent *control* as the object, and *uterine weights* as the endpoint.

Our goal in this paper is twofold. First, we describe a two-step automated process that leverages direct comparative sentences. The system is evaluated using more than 2 million sentences from full-text articles that appear in a sample of full-text articles in three journals: *Diabetes*, *Carcinogenesis*, and *Endocrinology*. Second, we provide a situated evaluation to illustrate how the facets that are extracted from comparison sentences can be employed to support the systematic review process. We did not select the focus of the situated evaluation in advance, but rather the focus emerged from the entities identified by the system. The situated evaluation of diabetes treatments provides insight into how the system would perform when embedded into the existing systematic review process.

1.1. Definitions

The system described in this paper identifies noun phrases that capture two entities that are being compared (called the agent, object) and the way in which those entities are compared (called the endpoints). This terminology borrows the agent and object terminology from the comparative claim described in Blake's Claim Framework [19], but we use endpoint rather than the basis of the comparison (or aspect) to capture how the entities are compared. This section provides the critical elements of a direct comparative sentence.

Comparative sentences are either gradable or not gradable. Gradable sentences enable the reader to order entities, for example the *TAM group* is lower than the *control group* with respect to *uterine weight* in sentence 1. In contrast, a non-gradable comparison does not provide enough information to order the reported entities, such as in sentence 2 where *tamoxifen* and *4-hydroxytamoxifen* cannot be ranked with respect to *uterine weight*. Non-gradable sentences are further characterized as *similar* or *different*, where sentence 2 is a *similar* type of non-gradable comparison.

(2) Since tamoxifen[agent] and 4-hydroxytamoxifen[object] had nearly identical effects on uterine weight[endpoint], this indicates that only a small proportion of the administered 4-hydroxytamoxifen reached the uterus. 10190564

Definition 1. Direct comparison sentences capture either gradable or non-gradable comparisons.

In addition to requiring that sentences be either gradable or non-gradable, a direct sentence comparison must include entities that play the role of an agent and an object as defined in Blake's Claim Framework [19]. Consider sentence 3 where two endpoints as well as the object of the comparison are explicitly mentioned but the sentence does not mention the agent. This sentence is considered out of scope because it does not contain an entity that plays the agent role. Agents and objects are typically different noun phrases, but in some sentences a single noun phrase can play both roles.

(3) LPO levels[endpoint] were slightly (596 ± 89 nmol/mg protein), but not significantly ($P > 0.05$) different from the normal group[object], and GSH levels[endpoint] remained significantly decreased ($P < 0.02$ vs. normal) (Fig. 1A and B). 12606525

Definition 2. Direct comparison sentences include entities that play an agent and object role.

Lastly, we are particularly interested in how entities are compared. For example the non-gradable comparison sentence shown in 4 provides noun phrases for the agent (men) and the object (women) roles, but does not provide information about how the authors established that men and women responded differently to hypoglycemia. Although this sentence may be useful to infer semantic types (i.e. that men and women have the same semantic type), the sentence does not contain an endpoint and thus would not be included in the system summary.

(4) Men[agent] and women[object] respond differently to an acute bout of hypoglycemia. 12829642

Definition 3. Direct comparison sentences include information about how entities were compared (the endpoint).

It is quite common for an author to report more than one comparison in the same sentence, such as in sentence 5 that first compares *bazedoxifene* with *ethinyl estradiol* and then compares *bazedoxifene* with *raloxifene*. The system should identify all noun phrases that play the agent, object or endpoint roles from each sentence.

Please cite this article in press as: C. Blake, A. Lucic, Automatic endpoint detection to support the systematic review process, J Biomed Inform (2015), http://dx.doi.org/10.1016/j.jbi.2015.05.004