Journal of Biomedical Informatics 52 (2014) 457–467

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Methodological Review

Text summarization in the biomedical domain: A systematic review of recent research

Rashmi Mishra [a], Jiantao Bian [a,b], Marcelo Fiszman [c], Charlene R. Weir [a,d], Siddhartha Jonnalagadda [e], Javed Mostafa [f], Guilherme Del Fiol [a,*]

[a] Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA
[b] Clinical Modeling Team, Intermountain Healthcare, Salt Lake City, UT, USA
[c] Lister Hill Center, National Library of Medicine, Bethesda, MD, USA
[d] VA Medical Center, Salt Lake City, UT, USA
[e] Department of Preventive Medicine-Health and Biomedical Informatics, Northwestern University, Chicago, IL, USA
[f] School of Information and Library Science (SILS), University of North Carolina, Chapel Hill, NC, USA

ARTICLE INFO

Article history:
Received 26 February 2014
Accepted 23 June 2014
Available online 10 July 2014

Keywords:
Text summarization
Intrinsic evaluation
Language processing
Machine learning
Biomedical domain

ABSTRACT

Objective: The amount of information for clinicians and clinical researchers is growing exponentially. Text summarization reduces information as an attempt to enable users to find and understand relevant source texts more quickly and effortlessly. In recent years, substantial research has been conducted to develop and evaluate various summarization techniques in the biomedical domain. The goal of this study was to systematically review recent published research on summarization of textual documents in the biomedical domain.
Materials and methods: MEDLINE (2000 to October 2013), IEEE Digital Library, and the ACM digital library were searched. Investigators independently screened and abstracted studies that examined text summarization techniques in the biomedical domain. Information is derived from selected articles on five dimensions: *input*, *purpose*, *output*, *method* and *evaluation*.
Results: Of 10,786 studies retrieved, 34 (0.3%) met the inclusion criteria. Natural language processing (17; 50%) and a hybrid technique comprising of statistical, Natural language processing and machine learning (15; 44%) were the most common summarization approaches. Most studies (28; 82%) conducted an intrinsic evaluation.
Discussion: This is the first systematic review of text summarization in the biomedical domain. The study identified research gaps and provides recommendations for guiding future research on biomedical text summarization.
Conclusion: Recent research has focused on a hybrid technique comprising statistical, language processing and machine learning techniques. Further research is needed on the application and evaluation of text summarization in real research or patient care settings.

1. Introduction

The amount of information available for clinicians and clinical researchers is growing exponentially, both in the biomedical literature and patients' health records [1,2]. To provide optimal patient care, clinicians need to efficiently and effectively retrieve, interpret, and integrate relevant information from multiple source [2]. Likewise, researchers need to navigate a vast amount of information from the biomedical literature for tasks such as generating new hypotheses and understanding the state-of-the-art in a given area. Electronic resources such as online literature databases and electronic health record (EHR) systems have been designed to help clinicians and researchers with their information management needs. However, the more resources grow, the harder it becomes for users to access information efficiently. Advances in information retrieval technology have shown some value in helping clinicians manage information overload [3]. Yet, information seekers often need to screen several documents and scan several pages of narrative content to find information that is relevant to their information needs [2].

* Corresponding author. Address: Department of Biomedical Informatics, University of Utah, 421 Wakara Way, Salt Lake City, UT 84108, USA. Fax: +1 (801) 581 4297.

E-mail address: guilherme.delfiol@utah.edu (G. Del Fiol).

http://dx.doi.org/10.1016/j.jbi.2014.06.009
1532-0464/

Automatic text summarization is a promising method for helping clinicians and researchers seeking information to efficiently obtain the "gist" in a given topic by producing a textual or graphical summary from one or multiple documents. A summary is "a reductive transformation of source text to summary text through content reduction selection and/or generalization on what is important in the source" [4]. The goal of text summarization is to present a subset of the source text, which expresses the most important points with minimal redundancy. The reduction of data accomplished by text summarization aims to allow users to identify and process relevant information more quickly and accurately. Thus, text summarization may become an important tool to assist clinicians and researchers with their information and knowledge management tasks.

Important advances have been achieved recently in text summarization. As a result, several applications that leverage text summarization techniques have become available to the general public [5]. There has been a growing interest in researching text summarization techniques in the biomedical domain. An informal literature survey conducted by Afantenos et al. identified ten biomedical text summarization studies published between 1999 and 2003 [6]. Since then, there have been significant advances in the summarization tools and techniques employed in the biomedical domain. However, no systematic review on this topic has been conducted to date. A systematic review will promote improved understanding of the literature on this topic, identify gaps, and provide directions for future research. In the present study, we conducted a systematic review on text summarization methods applied to the biomedical literature and EHR systems. The systematic review is aimed at: (1) identifying the different techniques, areas of application, and evaluation methods over the last decade; (2) identifying research trends; (3) identifying research gaps; and (4) proposing recommendations to guide future research.

2. Methods

We based the methodology of our study on the *Standards for Systematic Reviews* set by the Institute of Medicine [7]. The study protocol was iteratively designed and refined with input from the study co-authors. The following subsections describe each of the steps that were performed to identify, screen, and abstract data form the included studies.

2.1. Data sources and searches

The search strategies were developed with the help of the expert review committee and a medical librarian. The strategies were further tested and refined against a list of relevant citations from previous reviews on the topic. Three databases were searched: PubMed, IEEE, and ACM digital library. Searches were limited to the period between Jan 1st 2000 and October 16th 2013. The overall search strategy was to retrieve articles that included terms related to *text summarization*, such as *"medical text summarization", "clinical text summarization", and "biomedical summarization"*. The search time period was limited to avoid overlap with the review by Afantenos et al. [6]. The search strategies applied are provided in the online supplement. In addition to searching literature databases, we inspected the citations of included articles with a special focus on previous relevant reviews. Finally, we requested input from the study co-authors for potentially relevant references that could have been missed by the literature search.

2.2. Study selection

We included original research studies that developed and evaluated text summarization methods in the medical domain, including summarization of the biomedical literature and electronic health record documents.

We excluded studies that met any of the following criteria: (1) Summarization of content outside the biomedical domain; (2) summarization of the basic science literature, such as molecular biology; (3) not original research, such as editorials and opinion papers; (4) emphasis placed on text summarization tools, but without an evaluation component; (5) related techniques (e.g., text mining) that can be used to support text summarization, but that did not produce a summary; (6) not written in English; (7) image and multimedia summarization without a text summarization component; and (8) articles included in the survey by Afantenos et al. [6].

2.2.1. Abstract screening

The title and abstract of each article retrieved were reviewed independently by two of the study authors (JB, RM). Articles were labeled as "not relevant" or "potentially relevant." For calibration and refinement of the inclusion and exclusion criteria, 50 citations were randomly selected and independently reviewed. Disagreements were resolved by consensus with a third author (GDF). In a second round, another set of 50 articles was reviewed in a similar way. In a third round, 815 abstracts were independently reviewed achieving a strong level of agreement (kappa = 0.82). In a final round the remaining citations (7871) were evenly assigned between the two reviewers and screened.

2.2.2. Article selection

Two authors (JB, RM) independently reviewed the full-text of a subset of 112 citations labeled as potentially relevant in the abstract screening phase. Disagreements between the two reviewers were reconciled with the help of a third reviewer (GDF). Since inter-rater agreement in this phase was high (kappa = 0.78), the remaining full-text articles (120) were evenly assigned between the two reviewers and screened.

2.3. Data extraction

A data abstraction spreadsheet was developed based on the text summarization categories described by Mani which are summarized below [8]. Two authors (RM, JB) independently reviewed the included articles [34] to extract the data into the data abstraction spreadsheet. Next, the data were compared and disagreements were reconciled through consensus with the assistance of a third reviewer (GDF).

The data abstraction tool was adapted from a classification of text summarization methods described by Mani and Maybury [9]. This classification consists of five dimensions: *input*, *purpose*, *output*, *method* and *evaluation*. The five classification categories are further described below.

2.3.1. Input

This dimension has been termed as "unit input parameter" or the "span parameter" by Sparck-Jones and Mani respectively [4,8]. We categorized the Input dimension according to four attributes: (1) *single* versus *multiple* document summarizations; (2) *monolingual* (input and output on the same language) versus *multilingual* summarization (input or output in multiple languages; (3) *abstract* versus *full-text*; (4) *biomedical research literature* versus *EHR documents*.

2.3.2. Purpose

Purpose denotes the stated main goal of the generated summary. This dimension was categorized according to two attributes: (1) *Generic* versus *user-oriented* summaries; and (2) *Broad spectrum* versus *Clinical decision support*.