Neurocomputing 135 (2014) 53–60

Contents lists available at ScienceDirect

Neurocomputing

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

Extracting answers from causal mechanisms in a medical document

A. Sobrino [a], C. Puente [b,*], J.A. Olivas [c]

[a] Faculty of Philosophy, University of Santiago de Compostela, Santiago de Compostela, Spain
[b] Advanced Technical Faculty of Engineering ICAI, Pontificia Comillas University, Madrid, Spain
[c] Information Technologies and Systems Department, University of Castilla-La Mancha, Ciudad Real, Spain

ARTICLE INFO

Article history:
Received 25 October 2012
Accepted 30 May 2013
Available online 7 February 2014

Keywords:
Causal questions
Mechanisms
Imperfect causality
Answering causal questions

ABSTRACT

The aim of this paper is to approach causal questions in medical documents eventually recovered from a search engine. Causal questions *par excellence* are *what*, *how* and *why*-questions. The 'pyramid of questions' shows this. At the top, why-questions are the prototype of causal questions. Usually why-questions are related to scientific explanations. Although cover law explanation is characteristically of physical sciences, it is less common in biological or medical knowledge. In medicine, laws applied to all cases are rare. It seems that doctors express their knowledge using mechanisms instead of natural laws. In this paper we will approach causal questions with the aim of: (1) answering what-questions as identifying the cause of an effect; (2) answering how-questions as selecting an appropriate part of a mechanism that relates pairs of cause-effect (3) answering why-questions as identifying central causes in the mechanism which answer how-questions. To automatically get answers to why-questions, we hypothesize that the deepest knowledge associated to them can be obtained from the central nodes of the graph that schematizes the mechanism. Our contribution is concerned with medical question answering systems, even though our approach does not address how to retrieve medical documents as a primary answer to a question, but how to extract relevant causal answers from a given document previously extracted by using a search engine. Thus, our paper deals with the automatic detection and extraction of causal relations from medical documents.

1. Introduction

Causality is an ancient topic that came from Aristotle [1,2], who distinguished between four types of causes:

- Material cause, involving the physical matter of which something is made; that is, the mass of which it consists.
- Formal cause, focusing on the way that a thing is intended and planned to be.
- Efficient cause, quoted as 'the primary source of the change'; the prior movement or the source energy that triggers the final effect.
- Final cause, as the end, goal or aim that a process leads to. The final cause is the teleology (from greek *telos*) that something is supposed to serve.

The Aristotelian view of causality traditionally offered a frame for providing answers to causal questions, as *what-q* or *why-q*.

In effect, Aristotle's typology serves to answer *what* and *for-what* questions. For example, in the presence of a statue, we can ask for the following queries, which belong to the types of cases aforementioned:

- 'What is it made from?' It is made of metal (material cause);
- 'What is its form?' A man in a praying attitude (formal cause);
- 'What produced it?' The sculptor (efficient cause);
- 'For what purpose?' To pay tribute to a virtuous person (final cause).

But Aristotle's typology enables to answer why-questions as well. Efficient causes seem to be the more appropriate for this task. In this paper we will follow this view.

Aristotle's efficient cause is intended as a way of performing explanations. Explanations are usually related to why-questions. A typical – although not academic – way to provide an explanation is to distinguish the components involved in a process identifying the first cause or impulse. In the aforementioned example, the sculptor is who acted in the first place.

More in depth, and referring to a medical context, we can argue that the efficient cause of a diagnosis it is not a doctor, but his

* Corresponding author.
E-mail addresses: alejandro.sobrino@usc.es (A. Sobrino), cristina.puente@upcomillas.es (C. Puente), joseangel.olivas@uclm.es (J.A. Olivas).

http://dx.doi.org/10.1016/j.neucom.2013.05.056
0925-2312

medical knowledge. As we previously said, Aristotle pointed out that efficient causes are the primary source of change. The transmission of such changes in a causal network is known as a mechanism. So, efficient causes are related to mechanisms. In this paper we deal with the analysis of *what*, *how* and *why*-questions focusing on efficient causes and mechanisms.

The following pyramid arranges interrogative particles depending on the potential complexity of their answers [3] (Fig. 1).

Ascending in the pyramid means the use of causal interrogatives, demanding complex answers instead of *yes, no* replies to questions, stimulating reflective and deepening thinking. At the top, why-questions ask for some kind of explanation.

Medical research provides mechanisms for explaining diseases, leading to new ideas about how the disease can be treated both for therapeutic and theoretic purposes. A disease explanation is better understood showing a causal mechanism, describing the interrelations among multiple factors involved in its origin and development. F. ex., the U.S National Institute of Health concluded that there is a correlation between bacterial infection and ulcers or, more specifically, a causal influence between infection with Helicobacter Pylori and the duodenal ulcer. But note that correlation is not the same as causation. Correlation can provide evidence for causes in terms of probabilistic contrast; that is, in terms of how much probable is an effect (e) with a cause (c) than without it (c). Cheng baptized that difference as the 'causal power of *c*' [4]. But whereas the probability with which *c* produces (when *c* is present or is absent), is an observable frequency, the causal power is a theoretical entity, like electrons induced from observations of traces. May be that *e* is due to alternative causes to *c*. Therefore, the causal power of *c* over *e* is better interpreted as a dispositional property of some entities to provoke, in the long run, other entities; i.e., the propensity – not probability – of *c* causing *e*. But propensity is a dispositional and imprecise concept.

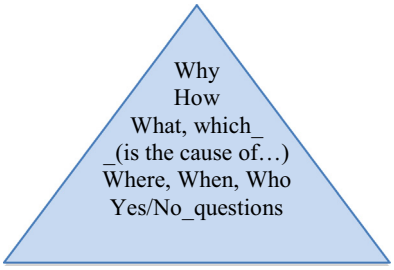

Fig. 1. Pyramid of questions' complexity.

In Medicine, diseases like schizophrenia, bulimia, or anorexia, frequently show fuzzy causal boundaries, as they present similar symptoms. Thus, causes frequently are complex and vague. That is, in medicine we should shift from the classic scientific paradigm of 'theory' to the more evasive of 'mechanism'. Causes are not single, but complex [5] and are not crisp, but imprecise [6].

This paper will approach how to reach answers to causal questions from mechanisms: (1) what-questions as identifying the cause of a mechanism; (2) how-questions as selecting the appropriate parts of a mechanism and (3) why-questions as extracting and summarizing highlighted paths in the answer of how-questions. Answers to why-questions match scientific explanations in a negative and a positive note: (−) as previously said, scientific explanations in medicine are based on mechanisms instead of natural laws; (+) scientific explanations are generally concerned with deepening or centrality, providing explanations as detailed as possible.

Medical question answering systems (MQ-AS) aim to provide the users with direct answers to the posed questions; instead of furnishing them with a large amount of relevant potentially documents. For seeking direct answers, MQ-AS need to go beyond the surface-level analysis of texts, providing lexical-syntactic and semantic resources, as well as reasoning capabilities, providing inference mechanisms to obtain more adequate answers. Traditional MQ-AS accept questions as inputs; those questions trigger a search engine providing relevant documents and, finally, the system operates over those documents in order to reach the direct answer to the question. In our approach, the step 2 is omitted. We act directly over a document previously recovered.

Although most MQ-AS have exploited syntactic or semantic resources, few approaches have scanned the utility of inference mechanisms. Our approach uses syntactic and semantic resources to perform an inference-based one, as aims to extract causal semantic relations from inferential mechanisms. Our approach is in the vein of the Girju's work about how to automatically detect and extract causal relations from texts [7].

Thus, our contribution is organized as follows: in section I we analyze the mechanism to provide an answer to what-questions. In Section 2 we will focus on the answer to how-questions. In Section 3 why-questions are addressed. Finally, a section of conclusions and references close this work.

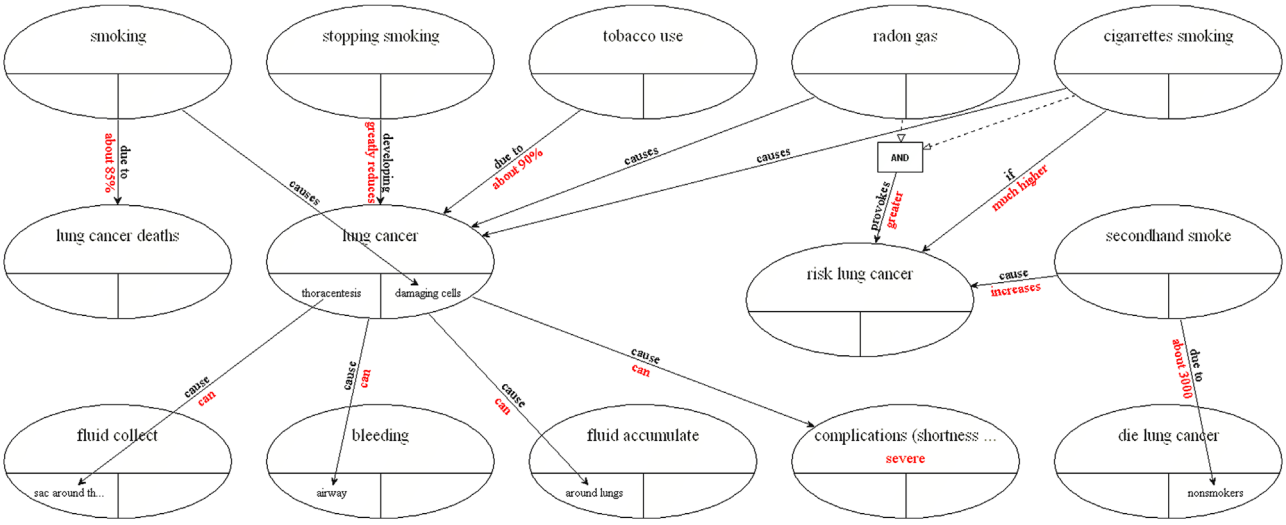

Fig. 2. Causal graph of the question 'What causes lung cancer?'