A Brief Comparison between Two Medical Question-Answering Systems: "Fushman-Lin system" and MEANS

Information on the Two Publications

To make a comparison study, this essay takes two publications: *Answering Clinical Questions with Knowledge-Based and Statistical Techniques* written by Dina Demner-Fushman and Jimmy Lin, published in 2007; *MEANS: A Medical Question-Answering System Combining NLP Techniques and Semantic Web technologies* written by Asma Ben Abacha and Pierre Zweigenbaum, published in 2015. Both of the two publications discuss about the construction of medcial question-answering system with computational methods.

ABSTRACT

This essay makes a comparison between two medical question-answering systems: the clinical question answering system designed by Dina Demner-Fushman and Jimmy Lin, the "Fushman-Lin system", and the medical question-answering system (MEANS) designed by Asma Ben Abacha and Pierre Zweigenbaum. Comparison shows that these two systems are all the product of modern medicine and information technology. They serve physicians with advanced medical decision-making assistant. These two system are both established on knowledge-based starting point. These two systems share very similar framework, which contains three elements: enquiry input, medical text input, and answering composition. With the help of originally designed ontology, MEANS can process more types of enquiry input, more types of medical text input, thus presenting more flexibility and adaptability than Fushman-Lin system. Some personal comments on their shortcomings and prospective improvements are also included at the end of the essay.

KEY WORDS: Medicine; Question-answering; Information technology; Computational linguistics

Content

This essay made a general comparison between two medical question-answering systems: one designed by Dina Demner-Fushman and Jimmy Lin in the article named *Answering Clinical Questions with Knowledge-Based and Statistical Techniques*, and the other one designed by Asma Ben Abacha and Pierre Zweigenbaum in the article named *MEANS: A Medical Question-Answering System Combining NLP Techniques and Semantic Web technologies*. Because Fushman and Lin did not give their system a name, for the discussion convenience, their system is called "Fushman-Lin system" in this essay. This essay firstly compares the background and starting point between the two systems; then compares their general framework and discusses their similarities and differences in detailed sub-procedures; and finally points out the author's own opinion about the shortcomings and delivers the prospect of future medical question-answering system.

The design of the two medical question-answering system shares similar background. With the development of modern medicine research, thousands of professional medical publications are produced every year. How to deal with medical problems with the most advanced solutions from various sub-fields at any time has become a difficult question for medical workers around the world. Physicians dealing with problems in one field could obtain significant inspirations from other related fields. But it is impossible for one physician to view all the new and old text which might contain valuable content in person. The development of computer science and information technology, especially computational language processing, provides possible medical question-answering methods powered by machine work. With the help of computer, people could get updated solutions to medical problems through very quick machine powered retrieval from millions of textual materials exactly at the moment of demand. Both Fushman-Lin system and MEANS are promoted for this purpose.

Because of the similar task for searching answers to medical questions from existing medical documents, both of the two systems are called knowledge-based question-answering system. Knowledge-based means that the substantial property of the Q-A system stands on a huge collection of existing digital medical materials, a huge medical document corpus. Any potential answer should be directly searched and griped from the corpus, or be generated from the primitive retrieval results from the corpus. It implies that the quality of the corpus being used can apparently influence the effectiveness and intelligence of the whole Q-A system. Knowledge-based also means that the composing of enquiries, the mechanical method to process the question and answering analysis should be designed according to a certain kind of human knowledge pattern. Not only are the knowledge facts required, but also "the knowledge of knowledge", which is a kind of quite complete ontology of a certain domain. Obviously, a medical ontology is needed here in order to organize, to synthesize, to control, and to operate all the "knowledge points" in the medical document corpus.

In general, Fushman-Lin system and MEANS share similar features in the basic framework of medical question-answering solution. They have same underlining procedure which can be summed up into 3 elements: the first one is to make human's medical enquiries understandable by machine; the second one is to make existing medical materials understandable by machine; the last one is, with the help of machine work, to form human's ideal answers based on enquiries and related medical materials. For the first part, both of the two systems try to convert natural language into specially featured machine-readable language. The common problem is how to design the mechanism to ensure that the computer can easily recognize the natural language for further information extraction. For the second part, both of the two systems try to convert the knowledge base, or medical document corpus, into the format that can be processed by machine language. The third part is even more complicated and contains more functions: how to make it possible that the converted users enquiry can be directly connected with the converted medical document corpus; how to extract the key points, what users really mean and what they really

want, from converted users original enquiries; how to extract key points which could contain the expected answers from the converted medical corpus; how to evaluate the significance of different groups of possible information; and how to generate users-readable answers from the extracted information. These three very basic steps of procedure seem to be solid and firm since MEANS was published after a lapse of many years behind Fushman-Lin system, which could imply that the basic elements of knowledge-based medical question-answering system and their relative detailed problems have not changed until now. These three basic points are backbones for both Fushman-Lin system and MEANS, and perhaps all the question-answering systems of this kind in the world. Although these two systems are generally similar in the very basic framework, they still have many different features in the 3 elements above.

In the medical enquiry production part, both two systems face to physicians, rather than patients. The same subject to physicians may be cause by the immature feature of the two systems. Though medical question-answering system has been developed for many years, it is still not well-rounded enough to achieve large-scale popularization. Immature system means that the enquiries imputed for the machine recognition are still limited, constrained and need a strong professional logic in a special domain. In Fushman-Lin system, this limit is more obvious: only WH questions are considered in it, and input questions should be well-organized by users. Instead of using NLP methods to translate original enquiries into machine language, this system has a high requirement for users to input their enquires more likely to be a cross-word puzzle in which any keyword imputed should be well considered to fulfill a special position with special a meaning. This design is to reduce as much as possible the mistakes might be caused by user's casual language expression. In some way, users shoulders a part of machine work for original language translation. Though this kind of design does reduce the mistakes of original meaning recognition, it is not like an ideal modern natural language processing, which should be more intelligent and more convenient to approach. But, MEANS takes a very different way of nature language recognition. It implements NLP to translate the original human language into machine-readable language. In order to reduce mistakes caused by the weakness and probable errors in language processing, it takes a special technique, called "query relaxation approach", to generalize the meaning of original language, thus making the meaning domain processed by NLP large enough to contain the original meaning of natural language. MEANS also extends the input question type: not only WH questions can be processed, but also Boolean Yeas/No questions. People are not required to form seriously professional enquires with special format. Instead, they can ask more natural questions to the system. This change significantly improves the users experience when people are cooperating with computer: physicians never need to waste time thinking hard about how to make up their enquiries in a special and serious format, or they would not get sufficient answers. The implement of advanced NLP techniques also makes the medical question-answering system available for those who have never been trained before about how to make up enquires with special format. That makes it possible for every physician to use the question-answering system without tedious pre-training.

The improvement of enquiry input step in MEANS is substantially based on the improvement of inner ontology. The ontology of medical knowledge is necessary for machine to match the underlining connection between annotated user's enquiries with annotated medical text, because the ontology is also used to do the annotation of user's enquiries and medical corpus in order to make them readable for computer. The medical ontology generates a commonly-used annotation system implemented both on user's enquiry and medical corpus to ensure the accessibility of retrieval for related citations in medical corpus according to users enquires. MEANS takes the advantages of the newly designed ontology with more complicated conception and syntactic detection and taxonomy methods than the old Evidence-Based Medicine (EBM) ontology, which is used by Fushman-Lin system. The more complicated and well-rounded ontology ensures that MEANS can understand more kinds of input, or in other words, more kinds of different question types than Fushman-Lin system.

The EBM ontology is only used by MEDLINE and its search engine, PubMed. Due to the lack of originally made ontology, Fushman-Lin system has to limit its accessible medical corpus on MEDLINE and its extended parts, which seriously constricts its scalability. The old EBM system can only annotate pure professional medical concepts without any personal information about patients. That makes it impossible to provide targeted personal medical treatment for individuals. The individuality is omitted from medical suggestion process. But for MEANS, the more complicated ontology contains annotation method for patient's personal medical record. It means that patient's medical record can also be added into the whole corpus and be recognized by the system, thus making it possible to retrieval medical treatment with the preference of individual conditions. The implement of patient's medical record is also pointed out in the "discussion" part of Fuman and Lin's article to be a direction of future improvement of their system. And it has been achieve in MEANS with the help of novel ontology.

With the help of originally designed ontology, MEANS is not limited with any specific medical document corpus or any specifically attached searching engine, but various of medical documents and publications. Any medical text can be processed by MEANS, so, theoretically, it has no boundaries for new medical documents without specially designed standard format. The universality could make MEANS much more powerful than Fushman-Lin system, since they are both knowledge-based question-answering system, the size of corpus, which is the stand of the knowledge variety, is vitally important for the effectiveness of the whole system. Different countries and organizations have different composing standard for medical text. But the primary intention of the design of medical question-answering system has no boundaries in different standards or formats. Without the universal-efficient ontology and related annotation system, Fushman-Lin system is not accessible to medical documents outside MEDLINE. Documents should be composed and synthesized with the guideline of MEDLINE, or they would not be processed by Fushman-Lin system. One the country, since MEANS is originally designed with the corpus of different sources of scientific articles and clinical texts, its adaptability to universal medical documents is much better than Fushman-Lin system.

When griping key points which are attached with potential answers to users enquires, both of Fushman-Lin system and MEANS take the use of a combination of semantic matching and probabilistic statistical method to select the best fitted answer pattern from several candidate answers. The semantic rule-based question-answering matching would not be that precise and accurate without the collaboration with statistical method. In order to select the best answer from many possible candidates, both of the two systems implement a scoring or evaluation method to measure the fitness among different answers. Fushman-Lin system uses the automatic scoring method to calculate the conformity between answer candidates and the users-demanded answer pattern. MEANS provide a certain part of context with the answer for users to decide whether to admit or to reject the answer conveniently.

Both the Fushman-Lin system and MEANS still have some similar shortcomings: they are designed to physicians and cannot be used by the general public as a popular medical recommendation service; both of them rely on knowledge-based ontology and corpus which cannot easily adapt changes in medical theory and novel medical concepts; they only take English as the working language, which could limit the universal use of the medical Q-A system; they only work in single-direction, which means that the whole Q-A process is the machine contribution to users, and users' selection of answers cannot help to improve the system. In order to popularize the medical Q-A system, the input terminal has to be sensitive to all kinds of questions from not only physicians, but also from patients. It is a great problem that in most cases, patients tend to search the internet by themselves before they go to see the doctor. They may come up with not professional questions with some high value information about their actual feelings. It requires the Q-A input terminal to develop a better NLP method to catch key points from more casual and more obscure, or in other words, more natural human language from different groups with different background. Because the very complex ontology in Q-A system is pre-designed according to the existing medical knowledge, it may be not that efficient towards new development in medical theory. And the number of

new medical concepts also increases at any time that there must be some vital mistakes in automatic recognition and annotation towards these new concepts. How to maintain a machine-human collaboration mechanism to ensure the timeliness and accuracy of medical corpus is still a problem. Only one language in medical Q-A system, though has the advantage of high-level unity, is still inefficient when dealing with medical documents written in different languages. The solution may be to establish an international standard for all the medical text, or to build a parallel corpus coordinating mechanism to bridge texts written in different languages, or to bridge users with different language background. Finally, it is obvious that the selection and decision-making part from users contains quite valuable information about the accuracy of machine-made answers. Users' decisions or comments deserve to be carefully collected and analyzed in order to make a great improvement of the Q-A system, or in other words, to broaden and deepen the knowledge in medical Q-A system. A feedback improvement method is supposed to be developed.

To conclude, Fushman-Lin system and MEANS share the same background, the same starting point, the same service target, similar backbone framework, but different detailed methods. Both of them are the production of modern medicine research and information technology. They are both knowledge-based Q-A system serving advanced auxiliary medical decision-making for physicians. They share the similar framework which contains the enquiry input, the medical text input and the matching process to extract answers for enquires from corpus. With the help of originally designed medical oncology, MEANS can process more types of enquires and medical text than Fushman-Lin system. Both of the two systems are not perfect. A more popular-used cross-language medical Q-A system with self-refresh function and user's feedback improvement method may be the future direction.

References

Ben Abacha, A., & Zweigenbaum, P. (2015). MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies. Information Processing & Management, 51(5), 570-594. doi: https://doi.org/10.1016/j.ipm.2015.04.006

Demner-Fushman, D., & Lin, J. (2007). Answering Clinical Questions with Knowledge-Based and Statistical Techniques. Computational Linguistics, 33(1), 63-103. doi: 10.1162/coli.2007.33.1.63