Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 33 (2007) 251–255
Expert Systems with Applications
www.elsevier.com/locate/eswa

Internet as a knowledge base for medical diagnostic assistance

Aviv Segev [a,*], Moshe Leshno [b], Moshe Zviran [b]

[a] Faculty of Industrial Engineering and Management, Technion, Haifa 32000, Israel
[b] Faculty of Management, Tel Aviv University, Tel Aviv 69978, Israel

Abstract

The paper addresses the determination of the context of medical analysis case studies. A model of context recognition was used to extract information from actual medical cases. The goal of the research was to examine a method for encapsulating a patient's medical history and current situation into keywords for the physician performing the analysis. The algorithm yielded good results in the analysis of the medical case studies and the model was able to determine the correct diagnosis in some of the cases. An advantage of the model is the use of the Internet as an existing database that is constantly updated for possible symptoms and diagnoses. The model can serve as a decision support system for a physician presented with a patient's medical record. The model can assist in identifying some of the key issues in a patient's medical records or can suggest a possible diagnosis. The model can therefore assist the physician in his review of a patient's medical records.

Keywords: Computer-assisted diagnosis; Case-based analysis; Knowledge discovery; Decision support system

1. Introduction

Clinical decision-making is a complicated process based on experience, judgment, and reasoning. This process should simultaneously integrate information from the medical literature and a variety of other sources, including quantitative results of clinical trials and, most importantly, diagnostic test results. Kukar discussed how reliability of diagnoses is assessed in medical decision-making and proposed a general framework for reliability estimation in machine learning, based on transductive inference (Kukar, 2003).

Abu-Hanna et al. suggested a method for studying and improving the performance behavior of current state-of-the-art intensive care prognostic models. The method was based on machine learning and statistical ideas and relied on exploiting information that underlies a score variable. This underlying information was used to construct a classification tree whose nodes denote patient subpopulations (Abu-Hanna & De Keizer, 2003).

A constrained-syntax genetic programming (GP) algorithm for discovering classification rules was implemented in medical data sets. The proposed algorithm includes several syntactic constraints to be enforced by the system using a disjunctive normal form representation, so that individuals represent valid, easily interpreted rule sets (Bojarczuk, Lopes, Freitas, & Michalkiewicz, 2004).

A two-phase hybrid evolutionary classification technique for extracting classification rules can be used in clinical practice for better understanding and prevention of unwanted medical events. In the first phase, a hybrid evolutionary algorithm (EA) was utilized to restrict the search space by evolving a pool of good candidate rules; e.g. genetic programming (GP) was applied to evolve nominal attributes for free structured rules and genetic algorithm (GA) was used to optimize the numeric attributes for concise classification rules without the need of discretization (Tan, Yu, Heng, & Lee, 2003).

McCarthy (1987) in his paper *Generality in Artificial Intelligence* mentioned some of the main problems existing in the field. The formalization of the notion of context was defined as one of the main problems. McCarthy argues that a most

* Corresponding author. Tel.: +972 4 8294436.
E-mail address: asegev@technion.ac.il (A. Segev).

0957-4174/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2006.04.013

general context does not exist. Consequently, the formalization of context and a formal theory of introducing context as formal objects were developed (McCarthy & Buvac, 1997).

At its most broad, the figurative sense of context implies a generalized set of relationships believed in some way or other to be understood as relevant to the object or event under discussion (Dilley, 1999). In this paper context is used as an analysis of a case study detecting keywords that could describe the patients medical symptoms, possible diagnoses or even determine the accurate diagnosis.

The effective recognition of context is essential in the common sense knowledge problem, considered to be one of the primary research areas in artificial intelligence (AI) (Ein-Dor, 1999).

Information seeking is the process in which people turn to information resources in order to increase their level of knowledge in regards to their goals (Modica, Gal, & Jamil, 2001). Information seeking has influenced the way modern libraries operate (using instruments such as catalogs, classifications, and indexing) and has affected the World Wide Web in the form of search engines.

Among the medical fields in which agents have already been considered is the retrieval of medical knowledge from the Internet (Baujard, Baujard, Aurel, Boyer, & Appel, 1998). The use of the Internet as a database provides immediate access to a nearly infinite amount of data in a multiplicity of fields, including medicine. This paper presents a model of context recognition that uses the Internet as a knowledge base to attempt to achieve information on a patient's medical condition (Segev, Leshno, & Zviran, in press).

2. Method

The Clinico Pathological Conference (CPC) of the Johns Hopkins University School of Medicine is held monthly. Every month, according to the topics covered in the second-year curriculum, an expert clinician is presented with the clinical course, radiological findings, and relevant laboratory results for a particular patient. The clinical expert puts a case together in a way that should be educational to all members of the Hopkins medical community, from medical student to senior clinician (Johns Hopkins University School of Medicine, 2004).

Each case study includes initial information with clinical course, radiological findings, and relevant laboratory results for a particular patient. The cases are posted on-line and the readers can suggest the next best step and their best guess for a diagnosis. The clinician's and the pathologist's presentations and the answer are posted on the Web after the conference.

The algorithm was implemented on the CPC case study initial information in order to obtain the context of the initial information. The topics selected by the algorithm were compared with the answers and the presentations to see whether the algorithm is of value in summarizing information for physicians or whether it achieves a competent level of diagnosis.

The method used for the medical analysis of the case studies is based on the context recognition model (Segev et al., in press). The model is based on the streaming in text format of information that represents input from different sources. The context recognition model output is a set of contexts that attempt to describe the current scenario most accurately. The set of contexts is a list of words or phrases, each describing an aspect of the scenario. The algorithm attempts to reach results similar to those achieved by the human process of determining the set of contexts that describe the current scenario.

The model consists of four major processes: collecting data, selecting contexts for each text (keywords), and ranking the contexts.

2.1. Collecting data

Each text is decomposed into single words, when words are letter strings separated by spaces. All punctuation is removed from the text. The words are then checked according to a set of dictionaries. The first dictionary is a "Stop List", consisting of words that do not add to the understanding of the context, such as I, me, in, are, the. All words that appear in this dictionary are ignored. The next step uses a set of dictionaries according to fields of knowledge to sieve the words that are not related to the specific field of knowledge. If the word appears in the field of knowledge dictionary, then it is added to the list of keywords that are searched in the context database, otherwise it is ignored. The dictionary used for the medical cases was the On-Line Medical Dictionary (2004). This process continues for each word in the text. After each text passes through this module, the algorithm sends a list of words to be checked for a possible set of contexts.

2.2. Selecting contexts for each text (keywords)

The selection of the current context is based on a search through the database for all relevant documents according to keywords and on the clustering of the results into possible contexts. Once a list of keywords exists, each keyword is searched in the context database – the Internet – and a set of contexts is extracted. This creates a list of preliminary contexts for each keyword. The contexts were represented by words or sets of words, which can be viewed as meta data created for each set of Internet web pages. The Internet can then be viewed as an immense set of words that represent different possible contexts, each associated with its respective web page. The full list of preliminary contexts for all the keywords includes all the possible contexts for this current text.

2.3. Ranking the contexts

New preliminary contexts can now be created according to textual sub-strings of existing preliminary contexts. They can also be united according to singular and plural form or similarities in meaning – synonyms. Similarly, this step sums the number of documents referring to the preliminary contexts. Multiple reference pages from similar web sources are counted as one instance. Each document usually refers to multiple contexts, consequently creating a long list of preliminary contexts.

The last step involves ranking the set of preliminary contexts according to both the number of references from the documents and the number of appearances in the text. This step maps all the preliminary contexts to a two dimensional graph, allowing the contexts that receive very high ranking in both characterizations to be located.

After each session of ranking, the list is used for two purposes – resetting the set of preliminary contexts and declaring the current context. The current list of contexts joins the new preliminary contexts arriving from the continuously streaming text. The lists are united and the ranking process is repeated. In parallel to the repetition of the ranking algorithm, the set of ranked preliminary contexts is forwarded to the next module to determine the current contexts.

2.4. Declaring the current contexts

The output of the ranking stage is the current context or a set of highest ranking contexts that differ essentially. The algorithm then returns to the first step to collect more texts and feed them again to the database. The set of preliminary contexts that has the top number of references, both in number of Internet pages and in number of appearances in all of the texts, is defined as the highest ranking and is declared to be the current contexts.

The process of determining the current context continues as long as information about the current scenario continues to stream in. The system constantly confirms and disconfirms the possible contexts as related to the different items of information received.

3. Results

The context recognition algorithm was used with twenty case studies. In 25% of the cases, the algorithm achieved the correct diagnosis. In 30% of the cases, the algorithm did not result in any valuable information. In the remaining 45% of the cases, the algorithm resulted either a possible diagnosis, which was mentioned in the case study but was not the correct diagnosis, or one of the symptoms. Table 1 presents the breakdown of the algorithm performance.

The performance of the algorithm is comparable to that of a physician. Research supports the assertion that errors are widely found in the practice of medicine. Diagnostic error rates of 40–60% have been documented, when this rate seems to be stable across hospitals, countries, and even eras of practice. The rate of errors with a significance influence on patient outcome is around 10%. The rate has hardly been influenced by improvements in diagnostic technology (Goldberg, Kuhn, Andrew, & Thomas, 2002).

Table 1
Algorithm performance

Category of algorithm performance	Cases	% with overlap[a]	% without overlap
Identifies the correct diagnosis	C1, C6, C7, C8, C14	25	25
Supplies information about a symptom	C1, C2, C7, C11, C15, C16, C17, C18, C19, C20	50	15
Supplies information about a possible diagnosis	C1, C2, C6, C7, C8, C11, C14, C15, C16, C19, C20	55	30
Does not supply any valuable information	C3, C4, C5, C9, C10, C12, C13, C15	30	30

[a] Overlap refers to the situation in which each result of the algorithm can belong to more than one category. Thus, when there is no overlap, the result belongs only to the particular category – the best classification.

The significance of the results was analyzed using the identical populations test. The test for homogeneity is designed to test the null hypothesis that two random samples are drawn from the same population or from different populations, according to some criterion of classification applied to the samples.

The Chi-square Pearson Test for Association is a test of statistical significance. Typically, the hypothesis is whether or not two populations are different in some characteristic or aspect of their behavior based on two random samples. The algorithm success rate (25%) is compared to the physician success rate (Table 2).

The results of the identical populations tests comparing the algorithm rate of the detection of the correct diagnosis with the physician rate of the detection of the correct diagnosis showed that the algorithm performance was very similar to the physician performance.

The next two examples present the findings of the algorithm for each case study and compare them to the actual diagnosis or the possible diagnoses of the case, as indicated in the answer of the case study.

The following results for case study 14 were:

Health, Test, History, Medical, Drug, Search, Children, Lab, Cushing, Book, Cancer, Cortisol, Treatment, Research, Other Topics, Journal.

Answer to Case 14 (Johns Hopkins, University School of Medicine, 2004):

> "Hence, the diagnosis was that of a pulmonary carcinoid tumor ectopically producing ACTH, leading to Cushing's syndrome."

Table 2
Identical populations test

H0: The populations are almost identical. There is no significant difference between the populations			
Physician diagnosis at 60%	$\chi^2 = 0.2506393$	P-value = 0.617	Little or no real evidence against the null hypothesis
Physician diagnosis at 40%	$\chi^2 = 0.051282$	P-value = 0.821	Little or no real evidence against the null hypothesis

The example shows that the algorithm reached the right diagnosis. The results include a set of words, some of which are irrelevant. However, there is a clear match between the words the algorithm detected and the actual diagnosis mentioned in the case study.

The following results for case 16 were:

Health, Medicine, History, Medical, Clinical, Disease, Care, Drug, Social, Infection, Asthma, FAQ, Child, Pulmo, Patient, Treatment.

Answer to Case 16 (Johns Hopkins University School of Medicine, 2004):

> "This patient's past medical history is remarkable for late onset but clinically severe asthma... Importantly, the lungs did not show significant hyperinflation, indicating that the patient did not die of an asthma exacerbation, and there was no evidence of pulmonary emboli... The actual cause of death in this patient is difficult to determine... One may hypothesize that immunosuppression from corticosteroids led her to multifocal underlying CMV pneumonitis, and that her superimposed acute bronchopneumonia and acid pneumonitis from aspiration (Mendleson's syndrome) produced hypoxia which pushed her "over the border" and led to her demise."

In this case it can been seen that the algorithm supplied a diagnosis – asthma – that was the wrong diagnosis. However, this diagnosis was considered as a possible diagnosis also by the physicians analyzing the case study. The diagnosis was based on previous symptoms that were mentioned in the case itself.

4. Discussion

Clinical decision making is viewed as a decision making process that integrates the physician's experience and knowledge with information from the medical literature. The paper presents a model that can serve as a decision support system for the analysis of case studies reviewed by a physician.

The results show that the algorithm provides the physician with a high percentage of relevant information that can assist in the retrieval of key information from the medical records or that can suggest a possible diagnosis. The research found that the algorithm performance is similar to the physician performance. The algorithm performs much better than other simple key phrase extraction models based on the number of words appearing in the text (Segev et al., in press). In addition, the analysis of the number of words in a text would fail to detect critical symptoms that appear only once in the text. Furthermore, a simple algorithm would not be able to suggest a possible diagnosis that is not mentioned in the text.

An interesting finding is the number of cases in which the algorithm was able to achieve the correct diagnosis in comparison to the case analysis expert opinion. These are good results, since a medical knowledge base was not created for the system.

Examination of the results of the algorithm shows that some of the correct diagnoses were found further down in the list. Thus, increasing the number of keywords presented to the physicians to approximately fifty would have multiplied the number of cases correctly diagnosed by the algorithm by two. However, it seems that it is not possible to ask a physician to review fifty words.

The implementation of the algorithm based on the use of the Internet as a knowledge base allows an existing pool of data to be exploited. Other decision support methods for medical systems require building a database for possible symptoms and possible diagnoses and connecting them. This type of database requires constant updating as new medical research is published. Thus, the algorithm has an advantage since it is based on preexisting data inferenced from the Internet. The data is being constantly updated as new information is published on the Internet.

The algorithm can provide a second opinion for the physician or it can direct to initial directions of inquiry. Moreover, the algorithm requires minimal resources for implementation and thus can be implemented widely at a low cost.

One limitation of the research is the number of case studies it analyzed. Future analysis of the model should focus on a greater number of case studies. Another limitation of the research is the inability of the algorithm to analyze any image data presented to the physician. There are existing image analysis technologies that can analyze an image and detect an object displayed with a very high accuracy. However, these systems require prior knowledge – what is the object of the search, or in other words, what part of the body is being analyzed.

Recommendations for further research include the integration of alternative methods for the identification of the correct medical diagnosis, to improve the results. One possible method might be the integration of an existing diagnosis system with the current model. This integration would allow the resultant system to work on free text describing a case study while using pre-defined medical diagnosis information. Other research can examine whether and how over time the results of the model change as the Internet information evolves.

References

Abu-Hanna, A., & De Keizer, N. (2003). Integrating classification trees with local logistic regression in Intensive Care prognosis. *Artificial Intelligence in Medicine, 29*, 5–23.

Baujard, O., Baujard, V., Aurel, S., Boyer, C., & Appel, R. D. (1998). MARVIN, a multi-agent softbot to retrieve multilingual medical information on the Web. *Medical Informatics, 23*(3), 87–91.

Bojarczuk, C. C., Lopes, H. S., Freitas, A. A., & Michalkiewicz, E. L. (2004). A constrained-syntax genetic programming system for discovering classification rules: application to medical data sets. *Artificial Intelligence in Medicine, 30*, 27–48.

Dilley, R. (1999). The Problem of Context. Berghahn Books.

Ein-Dor, P. (1999). Artificial intelligence: A short history and the next forty years Invited chapter. In Kenneth E. Kendall (Ed.), *Emerging Information Technologies*. Thousand Oaks, CA: Sage Publications.

Goldberg, R. M., Kuhn, G., Andrew, L. B., & Thomas, H. A. (2002). *Annals of Emergency Medicine, 39*(3), 287–292.

Johns Hopkins, University School of Medicine (2004). In: *Clinico-Pathological Conference*. Available from http://oac.med.jhmi.edu/cpc.

Johns Hopkins, University School of Medicine (2004). In: *Clinico-Pathological Conference*. Available from http://oac.med.jhmi.edu/cpc/cases/2002/cpc1_answer.html.

Johns Hopkins, University School of Medicine (2004). *Clinico-Pathological Conference*, Available from http://oac.med.jhmi.edu/cpc/caseFull.cfm?caseID=15&group=2003-2004&type=answer.

Kukar, M. (2003). Transductive reliability estimation for medical diagnosis. *Artificial Intelligence in Medicine, 29*, 81–106.

McCarthy, J. (1987). Generality in artificial intelligence. *Communication of ACM, 30*, 1030–1035.

McCarthy, J., & Buvac, S. (1997). *Formalizing Context*. *Computing Natural Language*. Stanford University, pp. 13–50.

Modica, G., Gal, A. & Jamil, H. M. (2001). The use of machine-generated ontologies in dynamic information seeking. In: *Proceedings of sixth international conference on cooperative information systems* (*CoopIS 2001*), Trento, Italy, September 2001.

On-Line Medical Dictionary (2004). Available from http://cancerweb.ncl.ac.uk/omd.

Segev, A., Leshno, M. & Zviran, M. (in press). Context recognition using Internet as a knowledge base. *Journal of Intelligent Information Systems*.

Tan, K. C., Yu, Q., Heng, C. M., & Lee, T. H. (2003). Evolutionary computing for knowledge discovery in medical diagnosis. *Artificial Intelligence in Medicine, 27*, 129–154.