www.wordlyknowledge.uz Volume: 1, Issue: 2, 2023

PROBLEMS OF MEDICAL DIAGNOSIS BASED ON INTELLECTUAL ANALYSIS OF DATA

Kalbaev Allambergen Markabaevich

Joldasbaev Dauran Rustem ugli

ANNOTATION: This article covers medical diagnostic problems based on intellectual data analysis.

KEY WORDS: Analysis methods,medical diagnosis, classification.

"Now that we have gathered so much data, what do we do with it?" This is the opening statement of the editorial by Usama Fayyad and Ramasamy Uthurusamy in the Communications of the ACM, Special issue on Data Mining [17]. Recently, many statements of this kind appeared in journals, conference proceedings, and other materials that deal with data analysis, knowledge discovery, and machine learning. They all express a concern about how to "make sense" from the large volumes of data being generated and stored in almost all fields of human activity.

Especially in the last few years, the digital revolution provided relatively inexpensive and available means to collect and store the data. For example in medicine, still in the mid-nineties one of the fathers of Artificial Intelligence in Medicine, Edward H. Shortliffe, partially blamed the underdeveloped hospital infrastructure for the failure to fulfill the initial promise of the field [81]. Recently, however, the situation is changing rapidly: modern hospitals are well equipped with monitoring and other data collection devices, and data is gathered and shared in inter- and intra-hospital information systems. In fact, medical informatics has become a must and an integral part of every successful medical institution [83].

The increase in data volume causes greater difficulties in extracting useful information for decision support. The traditional manual data analysis has become insufficient, and methods for efficient computer-based analysis indispensable. From this need, a new interdisciplinary field of Knowledge discovery in databases (KDD) was born [21]. KDD encompasses statistical, pattern recognition, machine learning, and visualization tools to support the analysis of data and discovery of principles that are encoded within the data. KDD is frequently defined as a process [18] consisting of the following steps: understanding the domain, forming the dataset and cleaning the data, extracting of regularities hidden in the data thus formulating knowledge in the form of patterns, rules, etc. (this step in the overall KDD process is usually referred to as data mining (DM)), postprocessing of discovered knowledge, and exploitation of the results. In this paper we use the term intelligent data analysis (IDA) rather than KDD, despite the fact that it is hard to make the distinction between the two. IDA and KDD have in common the topic of investigation, which is interactive and iterative process of data analysis, and they share many common methods. A possible distinguishing feature is that the methodologies and techniques used in IDA are mostly (but not exclusively) knowledge-based (and therefore "intelligent" in the sense used in Artificial Intelligence): they either use the knowledge about the problem domain, of the underlying principles or the knowledge about the data analysis process itself. Another aspect involves the size of data: KDD is typically concerned with the extraction of knowledge from very large datasets whereas in IDA this is not necessarily the case. This also affects the type of data mining tools used: in KDD the data mining tools are executed mostly in batch mode (despite the fact that the entire KDD process is interactive), whereas in IDA the tools can either be batch or applied as interactive assistants.

Specific goals of applying IDA in medicine are:

www.wordlyknowledge.uz Volume: 1, Issue: 2, 2023

– the extraction (discovery) of medical knowledge for diagnostic, screening, prognostic, monitoring, therapy support or overall patient management tasks (i.e., data mining), and

– the intelligent interpretation of patient data in a context-sensitive manner and the presentation of such interpretations in a visual or symbolic form (i.e., data abstraction); the temporal dimension in the representation and intelligent interpretation of patient data is of primary importance. As any other research in medicine is aimed at directly or indirectly enhancing the provision of health care, IDA research in medicine is no exception. As such, the benchmark tests for these methods and techniques can only be real world problems. Viable IDA proposals for medicine must be accompanied by detailed requirements that delineate the spectrum of real applications addressed by such proposals; in-depth evaluation of resulting systems thus constitutes a critical aspect. Another consideration is the role of IDA systems in a clinical setting. Their role is clearly that of an intelligent assistant that tries to bridge the gap between data gathering and data comprehension, in order to enable the physician to perform his task more efficiently and effectively. If the physician has at his disposal the right information at the right time, doubtless he will be in a better position to reach correct decisionsor instigate correct actions within the given time constraints. The information revolution made it possible to collect and store large volumes of data from diverse sources on electronic media. These data can be on a single case (e.g., one patient) or on multiple cases. Raw data as such are of little value since their sheer volume and/or the very specific level at which they are expressed make its utilization (operationalization) in the context of problem solving impossible. However such data can be converted to a mine of information wealth if the real gems of information are gleaned out by computationally intelligent means. The useful, operational information/knowledge, which is expressed at the right level of abstraction, is then readily available to support the decision making of the physician in imanaging a patient.

Important issues that arise from the rapidly emerging globality of data and information are:

– the provision of standards in terminology, vocabularies and formats to support mult linguality and sharing of data,

– standards for the abstraction and visualization of data,

– standards for interfaces between different sources of data,

– seamless integration of heterogeneous data; images and signals are important types data,

– standards for electronic patient records,

– reusability of data, knowledge, and tools.

Clinical data constitute an invaluable resource, the proper utilization of which impinges directly on the essential aim of health care which is “correct patient management”. Investing in the development of appropriate IDA methods, techniques and tools for the analysis of clinical data is thoroughly justified and this research ought to form a main thrust of activity by the relevant research communities.

Numerous intelligent data analysis methods have already been applied for supporting decision making in medicine (e.g., see [51]). These methods can be classified into two main categories: data mining and data abstraction. The majority of data mining IDA methods belong to machine learning and the majority of data abstraction methods perform temporal abstraction. This is the reason for machine learning and temporal abstraction to be the focus of investigation in this paper.

www.wordlyknowledge.uz Volume: 1, Issue: 2, 2023

The structure of this paper is as follows. In Section 2, the paper first sketches the history of research (Section 2.1) that led to the development of currentintelligent data analysis techniques. It then discusses the need for intelligent data analysis in medicine (Section 2.2), proposes a classification of intelligent data analysis methods (Section 2.3), and outlines some related characteristics of medical diagnosis problems (Section 2.4). The main scope of the paper are machine learning and temporal abstraction methods, and their application in medical diagnosis. A selection of methods and diagnostic domains is presented, and the performance or viability of such approaches is discussed in Sections 3 and 4. In Section 6 the paper concludes with the evaluation of selected intelligent data analysis methods and their applicability in medical diagnosis.

REFERENCES:

[1] Aikins, J. S. (1979). Prototypes and production rules: An approach to knowledge representation for hypothesis formation.

In Proc. Sixth International Joint Conference on Artificial Intelligence, pages 1–3.

[2] Barahona, P., and Christensen, J.P., editors (1994). Knowledge

and Decisions in Health Telematics, IOS Press.

[3] van Bemmel, J. H. (1996). Medical informatics, art or science? Meth. Inform. Med., 35:157–172.

[4] Bratko, I. and Kononenko, I. (1987). Learning diagnostic rules from incoplete and noisy data. In Phelps, B., editor, AI Methods in Statistics. Gower Technical Press, London.

22 Nada Lavrac˘∗ et al. / Intelligent data analysis for medical diagnosis: Using machine learning and temporal abstraction

[5] Bratko, I., Mozetic, I., and Lavra ˘ c, N. (1989). ˘ KARDIO: A Study in Deep and Qualitative Knowledge for Expert Systems. MIT Press, Cambridge, MA.

[6] Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification and Regression Trees. Wadsworth, Belmont.