Yuval Shahar

Section on Medical Informatics,
Stanford University,
Stanford, USA

Synopsis

Knowledge Processing

The late 1980s and the first half of the 1990s decade produced an increasing understanding and appreciation for the process of designing, developing, maintaining, and evaluating knowledge-based systems (KBSs). Projects such as the European KADS project [6] and the PROTEGE-II framework [8] promote the systematic development of knowledge-based decision-support systems, so as to facilitate the acquisition, representation, maintenance and, eventually, reuse and sharing of the knowledge these systems require and represent. These and similar approaches view knowledge acquisition as an active modeling task: An explicit modeling of the domain (e.g., infectious diseases), the task (e.g., diagnosis) and the problem-solving method or inference structure (e.g., heuristic classification). Furthermore, the increasing maturity of the area of clinical KBSs can be demonstrated by the growing number of studies evaluating the results of applying such systems or discussing methodologies for such evaluations.

The papers appearing in this section [1-5] demonstrate several features that can be related to the trend mentioned above. The first paper [1] discusses general development and evaluation methodologies of clinical KBSs; the second [2] analyzes the results of evaluating the same clinical KBS on epidemiologically different data sets and of applying different KBSs to the same data set; the third paper [3] presents a general task-modeling methodology and a particular logical reasoning method for diagnostic support; the fourth paper [4] discusses and evaluates an innovative technique for learning in competitive neural networks that enables diagnosis of multiple neurological disorders; and the last paper [5] demonstrates and evaluates the use of KBS techniques in a rather unusual setting for an expert system for diagnosis and treatment of chronic hepatitis based on Chinese-medicine diagnostic and therapeutic techniques.

Clarke et al. [1] present a comprehensive methodology for development and evaluation of knowledge-based systems within an application environment. The authors recommend an iterative development and evaluation cycle that includes four phases: (1) Preliminary exploration and/or development of an early prototype. The development style is expected to be somewhere between the fast, interactive, but potentially superficial rapid prototyping technique and a rigorous software-development methodology such as requirements specification. (2) Evaluation of the validity of the system, that is, the correctness of the output. Both the appropriateness of the inference mechanism chosen and the structure and contents of the knowledge base should be examined at that phase. (3) Evaluation of the functionality of the system, emphasizing the interaction with the user, the various responsibilities for maintenance of the system, and the degree to which the system retains its reliability when applied in another location and environment. (4) A field-trial evaluation of the system's impact on physicians' actions and on various measures of health care (such as length of stay), and a cost-benefit analysis (though such an analysis seems very difficult in most clinical systems). As other authors have suggested, the evaluation methodology is relevant from the initial development phase of a new KBS.

Clarke et al. emphasize the importance of conformity with six safety-first principles of data protection and integrity described by the C.E.C AIM Requirements board. Some of these principles are rather specific to the European community, but most are general to development of large software systems. The authors conclude that a formal evaluation methodology in which users, developers and experts collaborate, will contribute to the credibility of KBSs in clinical care and might alleviate some of the fears or apathy that is currently often felt by the medical profession towards such systems.

An interesting paper by Schioler et al. [2] examines in detail an issue, crucial for dissemination of knowledge-based technology and for enabling technologies such as telemedicine: the transferability of the clinical knowledge represented in clinical KBSs. The paper describes several experiments in knowledge transfer: testing different knowledge bases (KBs) created by different knowledge-acquisition techniques (implied by different reasoning methods) on the same database(DB), and testing the same KB on different DBs. The authors emphasize the importance of both knowledge-acquisition methodology and geographic differences on the transferability of expertise. The most robust system was not the most accurate one; a tradeoff exists.

The domain chosen by Schioler et al. was thyroid functional disorders. DBs included (1) a DB of patients and normal controls in Dublin, Ireland, documenting seven thyroid functional parameters; (2) a DB in Copenhagen, Denmark, of patients suspected of thyroid disorders, which was used as a training data set for some of the algorithms (and included five thyroid function parameters); and (3) another, similar, Copenhagen DB, which was used for evaluation of the trained KBSs.

The KBSs included (1) a rule-based system developed in Dublin for interpretation of all test-result combinations, (2) a manually constructed rule-based system from Copenhagen, based on the Copenhagen training DB, (3) a KBS using a KB derived from the Copenhagen KB using a recursive-partitioning induction algorithm for construction of binary classification trees, and (4) a probabilistic KBS developed from the Copenhagen training DB, whose KB was represented as a classification tree. Due to the multiple (three) diagnoses the performance measures included not sensitivity and specificity but correctness (ratio of correctly classified cases to all classified cases), coverage (fraction of cases classified), and test consumption (average number of test results used).

The two rule-based systems were tested on the Dublin DB. The correctness was similar, but the manual KB could not classify 35% (49/140) of the cases, due to missing required data, reflecting differences in laboratory practice among the two sites. (Unfortunately, a test of both KBSs on the Copenhagen training set was not described; similar results, reflecting inflexibility in data requirements, might be expected.) In a second experiment, the manually constructed rule-based system, the KBS derived by the induction algorithm, and the probabilistically derived KBS were trained on the Copenhagen training DB and tested on the Copenhagen evaluation DB. The main outlier amongst the KBSs was the probabilistic KB, due to a poor correctness measure in cases involving thyrotoxicosis. The poor performance might be due, as the authors point out, to the greater sensitivity of this algorithm to different thyrotoxic features in the training data set and in the testing data set. (Whether a probabilistic network can be or should be used as a static classification tree is another issue.) In the last experiment reported, the number of myxedema (hypothyroidism) cases in the Copenhagen training DB was doubled (increasing their proportion from 10% to 18%) and the two derived KBs were created from this new DB. This "epidemiological" transformation resulted in a dramatic improvement in the correctness of the probabilistically derived KBS, accompanied, however, by a significant reduction in the coverage. The induction algorithm seemed much more robust to population-distribution changes. In both KBS, the new DB caused an increase in the test consumption.

Huang et al. [3] consider both general design and specific reasoning considerations for clinical KBSs. The authors present a specific reasoning method for clinical diagnosis: a qualitative logic of argumentation, in which arguments for or against a proposition can either confirm, increase or diminish a belief in, or exclude a concluded proposition (e.g., a diagnosis). The framework, implemented in PROLOG within the European DILEMMA project, is somewhat similar to the KADS methodology [6] in distinguishing among different layers of knowledge. In particular, the knowledge base contains knowledge about domains (e.g., oncology), about tasks (e.g., diagnosis, therapy), and about patients (patient records). A task model defines which relation (e.g., causality) can play a specific role (e.g., refinement) in a particular task (e.g., diagnosis). This task modeling is similar to the mapping defined by Gennari et al. [7] between entities and relations in particular domains and the internal terminology of specific problem-solving methods, albeit in a much simpler format. (For instance, mapping might in general involve more complex functions, such as when a whole class is mapped into a single instance, or when procedural knowledge is mapped into a certain inference role.) The authors do not discuss automatic generation, from a task model, of a knowledge acquisition tool, as is the case in the PROTEGE-II project [8]. The underlying argumentation logic of Huang et al.'s system is described in other papers. Thus, the authors do not discuss in detail how a meaningful score is computed from different conflicting arguments (with the same equivalent absolute strength of support), which might reflect probabilistic dependencies of very different strengths, and which pertain to conclusions with very different prior probabilities. The authors acknowledge that the omission of facilities for probabilistic and other kinds of quantitative reasoning about uncertainty is controversial, but assume that it should be possible to introduce such methods into their system when necessary. However, as has been shown in the case of temporal and spatial reasoning, the addition of a new dimension (literally and conceptually) often forces the developers to change the whole framework.

The argumentation system described is at the prototype stage and remains to be generalized, implemented as a set of decision-support tools, and evaluated on clinical data and in a clinical setting. However, the intent to clarify the underlying roles of different types of knowledge is commendable and should facilitate the acquisition, maintenance, reuse, and sharing of clinical knowledge.

In a paper discussing and evaluating a very specific reasoning method, Cho and Reggia [4] present encouraging results for a new back-propagation rule for training competitive neural networks in the task of diagnosis of multiple neurological disorders. In standard neural networks, the activation mechanism utilizes only the static weight of the link among neighbor nodes; in contrast, nodes in competitive networks compete for the output of neighboring units, the ability to compete increasingly with the node's own activation level. The effect is to encourage several eventual winners, as opposed to a "winner take all phenomenon" that accompanies standard inhibition techniques (in which the lone winner inhibits other potential winners). The competitive property suggests better suitability for multiple-disorder diagnosis.

Cho and Reggia present a new learning rule that can be used in the context of competitive neural networks, and present persuasive, although very preliminary evidence that competitive networks trained on single-diagnosis cases do significantly better than standard networks when presented with completely new cases involving combinations of two disorders (in this case, the disorders involved different brain-damage localizations).

Given the typical complexity of real clinical cases, these results are encouraging with respect to the potential of diagnosis of multiple-disorder cases. The work also has knowledge-acquisition ramifications: in theory, competitive networks might need to be trained only on single-diagnosis cases in order to produce satisfactory behavior on multiple-diagnoses cases. However, the results are currently highly specific to the particular domain and training set described. Furthermore, due to well-known problems such as the fact that the interaction between different underlying disorders might emphasize unduly certain quantitative symptoms while "canceling out" others, caution is necessary regarding the clinical value of competitive networks. The work, however, is a significant step forward in the difficult diagnostic problem of multiple disorders.

In the last paper of this section, Zhao et al. [5] describe a rather different type of a clinical KBS: an expert system to diagnose and treat chronic hepatitis, based on Chinese medicine. The system is intended for use by physicians who are not necessarily knowledgeable in Chinese medicine. The need for the system occurs when the physician or the patient prefer this style of therapy, when conventional therapy is considered ineffective, or when a combination of Chinese and modern medicine is deemed desirable. The inference strategy is a hypothetico-deductive approach-findings suggest a set of hypotheses, that are then refined by referring additional questions to the user. In that respect the system is reminiscent of early systems such as INTERNIST-1 [9]. The authors do not specify whether there is any weight attached to the semantic links among findings and diagnoses, and how the system determines what is the most appropriate additional required piece of information. After deciding on a diagnosis, the system suggests one or more of a set of 42 herbal recipes, based on 96 herbs. The authors do not elaborate how that task is being carried out, and in particular, whether the system selects therapies indexed by the (single) final diagnosis, or attempts to "cover" the whole set of most likely disorders.

The Chinese-medicine chronic-hepatitis KBS has been evaluated in a preliminary fashion on 40 clinical cases, by comparison of the system's conclusions with the actual diagnosis and advice given. The results had been analyzed by a Chinese-medicine expert with respect to appropriateness of the diagnosis offered and the herbal recipes suggested; 87% (35/40) of the diagnoses and 82% (33/40) of the recipes recommended were judged as appropriate.

In summary, as several of the papers in the section emphasize, both a systematic design and a methodologically sound evaluation of clinical KBSs are crucial for successful dissemination and use of such systems in real clinical environments.

References

1. Clarke K, O'Moore R, Smeets R, Talmon J, Brender J, McNair P, Nykanen P, Grimson J, and Barber B. A methodology for evaluation of knowledge-based systems in medicine. Artificial Intelligence in Med 1994; 6:107-121.
2. Schioler T, Talmon J, Nolan J, and McNair P. Information technology factors in transferability of knowledge based systems in medicine. Artificial Intelligence in Med 1994; 6:189-201.
3. Huang J, Fox J, Gordon C, and Jackson-Smale A. Symbolic decision support in medical care. Artificial Intelligence in Med 1993; 5:415-430.
4. Cho S and Reggia JA. Multiple disorder diagnosis with adaptive competitive neural networks. Artificial Intelligence in Med 1993; 5:469-487.
5. Zhao YK, Tsutsui T, Endo A, Minato K, and Takahashi T. Design and development of an expert system to assist diagnosis and

treatment of chronic hepatitis using traditional Chinese medicine. Med Inform 1995; 19:37-45.
6. Special issue: The KADS approach to knowledge engineering. Knowledge Acquisition 1992; 4:1-161.
7. Gennari JH, Tu SW, Rothenfluh TE, and Musen MA. Mapping domains to methods in support of reuse. Int J Human-Computer Stud 1994; 41:399-424.
8. Tu SW, Eriksson H, Gennari JH, Shahar Y, and Musen MA. Ontology-based configuration of problem-solving methods and generation of knowledge-acquisition tools: Application of PROTEGE-II to protocol-based decision support. Artificial Intelligence in Med (in press).
9. Miller RA, Pople HE, and Myers JD. INTERNIST-1: An experimental computer-based diagnostic consultant for general internal medicine. New Eng J Med 1982; 307:468-476.

Address of the author:
Y. Shahar,
Section on Medical Informatics, Knowledge Systems Laboratory,
Medical School Office Building X215,
Stanford University, Stanford, CA 94305,
USA