

A. Jackson-Smale

Advanced Computation Laboratory
Imperial Cancer Research Fund
London, UK

Synopsis

Knowledge Processing

Introduction

This synopsis covers papers under the heading of "knowledge processing"; a term that covers the acquisition of knowledge and its validation, through representation of such difficult aspects as time, to the use of the knowledge in a reasoning system. The five papers in this section deal mainly with technical aspects of processing the particular type of data that we call "knowledge" and this is reflected by their original appearance in medical informatics (i.e., technical) journals rather than general medical journals. This is not to say that the medical content is not very high. On the contrary, the quality of the explanation in each of the papers of the medical domains used shows the degree of medical understanding required by knowledge engineers. Each paper uses very specific clinical examples with persuasive arguments that the methods used are general and re-usable. These knowledge-processing activities need to be shown to be re-usable not only by the same teams but by others. Therefore, knowledge engineers are increasingly required to formalise and explicitly document the methods used to process data and information into knowledge.

Knowledge Acquisition

Knowledge acquisition has long been acknowledged as a bottleneck in

building knowledge-based systems, and approaches to automate the process are always interesting. Automatic acquisition of knowledge from published literature is dogged by the problems of natural-language parsing, but given that libraries now classify scientific papers under subject headings, might not these classifications produce processable information? This is the pragmatic approach chosen by Cimino and Barnett [1]. They employed pattern matching to acquire "knowledge" from the MeSH (Medical Subject Headings) of the National Library of Medicine's MEDLINE database.

The paper describes the whole process of identifying key terms and then producing plausible relationships that are later validated by checking against the title and abstract of the spawning citation before final verification by "expert" clinicians. It is not clear how useful the knowledge thus gained will be, but several examples taken from the domain of cardiology demonstrate the potential for building a sizeable semantic network very quickly.

The authors state four types of general "rules" that can be defined between subject headings in the MeSH terms. The simple relationship "*X causes Y*" is the most common and useful, whilst more tenuous links such as "*C is related to D*" might also be useful if a reasoning system could deal with vague relationships. Indeed, the paper cries out for a theoretical reasoning system that can utilise the

knowledge to produce a useful end-product, or else the user is likely to be drowned beneath an avalanche of "links" with no selectivity.

One important advantage of this kind of knowledge acquisition is that all the "facts" have an identified published source. The authors deal nicely with the problems of later discrediting of information, and imply that it might be prudent to allow a reasonable "cooling off" period before including facts from MEDLINE in a knowledge base for real clinical use. There is large support for building accredited knowledge bases. The information may then possess higher or lower credence depending on the rigor of its methods and the reproduction of the results. So far, there is no automatic method of checking these criteria, although Cimino and Barnett suggest giving weights to facts based on how often the association occurs in the literature.

The research deserves to be extended to wider domains and perhaps other sources of information from individual institutional databases. I would also like to see some research into reasoning methods for utilising the knowledge effectively.

Knowledge Validation

How accurate should a diagnostic decision-support system be before it can become acceptable in everyday clinical practice? This interesting question is answered in the paper by Todd

and Stamper [2] who show that there are natural limits to diagnostic accuracy (termed "*diagnosability*") using any diagnostic methods. The limit appears to be around 75-80% for the domain of their study (abdominal pain of gynaecological origin). Abdominal pain has a central place in the history of computer-aided medical diagnosis thanks to the work of De Dombal [3] and Adams et al. [4]. Todd and Stamper concentrated on abdominal pain of possible gynaecological origin due to the special clinical challenge this provides. Results in this highly specific application will require confirmation in other application areas and it is gratifying to note that the experimental methods were designed to enhance reusability.

Training a knowledge-based system with Bayesian analysis as the underlying reasoning paradigm provides a problem for all but the commonest of conditions. A training database of 10,000 cases is necessary to approach the maximum diagnosability. The authors had only 500 cases in their database and describe various interesting methods for maximising their system's accuracy. These methods attempt to overcome the lack of validity in using the same cases for testing as for training. They include various probability-correction and parameter-smoothing factors. Simulation methods provided "*clinically reasonable*" sets of test cases and led to high levels of diagnostic accuracy (>80%) in other simulated cases. When the system was tested on real cases, the diagnostic accuracy decreases again to about 60%. The authors suggest that this could be due to "missing" data in the real cases. The implication of this is that thorough examination and full clinical investigation ought to achieve accurate diagnosis almost every time. They then go on to look at interactions between signs and symptoms and show that this limits the theoretically achievable diagnosability by about 5%.

The final conclusion gleaned from this work is that accurate (order of 80% accuracy) computer-assisted diagnosis using independence Bayesian analysis requires a large number of validated and complete training cases. As the authors themselves note, abdominal pain studies now have well over the number of cases needed for maximising diagnosability; how many other domains could benefit from similar studies? The simulation model used in the research is described in detail in a more recent publication [5].

Knowledge Authoring

Knowledge bases have traditionally been written by one (clinical) author. Several methods have been suggested for achieving consensus among multiple authors but existing studies often show a high degree of variation of opinion between experts. However, Giuse et al. [6] show that in the domain of acute perinephric abscess there was considerable agreement among a group of internists on the pertinent findings and the relevance of the clinical literature given them to review.

The QMR knowledge base, developed from INTERNIST, is a model of sound, controlled knowledge base authoring. The amount of effort that goes into producing the QMR disease profile brought about by this study is mind-boggling - over 100 published papers selected by a rigorous method were intensively reviewed by seven experts who produced their own individual QMR disease reviews. These were later analysed to produce a final profile containing over 200 pertinent findings related (by value) to the papers.

The amount of labour that went into the production of the final knowledge base for such a specific domain (by my reckoning about 4-6 man-months) points to the need for sharing of knowledge and reuse of knowledge bases by

different systems in order to save time and costs. Another paper in this section [7] is about the conversion of knowledge between representation schemes, but we also need to consider other fundamental problems about knowledge re-use. Questions not addressed in this paper include: 'How does knowledge cross national boundaries?', 'How does "consensus" knowledge change over time?', 'How do we deal with areas of real disagreement?' and, lastly, 'Where do we get the time to build knowledge bases in this way?'.

Knowledge Representation and Reasoning

Conversion of knowledge contained in one representation into another representation is often more difficult than it should be. Korver and Lucas discuss the conversion of the HEPAR rule base into a belief network [7]. They found that much of the knowledge they required to build the belief network could not be extracted from the rule base of HEPAR and they had to do significant amounts of extra knowledge acquisition.

It was interesting to note that the heuristically derived production rules of HEPAR could not be easily converted into causal belief networks (even given that "causes" can mean "influences" in this sense). One would think that causality underlies many diagnostic rules but apparently not in the right sense. Korver and Lucas showed that a number of extra vertices were needed in the belief network to fully represent all the HEPAR knowledge as well as some extensions to belief networks to represent cyclic influences.

The restricted expressive power of (acyclic) belief networks in representing cyclic influences may hinder natural representation of medical knowledge. Korver and Lucas describe how to extend belief networks to cope with

both positive and negative feedback loops and give an example of how to represent one negative feedback process in the renal regulatory mechanism by "cutting out" the loops. In order to reduce the redundancy of information and the inference time, it is desirable to minimise the number of vertices in the final network. Although I am in favour of applying "Ocham's razor" to any unnecessarily complex system, it does appear dangerous to chop out intermediate steps in a causal process. As the authors themselves point out, this is done with reluctance.

In the final section, the authors acknowledge the major long-standing problem with probabilistic systems - that probabilities are not available or are not valid for the target population. For such systems to achieve acceptability, probabilities computed from large numbers of real cases are inevitably needed. This paper shows a high quality depth of understanding of the medical field and also provides useful information for anyone intending to take rule-based systems and convert them into belief networks.

Temporal Reasoning

Practical, usable temporal reasoning systems are important in the growing use of protocols for treatment of cancer, AIDS and many other health problems. Planning and abstraction are dependent on a robust and theoretically sound temporal reasoning strategy. Clinicians naturally use temporal reasoning when they treat patients.

Computers are good at storing factual information and time-stamped

events. The AI community has come up with many methods for converting time-stamped "raw" data into natural concepts and intervals and contexts. Shahar and Musen [8] describe a temporal-abstraction system for patient monitoring called *RÉSUMÉ*. The general architecture of the *RÉSUMÉ* system exhibits a loosely coupled integration to an external patient database and the authors acknowledge that the system relies on a rich and comprehensive database of patient facts in order to utilise the system effectively.

Problems with truth maintenance of temporal data (when data that were previously unavailable become available for example) are covered within the CLIPS language used to build the system via the internal justification TMS system that creates temporal dependencies between data items and conclusions.

I think we can safely say that we have solved many of the problems of dealing with "clock" time (the *RÉSUMÉ* system is evidence of that) although problems of computational complexity and data capture remain difficult to crack. The next problem for AI could be the problems of dealing with what we might call "social" time. The urgency of a problem depends on the particular person with the problem and environmental factors. The differences in people of different ages and with different social backgrounds can be quantified only with sensitivity and considerable qualification. Getting computers to deal with the problems of social time is a possible new avenue for AI research, especially in medical informatics where time has many different meanings.

References

1. Cimino JJ, and Barnett GO. Automatic Knowledge Acquisition from MEDLINE. *Methods Inf Med* 1993;32:120-30.
2. Todd BS, and Stamper R. Limits to diagnostic accuracy. *Med Inf (London)* 1993;18:255-70.
3. De Dombal FT. Computer-aided decision support in acute abdominal pain, with special reference to the EC concerted action. *Int J Biomed Comput* 1990;26:183-8.
4. Adams ID, Chan M, Clifford PC, Cooke WM, Dallos V, De Dombal FT, Edwards MH, Hancock DM, Hewett DJ, McIntyre N, Sommerville PG, Spiegelhalter DJ, Wellwood J, and Wilson DH. Computer-aided diagnosis of acute abdominal pain: a multicentre study. *Br Med J* 1986;293:800-4.
5. Todd BS, Stamper R, and Macpherson P. The design and construction of a medical simulation model. *Comput Methods Programs Biomed* 1994;42:77-91.
6. Giuse NB, Giuse DA, Miller RA, Bankowitz RA, Janosky JE, Davidoff F, Hillner BE, Hripcsak G, Lincoln MJ, Middleton B, and Peden JG Jr. Evaluating consensus among physicians in medical knowledge base construction. *Methods Inf Med* 1993;32:137-45.
7. Korver M, and Lucas PJF. Converting a rule-based expert system into a belief network. *Med Inf (London)* 1993;18:219-41.
8. Shahar Y, and Musen M. *RÉSUMÉ*: A temporal-abstraction system for patient monitoring. *Comput Biomed Res* 1993;26:255-73.

Address of the author:
A. Jackson-Smale, Ph.D.,
Advanced Computation Laboratory,
Imperial Cancer Research Fund,
61 Lincolns Inn Fields,
London, WC2A 3PX,
United Kingdom