Proceedings of the MedInfo ´95 (North-Holland, Canada 1995) 989-993

A Case against Logic

Friedrich Steimann
Department of Medical Computer Science, University of Vienna
Währinger Gürtel 18–20, A-1090 Wien, Austria
e-mail: steimann@vm.akh-wien.ac.at

Abstract: Based on an example taken from Toxoplasma serology it is shown that employment of formal logic and its symbolic descendants are not always the best choice for supporting medical diagnosis.

1. Introduction

Despite a recent tendency towards multidisciplinary education, the vast majority of researchers working in the field of medical informatics is still more comfortable with one of the involved sciences than with the other. Even those who are equally skilled in medical and computer science may perceive their liaison as one of antagonists: medicine, surrendering to complexity and apparent randomness of human physiology, relies on elaborate models and abstractions acquired over centuries of observation and scrutiny, while computer science is founded on notions of computability, determinism, and provability. If a particular medical problem is to be solved using methods of informatics, a common medium must be found suitable to express the original problem in a form solvable by a computer and to translate the solution back to the medical domain.

Diagnosis has been the prevailing target of medical expert systems [1]. Evolution of such a diagnostic system commonly begins with a medical expert striving to verbalize a comprehensive specification of how (s)he is reasoning when diagnosing. This specification tends to be a generalization of typical cases (s)he is confronted with in everyday routine, supplemented by a rather loose collection of exceptions, anecdotal knowledge either from self-experience or learned from literature. The computer expert, on the other hand, is eager to find regular patterns in the diagnostic process that match one of the formalisms (s)he has at hand, which are traditionally logic-related.

Popularity and frequency of employment of logic-related formalisms such as rules are no coincidence: those who find the relation of natural language and formal logic far-fetched may consider that logic is an offspring of ancient philosophy dealing with natural language and grammar. As natural language is still the prime medium of human (including scientific) communication, employment of logic for formalization appears logical. It is hence no surprise that the first medical expert system MYCIN [2] was indeed rule-based.

Soon after MYCIN's introduction, the crucial role of time in literally all medical reasoning processes was recognized [3–8]. Interestingly, until then this matter had largely been ignored by the theoretic community — only the practical employment of formal logic made it very clear that adequate treatment of the temporal dimension was inevitable if realistic problems were to be solved. However, as it turned out, inclusion of time could not be accomplished by just adding another variable to the problem space: the semantics of time, which is inherently linked to notions of causation and change, requires its own special treatment. Many solutions, theoretical and practical, have been suggested since [3, 6–14]; one that has proven generally applicable has not yet been identified.

2. A Practical Problem

When acquired during pregnancy, primary infection with *Toxoplasma gondii*, a widespread parasite, presents a serious threat to the unborn child: although the immune system of a healthy mother can control the infection, pathogens may cross the placental barrier and afflict the unborn [15]. Depending on the gestational age, clinical symptoms range from fetal death and stillbirth to clinically healthy newborns with an 80% chance of developing blindness in adulthood. Because early treatment can prevent fetal damage, safe and reliable detection of postconceptional onsets of infection is a desirable goal.

General screening for toxoplasma infection requires a series of blood samples to be drawn from every mother and to be tested for the presence of specific IgG and IgM antibodies. Depending on the employed tests, the height and course of antibody concentrations give clues about the stage of infection: rising or extremely high titers hint at acute infection, while low or falling titers suggest that infection is chronic. Figure 1 depicts typical courses of infection as observable through the dye test (DT) measuring mostly IgG antibodies and through the IgM immunosorbent agglutination assay (ISAGA).

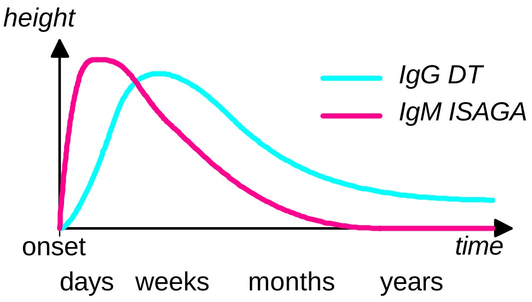

Figure 1: typical course of infection as observable through IgG DT and IgM ISAGA

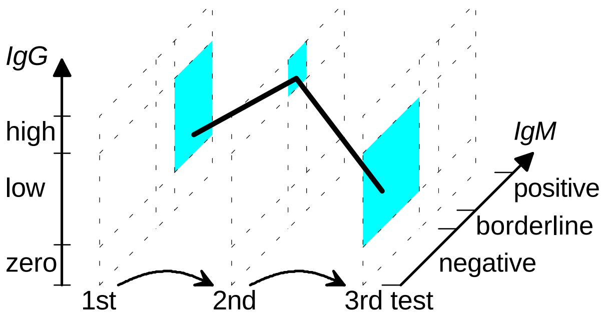

Figure 2: reduced symbolic problem space of Toxoplasma serology including a typical course

Diagnosis of acute infection is handicapped by several factors. Firstly, for practical reasons, patients cannot be monitored continuously; in fact, sampling is usually so sparse that characteristic properties of the curve such as the peak titer (cf. Fig. 1) are frequently missed. Secondly, courses of immune response vary widely: some individual's IgG reaction to acute infection is so moderate that it may be mistaken as chronic, while other's high readings of IgM may persist for several years [15]. It therefore comes as no surprise that only experienced clinicians feel confident in associating one out of the following five diagnostic classes with any given sequence of findings:

- *acute*, here meaning postconceptional infection,
- *chronic*, here meaning preconceptional infection,
- *suspected acute*, i.e., either chronic or acute infection in the above sense,
- *seronegative*, i.e., no infection at all, or
- *inconsistent data*, i.e., constellation does not occur naturally.

Results of testing are inherently discrete: the IgG DT result is obtained in steps of fourfold dilution ranging from 1:4 to 1:65536 including 0, IgM ISAGA is obtained as an index ranging from 0 to 12, and the time of testing is stated as the gestational week ranging from 1 to 40.

A few simple calculations give an idea of the complexity of the problem. If samples were taken at regular intervals, say, once per trimester (one third of a pregnancy), then more than 1.6 million ($9^3 \times 13^3$) different combinations of test results would theoretically be possible. In clinical practice, however, sampling times are determined by uncontrollable factors, which basically means that samples arrive at almost arbitrary intervals. Because adequate diagnosis cannot neglect the time interval between samples, the total number of possibilities for three consecutive samples during 40 weeks of pregnancy is boosted to approximately 16 billion, given that samples are at least one week apart. Even if natural constraints reduce the number of practical cases, the problem remains orders of magnitude too high for human comprehension. Despite this combinatorial complexity, an experienced clinician is sufficiently accurate when discriminating acute from chronic infections exclusively based on objective data. This hints on the availability of extremely effective abstractions of the problem.

It should be noted that the problem of determining the onset of infection is resistant to most analytical methods: sampling is neither regular nor dense enough to allow for signal reconstruction techniques or effective time series analysis. Methods of artificial intelligence seem therefore indicated.

3. Logical Solutions

Conversion of quantitative findings to qualitative terms (*number-to-symbol conversion*) is common practice in expert systems, as it reduces the number of cases to be considered and renders subsequent reasoning independent of numeric thresholds [3]. In the given example, practical ranges for IgG DT are *zero* (0–1:4), *low* (1:16–1:4096), and *high* (1:16384–1:65536), as well as *negative* (0–5), *borderline* (6–8), and *positive* (9–12) for IgM ISAGA. If, at the same time, sampling intervals can be quantified to whole months, the number of cases to be considered is reduced to 61236 (see Fig. 2).

The price being paid for this kind of abstraction is discontinuity and loss of information. For example, two subsequent DT samples (1:16, 1:4096) taken from a patient are converted to (*low*, *low*) and hence regarded as constant, while the same patient's samples (1:256, 1:16384), taken a little later, would be interpreted as (*low*, *high*), which is considered indicative of acute infection and hence leads to a different diagnosis. In addition, the former sample is indistinguishable from, say, (1:1024, 1:256), which is an entirely different case. An established remedy to preserve information lost in number-to-symbol conversion is feature extraction, in this case the derivation of the direction and speed of change from consecutive samples and their inclusion in the decision process. However, doing so results in more variables to be considered and thus adds to complexity.

Diagnostic knowledge can now be formalized by associating each possible constellation of findings with one of the five diagnostic classes. Quite clearly, this assignment would hardly be done by hand, i.e., by enumerating all constellations explainable with one diagnosis; instead, some kind of characteristic function is sought that decides for each constellation to which class it belongs. This is where logic comes into play: a rule is a partial implementation of such a characteristic function mapping a set of findings specified in its conditional part to an (intermediate) diagnosis provided in the conclusion. However, with a problem space of the above size, identifying appropriate rules and specifying a ruleset that is both consistent and complete is intrinsically difficult.

Alternatively, a decision tree can be employed to formalize the diagnostic process. Each path in the decision tree represents a nested rule of the form *if-then if-then* ... Complexity is reduced by different paths sharing the same nodes, i.e., by factoring out common subseries of findings that lead to the same (intermediate) result. In this approach, each consecutive finding adds to differentiating possible from impossible diagnoses; the diagnosis is derived incrementally. It exploits the fact that sera are drawn strictly sequentially; it also has the advantage of leaving the number of sera required for diagnosis variable, which meets clinical practice.

TOXOPERT [16] is a decision-tree-based expert system for Toxoplasma serodiagnosis equipped with an initial knowledge base of 45 nodes and 71 vertices specified by a group of experts. Although TOXOPERT works well for all standard situations, current testing shows that practical cases occur for which no decision path can be found although a trained clinician has no problems with diagnosing these cases.

4. A Picture is Worth a Thousand Words

In his pioneering work of Toxoplasma gondii serodiagnosis, Thalhammer [17] observed courses of acute infections in Austria for more than 30 years. Unaware of any of the formal methods mentioned above, he based his diagnosis on graphical considerations: "*... The titer gradient of all acute infections (...) proves to be nearly identical. This confirms that maximum serum antibody concentration is reached approximately two months after primary infection. In combination with the gestational age one can thereby differentiate pre- from post-conceptional infection with high certainty ...*" [17]. Ironically, he used words not graphs to communicate the picture he had in mind.

Rather than yielding to the temptation of extracting the regularities implicit in his description and putting them into rules, one should realize that there is a graphical solution to the problem: if an infection takes typical courses and these courses are known in sufficient detail, then a diagnosis can be derived by directly matching them with individual findings. The obvious gain is that the graphical representation can provide a complete specification of diagnostic expertise suffering from none of the aforementioned deficiencies.

Translating Thalhammer's observations — supplemented by data of evident acute infections extracted from a toxoplasmosis database — to fuzzy relations and their implementation in a computer program called ONSET

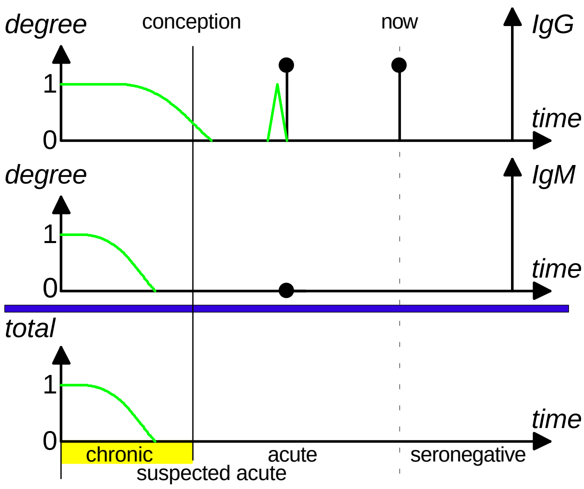

Figure 3: ONSET's method

ONSET derives the serodiagnosis graphically. For each performed test, the findings are plotted in a respective diagram (dots and scales on the right). For each possible time of onset, its degree of compatibility with the findings (obtained by matching the findings with the fuzzy courses starting at that time) is recorded (grey curves and scales on the left). The combined distribution (bottom), relative to the date of conception and current time, yields the diagnosis, which is *chronic* in the given example.

turned out to be straightforward [18]. To assess possible times of onset of infection for a given case, ONSET matches findings repeatedly with the fuzzy courses for varying temporal offsets and records the degree of match along the time line (Fig. 3). By relating the so derived possible times of onset to the date of conception, one of the five above diagnoses can automatically be derived. In a first retrospective study involving 1000 cases, ONSET achieved an accuracy of 91.4% [18].

5. Discussion

Although the initial problem statement seemed manageable, the logic-based solution of TOXOPERT is not only considerably complex, but also flawed by the undesirable properties of incompleteness and discontinuity. Strictly speaking, these inadequacies are no artefact of formal logic — rather, they can be traced back to the initial verbal specification of the diagnostic process and reflect typical shortcomings of natural language.

Taking a graphical approach, the ONSET program suffers from none of the addressed deficiencies. Nevertheless, rule-based systems have been employed for interpretation of serological test results (e.g., [19–21]), but their accounting for time is only poorly developed.

Despite considerable effort in the field of temporal logic, only few ideas contribute directly to the practical problem of interpreting clinical time series, sequences of explicitly time-stamped data such as consecutive serological findings. (Among those that do is TrenDx [22] utilizing the temporal utility package TUP [6].) Instead, authors prefer to focus on deriving temporal abstractions holding over intervals [8, 12, 23] and applying qualitative time measures and relations [7, 9, 24]. However, this approach is not only unsuited for the given problem (extremely few samples leave nothing to abstract), it also suffers from the mismatch of bottom-up temporal abstraction and requirements of subsequent symbolic reasoning[1].

Solutions regarding the quantitative nature of the problem can be found in the field of time series analysis and trend detection (cf., e.g., [25–28]). However, these approaches usually make strong assumptions on data and problem such as periodic sampling and availability of stochastic models, thus impeding their realistic employment in the medical field.

6. Conclusion

It is the author's personal experience that co-operative clinicians, well aware of the ambiguity in their discourse, tend to believe they only have to adopt what is commonly regarded as "computer language" in order to turn their soft constraints into hard definitions.[2] It is our responsibility not to make them believe that softness is due to their lack of logical training, neglecting that medicine, compared to the standards of formal logic, is inherently "soft". In fact, precise definitions are not only hard to obtain, they are also frequently medically inadequate.

Many past contributions of informatics to soft sciences such as medicine have suffered from overformalized notations and concepts resulting in inadequate problem statements and restrictedly — if at all — applicable solutions. Particularly in the area of medical expert systems where joint effort of medical and informational expertise is a prerequisite for success, the intricacy of "computer language" (which is not even always tamed by practising logicians[3]) has turned out to be a considerable obstacle and led to much disappointment and frustration.

This paper wants to encourage computer scientists to leave — the next time they encounter a new problem — their familiar tools aside and spend some time thinking about the underlying nature of the problem. If the apparent discrepancy of medical and computer science is understood as a chance rather than a nuisance, medical informatics may become the leading science of decision making and problem solving.

7. Acknowledgements

I would like to thank Kazem Sadegh-Zadeh and Klaus-Peter Adlassnig for their encouragement and the clinicians of the University of Vienna Medical School for their interest and co-operation.

[1] Buttom-up temporal abstraction is very much like feature extraction — it basically derives a new variable that needs to be interpreted.

[2] An anecdotal example: In defining the development of the adult respiratory distress syndrome (ARDS), two clinicians struggled whether to join a list of preconditions and risk factors with logical *and* or *or*. As it turned out, medical literature had already developed a scoring system that implements a compensating operator equally easy to understand and to implement in a computer [29].

[3] The readers confident that logic presents no problem to them may reflect how much time they have spent on debugging nested *if-then-else* clauses involving multiple variables and logical connectives.

8. References

[1] Heathfield HA, Wyatt J: Philosophies for the design and development of clinical decision-support systems. *Methods of Information in Medicine* 32, 1–8, 1993

[2] Shortliffe EH: *Computer-Based Medical Consultations: MYCIN*. Elsevier, New York 1976

[3] Fagan LM, Kunz JC, Feigenbaum EA, Osborn JJ: Extensions to the rule-based formalism for a monitoring task. In: Buchanan, Shortliffe (eds.): *Rule-Based Expert Systems*, 397–423, 1984

[4] Kahn MG: Modelling time in medical decision support programs. *Medical Decision Making* 11, 249–264, 1991

[5] Keravnou E: Guest editorial: Medical temporal reasoning. *Artificial Intelligence in Medicine* 3, 289–290, 1991

[6] Kohane IS: Temporal reasoning in medical expert systems. In: *MEDINFO*-86, Amsterdam, 170–174, 1986

[7] Long WJ: Reasoning about state from causation and time in the medical domain. In: *AAAI*-83, Washington, Wiliam Kaufmann, 251–254, 1983

[8] Russ TA: A system for using time dependent data in patient management. In: *MEDINFO*-86 Amsterdam, Elsevier, 165–169, 1986

[9] Allen JF: Towards a general theory of action and time. *Artificial Intelligence* 23, 123–154, 1984

[10] Dean TL, McDermott DV: Temporal data base management. *Artificial Intelligence* 32, 1–55, 1987

[11] McDermott DV: A temporal logic for reasoning about processes and plans. *Cognitive Science* 6, 101–155, 1982

[12] Kahn MG: Combining physiological models and symbolic methods to interpret time-varying patient data. *Methods of Information in Medicine* 30, 167–178, 1991

[13] Kowalski R, Sergot M: A logic-based calculus of events. *New Generation Computing* 4, 67–95, 1986

[14] Shoham Y: Temporal logics in AI: semantical and ontological considerations. *Artificial Intelligence* 33, 89–104, 1986

[15] McCabe RE, Remington JS: The diagnosis and treatment of toxoplasmosis. *European Journal of Clinicial Microbiology* 2(2), 95–104, 1983

[16] Nagy S: *Eine Shell für medizinische Expertensysteme in der Serologie mit einer Wissensbasis für das Toxoplasmose-Screening.* Diplomarbeit, Depart. of Medical Computer Sciences, Univ. of Vienna 1994

[17] Thalhammer O, Heller-Szöllösy E: Erfahrungen mit routinemäßigem Toxoplasmose-Screening bei Schwangeren zwecks Verhütung angeborener Toxoplasmose. Eine prospektive Untersuchung. *Wiener klinische Wochenschrift* 91, 20–25, 1979

[18] Steimann F, Hayde M, Panzenböck B, Adlassnig KP, Pollak A: Fuzzy support of serodiagnosis: the ONSET program. *IEEE Engineering in Medicine and Biology Magazine* 13(5), 1994, in press

[19] Adlassnig KP, Horak W: Routinely-used automated interpretive analysis of Hepatitis A and B serology findings by a medical expert system. In: *Proc. MIE'*90, Springer-Verlag, Berlin, 313–318, 1990

[20] Pohl B, Trendelenburg C: Pro.M.D. — A diagnostic expert system shell for clinical chemestry test result interpretation. *Methods of Information in Medicine* 27, 111–117, 1988

[21] Schröder JP, Groß U, Heesemann J, Pohl B, Trendelenburg C: Entwicklung eines wissensbasierten Systems für die Toxoplasmose-Diagnostik. *Klinisches Labor* 40, 83–85, 1994

[22] Haimowitz IJ: *Knowledge-Based Trend Detection and Diagnosis.* PhD thesis, MIT, 1994

[23] Shahar Y, Musen MA: A temporal-abstraction system for patient monitoring. In: *Proc. of the 16th Annual Symposium on Computer Applications in Medical Care,* IEEE press, 121–127, 1992

[24] Nejdl W, Gamper J: Qualitative temporal behavior description and temporal diagnosis using interval algebra. *Fourth International Workshop on Principles of Diagnosis*, 91–104, 1993

[25] Avent RK, Charlton JD: A critical review of trend-detection methodologies for biomedical monitoring systems. *Critical Reviews in Biomedical Engineering* 17(6), 621–659, 1990

[26] Challis RE, Kitney RI: Biomedical signal processing (in four parts) part 1: time domain methods. *Medical & Biological Engineering and Computing* 28, 509–524, 1990

[27] Gordon K, Smith AFM: Modelling and monitoring biomedical time series. In Spall J (ed.): *Bayesian Analysis of Time Series and Dynamic Models*, New York, Dekker 1988

[28] Sittig DF, Factor M: Physiologic trend detection and artifact rejection: a parallel implementation of a multi-state Kalman filtering algorithm. *Computer Methods and Programs in Biomedicine* 31, 1–10, 1990

[29] ARDS. In: Niemer M, Nemes C, Lundsgaard-Hansen P, Blauhut B (eds.): *Datenbuch Intensivmedizin*, Gustav Fischer Verlag, New York, 131–170, 1992