F. Borst

Division of Medical Informatics,
Hôpital Cantonal, Hôpitaux
Universitaires de Genève,
Geneva, Switzerland

Commentary

Sequential Diagnosis by Computer

Reflections on G.A. Gorry and G.O. Barnett's paper:
Sequential Diagnosis by Computer

1. Introduction

Gorry and Barnett [1] described a computer-aided diagnosis system based on a sequential approach, mimicking the physician who successively evaluates the current view of the problem (the patient), chooses to perform a test in the hope of gaining additional information, and evaluates again - up to the point of ceasing - testing and making his diagnosis.

The system was based on:
- The information structure constituting the medical experience, based on a set of probabilities linking signs and symptoms;
- The inference function, applying Bayes' conditional probability rules to the information structure;
- The test selection function invoked by the program to select a test for the patient. Tests are sequentially selected according to the «current view» on the patient, the costs of tests, and the costs associated with possible misdiagnoses.

This paper presents a pioneering study in the field of computer-aided diagnosis and raises several issues largely still unsolved after 30 years:
- The "information structure" refers to the knowledge representation that is, the way of storing the knowledge. Thirty years of computerized decision-making still leave this matter open, with a variety of possible solutions ranging from numerical data representation to production rules, as well as other means, including frames, heuristic methods or representations of cases (Section 2).
- The "inference function" describes how this knowledge is handled by the computer to provide the appropriate diagnosis on a given patient. Here, too, multiple solutions depend on the knowledge representation. Gorry and Barnett enlarge this function by adding a time dimension, i.e., a sequential approach in contrast to most of the existing decision-making systems which perform an «all at once» approach. Behind the sequential approach lies the problem of the usefulness and costs of tests, which is a challenging issue (Section 3).
- How does the program interact with the user; globally or sequentially? This question is closely linked to the notion of "clinicians' acceptance" (Section 4).

2. Information structure

The "information structure" refers to the knowledge representation, i.e., the way of storing the knowledge such that it can be handled by the "inference function". Gorry's system of information structure is built around Bayes' rules, which is a set of "a priori probabilities for the diseases and conditional probabilities for various signs and symptoms, given these diseases".

Historically, Bayes' rule was first applied in health care by Homer Warner in the early sixties [2], followed by this paper in 1968. Warner and Gorry are, thus, pioneers of this approach. Then came De Dombal, in 1972, in the field of abdominal pain [3], together with Lusted [4]. Bayes' rule systems are still used in medicine today, even though the problems of using them are well known:
- Bayes' approach is based on real observations reduced to averages. A patient having two diseases simultaneously may not be recognized by the system, if being in-between two disease profiles;
- Associations of signs or symptoms cannot be handled as Bayes' rule considers all findings independent, which is an uncommon situation in medicine.

After thirty years of computerized decision-making, the discussion on knowledge representation remains open, with a wide variety of solutions:
- Systems based on numerical representation of knowledge, other than Bayes, i.e., decision analysis [5, 6] and discriminant analysis [7];
- Systems based on production rules [8], alerts or critiquing systems (Sec-

tion 4),
- Heuristic systems, such as QMR, the largest system in the field of internal medicine [9], or systems based on heuristic classification [10];
- Neuronal networks [11] or learning-based systems [12];
- Semantic networks or frames [13, 14];
- Systems based on case representation [15, 16].

These systems also have their drawbacks; most of them are confined to a restricted domain and their knowledge is difficult to maintain and update. Gorry admits that the determination of this knowledge is difficult: "The determination of these costs (on some common scale) is undoubtedly extremely difficult in any actual problem area". If the initial determination of parameters is difficult, updating will also be complex and time consuming, with the constant risk of destroying the coherence of the previously recorded knowledge. QMR, for example, requires an effort of more than 35 man-years for building and maintaining the knowledge. The process of adding a new disease or syndrome is time consuming and complex. Every two or three years, an entire set of "classical" cases is analyzed with the aim of checking the coherence of the knowledge base [17].

In this context, critiquing systems are emerging, which are increasingly used for decision-making (Section 4).

The origin of knowledge is an important issue in decision-making programs. Either the knowledge comes from experts and is represented as "disease knowledge", building an average ideal case for each diagnosis, or the knowledge originates from a set of individual cases. Should the knowledge be represented as diseases or as cases?

Most systems presented above represent their knowledge in the form of an average case for each diagnosis, the memory of original cases having been reduced to average probability factors (i.e., the probability of having such diagnosis, on the basis of a given finding). As long as the reference population is homogeneous this approach is adequate, but in the presence of sub-populations the average case (the theoretical case) diverges from reality, each sub-population having its own characteristics and associations of findings. The system will then look for an average case that does not exist.

This issue was raised by Gorry himself some years later: "... diagnostic approach can be considered as solving a problem which consists of classifying a patient with the goal of comparison with previously known patients from which therapeutics and prognostic implications are known" [18]. Kassirer added that this classification process has to be based on numerous data: "a large number of data have to be gathered on many aspects of the patients before the physicians can be convinced of the exactitude of the diagnosis" [19].

This leads to another comment: one can notice in the presented results a relationship between the number of studied cases (column 3 of Table 2 in the article) and the average probability assigned to correct diagnoses (column 4): the correlation coefficient R is 0.59. In other words, the mean probability assigned to correct diagnoses is 0.26 (median 0.04) when the number of cases is lower than 11, and 0.69 (median 0.75) when greater than 10 (t-test p-value < 0.001, Mann-Whitney p-value $= 0.01$). This may be due to a random effect or to the fact that the disease knowledge is better when the prevalence is higher; this better knowledge is reflected in the "information structure" (the set of probability relations between 35 diseases and 57 signs and symptoms available in the program). There is also a significant relationship between the average probability assigned to the correct disease (column 4 of Table 2) and the a priori probability of the disease (column 2). This possible relationship between the number of cases and the diagnostic power suggests that the system may be improved by increasing the number of cases.

3. Sequential Approach and Usefulness of Tests

The inference function represents the way in which the computer reaches a diagnosis according to its knowledge representation and the patient data. If Bayes' rule has become a classic since 1968, the way of applying it, by a sequential approach, remains original to this day.

The potential value of a test is influenced by the current view on the patient, i.e., his specific context. For example, a chest X-ray will provide more information in suspicion of tuberculosis than in suspicion of appendicitis. The more information the physician obtains about the patient, the less risk of a possible misdiagnosis. On the other hand, the tests available are not without some costs in terms of patient discomfort, time of skilled persons, money, etc. There is a conflicting tendency to keep the number of diagnostic tests to a minimum. The physician resolves this conflict by performing a sequential diagnosis. When the results of the test are known and are incorporated into his current view, he can choose to perform another test in the hope of gaining additional information or to cease testing and make his diagnosis. Gorry and Barnett's paper is centered around this conflict and around the question: "what is the best test to perform for my patient, using what I know about him or her, to have the best chance of getting a definite diagnosis?" This approach underlines the importance of the context of the patient, the "current view", and leads to a sequential diagnosis by computer, reproducing the physician's reasoning.

Gorry and Barnett's approach of the cost problem remains one of the most interesting aspects of this paper. Already in 1968, this problem was considered as central, even though at that time economic constraints on health care were not as strong as today. Their notion of cost contains "the costs of tests and the costs of the misdiagnoses". During the physician-computer session, tests are sequentially selected according to the current view of the patient, their own costs and the costs associated with possible misdiagnoses. Each pair of diagnoses is associated with a cost of misdiagnosis, e.g., 1,000,000 for diagnosing a malign tumor as benign, or 100,000 for the opposite. "That is, decisions about the patient reflect not only the likelihoods of the possible diagnoses, but the potential cost of misdiagnoses as well".

Costs of test are also present in other decision-making systems, such as QMR [9], which provides for each proposed test an estimation of its, cost e.g., "Simple, Inexpensive Laboratory Tests" (such as "ECG: ST Segment Depression With Reciprocal Elevation During Substernal Pain") or "Moderately Expensive and/or Invasive Laboratory Tests" (such as "ECG: Ventricular Premature Contractions Exertional"). Decision analysis also contains the notion of utility of tests, in the management, for example, of pulmonary embolism where there are risks of complications from both pulmonary arteriography (a potentially harmful invasive test) and long-term anticoagulation [5,6]. The article definitely Gorry clearly opened up a path by the clear notion of "misdiagnosis".

One may argue that cost is associated with time, as the cost of one day of hospitalization is probably higher today than most of the performed tests, and the rapid performance of a powerful test at the beginning of hospitalization can save time and thus cost, by ruling out several diagnoses. This observation could lead to the need for updating cost data within the model.

A dynamic approach can be seen in more recent models, such as "Dynamic decision analysis", which are able to incorporate a Bayesian learning system to automatically learn the probabilistic parameters from large medical databases [20].

4. Clinicians' Acceptance

Shortliffe, in 1976[8], quoting Croft [21], notes that three basic problems remain to be solved before developing diagnostic problems:

- Lack of standard medical definitions;
- Lack of large, reliable medical data bases;
- Lack of acceptance of computer-aided diagnosis by the medical profession.

User acceptance was already a problem in 1972 and remains so today.

User acceptance, especially by clinicians, plays a major role in the dissemination of decision-making software. For acceptance by clinicians, a computer program should:

- Avoid silly questions (it should contain basic knowledge and be context-dependent);
- Be able to explain its reasoning (not act as a black-box);
- Contain an undetermined factor that will ensure its success.

Unfortunately, the most important factor is probably the last one. Response time is also important, because physicians are always under pressure of time.

User acceptance certainly plays a role in Gorry's assertion: "the fundamental role that sequential decision-making plays in the process. It seems clear that it will be necessary for a computer program to exploit an analogous capability". Another detail shows that the author is well aware of this problem; in his example session (Table 1), he quotes successive dialogues with the computer, which uses at least three synonyms to incite the user to interact: "Please continue", "OK, please go on" and "All right, go on please". This detail strongly suggests that Gorry considered the form of the man-machine dialogue as important.

In terms of clinical acceptance, it is worth mentioning "critiquing systems" that have been well accepted. These systems are often simple, if based on only a few rules, and are integrated into existing information systems, such as hospital information systems or electronic patient records. For example, The HELP system integrated a computer system to minimize overtransfusion by prompting physicians when orders were made that did not meet accepted criteria [22]. Safran has integrated such alerts in the electronic patient history, for example, that "your patient's white blood cell count has dropped and you should consider adjusting the AZT dose", or "your patient's CD4 count has been below 200 on two occasions and you should consider prophylaxis for Pneumocystis carinii pneumonia" [23].

5. Conclusion

This paper contains the germs of what would become a period of 30 years of decision-making systems, knowledge representation, and how to deliver this knowledge to the physician. Gorry and Barnett point out a central question: the usefulness and costs of tests and the risk of misdiagnosis. As they say, "Any measure of diagnostic performance should be based on total cost, both the cost of testing and the cost of misdiagnosis". It should become a reflex for any physician before requesting a test to ask: "What happens if the test is positive; will I change my therapeutic attitude?", "What if the test is negative?" and, to add a third question: "To what extent could the test result be misleading, i.e., leading to a misdiagnosis?" Today's economic constraints begin to promote such questions as common behavior, involving "practice guidelines" or "evidence-based medicine".

Looking at 30 years in retrospect, with the ongoing problem of updating knowledge representation, one may take advantage of today's progress in information technology and hospital information systems that allows to integrate data collection in a *continuous process* for *all cases*, not only during some periods on some samples. Applying this advantage to Gorry and Barnett's approach to the utility of tests, then leads to a continuous monitoring of the performance of tests. One could imagine, for example, a system where the test performer, let's say the radiologist, states for each report the suspected diagnosis (a priori), the diagnosis proven or not after the test (a posteriori), and the discriminant power of the examination (in terms of "normal result", "unspecific abnormalities", "pathological findings unrelated to the principal diagnosis" or "examination leading to diagnosis"). Such a method could bring Gorry and Barnett's approach into daily practice and thus make a link between old and recent techniques.

References

1. Gorry GA, Barnett GO. Sequential diagnosis by computer. JAMA 1968;205:141-6.
2. Warner HR, Toronto AF, Veasy LG, Stephenson R. A Mathematical approach to medical diagnosis: Applications to congenital heart disease. JAMA 1961;177:177-83.
3. De Dombal FT, Leaper DJ, Staniland J.R. Computer-aided diagnosis of acute abdominal pain. Brit Med J 1972;2:9-13.
4. Lusted LB. Some roots of clinical decision making. In: Blum B, Duncan K, eds. *A History of Medical Informatics*. Reading, MA: Addison-Wesley, 1990: 385-425.
5. Pauker SG, Kassirer JP. Clinical applications of decision analysis. Semin Nucl Med 1978;8:324-35.
6. Eckman MH, Levine HJ, Pauker SG. Making decisions about antithrombotic therapy in heart disease. Decision analytic and cost-effectiveness issues. Chest 1995;108(suppl):457S-70S.
7. Boom RA. Trends in computerized diagnosis. In: Van Bemmel JH, Ball MJ, Wigertz O, eds. *Proceedings of the Fourth World Conference on Medical Informatics Medinfo83*. Amsterdam: North-Holland, 1983:440-2.
8. Shortliffe EH. *Computer-Based Medical Consultation: MYCIN*. New York: Elsevier, 1976.
9. Aliferis CF, Cooper GF, Buchanan BG, Miller RA, Bankowitz R, Giuse N. Temporal reasoning abstractions in QM. In: Greenes RA et al., eds. *Proceedings of the MEDINFO 95*. Amsterdam: Elsevier Science Publ (North-Holland), 1995:847-51.
10. Musen MA. Modelling for decision support. Chapter 28. In: Van Bemmel JH, Musen MA, eds. *Handbook of Medical Informatics*. Houten/Diegem: Bohn Stafleu Van Loghum, 1997:431-48.
11. Abdi H, A generalized Approach for Connectionist Auto-Associative Memories: Interpretation, Implication and Illustration for Face Processing. In: Demongeot J,, Hervé T, Rialle V, Roche C, eds. *Artificial Intelligence and Cognitive Sciences*. Manchester: Manchester University Press, 1988:149-65.
12. Demongeot J, Robert C. A study of different uncertainty coefficients used in artificial intelligence: towards a natural definition of weights in semantic networks.In: Demongeot J,, Hervé T, Rialle V, Roche C, eds. *Artificial Intelligence and Cognitive Sciences*. Manchester: Manchester University Press, 1988:177-86.
13. Minsky M. A Framework for Representing Knowledge. In: Winston PH, ed. *Psychology of Computer Vision*. New York: McGraw-Hill, 1975:211-81.
14. van Bemmel JH, Musen MA, Miller RA. Methods for decision support. Chapter 15. In: Van Bemmel JH, Musen MA, eds. *Handbook of Medical Informatics*. Houten/ Diegem: Bohn Stafleu Van Loghum, 1997:233-60.
15. Shortliffe EH. The Networked Physician: Practitioner of the Future. In: Ball MJ, Douglas J, O'Destry RI, Albright JW, eds. *Healthcare Information Management Systems*. New York: Springer-Verlag, 1991:3-18.
16. Safran C, Porter D, Lightfoot I, et al. Clinquery: A system for on-line searching of data in a teaching hospital. Ann Intern Med 1989;111:751-6.
17. Giuse DA, Miller RA, Giuse NB. Strategies for medical knowledge acquisition. Chapter 17. In: Van Bemmel JH, Musen MA, eds. *Handbook of Medical Informatics*. Houten/Diegem: Bohn Stafleu Van Loghum, 1997:277-92.
18. Gorry GA. Modelling the diagnostic process. J Med Educ 1970;45:293-302.
19. Kassirer JP, Gorry GA. Clinical Problem Solving : A Behavioral Analysis. Ann Intern Med 1978;89:245-55.
20. Tze-Yun Leong TY, Cao C, Modelling medical decisions in DynaMoL: A new general framework of dynamic decision analysis. In: Cesnik B et al., eds. *Proceedings MEDINFO 98*. Amsterdam: IOS Press, 1998:483-7.
21. Croft DJ. Is computerized diagnosis possible? Comput Biomed Res 1972;5:351-67.
22. Gardner RM, Christiansen PD et al. Computerized continuous quality improvement methods used to optimize blood transfusions. In: Safran C, ed. *Proceedings 17th Annual Symposium on Computer Applications in Medical Care*. New York: McGraw-Hill, 1993:166-70.
23. Safran C, Rind DM et al. An electronic medical record that helps care for patients with HIV infection. In: Safran C, ed. *Proceedings 17th Annual Symposium on Computer Applications in Medical Care*. New York McGraw-Hill, 1993:224-8.

Address of the author:
F. Borst,
Division of Medical Informatics,
Hôpital Cantonal ,
HUG - Hôpitaux Universitaires de Genève,
1211 Geneva 14 ,
Switzerland