Automated Test Selection in Decision-Support Systems: 
* a Case Study in Oncology

Danielle Sent and Linda C. van der Gaag 
Department of Information and Computing Sciences 
Utrecht University 
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands

Abstract. Decision-support systems in medicine should be equipped with a facility that provides patient-tailored information about which tests have to be performed in which phase of the patient’s management. A decision-support system that is equipped with a good test-selection facility may result in ordering fewer tests, decreasing financial costs, shortening the waiting lists, improving the patient’s quality of life and an improvement of medicine in general. In close cooperation with two experts in oncology, we designed such a facility for automatic selection of tests for a decision-support system for cancer of the oesophagus. The facility selects tests taking into account the health status of the patient, the sensitivity and specificity of the tests and the sequence of goals for which the physician wants to perform tests; closely matching current routines, the facility moreover selects multiple tests at the same time. We feel that by extending our decision-support system with the facility, the system provides further support for a patient’s management and has become even more interesting for use in daily medical practice. We describe our decision-support system in the field of oesophageal cancer. We ran experiments with our facility and presented those results to our domain experts. In this paper, we will discuss the results.

Keywords: decision-support system, oesophageal cancer, diagnostic test selection, patient-tailored management

1. Introduction

Since over the last decades researchers have come to understand more and more of diseases and their management, it nowadays is hard, even for medical specialists, to keep up-to-date with medical literature and with new insights of disease, new drugs and new procedures. Furthermore, physicians have to be more and more aware that any mistake they make, not only will affect the patient under their care, but can also result in professional sanctions and financial compensation. On the other hand, the costs of treatments, tests and procedures are increasing. Hospitals have to provide care with a limited budget and yet have to keep care at the highest possible level. To cope with the increasing complexity of medical practice and to provide for a constant level of care, medical procedures are becoming more and more standardised. Various different medical protocols and guidelines have been developed, for example, in which the physicians’ daily problem-solving strategies have been caught. Also, decision-support systems can assist physicians in their complex problem-solving tasks, by providing support that is tailored to individual patients.

To be accepted and integrated with other hospital information systems, a decision-support system should have the confidence of the physicians and prove to be of added value to their daily practice, for example by resulting in lower costs, shorter waiting lists, less discomfort for patients, and better care. The conclusions reached by the system should be in accordance with the state of the art in medical practice. These conclusions should in addition really support its users, rather than, for example, only provide information they readily consider by themselves. To support the entire process of a patient’s management, moreover, a decision-support system should not only provide information about the most probable diseases or best suitable therapy, it should also provide its user with patient-tailored information about which tests should be performed in the various phases of the patient’s management. In this paper, we describe the design of such an automated test-selection facility.

Decision-support systems in medicine, where reasoning with uncertainty is quite prominent, are often based upon techniques from the field of statistics. More specifically, these systems increasingly build upon a Bayesian network for their reasoning [1,2]. A probabilistic network is a concise representation of a joint probability distribution on a set of statistical variables.

The test-selection algorithms currently in use in such decision-support systems are based upon the mathematical principles of decision theory. Most of these algorithms serve to select diagnostic tests in a myopic fashion. In each
step of the algorithm, the most informative test is selected from among all possible tests. The most informative test is computed by a measure based upon information theory. The user is prompted for the result of the test. The result is entered into the decision-support system. Then the next most informative test is selected. This process continues until results for all available tests have been entered into the system.

Over the past decade, a decision-support system has been developed at Utrecht University that is aimed at a better patient-specific management of oesophageal cancer. A myopic test-selection facility was added to this system. Based upon our knowledge of the management of patients as well as our experiences upon working with our system, we feel that a myopic test-selection strategy is an oversimplification of the experts’ problem-solving strategy in many medical fields of application. From the three interviews we had with our domain experts, we learned that tests are ordered in packages and tests provide information for various different subgoals that are considered one by one. Based upon these findings, we designed a test-selection facility for a decision-support system in the field of oncology. Our facility induces a test-selection strategy that is also based upon the mathematical principles of decision theory, yet more closely fits in with the test-selection strategy employed by physicians in their daily practice. To study the practicability of our facility, we evaluate the test-selection facility in the context of a decision-support system in the domain of oesophageal cancer.

This paper is organised as follows. In Section 2, we provide some preliminaries from the field of oesophageal cancer and describe the decision-support system for which we designed a test-selection facility. In Section 3 we describe the various different components of our facility and their background. Section 4 presents the test-selection strategy resulting from our facility in practice. We conclude the paper with our conclusions in Section 5.

2. Preliminaries

With the help of two experts in gastrointestinal oncology from the Netherlands Cancer Institute, Antoni van Leeuwenhoekhuis, a decision-support system for the field of oesophageal cancer has been developed [3,4]. We briefly describe this field of expertise and the decision-support system under study.

2.1. Domain of Application

Cancer of the oesophagus has a low incidence in the Netherlands and is often only diagnosed in a later stage of the disease. The tumour invades the oesophageal wall and may, in time, invade neighbouring organs beyond the oesophagus. When the tumour has invaded lymphatic vessels and blood vessels more specifically, it may give rise to secondary tumours, or metastases, in lymph nodes and in such organs as the liver and lungs. The depth of invasion of the oesophageal tumour and the spread of its secondary tumours indicate the severity of the disease, which is summarised in the cancer's stage. In order to assess the stage of a patient's oesophageal cancer, generally a number of diagnostic tests are performed, such as a gastroscopy and a CT-scan of the abdomen; These tests serve to give insight in the different aspects of the cancer that determine its stage.

For patients suffering from oesophageal cancer, various different treatment alternatives are available. These alternatives include surgical removal of the primary tumour, administering radiotherapy, and positioning a prosthesis. Providing a therapy aims at removal or reduction of the patient's primary tumour to prolong life expectancy and to improve the passage of food through the oesophagus. The therapies differ in the extent to which these effects can be attained, however, dependent upon the depth of invasion of the primary tumour and the extent of its metastases.

2.2. The Oesophageal Cancer Network

Over the past decade, a decision-support system has been developed that is aimed at a better tailored patient-specific management of oesophageal cancer. The system has a Bayesian network at its kernel that has been constructed with the help of two domain experts from the Netherlands Cancer Institute. We refer to this network as the oesophageal cancer network. The network in essence provides for the staging of the cancer of the oesophagus. It models the presentation characteristics of an oesophageal tumour and describes the pathophysiological processes that influence its growth and metastasis. The network further represents the various diagnostic tests that are commonly used to gain insight in the properties of the tumour and the extent of its advance. The main diagnostic variable is the variable Stage that captures the extent to which the cancer has progressed. The network currently includes 42 stochastic variables, for which almost 1000 parameter probabilities have been specified. Of the 42 variables included in the network, 25 variables serve to model the results of 12 available physical tests. Using data from 185 patients, an
evaluation study of the oesophageal cancer network was conducted. It was found that for 86% of the patients, the network established the stage of the patient’s cancer correctly [5]. Figure 1 depicts the oesophageal cancer network.

3. The Test-selection Facility

A test-selection facility should be able to provide information about which tests are expected to provide the most valuable information in the present phase of the diagnostic process. The facility should induce a test-selection strategy that is based upon the mathematical principles of decision theory and the mathematical basis of a probabilistic network, yet closely fits the test-selection strategy employed by physicians in their daily practice.

The test-selection process should further result in a strategy that does not suffer from overtesting, that is, it should not select too many tests. Even more importantly, the resulting strategy should not halt the test-selection process too soon, since this may lead to misdiagnosis with possibly major consequences for both patients and physician. Automated test selection thus consists of three basic elements: a measure for establishing the usefulness of performing a particular test, a test-selection loop, and a criterion for deciding when to stop gathering further information.

Most algorithms currently in use in practical decision-support systems serve to select diagnostic tests co-called myopically. In each step of the test-selection loop, the most informative test variable, the variable for instance representing the partial result Gastro-shape of a gastroscopic examination of the oesophagus, as can be seen in Figure 1, is selected from among all possible test variables to indicate the next test to perform. The user of the decision-support system is then prompted for the value of the selected variable. The result is entered into the decision-support system and propagated to establish the posterior probabilities for all variables. From the set of test variables still available, the next variable is selected. We feel that the test-selection strategy that is induced by such a myopic algorithm is an oversimplification of the physicians’ problem-solving strategy.

A non-myopic test-selection algorithm selects at each step multiple tests of test variables as the most informative cluster of test variables. Such a test-selection algorithm might become computationally too demanding, especially if a meaningful clustering of test variables results in clusters of relatively large size. A blood test, for example, may
easily yield results for tens of variables in a network. To save computation time yet retain from the idea of non-myopia, we designed an algorithm that implies a semi-myopic test-selection strategy. In pseudo code, our algorithm equals

*Input:* List of subgoals
List of test variables belonging to one physical test, organised per subgoal
Stopping criterion subgoal = false
Stopping criterion overall = false

**WHILE** list of subgoals is not empty and stopping criterion overall is false **DO**
Select a subgoal
Remove that subgoal from the list of subgoals

**WHILE** the list of test for that subgoal is not empty, stopping criterion subgoal is false and stopping criterion overall = false **DO**
Compute most informative test variable
Prompt for evidence for this test variable and all variables belonging to the same physical test as the most informative test variable
Propagate the evidence
Remove the physical tests and belonging test variables from the list of test variables
Compute stopping criterion subgoal
Compute stopping criterion overall

**OD**

In this section, we will discuss the three elements of our automated test-selection facility and the ingredients of the algorithm as presented above.

3.1. Test-selection Measure

The usefulness of performing a diagnostic test differs per patient and for a specific patient; it is also influenced by the stage of the diagnostic process the patient is in. The usefulness of a test should thus be patient-tailored and dynamic when it is used for test-selection purposes. The usefulness is typically expressed using a so-called test-selection measure. Such a measure can be based solely upon the expected decrease in diagnostic uncertainty after performing a test, that is, the test that is likely to induce the largest decrease in diagnostic uncertainty is selected as the most useful. We note that the diagnostic uncertainty is the smallest when a disease has been established with certainty and the largest when all options are equally likely. In addition such a measure can incorporate the costs of the test. The term costs of a test is generally used in a very broad sense, including numerous aspects like the costs associated with hospitalisation, nursing care, physicians’ fees as well as the risk involved, the discomfort to the patient and the possible consequences of the lag time between ordering a test and obtaining its result.

In test-selection, information measures are used for expressing diagnostic uncertainty. These measures serve to assign a numerical value to the probability distribution over the main diagnostic variable, which basically captures the amount of information that is required to resolve any uncertainty about the diagnosis. The numerical value attains its maximum when the diagnostic uncertainty is the highest, that is, when all possible diagnoses are equally likely, and its minimum when the diagnostic uncertainty is the lowest, that is, when one of the values of the diagnostic variable has been established with certainty. The three most commonly used information measures are the Shannon entropy, the Gini index and the misclassification error [6,7]. We experimented with the three measures with our decision-support system for the staging of oesophageal cancer. We concluded that the misclassification error should not be used for test-selection purposes due to its tendency to select tests randomly when all possible shifts in certain cases. The Shannon entropy and the Gini index are equally suited for test selection. For computational reasons, we selected the Gini index as the information measure we would be using for our test-selection facility. This measure can easily be extended to be able to take the costs of a test into account as well. Note that we did not take the costs involved into consideration, since we did not have the appropriate information available.

3.2. Test-selection Loop

With the advance of protocols for many fields of medicine, guidelines are becoming available that prescribe, to least some extent, the diagnostic tests that physicians should perform at various intervals in their reasoning processes. Such a protocol could constitute the basis for an automated test-selection facility that would closely fit in with medical practice. Unfortunately, such a detailed protocol is not available yet for the management of patients.
with cancer of the oesophagus. We decided to design a test-selection facility for our system that would build upon the arguments used by the experts for deciding whether or not to order specific tests. The resulting facility would thus more closely fit in with the strategies for test selection currently used in the domain than a standard sequential test-selection facility.

For eliciting the arguments underlying the experts’ test-selection strategy, we used an elicitation method that was composed of three focused interviews: an unstructured interview, during which we elicited the general test-selection strategy employed by our domain experts, followed up by a structured one during which we discussed eight patient cases using vignettes and an interview for refinement purposes [8]. From these interviews, we learned that the test-selection strategy employed by our domain experts in their daily practice is aimed at deciding between different therapies rather than at establishing the stage of the disease. In their strategy, moreover, different subgoals can be identified that are addressed sequentially. We feel that a more involved test-selection facility should be able to take such subgoals into account in order to result in a test-selection strategy that feels natural to the physicians to work with. Moreover, our experts have been found to order tests in packages, that is, tests that serve to provide information for a certain subgoal are ordered all at once, to reduce the length in time of the diagnostic phase of a patient’s management. For the latter purpose, a myopic algorithm should not be used. A non-myopic test-selection algorithm, which selects a combination of tests variables, rather than just one test variable, prompts the user for its results and selects the next combination of tests, should be preferred. However, such a test-selection algorithm might become computationally too demanding, especially if a meaningful clustering of test variables results in clusters of relatively large size. In the domain of oesophageal cancer, multiple test variables serve to model the results of a single physical test. We designed a so-called semi-myopic algorithm that selects test variables myopically, however prompt the user not only for the result of the most informative test variable, but for the results of all test variables that belong to the same physical test as the physical test the most informative test variable pertains to. To be able to implement our test-selection facility, we had to provide our decision-support system with detailed information about the various subgoals and test variables and physical tests involved.

3.3. Stopping Criterion

Automated test selection not just includes selecting appropriate tests in the different steps of a patient’s management, it also involves deciding when to stop gathering further information. For this purpose, a test-selection facility is equipped with a stopping criterion. With this criterion, it decides after processing the findings entered by the user, whether or not the diagnostic uncertainty has sufficiently decreased and no further tests should be ordered for the patient. A good stopping criterion can thus leads to fewer tests being performed, which results in lower financial costs, less discomfort for the patient, and perhaps even shorter waiting lists. More importantly, however, a good stopping criterion should prevent the test-selection algorithm to halt too early.

Most test-selection facilities use a stopping criterion that is based upon using a threshold value on the uncertainty. Such a stopping criterion for instance builds upon the probability of the most likely value of the main diagnostic variable. In general, we adopt this principle, however, we propose to be able to apply this criterion to the various subgoals distinguished. We further observe that, since we use the expected decrease of diagnostic uncertainty based upon the value of the Gini index over a specific goal variable as the basis of our test selection, this expected decrease can be taken for a stopping criterion as well: a small expected decrease indicated that there is not much uncertainty left with regard to the most likely value. It may then very well be that a test result exists, for a not yet performed test that serves to again increase the diagnostic uncertainty. To forestall halting too early, we have designed our stopping criterion to apply a restricted look-ahead. Suppose that after the result for a test variable has been entered and processed the value of the Gini index over the subgoal is smaller than or equal to a prespecified threshold. Before the test-selection is actually halted, the algorithm examines whether or not a next test result could induce a shift in diagnostic uncertainty to a value above the threshold.

Note that when the costs of the various diagnostic tests are taken into consideration, these pose a more or less natural stopping criterion: the selection of diagnostic tests is continued as long as the expected benefits of a test outweigh its costs. Since our test-selection facility at this moment does not take the costs of diagnostic tests into consideration, we will focus on a different stopping criterion.

4. Preliminary Results

We ran some experiments with our decision-support system for oesophageal cancer using eight patient cases. We were interested mainly in whether or not our stopping criterion would indicate to stop testing at the exact step in the test-selection process at which the experts would want it to stop. We were also interested in whether or not the
algorithm would result in sequences of tests that would fit the test-selection strategy employed by our experts; we were aware, however, that in our experiments we did not take the costs involved into consideration and as a consequence we did not expect a perfect match.

We presented the sequences of tests and the moments at which the test-selection process halted, as generated by our test-selection facility, to our domain experts. The experts indicated that they felt quite comfortable with the results. The sequence of tests felt natural and the moment at which the test-selection process halted was considered correctly. The experts indicated, however, that in their daily routines they would prefer to perform ‘simple’ tests first, that is, tests for which the results are obtained within a short period of time. We would like to note that preferring simple tests can be readily accommodated by our algorithm by including a notion of time.

5. Conclusions

Upon working with a decision-support system in oncology, we noticed that the commonly used test-selection facility presented an oversimplification of medical practice. In this paper, we presented a new facility that offers patient-tailored test selection in a way that closely fits in with the daily routines of our experts. Our new facility is based upon mathematical principles, as well as on knowledge of the arguments underlying the daily test-selection strategies employed by physicians. Our algorithm selects the most informative variable describing a single result from a physical test, but prompts the user for all results from the test. Our algorithm further used the subgoals that our domain experts use in their test-selection strategies. We presented the test-selection sequences generated by our algorithm to the experts. They indicated that they felt quite comfortable with the sequences generated by our algorithm for a number of patient cases.

We would like to note that the current test-selection routines employed by our experts are based upon general guidelines. These guidelines, however, are not patient-specific. Patient-specific guidelines would easily become impractical to work with. With our test-selection facility, however, the experts are supported in working far more patient specific, which is likely to induce a decrease in the amount of money spend and an increase in the quality of patient management.

Although further research on including for example the time it takes to obtain a test result should still be performed, we feel that we have taken an important step towards the acceptance of decision-support systems as valuable assistants for physicians.

Acknowledgements

This research was partly supported by the Netherlands Organisation for Scientific Research. We are most grateful to Babs Taal and Berthe Aleman from the Netherlands Cancer Institute, Antoni van Leeuwenhoekhuis, who spent much time and effort in the construction of the oesophageal cancer network and provided information necessary for the design of our test-selection facility. We would furthermore like to thank Arend-Jan de Groote, Martijn Schrage and Arjan van IJzendoorn for implementing the test-selection algorithm in our Bayesian-network toolbox.

References