On the Evaluation of Probabilistic Networks

Linda C. van der Gaag and Silja Renooij

Institute of Information and Computing Sciences, Utrecht University,
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
{linda,silja}@cs.uu.nl

Abstract. As more and more probabilistic networks are being developed for medical applications, the question arises as to their value for clinical practice. Often the clinical value of a network is expressed as the percentage correct of predicted overall outcome, based upon an evaluation study using real-life patient data. In this paper, we propose another method of evaluation that focuses on intermediate outcomes of interest. We illustrate this method for a real-life probabilistic network for the staging of oesophageal cancer and show that it can provide valuable information in addition to a percentage correct.

1 Introduction

In various fields of clinical medicine, probabilistic networks are being developed to support physicians in the difficult tasks of diagnosis and prognostication. A probabilistic network basically is a statistical model comprised of a graphical structure and an associated set of probability distributions [1]. The graphical structure models the statistical variables that are relevant in the field of application, along with the influential relationships between them; the strengths of the relationships are captured by conditional probabilities.

To establish the value of a probabilistic network for clinical practice, it generally is subjected to an evaluation study using real-life patient data. For each patient, the available data are entered into the network whereupon the network computes the most likely diagnosis, or another outcome of interest; the computed outcome is then compared against a given standard of validity. The results of the study are often summarised in the *percentage correct*, or *accuracy*, of predicted outcome, that is, in the percentage of patients for whom the outcome is correct.

An evaluation study as described above typically focuses on a single *overall outcome*. In this paper we investigate another, more elaborate, method of evaluation that focuses on *intermediate outcomes* of interest. We illustrate this method of evaluation for a real-life probabilistic network for the staging of oesophageal cancer and show that it can provide valuable additional information for assessing the clinical value of a network. In Sect. 2, we briefly describe the oesophagus network and available patient data. We present our evaluation method in Sect. 3. The paper ends with some concluding observations in Sect. 4.

S. Quaglini, P. Barahona, and S. Andreassen (Eds.): AIME 2001, LNAI 2101, pp. 457–461, 2001.

2 The Oesophagus Network and the Patient Data

With the help of two experts in gastrointestinal oncology, we developed a probabilistic network that captures the state-of-the-art knowledge about oesophageal cancer [2]. The network describes the characteristics of an oesophageal tumour and the pathophysiological processes of invasion and metastasis. The depth of invasion and extent of metastasis are summarised in the tumour's *stage*, which is either I, IIA, IIB, III, IVA, or IVB, in the order of advanced disease. The network includes 42 variables, of which 25 are observable, and almost 1000 probabilities.

For investigating the clinical value of the oesophagus network, we have available the medical records of 156 patients diagnosed with oesophageal cancer. For each patient, various observations, for example obtained from diagnostic tests, are recorded. In addition, the tumour's stage and the values of various intermediate, unobservable variables are stated; these values basically are conjectures of the attending physician. The three most important intermediate variables pertain to the presence or absence of haematogenous metastases (the variable *Haema-metas*), to the extent of lymph node metastases (*Lymph-metas*, with the values N0, N1, and M1), and to the depth of invasion of the tumour into the oesophageal wall (*Invasion-wall*, with the values T1, T2, T3, and T4).

3 Evaluation of Intermediate Outcomes

The method of evaluation that is commonly employed for establishing the clinical value of a probabilistic network, amounts to entering patient data into the network, computing the most likely outcome, and comparing it against a given standard of validity. Such an evaluation typically focuses on a single predicted outcome. Another method of evaluation is to subsequently enter all possible outcomes, compute the expected distributions for the observable variables, and compare these against the patient data [3]. For probabilistic networks that model a large number of observable variables for which relatively few data are available, such an evaluation is infeasible. The basic idea of this evaluation method, however, can be used not just for observable variables, but also for crucial intermediate variables. We illustrate this observation for the oesophagus network.

The stage of an oesophageal tumour is defined by the values of the intermediate variables *Haema-metas*, *Lymph-metas* and *Invasion-wall*. The ability of the oesophagus network to distinguish between the various stages thus depends on how much the probability distributions for the three intermediate variables differ per stage. Fig. 1 shows the prior distributions for these variables for the six different stages. We observe, for example, that the probability distributions for *Invasion-wall* are rather similar for the stages IVA and IVB.

In the field of statistics, various measures have been developed for expressing the difference, or *distance*, between two probability distributions. An example of such a measure is the *Kullback-Leibler information divergence* [4]. For a statistical variable with m values, we consider two different distributions p and p', with probabilities p_j and p'_j, $j = 1, \ldots, m$, respectively. The Kullback-Leibler information divergence $I(p, p')$ of p from p' is defined as

	Haema-metas		*Lymph-metas*			*Invasion-wall*			
	no	yes	N0	N1	M1	T1	T2	T3	T4
stage I	1	0	1	0	0	1	0	0	0
stage IIA	1	0	1	0	0	0	0.4267	0.5733	0
stage IIB	1	0	0	1	0	0.1108	0.8892	0	0
stage III	1	0	0.0345	0.9655	0	0	0	0.8665	0.1335
stage IVA	1	0	0	0	1	0.0231	0.1234	0.6395	0.2141
stage IVB	0	1	0.4480	0.3929	0.1591	0.0089	0.2166	0.6734	0.1011

Fig. 1. The probability distributions for the three intermediate variables per stage.

$$I(p, p') = \sum_{j=1}^{m} p_j \cdot \ln \frac{p_j}{p'_j}$$

where $0 \cdot \ln \frac{0}{x} = 0$ for all values of x. The divergence $I(p, p')$ equals infinity whenever $p'_j = 0$ and $p_j > 0$; the measure is therefore not symmetric in its arguments. We illustrate the Kullback-Leibler divergence for the variable *Invasion-wall*. Writing p_i for the distributions for stage i, we find from Fig. 1 that, for example, $I(p_{\text{IVA}}, p_{\text{IVB}}) = 0.0802$, $I(p_{\text{IVB}}, p_{\text{IVA}}) = 0.0723$, $I(p_{\text{I}}, p_{\text{IVB}}) = 4.7217$, and $I(p_{\text{IVB}}, p_{\text{I}}) = \infty$. We note that the probability distributions for the stages I and IVB diverge much more than those for IVA and IVB. The network is therefore more likely to confuse the stages IVA and IVB than it is to confuse I and IVB.

Using the Kullback-Leibler divergence, we can compare the probability distributions for the three intermediate variables per stage, with the same distributions given a patient's data. As an example, Fig. 2 shows the distributions for a specific patient. Based upon these distributions, the network concludes that stage III is the most likely stage for the patient's tumour. The medical record, however, states stage IVA, which is the second most likely in the network's prediction. Upon comparing the probability distributions p_i for stage i from Fig. 1 with the distributions p for the patient, we find the Kullback-Leibler divergences shown in Fig. 3. From these divergences, we note that the probability distribution computed for the patient for the variable *Lymph-metas* points to the stages IIB and III, whereas the distribution for *Invasion-wall* favours the stages IIA, III, and IVB. The divergences for the variable *Haema-metas* reveal stage IVB to be rather unlikely. The distributions computed for the patient, therefore, do not unambiguously point to a single specific stage. The network's confusion, however, is not taken into consideration in establishing the percentage correct.

The prior divergences between the probability distributions of the intermediate variables per stage will change when a patient's data are entered into the network. The distance between the distributions for two different stages can become smaller, or the distributions can become more divergent. For example, without taking patient data into consideration, we find for the variable *Invasion-wall* that the divergence of the distribution given stage III from the distribution given stage IVA equals 0.2002. After the data of the patient described above have been entered, this divergence reduces to 0.1454: the network has become

Haema-metas		*Lymph-metas*			*Invasion-wall*			
no	yes	N0	N1	M1	T1	T2	T3	T4
0.9071	0.0929	0.0334	0.6432	0.3234	0.0088	0.1285	0.8590	0.0037

Fig. 2. The probability distributions for the three variables given a patient's data.

$I_{Haema}(p_{\rm I}, p) = 0.0975$ $I_{Lymph}(p_{\rm I}, p) = 3.3992$ $I_{Invasion}(p_{\rm I}, p) = 4.7330$
$I_{Haema}(p_{\rm IIA}, p) = 0.0975$ $I_{Lymph}(p_{\rm IIA}, p) = 3.3992$ $I_{Invasion}(p_{\rm IIA}, p) = 0.2803$
$I_{Haema}(p_{\rm IIB}, p) = 0.0975$ $I_{Lymph}(p_{\rm IIB}, p) = 0.4413$ $I_{Invasion}(p_{\rm IIB}, p) = 2.0007$
$I_{Haema}(p_{\rm III}, p) = 0.0975$ $I_{Lymph}(p_{\rm III}, p) = 0.3933$ $I_{Invasion}(p_{\rm III}, p) = 0.4862$
$I_{Haema}(p_{\rm IVA}, p) = 0.0975$ $I_{Lymph}(p_{\rm IVA}, p) = 1.1289$ $I_{Invasion}(p_{\rm IVA}, p) = 0.6974$
$I_{Haema}(p_{\rm IVB}, p) = 2.3762$ $I_{Lymph}(p_{\rm IVB}, p) = 0.8566$ $I_{Invasion}(p_{\rm IVB}, p) = 0.2837$

Fig. 3. The Kullback-Leibler divergences for the various stages, given a patient's data.

even more likely to confuse the stages III and IVA. We have found similar results for most patients for whom the network yields an incorrect stage.

4 Conclusions

To establish the value of a probabilistic network for clinical practice, it is generally subjected to an evaluation study that amounts to computing the most likely outcome for a patient from the network and comparing it against a given standard of validity. We discussed another method of evaluation that serves to gain insight in the probability distributions that are computed for crucial intermediate variables. We suggested the use of the Kullback-Leibler information divergence to investigate the distances between distributions. As it uncovers a network's degree of confusion, we feel that this method of evaluation provides valuable information in addition to a percentage correct.

Acknowledgements. This research has been (partly) supported by the Netherlands Computer Science Research Foundation with financial support from the Netherlands Organisation for Scientific Research. We are grateful to Babs Taal and Berthe Aleman from the Netherlands Cancer Institute, Antoni van Leeuwenhoekhuis, who spent much effort in the construction of the oesophagus network.

References

1. F.V. Jensen (1996). *An Introduction to Bayesian Networks*. UCL Press, London.
2. L.C. van der Gaag, S. Renooij, C.L.M. Witteman, B.M.P. Aleman, B.G. Taal (2001). Probabilities for a probabilistic network: A case-study in oesophageal carcinoma, submitted for publication.

3. S. Andreassen, F.V. Jensen, K.G. Olesen (1991). Medical expert systems based on causal probabilistic networks. *International Journal on Biomedical Computing*, vol. 28, pp. 1 – 30.
4. S. Kullback, R.A. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics*, vol. 22, pp. 79 – 86.