



Discovering interobserver variability in the cytodiagnosis of breast cancer using decision trees and Bayesian networks

Nicandro Cruz-Ramírez^{a,*}, Héctor-Gabriel Acosta-Mesa^a,
Humberto Carrillo-Calvet^b, Rocío-Erandi Barrientos-Martínez^a

^aFacultad de Física e Inteligencia Artificial, Universidad Veracruzana, Sebastián Camacho # 5, Col. Centro, C.P. 91000, Xalapa, Veracruz, Mexico

^bFacultad de Ciencias, Universidad Nacional Autónoma de México, Circuito Exterior Ciudad Universitaria, México, D.F., Mexico

ARTICLE INFO

Article history:

Received 1 October 2007

Received in revised form 22 May 2009

Accepted 28 May 2009

Available online 6 June 2009

Keywords:

Breast cancer

Classification

Decision trees

Bayesian networks

Interobserver variability

ABSTRACT

We evaluate the performance of two decision tree procedures and four Bayesian network classifiers as potential decision support systems in the cytodiagnosis of breast cancer. In order to test their performance thoroughly, we use two real-world databases containing 692 cases and 322 cases collected by a single observer and 19 observers, respectively. The results show that, in general, there are considerable differences in all tests (accuracy, sensitivity, specificity, PV+, PV– and ROC) when a specific classifier uses the single-observer dataset compared to those when this same classifier uses the multiple-observer dataset. These results suggest that different observers see different things: a problem known as interobserver variability. We graphically unveil such a problem by presenting the structures of the decision trees and Bayesian networks resultant from running both databases.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

One of the most common types of cancer that affects women in North America, Europe and the Antipodes, is breast cancer [1]. The usual practice in the United Kingdom to accurately diagnose such a disease is to use three different methods: the surgeon's clinical opinion, a mammography and the cytological studies. That is why this diagnosis process is known as the triple approach [1]. Because mammographic studies are not decisive in the diagnosis of breast cancer [1,2], an alternative confirmatory method is needed to support or reject these findings. The most common confirmatory method used in the United Kingdom for this purpose is that of fine needle aspiration of breast lesion (FNAB) [1,3,4]. Such a technique involves a process where a syringe sucks cells from breast lesions using a fine bore needle similar to those used for blood samples. Once this is done, these cells are transferred to a transport solution and sent to the pathology laboratory for a microscopic study carried out by a trained cytopathologist [1]. In this paper, we focus on the discovery of inconsistencies in the interpretation of a sample by building classifiers, based on decision trees and

Bayesian networks, from a cytopathological database retrospectively collected by a single observer and a database prospectively collected by 19 observers.

The time it normally takes to a medical doctor to become an independent practicing cytopathologist in the UK is about 5 years as a minimum [1]. This fact can give an indication of the very complex learning process through which medical doctors have to pass. It is mainly for this reason that machine learning methods for decision-making may have two potential applications: (a) to accelerate the training process of learning by providing guidance to the trainee on what features are the most important ones to look at; and (b) to compare the final results to those of the trainee or even those of the expert so that the decision (whether the sample taken indicates a benign or a malignant output) can be made on more robust criteria. That is why we need to test potential decision support systems in this area to the extreme: we need to check whether it is possible to carry out an objective diagnosis rather than a subjective one; i.e., whether the interpretation of a sample by different pathologists is consistent. To this end, we use 692 consecutive adequate specimens collected by a single pathologist and 322 consecutive adequate specimens collected by 19 pathologists to train and test two decision tree procedures and four Bayesian network classifiers. The results show the presence of a difficulty known as the interobserver variability problem [5–7], which prevents the classifiers from generalizing well enough. The interobserver variability problem refers to the situation where there is no reproducibility of the results given by different experts

* Corresponding author at: Department of Artificial Intelligence, Universidad Veracruzana, Sebastian Camacho 5, Col. Centro, C.P. 91000, Xalapa, Veracruz, Mexico. Tel.: +52 228 817 29 57; fax: +52 228 817 28 55.

E-mail addresses: ncruz@uv.mx, nicandro.cruz@gmail.com (N. Cruz-Ramírez), heacosta@uv.mx (H.-G. Acosta-Mesa), carr@servidor.unam.mx (H. Carrillo-Calvet), erandi_bm@yahoo.com.mx (R.-E. Barrientos-Martínez).

of a determined area on the same set of cases. That is to say, in the specific case of the cytodiagnosis of breast cancer, pathologists do not agree on what cytological features are the most relevant for correctly diagnosing whether a patient has breast cancer or not.

Although the inconsistencies in the two datasets prevent the classifiers presented here from performing robustly, it can be argued that, with the aid of this kind of tools (decision trees and Bayesian networks), the main underlying principles, conditions, mechanisms and causes that lead to this problem could be gradually discovered since these tools graphically show the importance of each variable for the diagnosis of breast cancer (in the case of decision trees) as well as the probabilistic relations among these variables (in the case of Bayesian networks). It is important to mention here that some other works [1,4,8] have applied different classification methods, such as logistic regression, multilayer perceptron neural networks (MLPs) and adaptive resonance theory mapping neural networks (ARTMAPs) for analyzing these same datasets. However, these models (as well as others like symbolic rules, k -nearest neighbors and support vector machines) do not have the power of graphically unveil the mentioned, and central to this investigation, interobserver variability problem. By this we do not mean that such models lack the power for identifying this problem; we just mean that they originally lack the graphical power for unveiling it, since they are not graphical representations. We are not saying either that this graphical feature cannot be added to such models; we are saying that the traditional (original) formulation of these models does not consider this graphical feature [9–15]. For example, symbolic rules might be converted to decision trees but this step is not originally included in such an approach. It can also be argued that the graphical representations presented in this research provide a natural and intuitive framework to model interactions among variables. The terms natural and intuitive suggest that these graphical representations are, under certain conditions, easier to understand than other kinds of representations. A number of researchers from a wide range of scientific disciplines (cognitive psychology, developmental psychology, linguistics, anthropology and computer science) have given evidence that supports such a claim: Gattis [16], Liben [17], Tversky [18], Emmorey [19], Bryant and Squire [20], McGonigle and Chalmers [21], Hummel and Holyoak [22] and Larkin and Simon [23]. Some of them claim that these representations aid cognition because they are structured in such a familiar way that people can rely on them to structure memory, communication and reasoning [16]. Gattis [16] also points out that spatial representations are not merely metaphors that help understand cognitive processes but actual internal mechanisms that allow us to perform more abstract cognitive tasks. Larkin and Simon argue that a diagram can be superior to a verbal description because, when well used, the former “automatically supports a large number of perceptual inferences, which are extremely easy for humans” [16, p. 107]. Graphical representations are useful in reasoning tasks because, through their structure (which can represent order, directionality and relations) and the partial knowledge about their elements and the relations among them, it is possible to infer the values of the elements and their relationships that are unknown [16]. In a similar vein, Larkin and Simon [23] also claim that these representations have the power to group together all the information that is used together, which avoids the problem of searching large amounts of data to find the elements needed for performing inference. As Tversky [18] points out, graphical representation can be used to reveal hidden knowledge, providing models that facilitate inference and discovery. It is very important to remark that, in words of Tversky, “long before there was written language, there were depictions, of myriad varieties” [11, p. 80]. In sum, graphical representations can represent abstract concepts and information in such a way that this

information can be accessed and integrated quickly and easily. They also facilitate group communication [18].

It is also important to justify why these two specific models (decision trees and Bayesian networks) can be regarded as Soft Computing (SC) methods: it is mainly because such models can learn from experimental data and have the power for suitably representing human knowledge [24–27]; enterprises on which the core SC techniques (support vector machines, artificial neural networks and fuzzy logic models) mainly focus their attention. According to Kecman [24], there are additional techniques that can be considered as extensions of SC: evolutionary algorithms, probabilistic reasoning, fractals and chaos theories and belief networks (also known as Bayesian networks). Kecman points out too that IF–THEN rules can express almost all structured human knowledge: decision trees can be easily converted into this kind of rules [13,15,27]. Finally, classification lies within the range of problems solved by SC methods: this is one of the central topics in the present work as well as the transferring of human knowledge into analytical models [24].

The remainder of this paper is organized as follows. In Section 2 we give the rationale for decision trees and Bayesian networks in order to better understand how these models work. In Section 3 we present the materials and methods used for this research. In Section 4 we present the performance of two algorithms that build decision trees and four algorithms that construct Bayesian network structures from a database collected by a single observer and a database collected by multiple observers. In Section 5 we discuss the results given by these procedures and finally, in Section 6, we present some conclusions and propose some sensible directions for future research.

2. Decision trees and Bayesian networks

2.1. Decision trees

The motivation of this type of model is that of inductive learning from observations. A decision tree [15,26,27] is a graphical representation of a function that maps an input (described by a set of attributes) to a specific output (a decision that represents the predicted value for that input). Table 1 shows a sample of observations [28] from which we want to learn the unknown function that maps inputs to outputs.

This example refers to a well-known database found in the pattern recognition literature [28,29]. In this case, the input of the function is the combination of values for the first four columns (sepal_length, sepal_width, petal_length and petal_width; all measured in centimeters), and the output is the decision of the type of iris plant. It is important to make a distinction here: in classification terminology, the variables that serve as input are usually called the attributes and the variable for which we want to make the prediction is usually called the class. Then, our goal is to discover the relation between the attributes and the class and represent it in the form of a tree. One of the most popular algorithms for building decision trees from data is ID3 and its descendants (C4.5 and C5.0) [11,26,27]. The main idea is to recursively choose the attributes that partition the data into homogeneous classes; i.e., no examples with the same values for the attributes and different values for the class should be in the same leaf (see Fig. 1; circles represent attributes, rectangles or leaves represent the class).

We use the tree of Fig. 1 in the following way: given an example for which the class is not known, we traverse each branch of the decision tree accordingly in order to label that example. For instance, if we have the following values for the attributes, sepal_length = 6.5, sepal_width = 3, petal_length = 5.8 and petal_width = 2.2, the decision tree will label this special case as ‘Iris-

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات