Application of the Naïve Bayesian Classifier to optimize treatment decisions

Joanna Kazmierska\textsuperscript{a,*}, Julian Malicki\textsuperscript{b}

\textsuperscript{a}Department of Radiotherapy, \textsuperscript{b}Department of Medical Physics, Great Poland Cancer Centre, Poznan, Poland

Abstract

\textbf{Background and purpose:} To study the accuracy, specificity and sensitivity of the Naïve Bayesian Classifier (NBC) in the assessment of individual risk of cancer relapse or progression after radiotherapy (RT).

\textbf{Materials and methods:} Data of 142 brain tumour patients irradiated from 2000 to 2005 were analyzed. Ninety-six attributes related to disease, patient and treatment were chosen. Attributes in binary form consisted of the training set for NBC learning. NBC calculated an individual conditional probability of being assigned to: relapse or progression (1), or no relapse or progression (0) group. Accuracy, attribute selection and quality of classifier were determined by comparison with actual treatment results, leave-one-out and cross validation methods, respectively.

Clinical setting test utilized data of 35 patients. Treatment results at classification were unknown and were compared with classification results after 3 months.

\textbf{Results:} High classification accuracy (84%), specificity (0.87) and sensitivity (0.80) were achieved, both for classifier training and in progressive clinical evaluation.

\textbf{Conclusions:} NBC is a useful tool to support the assessment of individual risk of relapse or progression in patients diagnosed with brain tumour undergoing RT postoperatively.

© 2007 Elsevier Ireland Ltd. All rights reserved. Radiotherapy and Oncology 86 (2008) 211–216.

Keywords: Education; Artificial Intelligence; Naïve Bayesian Classifier; Radiotherapy; Brain tumours

Artificial Intelligence systems supporting treatment decisions still require careful study and remain a challenge. A simple, user friendly classifier can assist clinician in decision making process.

A decision support system or an expert system is a computer program facilitating problem solving in a given field by drawing inference from a knowledge base developed from human expertise. Such systems may mimic human type of learning and inferring but their additional virtue is an ability to search for rules quickly when the amount of data is considerable. Theoretical foundations of computer expert systems have been known for years now [1,2]. It is also accepted that their broader use may lead to improved patient safety, improved quality of care, and efficiency in health care delivery [3,4]. Despite obvious benefits such systems are not yet widely used in everyday clinical practice [6]. Coiera, looking for possible reasons, pointed to the dependence on an electronic medical record system to supply data, poor human interface design and failure to fit naturally into the routine process of care [7]. Many of these systems are experimental in their nature and often difficult to work with [2,8].

On the other hand an expert system may bring additional value, e.g. in radiotherapy of brain tumours. Many trials [5,9] have demonstrated significant survival advantage for patients who received radiotherapy, which is especially true for high grade gliomas. However, on the other hand radiotherapy may result in deterioration of treatment results due to delayed side effects.

Moreover, despite radiotherapy, a relapse or progression of the disease is observed, especially in glioma but also in other tumours [10]. Therefore, the treatment decision for an individual patient remains a challenge.

Clinical studies on brain tumours provided with pre-treatment patients’ characteristics may be of help in making such decisions but only in high grade glioma cases. RTOG trials identified a set of prognostic factors by applying nonparametric recursive partitioning technique [11].

There is still a wide variety of clinical situations where physician’s decision must be based on accessible data and own experience only. Therefore it would be of value to construct a model helping in an assessment of the individual risk of relapse or progression in an individual allowing for individualized treatment as opposed to the current
Application of the Naïve Bayesian Classifier

Modern patient management should be based on scientific evidence but it is often difficult for a clinician to see patterns in the available data.

Purpose
To study the accuracy, specificity and sensitivity of the Naïve Bayesian Classifier in the assessment of individual risk of cancer relapse or progression after radiotherapy.

Materials and methods
In light of the above, the aim of our work is to create and test an expert system which would combine simplicity of operation, credibility and user friendly interface.

Review of various Artificial Intelligence (AI) methods, most often used for construction of expert systems, was subject of many publications and thus will not be described here [2,8,12]. In this paper we will focus on the Naïve Bayesian Classifier, which in our opinion offers the most of the virtues mentioned above.

Naïve Bayesian Classifier (NBC)
Classifier means all mathematical processes leading to computing the probability of being categorized (classified) to a given class. Due to the relative clarity of data processing NBC enables every clinician to understand how the results of classification are reached. This method allows avoiding “black box” type processes which are not transparent and understandable for the system’s users [6].

Construction of the classifier is based on the Bayes Theorem, discovered in the 18th century by Thomas Bayes and rediscovered a century later by the great French mathematician Laplace (see Fig. 1). Bayes Theorem provides a mathematical method that could be used to calculate, given occurrences in prior trials, the likelihood of a target occurrence in future trials. According to Bayesian logic, the only way to quantify a situation with an uncertain outcome is through determining its probability. Bayes Theorem is a means of quantifying uncertainty. Based on probability theory, the theorem defines a rule for refining a hypothesis by factoring in additional evidence and background information, and leads to a number representing the degree of probability that the hypothesis is true.

Although the formula itself may look simple, in reality it is hard to calculate the probability for more than one attribute without the help of a computer (see Fig. 2). 

NBC is tolerant for noisy and incomplete data. Attributes set can be easily expanded without the need to construct a new model. Such flexibility allows for easy incorporation of new data coming as a result of progress in medical sciences. The only potential drawback of NBC comes from an assumption of independence of attributes — achieving which can be difficult in medical datasets [13,14]. Despite ongoing debate many researchers argue that NBC is relatively robust and that management of attribute independence is not an important problem [13].

All calculations in this study were made with Naïve Bayesian Classifier implemented in WEKA environment (http://www.cs.waikato.ac.nz/ml/weka/). WEKA is a collection of machine learning algorithms for solving real-world data mining problems. It is written in Java and runs on almost any platform, enables testing databases applying different AI systems.

Patients and attributes
Retrospective data of 142 patients diagnosed with brain tumour, treated in our centre from January the 1st 2000 to December 31st 2005, were analyzed. The spectrum of actually diagnosed tumour types ranged from meningiomas, with good survival prognosis, to glioblastomas, characterized by worse outcomes.

Such a diverse patient group allows the system to be tested in different clinical situations.

Such a group is homogeneous enough to be characterized by common attributes such as location of tumour, symptoms, diagnosis, treatment and supportive care used. Moreover, relapse of tumour can be easily recognized in a widely available radiological test such as magnetic resonance imaging (MRI) or computed tomography (CT).

Treatment outcome was evaluated 3 months after the completion of radiotherapy. All patient cases considered were very well documented. Sixteen patients were excluded from the analysis due to incomplete data.

All patients underwent surgery. Complete resection, subtotal resection and biopsy only was performed in 64 (45%), 66 (46.5%) and 12 (8.5%) patients, respectively. All patients were irradiated postoperatively, with radical or palliative intention.

Some patients did not complete radiotherapy, which resulted in low dose delivered in a few cases. Six and 9 MV photons and gamma rays were used. The use of Cobalt and photons 9 MV has been stopped in favour of using 6 MV. At the end of December 2005 the retrospective database was closed and collection of prospective data started to create the patient test set to prove the classifier in practice. In the meantime, new techniques of irradiation were implemented (e.g. IMRT or stereotactic irradiation) and the system was updated to reflect these advancements. In this study patients with relapse and progression comprised

\[
P(H \mid d) = \frac{P(d \mid H) \times PH}{Pd}
\]

Fig. 1. Bayes–Laplace Theorem. \(P(H|d)\) — posterior probability; \(P(d|H)\) — likelihood of evidence given the hypothesis; \(PH\) — prior probability of the hypothesis; \(Pd\)-serving normalization role.

\[
p(h \mid d) = \frac{p(d \mid h)p(h)}{\sum_{h' \in H} p(d \mid h')p(h')}
\]

Fig. 2. Bayes formula to compute posterior probability with more than one attribute. \(p(h|d)\) — posterior probability; \(p(d|h)\) — likelihood of evidence given the hypothesis; \(p(h)\) — prior probability of the hypothesis; \(Ep(d|h')p(h')\) — sum over spaces of hypotheses.
class 1, and patients with no relapse or stabilization comprised class 0.

Selection of relevant attributes is one of the most important steps in classifier construction.

For the purpose of this study 96 attributes were selected based on a literature search [5,9–11,15–25] and authors’ expertise. Continuous attributes were discretized, following accepted norms or arbitrary ranges; e.g. for the attribute “age” 10-year ranges were applied [26].

Every patient was described by a set of attributes, presented in binary (0 or 1) form.

Attributes selected can be categorized into the following groups:

1. Histology (histological type of tumour).
2. Cranial imaging.
3. Location of primary tumour.
7. Coexistent diseases.
8. Laboratory findings.
9. Supportive care.

The following section describes patient characteristics arranged by attribute groups.

**Histology**

The majority of patients were diagnosed with high grade astrocytoma. Patients with meningiomas and low grade astrocytoma were irradiated in case of recurrence after surgery, or in presence of neurological symptoms, when surgery was impossible. Rare primary brain tumours, such as ependymoma, were not included into the dataset because there were only two such cases with highly incomplete data.

**Radiological findings**

Radiological features were evaluated by a radiologist in MRI or CT. Tumour size was assessed according to preoperative images. High contrasting of tumour in CT or MRI is a feature of its higher malignancy [25] and presence of a cyst in the tumour can be a prognostic factor for better outcome [22].

**Location of tumour**

Locations presented include all location patterns discovered in the database. When recognition of a tumour’s primary location was impossible it was attributed to two lobes.

**Symptoms before treatment**

Attributes describing neurological symptoms were chosen based on a literature search [9] and interviews with patients and relatives.

**Surgical treatment**

Most patients underwent tumour resection. Only 12 of them underwent biopsy due to location of tumour not allowing for even partial removal.

Some patients, especially with meningioma, were irradiated after removal of first or next recurrence.

**Radiotherapy**

The use of Cobalt and photons 9 MV has been stopped in favour of using 6 MV.

Daily doses of 1.8–2.5 Gy were dedicated to the treatment with radical intention whereas doses of 3 and 4 Gy were used only for palliative treatment. Total dose less than 40 Gy was used in palliative treatment. Some of the patients did not complete the full planned dose due to worsening of general status.

Symptoms occurring during irradiation such as seizures and severe headache, as well as episodes of fever over 38 °C without any proof of infection, have also been included in the evaluation.

During the test set data collection period new modes of treatment were launched in our centre including IMRT and stereotactic boost. One patient was treated with temozolomide concurrently with radiotherapy. All other patients in the training set were treated using 3DCRT treatment plans.

It was necessary to update the classifier with those new attributes, and include them in calculation of probabilities.

Different periods of time between surgery and radiotherapy were caused by multiple factors such as surgery complications, the time of wound healing, patient’s general condition and radiotherapy department capacity.

**Coexistent diseases**

Diabetes and hypertension, conditions with potential to influence outcome, were the most frequent coexistent diseases.

**Laboratory test results at the time of radiotherapy start**

Discretization of haematological attributes was performed according to standard ranges of haemoglobin (Hb) concentration and leukocyte count, used in our centre: normal range Leukocytes 4.6–10.2 G/L, Hb 7.6–10.0 mmol/L.

**Supportive care**

Mannitol and steroids were the most common drugs used in treatment of cerebral oedema during radiotherapy. Mannitol administration for more than 7 days was avoided, but was necessary in some cases.

Steroids such as dexamethasone were administered most often in doses ranging from 4 to 8 mg. Higher doses were necessary in some cases. All patients with paresis were rehabilitated.

**Patient**

Attributes reflect the patient’s general status on the Karnofski Performance Status Scale, social information about an individual patient. The year of treatment can have a potential impact on its outcome due to the availability of new treatment facilitates, treatment planning system or better positioning of patients.

This set of attributes does not include results of genetic testing, since it is not yet routine practice in our centre. The classifier however allows for easy expansion of the attribute set to include these valuable information as soon as they are available.
Records of 142 patients consisted of the training set for NBC learning.

Evaluation of classifier – training set

To evaluate classifier’s quality we performed an assessment of: classification accuracy, importance of attributes selected, specificity and sensitivity.

Accuracy of classification

Classification accuracy was determined by cross validation method [26]. In this method a record or group of records (ideally in all possible combinations) is eliminated from the training set and then the classifier is re-trained in this reduced training set. Excluded records were subsequently presented to the classifier as new, unseen cases and were subject to classification. This process was repeated as many times as many records were in the training set. We believe that such procedure is the most appropriate for medical records which contain data of individual patients.

Cross validation allows determining whether the classifier is able to generalize the knowledge, looking for general rules and applying them to unknown records (cases). This also protects the system from “learning by heart” or “over fitting”, i.e. looking for identical cases rather than general rules.

Evaluation of attribute selection

Accuracy of learning algorithms decreases if the set of data contains irrelevant attributes [27,28]. Initially relevance of attributes was assessed by leave-one-out method — elimination of one attribute at time from the training set, followed by recalculating of probabilities for the dataset reduced by eliminated attributes. Reduced classification accuracy following elimination of an attribute from the set reflected its relevance (importance) in an assessment of risk relapse. Further optimization of the classifier was performed with use of eight different evaluators implemented in WEKA environment. These evaluators give the opportunity of choosing a smaller subset from all attributes and determining the relevance of the attributes by searching in all possible subsets of attributes and finding the one subset that works best for classifying. These methods also allow to show the most relevant attributes and their relevance rank. While assessing different evaluators we were looking for the one providing the highest combined accuracy, specificity and sensitivity.

The test set

After evaluation of the classifier using retrospective data, we tested the system with the prospective test dataset. The test set consisted of the data from 35 patients who had completed treatment and by the time the data collection was closed remained under follow-up. Treatment results at the time of classification were unknown. After 3 months actual treatment outcome was compared with classification results, and records of these patients were included in the training set — enriching the database. After adding the data of 35 new patients to the training set parameters of classifiers did not change. High accuracy, specificity and sensitivity were achieved not only for retrospective but also for prospective data.

Result

High classification accuracy was achieved, both for training and test sets. Initial classification accuracy was 82%. This means 116 accurately classified patients, and 26 wrongly classified patients out of 142 cases. Sensitivity of the system including the test set was 0.79 and specificity 0.83.

After optimization with WEKA built-in evaluators accuracy increased to 84% (119 accurately and 23 wrongly classified patients) with sensitivity 0.80 and specificity 0.87. These best results were obtained with evaluator named ReliefFAttributeEval and for the subset of 43 attributes, the highest ranked of them falling into histology, radiotherapy, surgical treatment and general status attributes groups.

Discussion

Our aim was to construct a decision support system combining high performance accuracy with simplicity of use which will allow for wide application.

During the course of our research we have also used other systems in order to compare results and look for potentially the best system to suit our needs. We have performed tests using: decision trees and Bayesian Net — which appear to be one of the most popular systems used in medicine [29,30]. Detailed description of the results would go beyond the scope of this paper; however, key findings are presented in Table 1. After initial assessment we discontinued further studies with other systems. Classifiers like decision trees or neural networks would require clinician to put much more effort in working with the system. It is necessary to “prune” decision trees to eliminate unreal rules generated by the tree. Expanding trees with new attributes is also work and time consuming. Similarly, using neural networks in daily clinical context might be more troublesome. Evaluation of the importance of attributes in neural network is not possible, what makes optimization and generalization process difficult. Thus we focused on NBC for its best accuracy, sensitivity, specificity and simplicity in everyday maintenance.

One of the most important potential limitations of NBC use was the necessary assumption of independence of attributes. However, as noted by Jakulin et al. [31] although in practice this assumption is not quite true, experience shows that the NBC approach in medical applications is effective.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Results of evaluation for some systems often used for medical purposes</th>
</tr>
</thead>
<tbody>
<tr>
<td>System</td>
<td>Accuracy</td>
</tr>
<tr>
<td>NBC</td>
<td>82%</td>
</tr>
<tr>
<td>Bayesian Net</td>
<td>80%</td>
</tr>
<tr>
<td>Decision trees</td>
<td>71%</td>
</tr>
</tbody>
</table>
and gives relatively good classification accuracy in comparison with other, more elaborate learning methods.

Eisenstain [13] points out that one group of researchers has stated that the problems encountered in managing conditional dependence have been an impediment to the acceptance of Bayesian analysis. In contrast, other researchers have asserted that the simple Bayes model is relatively robust and that the management of conditional dependence is not an important problem. The high accuracy of our model seems to confirm the later conclusion.

Unlike more complex systems, NBC allows for easy inclusion of additional attributes. Ability to add attributes in an easy way is a great benefit when operating in fast changing fields like medicine, where new diagnostic and treatment tools are constantly introduced. Since the system’s learning process mimics the human one, adding new attributes or data unknown to the system may result in temporary deterioration of classification accuracy. We have observed this phenomenon too, but it was transient and results improved with increasing number of sample cases including new attributes. This potential disadvantage causing periodical fluctuations in the system’s accuracy might be overcome by multicentre co-operation. This would allow for gathering relatively quickly an amount of samples sufficient to maintain constant and high level of classification accuracy. On the other hand NBC is relatively robust as compared to other systems, which means resistant to missing data. In clinical practice missing data is a common problem and a challenge to many research projects, and thus system robustness should be a sought-after feature.

One of the reasons for limited use of expert systems is dependence on an electronic medical record system to supply their data. While it remains true for NBC, this system does not bring additional constraints in this respect and data collection is possible without costly systems, provided it is done on a regular basis and the amount of data is not overwhelming. The system designed for the purpose of this work can be managed by a single person without devoting a significant amount of additional time even if it is not possible to connect the system to a general database for easy data retrieval.

The Naive Bayesian Classifier proved to be an accurate and simple-to-use tool. Using NBC does not require extensive training and the system is ready to be expanded with new data. We have developed a simple and reliable system able to provide help in everyday work and not requiring extensive training prior to its use.

NBC can be applied to any patient and disease group, provided we are able to select the right attributes and create a reliable database. Using this expert system would allow for the assessment of risk of failure in the context of an individual set of attributes of a single patient and his or her particular diagnosis and status. So far, only general risk factors determined by evaluation of large patient groups using classical statistical methods have been used. In contrast NBC allows for an individualized approach. Also it could be of interest to assess how results of classification are influenced by different modes of treatment, especially in clinically difficult, borderline cases. It may allow the physician to simulate the results of different treatment procedures (classification) and to apply the best procedure in a particular case. Classification result could offer valuable tips for the decision making process in such difficult cases. Such a system will also allow better information to be offered to patients and their families about their possible treatment outcome and prognosis.

Every system designed for daily use in a clinic must be user friendly, not time consuming, with clear and understandable rules of operation. Only such a system would be accepted by clinicians and widely used in hospitals. Having in mind these conditions, NBC seems to be useful and deserves further studies.

Conclusions

The verification of the studied and trained Naive Bayesian Classifier to assess individual risk of patient relapse after radiotherapy of brain gliomas revealed accuracy of 84%, specificity 0.87 and sensitivity 0.80. These values showed the classifier’s credibility and thus its potential in supporting individual treatment decisions.

Technical requirements were simple, entering the data required only basic computer skills and was not time consuming. Connecting the system to hospital database and retrieving from there routinely collected data would allow for further reducing the amount of work related to system maintenance.

Results achieved are very encouraging and justify further development of NBC systems as valuable tool for supporting everyday clinical decisions.

Acknowledgement

The authors would like to thank Dr Michal Kazmierski for his support in preparing the manuscript.

*Corresponding author. Joanna Kazmierska, Radiotherapy Department, Great Poland Cancer Centre, ul. Garbary 15, 61-688 Poznan, Poland. E-mail address: joanna.kazmierska@wco.pl

Received 1 July 2007; received in revised form 8 October 2007; accepted 11 October 2007; Available online 26 November 2007

References


