Research on clinical decision support systems development for atrophic gastritis screening

Sergei Parshutin*1, Arnis Kirshners

Abstract

The paper presents a pilot research on the application of clinical decision support systems in a atrophic gastritis screening task. Two different DSS learning strategies have been tested – a standalone classifier and classifier ensemble application. Such classification algorithms as C4.5, CART, JRip and Naïve Bayes were used as base classifiers. The classifiers were evaluated on the respondent medical data from an inquiry form, containing 28 attributes and 840 records. The dataset was preprocessed using simple methods in initial data analysis as well as more complex data mining methods for feature selection. The obtained results are summarized and discussed in order to summarize an information on what learning strategies are more applicable to the present dataset and should be studied in more detail in primary research.

1. Introduction

Cancer is the worldwide problem in social health and one of the leading causes of death. Nevertheless it is known that most of cancer types are treatable. Referencing the World Health Organization data, at least 40% of all local cancer types are treatable and can be prevented, avoiding the risk factors, common not only for cancer, but also for the most chronic diseases. These risk factors are well known and the most important of them are smoking, alcohol and other pernicious habits, activity shortage, adiposis (excessive weight) and different infectious agents. New medical technologies, new medicaments, vaccines, screening systems are continuously developed and introduced, all aimed at the identification and treatment of cancer at initial stages and the improvement of life quality and life length for patients with cancer.

Even though globally the gastric cancer incidence is declining and in many Western countries the disease is not considered among the major health issues any more, globally the cancer of the stomach is still continuing to be an important healthcare problem. Gastric cancer is remaining the second leading cause of mortality worldwide within the group of malignant diseases after the lung cancer, and is accounting for almost 10% of cancer related deaths. Among men gastric cancer is the second (after lung cancer), but among women – the third leading (after breast and lung) cause of cancer-related deaths (Su et al., 2007; WHO, 2013).

Gastric cancer is a very challenging malignancy given that it presents late, has complex pathogenetic mechanisms with multiple carcinogenic processes implicated, and is only moderately sensitive to chemotherapy and radiation. Gastric cancer presents mostly in an advanced stage and is lethal unless diagnosed early (Crew & Neugut, 2006; Miranda, Abelha, Santos, Machado, & Neves, 2009; Varadhachary & Ajani, 2005).

The present paper discusses a possibility of application of CDSS – Clinical Decision Support Systems in order to give an expert additional information on probable disease; an atrophic gastritis in our case. Section 2 gives a look into CDSS, defines the main objectives of the system and reveals methods used in the pilot research. Section 3 presents the system evaluation results, which are summarized and discussed in Section 4.

2. Clinical decision support system

Clinical decision support systems (CDSS) are computer systems designed to support clinician decision making about specific patients at the point of time these decisions are made. Decision support systems have been incorporated in healthcare information systems for a long time, but these systems have usually supported retrospective analyses of both administrative and clinical data (Intarajak & Kang, 2009; Sauter, 2011). Recently, sophisticated data mining approaches have been proposed for similar retrospective analyses of both administrative and clinical data (Intarajak & Kang, 2009; Sauter, 2011), bringing new possibilities in designing efficient clinical decision support systems.

Barner and La Lande in Sauter (2011) propose a simple CDSS classification as knowledge-based and nonknowledge-based systems.
2.1. Concepts and objectives

As our pilot research focuses on a specific disease – atrophic gastritis (AG), Bagging (bootstrap aggregating) is a technique that repeatedly creates subsets of training data in order to improve the accuracy of the classiﬁcation. This classiﬁcation is still applicable, but it should be pointed out that modern data mining approaches provide a wide range of possibilities to generate rules using ANN and GA (Galina & Parshutin, 2011; Quteishat & Lim, 2007). With respect to this classiﬁcation the knowledge-based CDSS has been chosen for the present pilot research.

2.2. Methods

Medical informatics research has employed traditional statistical methods, such as logistic regression and discriminant analysis, support vector machines, as well as learning methods, such as decision trees, neural networks and case-based reasoning (Chun, Kim, Haehn, Park, & Chun, 2008; Flouri & Duffy, 2006; Miranda et al., 2009; Lee, Lin, & Lee, 2006; Su et al., 2007; Vercellis, 2009). These methods are used individually or combined with other data mining methods.

There are two learning strategies compared in the present research. The ﬁrst one consists in building a base classiﬁer using a single classiﬁcation algorithm. The most popular classiﬁcation algorithms that are widely used are decision trees – C4.5 and CART; rule induction algorithms based on a Ripper algorithm and probabilistic algorithms as NB. Each of these algorithms will be evaluated and compared with others. An output of this learning strategy will be the prediction on the most probable disease for a speciﬁc descriptive data. An additional information can be obtained if the binary classiﬁer predicts class j and Lij if the binary function, returning 1 if a classiﬁer predicts class j and 0 otherwise. The weight wij is linked to the class j because such coefﬁcients as false negative rate differ for each class. In case if weight coefﬁcient does not differ upon the class, such as classiﬁcation accuracy, then Eq. (2) can be used.

\[
S_i = \frac{\sum_{j=1}^{k} w_{ij} \cdot L_{ij}}{\sum_{i=1}^{k} w_{ij}}
\]

The ﬁnal predicted class \( C \) will have the best score \( S' \), which is chosen using Eq. (3).

\[
S' = \begin{cases} \arg\max_{C} S_i, & w \rightarrow \max \\ \arg\min_{C} S_i, & w \rightarrow \min \\ \end{cases}
\]

3. System evaluation

This section provides an information about the dataset used in experiments, describes the data preprocessing steps taken to prepare data (Section 3.1), gives a look into the experiments performed to evaluate the learning strategies (see Section 2.2) and shows the obtained results (Section 3.2).

3.1. Data description and preprocessing

Learning strategies were evaluated using respondent medical data, who ﬁlled the inquiry form. The initial dataset contained 28 descriptive attributes and a binary target attribute denoting positive or negative atrophic gastritis diagnosis. The target attribute values were obtained using the golden standard histological analysis of a respondent tissue examples. In total, the initial dataset contained 840 records, having almost equal class proportions – 430 (51%) positive records and 410 (49%) negative records. Fig. 1 shows the absolute proportions for male and female patients.
دریافت فوری متن کامل مقاله

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