

Data Mining Applications in a Medical System: A Case Study

Morteza Bagherpour

Department of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran

Asma Erjaee

Pediatrics Department, Shiraz University of Medical Sciences, Shiraz, Iran

Amir Hossein Rasekh

Department of Computer and Electrical Engineering, Shiraz University, Shiraz, Iran

Seyed Mohsen Dehghani

Pediatrics Department, Shiraz University of Medical Sciences, Shiraz, Iran

INTRODUCTION

It is well known that helicobacter pylori (*H.pylori*) is a major cause of chronic active gastritis in both children and adults (Drumm, 1993). This infection remains mainly asymptomatic in children, however when it does become symptomatic it may present as gastrointestinal symptoms such as epigastric pain, heartburn, or dyspepsia like manifestations (Giacomo et al., 2002). According to guidelines published on diagnosis and treatment of *H. pylori* infection in children, interventional testing should be assumed in order to diagnose the cause of the gastrointestinal symptoms and not merely to determine the presence of *H.pylori* infection (Gold et al., 2000 and Bourke et al., 2005). However testing for *H.pylori* infection should be considered in patients with a positive family history of gastric cancer, those with refractory iron deficiency anemia, and before long term therapy with proton pump inhibitors (Bourke et al., 2005). There are a variety of tests for detection of *H. pylori* infection which can be classified as invasive vs. noninvasive.

In the invasive test a gastric specimen is obtained through endoscopy and further used for

culture, histopathology, PCR, and rapid urease test (RUT). Noninvasive tests include detection of *H.pylori* antigens in stool, detection of antibodies against *H.pylori* in serum, and urea breath test (Guarner et al., 2010). The best test for detection of *H.pylori* is one which is available, minimally invasive, greatly accurate, and inexpensive (Gold et al., 2000). However worldwide as it has been recommended by guidelines on diagnosis of *H. pylori* infection, endoscopy and biopsy are still the gold standard for diagnosis of this infection (Gold et al., 2000 and Bourke et al., 2005 and Guarner et al., 2010). Though, as it is known endoscopy is an invasive and unpleasant procedure, and also expensive.

Indeed in medicine, the only way to diagnose the existence of *Helicobacter Pylori* microbe is doing Endoscope but it's painful and insufferable for young children. To solve this problem, in this paper, some classifiers are employed to diagnose the existence of this infection.

Here we have proposed a data mining modeling, with a high sensitivity, as a noninvasive route for estimation of the possibility of *H.pylori* infection in children, according to the patient's history, gastrointestinal sign and symptoms. The

suggested approach will be cost-effective, easy and noninvasive, and will eliminate the need for endoscopy as an invasive test for diagnosis of *H.pylori* infection in children.

DATA COLLECTION

The patient related data was gathered through a randomized clinical trial study, where all children <18 year of age with possibility of *H. pylori* infection; according to their sign and symptoms, whom had referred to the Gastrointestinal clinic afflicted to Shiraz University of Medical Sciences from April 2011 till September 2011 were enrolled. First a questionnaire form was completed for each patient, including questions regarding the patient's symptoms (e.g. abdominal pain, nausea, vomiting, halitosis, GI bleeding ...) there duration, positive history of treatment with antacids (H₂ blockers and proton pump inhibitors), and any positive family history of acid peptic diseases in their first degree relatives. Also all patients were examined for tenderness in there epigastric area and if so this was entered in the form. The patient's weight and height were as well recorded in the questionnaire form. Questions regarding symptoms which could be possibly correlated to *H.pylori* infection in children were derived from previous studies on this concept (Drumm, 1993; Giacomo et al., 2002; Gold et al., 2000).

Further an endoscopy was performed for all subjects, through which an antral and corpus mucosal biopsy was obtained for histopathology and RUT. Biopsy specimens for histology were fixed in formalin and were sent to Shahid Motahari Pathology Laboratory of Shiraz University of Medical Sciences for analysis. Results regarding the histopathology and RUT were also entered in the form.

An informed consent was obtained from parents of all patients. The used dataset is collected through a six months period from those patients who need to perform the Endoscopy in order to diagnose the existence of *H.pylori* infection. The features of dataset are:

Male or Female, abdominal pain, Nocturnal awakening, Nausea, Vomiting, Halitosis, Heart Burn, Bloating, Belching, GI bleeding, Constipation, Diarrhea, Weight loss, Fatigue, Epigastric tenderness, Weight, height, Duration of symptoms, Previous treatment, Previous Endoscopy, Previous family H Acid peptic Dx, Rapid Urease test before therapy.

D

THE PROPOSED APPROCH

Here the proposed approach is presented by applying three data mining models. The models are then compared, to select the best alternative for prediction of *H.pylori* infection.

Supervised learning is the machine learning task for inferring a function from labeled training data. The training data consist of a set of training examples. In the supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. In this paper we used supervise learning algorithms (classification algorithm) such as Decision Tree, Naïve Bayes and Logistic Regression. Our Proposed in this paper predict are between two class types.

Bayesian Learning Method

Our Proposed approach in this paper predicts just two modes (Seen and Not Seen) that one of the best algorithms for this work is Naïve Bayes algorithm.

Bayesian learning methods (such as simple Bayesian classification) is able to provide useful practical solutions and forecasting features toward solving complicated problems. Bayesian methods offer hypotheses that are able to predict the most likely conditions (Lucas, 2001). Current applications can be classified with the combined weight of several hypotheses. Even in cases where Bayesian methods cannot be calculated, they can

5 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/data-mining-applications-in-a-medical-system/107263

Related Content

Skills and Competencies Required for Jobs in Business Analytics: A Content Analysis of Job Advertisements Using Text Mining

Linda A. Leon, Kala Chand Seal, Zbigniew H. Przasnyski and Ian Wiedenman (2017). *International Journal of Business Intelligence Research* (pp. 1-25).

www.irma-international.org/article/skills-and-competencies-required-for-jobs-in-business-analytics/182762

The Role of Analytics and Robo-Advisory in Investors' Financial Decisions and Risk Management: Review of Literature Post-Global Financial Crisis

Abhinav Pal, Shalini Singh Sharma and Kriti Priya Gupta (2021). *International Journal of Business Analytics* (pp. 46-62).

www.irma-international.org/article/the-role-of-analytics-and-robo-advisory-in-investors-financial-decisions-and-risk-management/276446

Big Data Mining and Analytics

Carson Kai-Sang Leung (2014). *Encyclopedia of Business Analytics and Optimization* (pp. 328-337).

www.irma-international.org/chapter/big-data-mining-and-analytics/107238

Business Intelligence in the Bayou: Recovering Costs in the Wake of Hurricane Katrina

Gregory Smith, Thilini Ariyachandra and Mark Frolick (2012). *Organizational Applications of Business Intelligence Management: Emerging Trends* (pp. 29-37).

www.irma-international.org/chapter/business-intelligence-bayou/63964

Analytics for Smarter Buildings

Young M. Lee, Lianjun An, Fei Liu, Raya Horesh, Young Tae Chae and Rui Zhang (2014). *International Journal of Business Analytics* (pp. 1-15).

www.irma-international.org/article/analytics-for-smarter-buildings/107066