Preface

Data mining is one of the technologies called to improve the quality of service in clinical medicine through the intelligent analysis of biomedical information. From the enunciation of evidence-based medicine in early 1990s [1], the need for creating evidence that could be quickly transferred to physician daily practice is one of the most important challenges in medicine. The use of statistics to prove the validity of the treatment over discrete populations; the creation of predictive models for diagnosis, prognosis, and treatment; and the inference of clinical guidelines as decision trees or workflows from instances of healthcare protocols are examples of how data mining can help in the application of Evidence Based Medicine.

The great interest that emerges from the use of data mining techniques has caused that there was a large amount of data mining books and papers available in literature. The majority of techniques or methodologies that are available for use are published and can be studied by clinical scientist around the world. However, despite the great penetration of those techniques in literature, their application to real daily practice is far to be complete. For that, when we were planning this book, our vision was not just to compile a set of data mining techniques, but also to document the deployment of advance solutions based on data mining in real biomedical scenarios, new approaches, and trends.

We have divided the book into three different parts. The first part deals with innovative data mining techniques with direct application to biomedical data problems; in the second part we selected works talking about the use of the Internet in data mining as well as how to use distributed data for making better model inferences. In the last part of the book, we made a selection of new applications of data mining techniques.

In Chapter 1, Fuster-Garcia et al. describe the automatic actigraphy pattern analysis for outpatient monitoring that has been incorporated in the Help4Mood EU project for helping people with major depression recover in their own home. The system allows the reduction of inherent complexity of the acquired data, the extraction of the most informative features, and the interpretation of the patient state based on the monitoring. For this, their proposal covers the main steps needed to analyze outpatient daily actigraphy patterns for outpatient monitoring: data acquisition, data pre-processing and quantification, non-lineal registration, feature extraction, anomaly detection, and visualization of the information extracted. Moreover, their study proposes several modeling and simulation techniques useful for experimental research or for testing new algorithms in actigraphy pattern analysis. The evaluation with actigraphy signals from 16 participants including controls and patients that have recovered from major depression demonstrates the utility to visually analyze the activity of the individuals and study their behavioral trends.

Biomedical classification problems are usually represented by imbalanced datasets. The performance of the classification models is usually measured by means of the empirical error or misclassification rate. Nevertheless, neither those loss functions nor the empirical error are adequate for learning from imbalanced data. In Chapter 2, Garcia-Gomez and Tortajada define the loss function of LBER whose associated empirical risk is equal to the balanced

error rate (BER). In these problems, the empirical error is uninformative about the performance of the classifier and the loss functions usually produce models that are shifted to the majority class. The results obtained in simulated and real biomedical data show that classifiers based on the LBER loss function are optimal in terms of the BER evaluation metric. Furthermore, the boundaries of the classifiers were invariant to the imbalance ratio of the training dataset. The LBER-based models outperformed the 0–1-based models and other algorithms for imbalanced data in terms of BER, regardless of the prevalence of the positive class. Finally, the authors demonstrate the equivalence of the loss function to the method of inverted prior probabilities, and generalize the loss function to any combination of error rates by class. Big data analysis applied to biomedical problems may benefit from this development due to the imbalance nature of most of the interesting problems to solve, such as predictive of adverse events, diagnosis, and prognosis classification.

In Chapter 3, Vicente presents a novel online method to audit predictive models using a Bayesian perspective. This audit method is specially designed for the continuous evaluation of the performance of clinical decision support systems deployed in real clinical environments. The method calculates the posterior odds of a model through the composition of a prior odds, a static odds, and a dynamic odds. These three components constitute the relevant information about the behavior of the model to evaluate if it is working correctly. The prior odds incorporates the similarity of the cases of the real scenario and the samples used to train the predictive model. The static odds is the performance reported by the designers of the predictive model and the dynamic odds is the performance evaluated with the cases seen by the model after deployment. The author reports the efficacy of the method to audit classifiers of brain tumor diagnosis with magnetic resonance spectroscopy (MRS). This method may help on assuring the best performance of the predictive models during their continuous usage in clinical practice.

What to do when we obtain underperformed expectations of the predictive models during their real use of predictive models? Tortajada et al. in Chapter 4 propose an incremental learning algorithm for logistic regression based on the Bayesian inference approach that may allow to update predictive models incrementally when new data are collected or even to perform a new calibration of a model from different centers. The performance of their algorithm is demonstrated by employing different benchmark datasets and a real brain tumor dataset. Moreover, they compare its performance to a previous incremental algorithm and a non-incremental Bayesian model, showing that the algorithm is independent of the data model and iterative, and it has a good convergence. The combination of audit models, such as the proposal from Vicente, with incremental learning algorithms, such as that proposed by Tortajada et al., may help on the assurement of the performance of clinical decision support systems during their continuous usage in clinical practice.

New trends like interactive pattern recognition [2] aim at the creation of human understandable data mining models allowing them the correction of the models to make a direct use of data mining techniques as well as facilitate its continuous optimization. In Chapter 5 new possibilities about the use of process mining techniques in clinical medicine are presented. Process mining is a paradigm that comes from the process management research field and that provides a framework that allows to infer the care processes that are being executed in human understandable workflows. These technologies allow experts in the understanding of the care process, and the evaluation of how the process deployment affects the quality of service to the patient.

Chapter 6 analyzes the patient history from a temporal perspective. Usually data mining techniques are seen from a static perspective and represent the status of the patient in a specific moment. Using temporal data mining techniques presented in this chapter it is possible to represent the dynamic behavior of the patient status in an easy human understandable way.

One of the worst problems that affect data mining techniques for creating valid models is the lack of data. Issues as the difficulty for achieve specific cases and the data protection regulations are barriers for enabling a common sharing of data that can be used for inferring better models that can be used for a better understanding of the illnesses and for improving the cares to final patients. Chapter 7 presents a model to allow feed data mining system from different distributed databases allowing them in the creation of better models using more available data.

Nowadays, the greatest data source is the Internet. The omnipresence of the Internet in our lives has changed our communication channels and medicine is not an exception. New trends use the Internet to explore new kind of diagnoses and treatment models that are patient centered covering them in a holistic way. From the arrival of web 2.0 human cybercitizens use the net not only to get information, but also, Internet is continuously feeding about us. For that, there is a great amount of information available about single humans. Usually cyberhumans write in the Internet its sentiments and desires. Using data mining technologies with this information it will be possible to prevent psychological disorders providing new ways to diagnosis and treat this using the Net [5]. Chapter 8 presents new trends of using sentiment analysis technologies over the Internet.

As we have pointed previously, Internet is used for gathering information. But, not only patients use the Internet to gather information about their and their relatives' health status [4], but also junior doctors trust in the Internet for being continuously informed [3]. However, their universality makes Internet not always trustable. It is necessary to create mechanism to filter trustable information to avoid misunderstandings in patient information. Chapter 9 presents the concept of health recommender systems that use data mining techniques for support patients and doctors for finding trustable health data over the Internet.

However, Internet is not only for persons, but also for systems and applications. New trends, as Cloud Computing, see Internet as a universal platform to host smart applications and platforms for continuous monitoring on patients in a ubiquitous way. Chapter 10 presents an m-health context aware model based on Cloud Computing technologies.

Finally, we end the book with four chapters dealing with applications of data mining technologies: Chapter 11 presents an innovative use of classical speech recognition techniques to detect Alzheimer disease on elderly people; Chapter 12 shows how data mining techniques can be used for detecting cancer in early stages; Chapter 13 presents the use of data mining for inferring individualized metabolic models for controlling chronic diabetic patients; Chapter 14 shows a selection of innovative techniques for cardiac analysis in detecting arrhythmias. Chapter 15 presents a knowledge-based system for empower diabetic patients and Chapter 16 presents how serious games can help in the detection of specific elderly people.

We hope that the reader find our compilation work interesting. Enjoy it!

Valencia, Spain

Carlos Fernandez-Llatas
Juan Miguel García-Gómez

References

1. Davidoff F, Haynes B, Sackett D, Smith R (1995) Evidence based medicine. BMJ 310(6987): 10851086. doi:10.1136/bmj.310.6987.1085. http://www.bmj.com/content/310/6987/1085.short
2. Fernndez-Llatas C, Meneu T, Traver V, Benedi JM (2013) Applying evidence-based medicine in telehealth: an interactive pattern recognition approximation. Int J Environ Res Public Health 10(11):5671–5682. doi:10.3390/ijerph10115671. http://www.mdpi.com/1660-4601/10/11/5671
3. Hughes B, Joshi I, Lemonde H, Wareham J (2009) Junior physician's use of web 2.0 for information seeking and medical education: a qualitative study. Int J Med Inform 78(10):645–655. doi:10.1016/j.ijmedinf.2009.04.008. PMID: 19501017
4. Khoo K, Bolt P, Babl FE, Jury S, Goldman RD (2008) Health information seeking by parents in the internet age. J Paediatr Child Health 44(7–8):419–423. doi:10.1111/j.1440-1754.2008.01322.x. PMID: 18564080
5. van Uden-Kraan CF, Drossaert CHC, Taal E, Seydel ER, van de Laar, MAFJ (2009) Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 74(1):61–69. doi:10.1016/j.pec.2008.07.044. PMID: 18778909

http://www.springer.com/978-1-4939-1984-0

Data Mining in Clinical Medicine
Llatas, C.F.; García-Gómez, J.M. (Eds.)
2015, XII, 270 p. 92 illus., 82 illus. in color., Hardcover
ISBN: 978-1-4939-1984-0
A product of Humana Press