2012 4th International Conference on Bioinformatics and Biomedical Technology
IPCBEE vol.29 (2012) © (2012) IACSIT Press, Singapore

Improving Efficiency of Telemedical Prevention Programs through Data-mining on Diagnostic Data

Petr Nálevka [1] , Vojtěch Svátek [1]

[1] University of Economics, Department of Information and Knowledge Engineering, Prague, Czech republic

Abstract. This article proposes application of data-mining methods on historic diagnostic data collected in telemedical prevention programs. The described techniques allow to improve efficiency of such programs by improving prediction accuracy and balancing trade-off between sensitivity and specificity which typically has a different cost in telemedicine. The proposed methods has been successfully applied to a state of the art telemedical program for schizophrenia prevention - the ITAREPS.

Keywords: Data-mining, Telemedicine, Temporal data, Sensitivity, Specificity, ROC curve, Schizophrenia prevention

1. Introduction

With growing cost of medical treatment, different telemedical prevention programs become more and more important. Automation in telemedicine can lower costs, required human resources and also increase speed in patient-doctor communication [1][2].

On the other hand, telemedicine brings new challenges especially for information and communication technology. The diagnostic methods or devices for in-home use operated by the patients may highly differ from in-person clinical diagnostics, so that traditional guidelines and best practices may be questioned.

Moreover, telemedical programs, tent to produce far greater amount of diagnostic data[1] than in-person diagnostics. Data-mining analysis performed on such data may not only reveal new information on the diagnosed diseases but also further optimize settings of the applied diagnostic tests.

Those are the reasons why this article proposes the application of data-mining methods to historic diagnostic data produced by telemedical programs in order to improve their efficiency.

2. Telemedical Prevention

Telemedical prevention programs are part of the *remote monitoring* category of telemedicine [3]. Typically various in-home diagnostic devices or patient's self-evaluations are communicated remotely.

On the remote servers different diagnostic streams are aggregated and analysed in order to detect warning trends in patients' development to allow timely intervention[2] and thus prevent further serious condition worsening. The following figure depicts the process of a typical telemedical prevention program.

[1] Some programs may for example produce a continuous stream of diagnostic data with continuous measurement of patient condition.

[2] Typically medication increase/decrease, life-style advices, or other remote or in-person consultations.

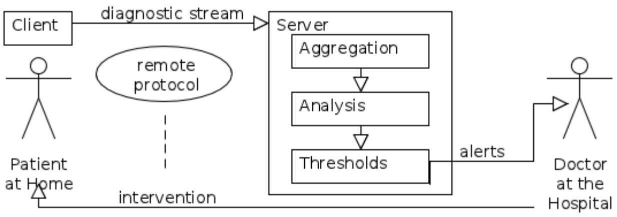

Fig. 1: Automated Telemedical Prevention.

3. Temporal Data-mining

The following specifics need to be considered when doing data-mining on temporal diagnostic data:

- *Temporal abstraction* — temporal abstraction needs to be applied in order to convert timestamped observations (*events*) into a standard feature matrix which may be directly used for data-mining [4].
- *Noise in data* — in-home diagnostic is typically highly noisy and incomplete. Noise in data needs to be addressed to train consistent and thus better performing classifiers.
- *Filtering* — high amount of various diagnostic data may over-fit the classifier and increase temporal complexity of the training process. Redundant or correlated data needs to be filtered or aggregated.
- *Rare-events* — typically in telemedical programs the predicted condition (relapse, worsening etc...) is relatively rare[3], thus imbalanced classes and lack of positive examples needs to be addressed [5][6].

3.1. Data Preparation

Correlation between various diagnostic variables may be addressed with principal component analysis (PCA). Another approach is to use factor analysis and aggregate the most correlated variables [7]

Temporal abstraction extracts features out of temporal events. Different methods are applicable for this purpose. The most simple method is to use the variable as a feature directly without any transformations. Another method is to build fixed length *temporal windows* either with a certain overlap or for each event. All events in each window are than used to apply different statistical measures (such as average, standard deviation, modus...) and form features.

Using various forgetting functions it is possible to abstract from temporal windows completely. In such case each feature is produced by applying a forgetting function to each single event and all its predecessors.

The following figure demonstrates the whole process on an example. Two correlated diagnostic streams are first aggregated using *sum* and later *average* is used to extract a feature value for each temporal window.

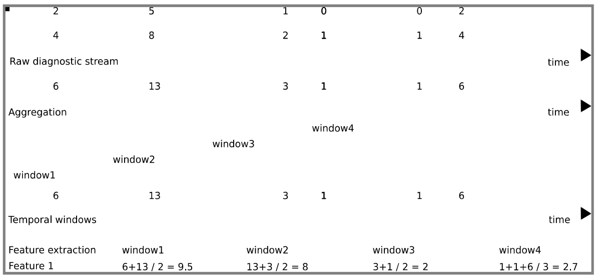

Fig. 2: Stream Aggregation and Temporal Abstraction using Temporal Windows.

An *intervention period* is a temporal window before each target even (the predicted event) long enough to potentially prevent the event through intervention. Feature values extracted from within the intervention period are considered positive examples. Values outside the intervention period are considered negative. Classifiers are than trained to recognize positive examples in order to predict target events. The related feature matrix construction is demonstrated in Fig. 3.

[3] This is especially true for chronic diseases which are especially suitable for telemedical prevention.

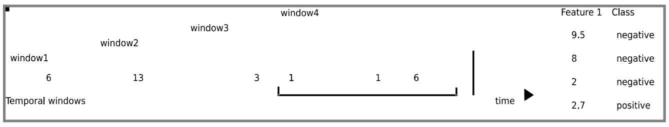

Fig. 3: Construction of the Feature Matrix with Target Class.

3.2. Modelling

This article suggests to test various data preparation approaches and eliminate methods which aren't suitable in a series of modelling experiments. A suitable model for prototyping (or even for production) is for example the Naive Bayes classifier. It works very well for many real life data sets and performs very fast as it does not search the whole state space [8][10].

Other models which has been successfully applied on temporal prediction problems include for example AdaBoost, SVM, various rule-based approaches, decision trees and others [11]. A good starting point is to test for example the top 10 data-mining algorithms identified at the 2006 ICDM Conference [12].

3.3. Evaluation

This article proposes the use of *sensitivity* and *specificity* as the measures to compare various models' performance on diagnostic data. Sensitivity and specificity are standard measures used in medical diagnostics. They are well understood by the doctors, they cope well with classes imbalance and they may be directly used to construct the Youden's index[4] in order to obtain a single comparable measure [13].

In addition, sensitivity and specificity is the base for ROC curves which allows model comparison for different classification thresholds. This is especially useful when there are requirements for a certain sensitivity or specificity thresholds in the telemedical program. In this case only the relevant segments of the ROC curves are mutually compared [15][16].

The proposed evaluation technique to estimate the model's performance on unknown data is leave-one-instance-out cross-validation which is especially useful when we have only a small amount of positive examples. It allows to use the maximum of the positive examples for training in each fold [13][14].

3.4. Noise in the Data

Noise in data causes inconsistencies and over-fits the model. The proposed approach is to establish thresholds to filter out noisy data in the training phase to obtain consistent models which perform better even on the noisy data[5] which had been left out.

3.5. Pitfalls

There are several pitfalls which needs to avoided when performing temporal data-mining on telemedical diagnostic data [17]:

- *Leave-one-instance-out* — classical leave-one-out cross-validation implemented in many different data-mining systems would give over optimistic performance estimates. The reason is, same patient's examples may lend in the validation set as well as in the training set and thus import additional knowledge into training and cause better results than what can be expected on unknown data.
- *Best model bias* — fine-tunning parameters of a single model would give over optimistic results. It is preferable to invent several models and data-preparation techniques which perform well on the data.
- *Discretization* — if used, discretization thresholds needs to be calculated for each training set again in each cross-validation fold. Calculating the thresholds from the whole data would again import additional knowledge from the validation set into training.

4. ITAREPS — the Case Study

[4] Youden's index = sensitivity + specificity - 1

[5] Noisy data are left out from training set but they are present in validation set.

ITAREPS[6] is a schizophrenia relapse prevention program invented in the Prague Psychiatric Centre by Dr. Filip Spaniel. It is a mature program operated since 2006 in Czech republic with over 400 patients. Over time it has been deployed in 7 additional countries[7] [18].

Each week, patients and their related carers[8] are asked via an SMS message to submit answers to a short questionnaire. The questionnaire consists of ten questions primarily targeted at detecting prodromal symptoms, such as: sleep problems, irritation, decrease in appetite, various dysphoric symptoms, behavioural changes, cognitive skill worsening, hearing voices etc...

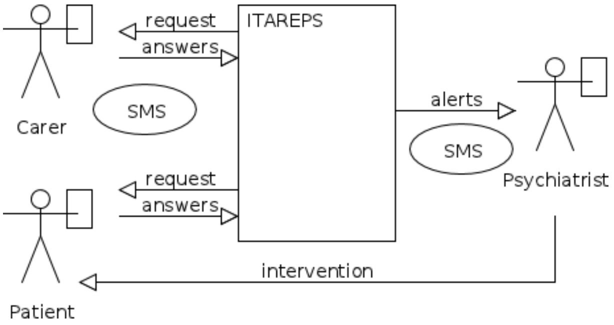

Fig. 4: the ITAREPS program.

Patient's state is evaluated relatively to the last evaluation. Answers range from 0 to 4, where 0 means no change or improvement and 1 to 4 means mild to extreme worsening respectively.

When the total sum of answers exceeds a given expert-defined threshold for a time period, ITAREPS sends an *alert* to the outpatient psychiatrist. S/he may than decide to increase medication to prevent consequent hospitalization which is not only costly but also significantly affects patients quality of life.

In a mirror design trial the program decreased the overall hospitalization rate of the participating patients by 70% after they entered the program [18]. But a recent one year double-blind trial revealed issues with a high false alert rate which discourages doctor to do proper interventions [19]. This issue is one of the motivations for the research described further.

4.1. Data-mining on ITAREPS Data

The history of patient and carer evaluation messages and the history of patient's hospitalizations has been used for a series of data-mining experiments. Experiments were performed on the passive branch of the double-blind trial which isn't contaminated with medication increases.

Factor analysis[9] has been applied to find suitable question clusters. Experiments revealed that clustering the 10 questions into 4 clusters outperforms other options. Related patient and carer clusters has been aggregated using *sum* which outperformed other approaches.

Various temporal abstraction techniques has been applied, but *exponential forgetting* or a simple *difference* between two last messages worked far better with the data.

Many different models have been applied to the resulting feature matrix but Naive Bayes, SVM and AdaBoost performed consistently better than other models.

Another technique which greatly improved relapse prediction performance was filtering of patient who produce noisy data. Some patients had issues to correctly self-evaluate their condition. The historic data contain examples of non-responders as well as patients who send bad scores all the time.

[6] Information Technology Aided Relapse Prevention in Schizophrenia

[7] Slovak republic, Great Britain, Japan, Taiwan, Malaysia, Saudi Arabia and Netherlands

[8] Appointed family members

[9] PCA was also tested, but did not perform well with the data.

A simple threshold on minimum average score in the intervention period and maximum score outside this period allowed to filtered noisy patients from the training set. This improved the overall performance as depicted in Fig. 5.

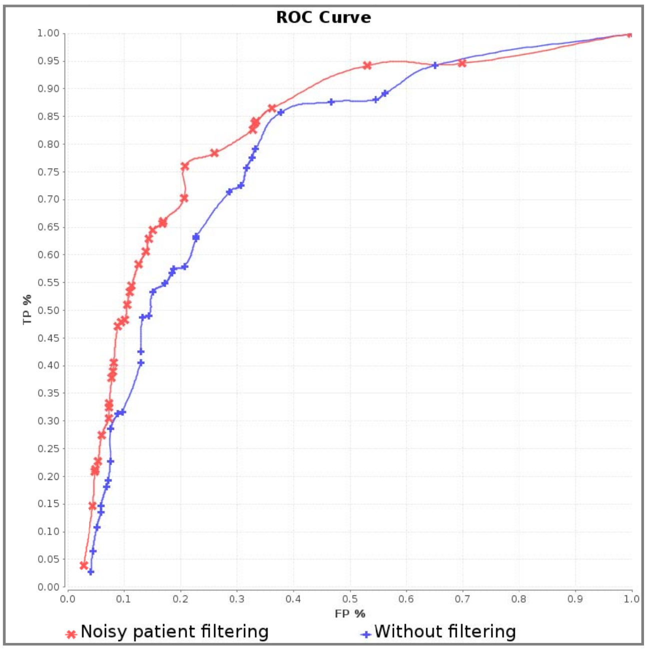

Fig. 5: Noise Filtering.

Various data preparation techniques and different models has been tested and evaluated using ROC curves similarly to Fig. 5 in order to identify a set of methods which perform better with the ITAREPS data.

4.2. Results

This section presents the results achieved using the standard Naive Bayes classifier and a simple last two message difference feature applied on 4 question clusters. The demonstrated results are performance estimates on unknown data calculated using the leave-one-instance-out cross-validation.

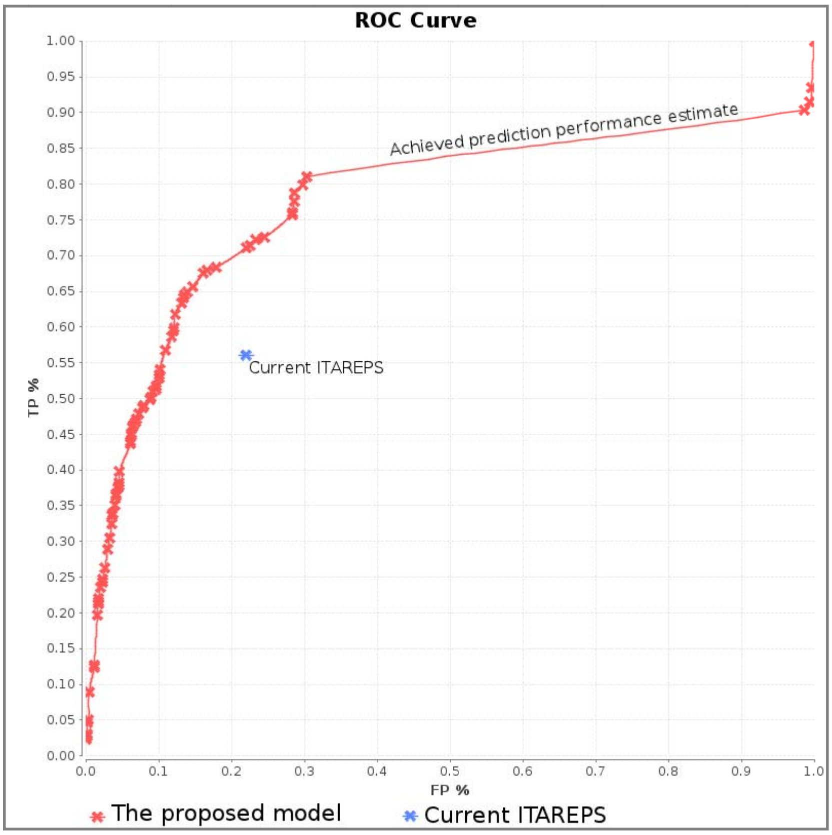

Fig. 4: Prediction Performance.

The proposed model shows a significant increase in prediction performance over the current ITAREPS program. The false alert rate has been decreased by 67.7% with the same sensitivity = 0.568.

Table 1: Prediction Performance

	Youden's index	**Sensitivity**	**Specificity**
ITAREPS	0.350	0.568	0.782
The Proposed Model	0.459[10]	0.568	0.891

The overall average false alert rate is 1.361 false alert per patient and year with 73.3% predicted hospitalizations[11] where ITAREPS performs at 4.208 false alerts with the same predicted hospitalizations rate. This satisfies *acceptance criteria* defined by experts prior to this research for the next generation ITAREPS with at least 70% predicted hospitalizations and max. 3 average false alerts per patient and year.

The proposed model performs at sensitivity = 0.568 and specificity = 0.891 which is comparable with widely used standard clinical diagnostic test in medicine.

The following figure shows prediction performance of the proposed model on the passive branch visualized on a time-scale. This visualization method can significantly help communicate the results with the doctors.

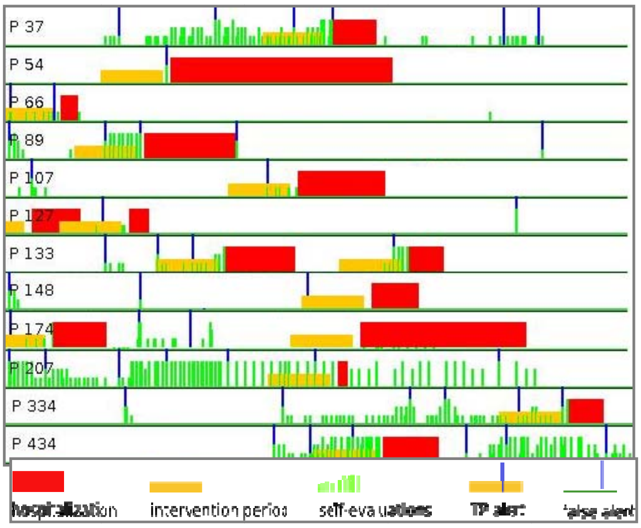

Fig. 5: Patient Time Scale Visualization[12].

In addition, the proposed model allows to trade-off sensitivity for specificity by selecting different classification thresholds. For example 80% of predicted hospitalizations may be exchanged for 1.778 false alerts per patient per year. This allows to optimize cost-efficiency of the program. To optimize prediction in ITAREPS we want to minimize the false alert rate within the acceptance criteria boundaries.

5. Applicability

The methods proposed in this article were successfully applied to schizophrenia relapse prediction but they may also be applicable to other telemedical prevention programs if the following criteria are met:

- Availability of historic data with a sufficient number of positive examples.
- Existence of a passive branch of a blind trial to obtain data not contaminated with interventions.

[10] Youden's index has been calculated for the same sensitivity as the ITAREPS's Youden's index.

[11] A hospitalization is considered as predicted when there is at least a single alert in the hospitalization's intervention period.

[12] Patients which weren't hospitalized in the observed period has been left out of the charts.

6. Conclusion

This article proposes the use of temporal data-mining methods to improve prediction performance in telemedical prevention programs and thus improve overall efficiency of such programs. Various data-preparation, modelling and evaluation techniques are first defined in general and later applied to a case study — the ITAREPS program for schizophrenia relapse prevention.

This article shows that improving prediction performance through data-mining on the diagnostic data history is a realistic goal. The proposed model performs significantly better than the expert-defined thresholds of the original ITAREPS reducing the overall rate of false alerts by 67.7%.

Similar approaches may be applicable for other telemedical programs which satisfy the applicability criteria defined previously.

7. Acknowledgement

This project was supported by Grant Agency of Czech republic, project 201/08/0802.

8. References

[1] Darkins A., Cary M.: Telemedicine and Telehealth: Principles, Policies, Performance and Pitfalls. Springer. 2000. ISBN: 0826113028

[2] Hersh W., Helfand M., Wallace J., Kraemer D., Patterson P., Shapiro S., Greenlick M.: Clinical outcomes resulting from telemedicine interventions: a systematic review. BMC Medical Informatics and Decision Making V1. 2001.

[3] Field M., Grigsby J.: Telemedicine and Remote Patient Monitoring. The Journal of the American Medical Association 2002, 288(4):423-425.

[4] Antunes C., Oliveira A.: Temporal data mining: An overview. KDD Workshop on Temporal Data Mining. 2001

[5] Weiss G., Hirsch H.: Learning to Predict Extremely Rare Events. Rutgers University New Brunswick. 2000.

[6] Vilalta R., Sheng M.: Predicting Rare Events In Temporal Domains. IEEE International Conference on Data Mining. 2002.

[7] Bryant F., Yarnold P., Grimm L., Yarnold P.: Principal components analysis and exploratory and confirmatory factor analysis. Reading and understanding multivariate statistics. American

[8] Psychological Association, IX. 1995. 99-136.Zhang H., The Optimality of Naive Bayes. FLAIRS conference. 2004.

[9] Seong-Pyo C., Sungshin K., So-Young, L., Chong-Bum L.: Bayesian networks based rare event prediction with sensor data. Knowledge-Based Systems Volume 22, Issue 5. 2009. 336-343.

[10] Greg H., Charles, E.: Bayesian approaches to failure prediction for disk drives. In Proceedings of the Eighteenth International Conference on Machine Learning (ICML'01), 2001.

[11] Murray F., Hughes G., Kreutz-Delgado K.: Machine learning methods for predicting failures in hard drives: A multiple-instance application. Journal of Machine Learning Research. 2005. 6:783–816.

[12] Wu X., Kumar V., Quinlan R., Ghosh J., Yang Q., Motoda H., McLachlan G., Ng A., Liu B., Yu P., et al.: Top 10 algorithms in data mining. Knowledge and information systems. Volume 14, 2006. 1-37.

[13] Rodriguez J., Perez A., Lozano J.: Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010. 569-575.

[14] Molinaro A., Simon R., Pfeiffer R.: Prediction Error Estimation: A Comparison of Resampling Methods. Bioinformatics. 2005. 21:3301–3307.

[15] Fawcett T.: ROC Graphs: Notes and Practical Considerations for Researchers, Machine Learning. 2004

[16] McClish D.: Analyzing a Portion of the ROC Curve. Med Decis Making vol. 9 no. 3. 1989. 190-195.

[17] Varma S., Simon R.: Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics. 2006. 7:91

[18] Španiel, F., et al.: ITAREPS: Information Technology Aided Relapse Prevention Programme in Schizophrenia.

Schizophrenia Research Volume 98. Issues 1-3. 2008. 312-317.

[19] Španiel, F., et al.: The Information Technology Aided Relapse Prevention Programme in Schizophrenia: an extension of a mirror-design follow-up, International Journal of Clinical Practice Volume 62, Issue 12. 2008. 1943–1946.