The data processing and decision support in healthcare systems

P.S. Stashevsky[a], I.N. Yakovina[a], T.V. Avdeenko[a], N.A.Bannova[a]

[a] *Novosibirsk State Technical University, 630073, Karl Marx Avenue 20, Novosibirsk, Russia*

Abstract

The article describes the characteristics of diagnostic and management tasks solved in healthcare information systems. We propose an approach based on the use of ensemble models (variant modeling) to solve problems of decision support. The variant modeling is a method that applies the set of models for solving one task. The models can have different structures and use different input features. Also, we propose the method called “Modeleteka” for structure store and automatically select models to solve tasks in healthcare information systems. This method allows append new models and train old from ensemble in the maintenance of information system.

Keywords: healthcare information system; decision support; ensemble models; variant modelling; modeleteka

1. Introduction

The healthcare information systems become very important instruments to solve the problems of research and prediction various diseases [1, 5, 7]. In this paper we consider a small part of such systems which store the patients with definite pathology on some geographic area (healthcare register of patients, HRP). It allows study structure disease and estimate effectiveness of medical and management decisions. Currently there is an important task to develop methods, which can use healthcare data stored in registers for preparation medical and management decisions.

2. The object of the study

In this article we explore methods for development and use diagnostic models in healthcare registers of patients. The paper discusses the following questions:

1) analyze stored data and structure of tasks solved in HRP;
2) development methods for the organization of the storage model and use them as an ensemble;
3) testing methods for the task of prediction the count of patients (on the example of patients with chronic kidney disease).

3. Methods

3.1. Healthcare register of patients: data and tasks

Today we have many products of HRP [3, 9]. The population-based registers may cover different levels from regional to international. For example, there is an international European and American registry of cystic fibrosis including 68 different indicators of the patients. There are some federal registers in Russian Federation, such as State register of patients with diabetes mellitus and State register of patients with cancer pathology. The federal registers, as a rule, keep one pathology (mono-register) and regional registries often keep information about multiple pathologies (poly-register). There are cancer register and register of arrhythmia in Novosibirsk region. Table 1 shows the analyze result of the data structure stored in the medical HRP about patients.

Table 1. The structure of stored data

Data group	Description
General data	General information about the patient, which is used for identification of the patient and does not apply to medical settings
Anamnesis data	Data on the life disease course and a patient (including comorbidities, treatment, etc.)
Clinical data	The results of examination of the patient experts describe the nature and extent of the pathology
Laboratory data	Results of laboratory and functional studies carried out for the patient.
Treatment data	Data on the purpose types of treatment used dosages, frequency and timing.

It should be noted, that not all HRP’s include all data groups and it effects on the methods and models for solving diagnostic and management tasks. Table 2 shows the result of compare data groups of HRP’s in Novosibirsk region.

Table 2. The data groups in HRP of Novosibirsk region

Data group	CKD	CR	AM	DT	AT
General data	full data	full data	full data	full data	Partly data
Anamnesis data	history of life and disease	history of life and disease	History of disease	History of disease	History of disease
Clinical data	Only med. exam.	full data	Only med. exam.	full data	Only med. exam.
Laboratory data	full	full	no	full	no
Treatment data	Spec. treatment	Full data	Full data (with hisory)	Spec. treatment	Spec. treatment

CKD – chronic kidney disease register, CR – cancer register, AM – asthma register, DT – diabetes register, AT – arrhythmia register.

In general solved by the medical registers tasks divide in two classes: diagnostic and management. According to the analysis of tasks proposed a general description (*Z*)*:*

$$Z_i = Z(P_{z1}, \ldots, P_{z5}),$$

$P_{z1},\ldots,P_{z5}$ (parameters of task): P_{z1} - type of task (classification, regression, optimization etc.); P_{z2} - durability of decisions (quick, long-term); P_{z3} – periodicity of decision; P_{z4} - class of problem (diagnostic, management); P_{z5} – priority of problem (high, medium, low).

As an example we consider the tasks solved in HRP's. There are problems of assessment and analysis of morbidity and mortality, predicting various indicators, as well as planning and optimization of medical care in the region. These tasks are different in speed, gravity, speed of decision-making and other parameters. Since registers contain heterogeneous data and many kinds of tasks, it is necessary to use a sufficiently large number of models to solve them. We use the next types of models:

1) expert assessment and modeling – using expert knowledge for preparation and making decision automatically,
2) OLAP and statistical analysis,
3) decision trees, based on machine learning,
4) probabilistic models, using analysis of patients survival.

3.2. The storage of models "Modeleteka"

A large number of models can be used to solve the tasks in HRP's for decision support. It should be noted that only a subset of models can be used in solving a specific problem. In this regard, there is a task of a structured storage and automatically select the models for solving task. We proposed a method "Modeleteka" to solve this problem.

Modeleteka (repository of models) is an ordered set of models that satisfy the requirement of completeness, minimum redundancy level of description and study for the application in the annex to the specific area [5,9]. The characteristics of modeleteka are ordering models, their completeness and lack of redundancy, the same complexity of models, and the similarity of detail in model descriptions, availability for each model set of features and the formulas described the relationship between elements.

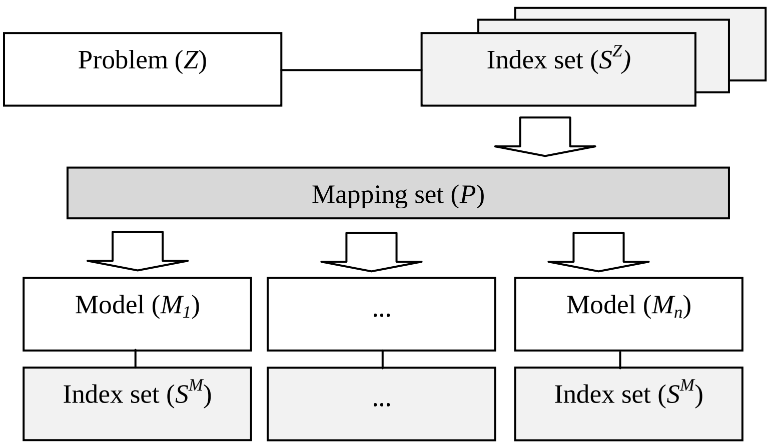

Fig. 1. The principe of work for select models.

The using of method is based on three basic concepts: ***problem*** (Z), ***model*** (M) and ***mapping space*** (P). The task is described ***index set*** S^Z, which defines a semantic, peculiarities and limitations of the task. We also use a set of indexes S^M for models, which describe a structure and specialty of using. The ***mapping space***s P is an auxiliary tool for search models and define the intersection of index spaces between task and models. The relationship between the concepts is described more detail in papers [1,7].

The method Modeleteka allows store and select set of models to solve various problems of healthcare information systems. In additional, the using of accuracy metrics allows applies evaluation procedures, when metrics is recalculated and models chsnge to new tasks.

3.3. The method of variant modelling

We obtain a set of models $M_1,...,M_n$ to solve task using Modeleteka. In this case, we can use ensemble models (variant modelling). It's a method, based on use of the models system (two and more models use together). So, we construct a vector-model (VM) – system of a minimal set of related on purpose, simple and close to the complexity models, that describes the features of decision support task.

Figure 2 shows the principle of the use Modeleteka and various modeling together. In the first step we select a set of models $M_1,...,M_n$ from modeleteka that are suitable for solving task. In the second step vector-model is formed and used. In the third step final decision R^* is calculated from the set of result values (the result of the vector-model $R_1,...,R_n$). As an example, there is a task of assessment efficiency of HRP, where we can use together a set of statistical models, survival analysis, expert model and other. The using of models combination allows get more information about object and makes more objective decision.

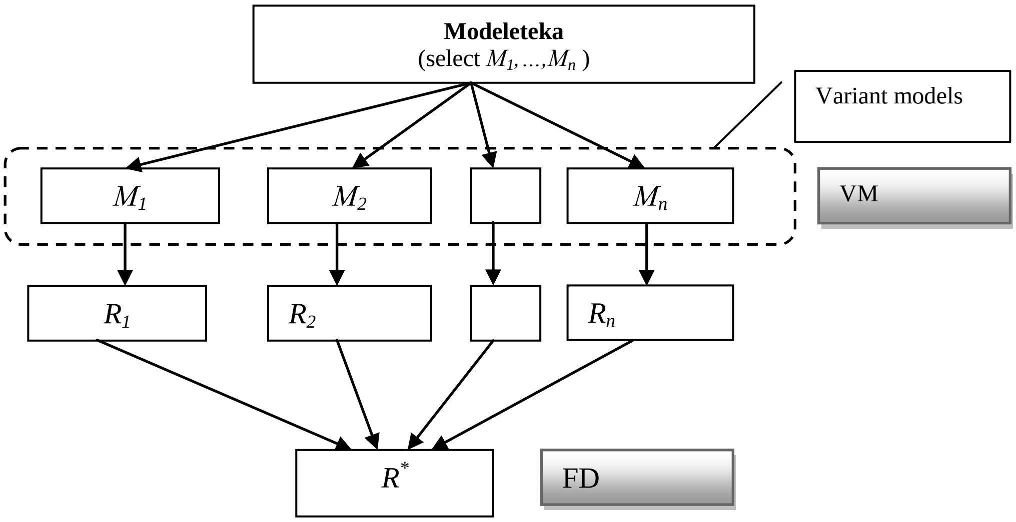

Fig. 2. The use Modeleteka and variant modelling (VM) to make final decision (FD).

4. Results and Discussion

We consider example for the proposed methods to solve the important task in HRP's. It is a prediction the count of patients of specific pathology. We use next indicators for solving this task:

1) incidence rate (Z_i)
2) mortality rate (Z_m)
3) survivability (Z_s)

Using the forecasted values of these three parameters, you can calculate the number of patients with specific pathology in the predicted year: $N = Z_i - Z_m + Z_s$.

We consider the task of forecasting the number of CKD (chronic kidney disease) patients in 2016, using historical data for 2010-2016 of mortality and incidence rates. We have a set of models for each parameters (vector-model) and specific model take into account their features for prediction. Table 3 shows the results of variant modeling for the calculation of each parameter (Z_i, Z_m, Z_s).

The final solution R^* can proposed by the method of ***diagnostic consultation***, which allows use the vector-modeling result for making final decision by analogy with a consultation of experts. The following ways use to get final result.

1) *Averaging the values for all models.*

$$R^* = \frac{1}{N}\sum_{i=1}^{N} R_i,$$

where R_i - solution obtained using the i-th model, N - number of models used in the vector-model.

2) *Use accuracy model.*

Another way is to use information about models accuracy: accuracy, true positive and false negative characteristics [9]. In this case, the final decision can calculate as maximum value of accuracy.

$$R^* = argmax(C_i, \dots, C_n),$$

where C_i-evaluation of the diagnostic accuracy of the i-th model.

3) *Weighted average of decisions*.

If the values are not exactly models are very different, it is possible to use a weighted average.

$$R^* = \frac{1}{N}\sum_{i=1}^{N} R_i * C_i / \sum_i^N C_i,$$

Table 3. The VM of Z_i, Z_m, Z_s

Models	**Incidence, number of person (Z_i)**	**Mortality, number of person (Z_m)**	**Survivability, number of person (Z_s)**
M_1 (regression)	88	-	-
M_2 (decision tree)	73	10	-
M_3 (expert)	62	-	219
M_4(approximation)	-	7	-
M_5 (probabilistic)	-	-	199

The results to apply the method of diagnostic consultation for solving the task of predict the incidence in 2016 is shown in Figure 3.

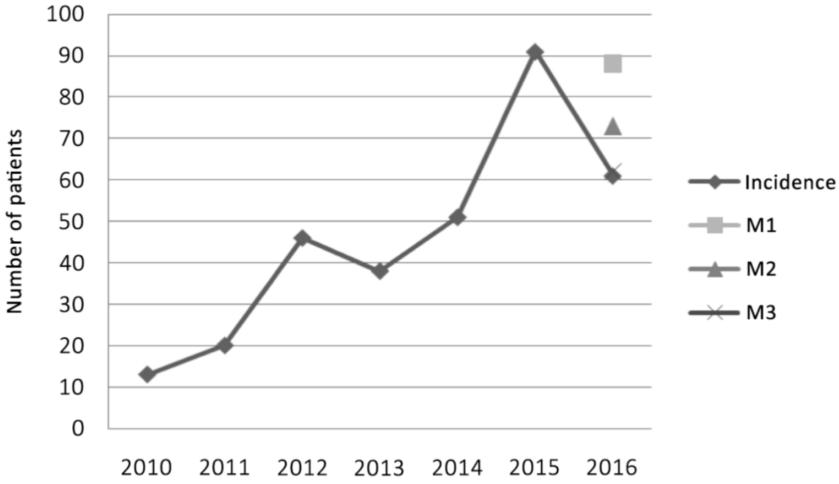

Fig. 3. VM of Incidence rate.

The figure shows that the model M_3 is the closest to the real value of incidence. Depending on the approach used, the final decision may be different, but in any case, closer to the actual value than in case separately use of the models. Figure 4 shows the calculation of the number of patients with CKD, using the method of variant model (VM) and specific models (M).

As seen from figure the use of variant modeling can improve accuracy of predict.

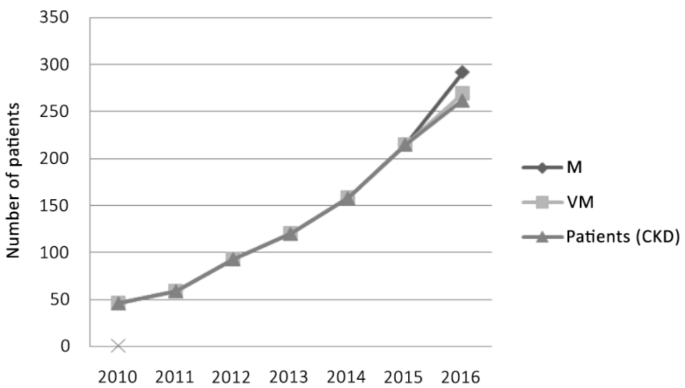

Fig. 4. VM of patient number with CKD.

5. Conclusion

In the article we present results of analyze data and tasks solved in the HRP's. We propose methods for storage the models and use it in ensemble, that allow to achieve the appearance of emergent properties by virtue of the joint application of a set of models. The methods tested on real data to solve the problem of prediction the count of patients with CKD in 2016 in Novosibirsk region.

References

[1] Abeku, T. Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia: a simple seasonal adjustment method performs best / T. A. Abeku, S. J. de Vlas, G. Borsboom, A. Teklehaimanot, A. Kebede, D. Olana, G.J. van Oortmarssen, J.D.F. Habbema // Tropical Medicine and International Health. — 2002. — Vol. 7, No 10. — P. 851–857.

[2] Araz, O. M. A pandemic influenza simulation model for preparedness planning / O. M. Araz, J. W. Fowler, T. W. Lant, M. Jehn // Proceedings of the 2009 Winter Simulation Conference. — 2009. — P. 1986–1995.

[3] Bao, L. A new infectious disease model for estimating and projecting HIV/AIDS epidemics. // Sexually Transmitted Infections. — 2012. — Vol. 88, Suppl. 2. — P. i58–i64.

[4] Cowling, B.J. Methods for monitoring influenza surveillance data / B.J. Cowling, I.O.L. Wong, L.M. Ho, S. Riley, G.M. Leung // International Journal of Epidemiology. — 2006. — Vol. 35, No 5. —P.1314–1321.

[5] Daley, D. J. Epidemic modelling: An introduction. / D. J. Daley, J. Gani // — Cambridge University Press, 1999. — 225 p.

[6] Gubarev, V.V. The flexible methods of making diagnostic decisions in medical information systems / V. V. Gubarev, P. S. Stashevsky, I. N. Shvaykova // Proceedings the 5-th IFOST-2010ю - p. 90-93 – (in Russian)

[7] Soebiyanto, R.P. Modeling Influenza Transmission Using Environmental Parameters / R. P. Soebiyanto, R. K. Kiang // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science. — 2010. — Vol. XXXVIII, part 8. — P. 330–334.

[8] Stashevsky, P.S. The technology of decision support in healthcare information systems / P. S. Stashevsky, I. N. Yakovina, O. V. Dunicheva // Perspektivi nauki. - 2015. – № 6 (69). – C. 109–113 – (in Russian)

[9] Stashevsky, P.S . A new approach to the analysis and evaluation of intelligent medical diagnostic systems / P.S. Stashevsky, I.N. Shvaykova // Modern Problems of Science: Proceedings of the international conference. - Tambov: Izd Pershin, RV, 2010 – p. 79-84 – (in Russian)

[10] Thompson, W. W. Mortality Associated With Influenza and Respiratory Syncytial Virus in the United States / W. W. Thompson, D. K. Shay, E. Weintraub, L. Brammer, Cox N., Anderson L. J., Fukuda K. // The Journal of the American Medical Association. — 2003. — Vol. 289, No 2. — P. 179–186.

[11] Vlasov, V. V. Accounting system and analysis of controlled clinical trials (Cochrane Collaboration) // Cardiology, 1998.- 7.- p. 51-53 – (in Russian)