THE USE OF INTELLIGENT JSM-TYPED SYSTEM FOR THE ANALYSIS OF CLINICAL DATA

L.V. Vinokurova, M. A. Agafonov, G. G. Varvanina, V. K. Fynn, E. S. Pankratova, D. A. Dobrynin

[1]*Central research institute of gastroenterology, Moscow, Russia*

[2]*RAS Russian institute of scientific and technical information, Moscow, Russia*

Key words: chronic pancreatitis, intelligent system, clinical analysis, JSM-method, diabetes

There has been an increasing number of patients with complicated forms of chronic pancreatitis (CP) in recent years (20-80 patients per 100000 of population), which leads to a permanent increase in the severity of these complications and makes them turn to the surgical treatment [9]. With differential diagnostic purposes, including the exclusion of pancreatic cancer, these patients undergo clinical, laboratory and instrumental examination [6, 10, 12]. However, the accuracy of the information about the clinical signs of the disease often depends on subjective parameters, such as the experience of the doctor and the patient's sociability. The use of computed tomography expanded diagnostic capabilities to identify complications in CP. The development of medical technologies has resulted in the emergence of new instrumental methods of diagnosis of diseases of the pancreatoduodenal zone in including endoscopic ultrasonography with fine-needle biopsy conduction. In this case, the subjective factor of evaluating the studies result is difficult to avoid too [11].

M-ANNHEIM classification-2007 includes an evaluation point system to determine the severity of the disease. Such systems have been developed for Crohn's disease and certain liver diseases in which this system helps to choose the right tactics of treatment and to evaluate the prognosis of the disease [13].

Gradation of pancreatic pain is conducted by the sum of peculiarities of pain and the need to describe different analgesics. For example, if the patient should be prescribed powerful narcotic analgesics (2 points), and he/she has periodic attacks of acute pancreatitis (3 points), the severity of pain is evaluated in 5 points. The

calculation of the index of CP severity includes all surgical pancreatic interventions and all the serious complications of pancreatitis, starting from the first manifestations of the disease and during medical history. If the patient had two or more severe complications, each of them separately should be included in the calculation of the index of CP severity. After filling out the M-ANNHEIM evaluation scoring system, it's necessary to summarize the points. The total score is the severity index of M-ANNHEIM. Mathematical analysis methods and mathematical modeling of pathological processes have increasingly helped the practitioners in recent years [1, 7].

Materials and methods

The intelligent JSM-typed[1] system (IntJSM), based on the JSM-method of automatic generation of hypotheses, was created in the sector of intelligent systems of RAS VINITI [3, 4, 5, 7]. Existing IntJSM includes: IntJSM = Solver of tasks + Information environment (base of the facts (BF) and base of knowledge (BK) + Intelligent interface (dialogue + representation of results + learning to work with system).

IntJSM solver is based on the JSM-automatic method of generating hypotheses realizing automated plausible reasonings. Developed plausible reasonings constitute the formal heuristic of the dependencies extraction of causal type from the database of structured facts.

Hypothesis of (±)-reasons extracted from BF, are generated by a specially designed induction procedure, based on the similarity of objects — the source or cause of the presence (or absence) of the studied effect. Predicting the effect is achieved by analogy, using the hypothesis of (±)-reasons contained in BK and generated by induction. And finally, JSM-reasoning ends with abductive procedure — an explanation of the initial state of BF, which is either a sufficient ground for hypotheses, or a means of expanding the BF to iterate JSM-argument if there are unexplained facts from BF.

[1] Named after a British philosopher John Stuart Mill.

Peculiarities of JSM-method:

1. Extracting knowledge of "cause–consequence" type is based on the principle of "similarity of facts gives rise to the similarity of effects and their frequency of occurrence" (this principle is different from the probabilistic approach to data analysis "repeatability of effects determines the similarity of the facts").
2. JSM-method, being non-statistical method of data analysis, is able to take into account the individual characteristics of the studied objects of study.
3. JSM-method is able to generate useful hypotheses on small datasets.
4. JSM-method works with open databases (not with the closed tables), indicating to the need of BF expansion, if necessary (this is carried out through abductive BF explanation).

Terms of application

JSM-method is applicable to the studied subject field, as the following conditions are required:

1. Possibility of data structuring and formal definition of similarity of facts (from BF).
2. Presence of positive and negative examples of (±)-examples in BF.
3. Presence of implicitly defined dependencies of causal type in BF ((±)-reasons of the studied effects).

Statement of problem

In this paper we describe the use of IntJSM for prediction of diabetes of pancreatogenic genesis in patients with CP.

Results of research

BF creation

The first step of a collaborative work was to develop a subsystem of knowledge presentation and BF creation — one of the IntJSM components. Attribute of this BF is the information about patients presented in medical records and described in accordance with the language of data presentation.

For the problem of predicting diabetes of pancreatogenic genesis we identified the following signs-attributes: patient's age, body mass index, bad habits (alcohol consumption and smoking), duration of disease, presence of calcificating CP, Wirsung's lithiasis, pancreatic hypertension, resection or drainage operations in the history, pancreatic necrosis in history, glycemic profile, C-peptide level; we evaluated the size of the pancreas, its contours, the presence of calcifications, the width of the main pancreatic duct, the concentration of gastrointestinal hormones in the peripheral blood (gastric inhibitory polypeptide, somatostatin, cholecystokinin, secretin, gastrin), the presence of the developed clinical picture of diabetes, and so on (39 signs in all). In BF the requirements of structuring data are implemented.

Study is based on the examination of 158 patients with complicated CP. Patients were divided into two age groups: from 20 to 60 (119 patients), from 61 to 74 (39 patients).

Customization of the subject area

To apply IntJSM to a specific task, there must be an adjustment, which includes the following:

1. Development of language of data presentation. Efficiency of analysis of the results of different studies depends on the fullest description of patient's examination. Analyzing these numerous medical facts presented in BF, f one type of data is selected from all the possible ones: one of the possible qualitative features is indicated from the list view.

   Example 1. Resections in medical history: At_1 — gPDR; At_2 — nPDR; At_3 — distal resection; At_4 — Beger operation.

   Example 2. Body mass index: At_1=16 and less — expressed mass deficit; At_2=16,1-18,5 — lack (deficiency) of body weight; At_3=18,5-25 — norm; At_4=25,1-30 — overweight, and so on.

   Example 3. Gastric inhibitory polypeptide: At_1 — less than 0.002 ng/ml; At_2 — 0,002-2 ng/ml; At_3 — 2-4 ng/ml; At_4 — 4-6 ng/ml.

   Thus, language of presentation of medical data was created, programs of health information input were developed. The resulting system provides

automated patient records. Medical data input form is very comfortable for the doctor. You can view BF, add information about the disease when it is received, include the results of various studies. If necessary, BF may be filled with new terms and concepts.

2. Definition of notion of "object" and "feature" in the terminology of JSM-method. In our task object is represented by ordered set of 39 clinical, instrumental and laboratory signs. The feature in this case one — diabetes of pancreatogenic origin.
3. Setting the operation of similarity. For the selected type of data, result of the operation of similarity is a coincidence of signs.
4. Setting the ratio of enclosure. Enclosure ratios are determined by the presence of specific signs.

Customization of experimental research

Customization of experimental research includes the following:

1. The choice of strategy: a simple method of similarity, a method of similarity with the prohibition of counterexamples (non-enclosure of hypotheses obtained by the method of similarity to the original examples of the opposite sign) separately for (+)- and (-)-examples.
2. Selection of the required number of parents[2].

A possible criterion for evaluating the selection of parameters and strategies of the experiment — the application of procedure "extension of a definition one by one": each object of sampling is sequentially given a value "uncertainty", object's extension of a definition is made by means of JSM-system with the selected parameters, and the extended value is compared with the existing one. Count the total number of correct and incorrect completions. Options are selected from paragraphs 1-2, in which upon application of the procedure "extension of a definition one by one" there will be the best result, i.e. the highest number of correct extensions of a definition and the least amount of wrong ones. These settings are logically used for

[2] Examples that formulate hypotheses.

the extension of a definition of patients newly recorded in BF, who need to be diagnosed.

Results of computer study

Data for computer studies were provided by employees of the Central Research Institute of Gastroenterology. The first computer research on the diagnosis of diabetes of pancreatogenic genesis was conducted on survey data of 81 CP patients, 21 of whom were diagnosed with diabetes of pancreatogenic genesis — "feature" in the terminology of JSM-method. Prognosis was based on the data of 42 patients.

To do this computer studies using the criterion of "extending a definition one by one", strategy and the parameters of the experiment were chosen:

1. a simple method with a prohibition of counterexamples;

2. number of parents «+» — 12;

3. number of parents «-» — 22.

At the first stage, by means of JSM-system by rules of a plausible conclusion of the 1st sort (induction), using the training sample of 81 patients, hypotheses of the 1st sort were generated: "Presence of a certain set of features in the patient was the reason for presence or absence of the diagnosis diabetes of pancreatogenic genesis". For example, set of the following signs: "presence of calcificating CP; resection of the pancreas in history; pancreatic necrosis in history; CP duration; glycemic profile indices; indices of gastrointestinal hormones, exocrine insufficiency, pain syndrome, presence or absence of mutations".

Generated hypotheses were the fragments of BK.

At the second stage, with the use of hypotheses of the 1st sort by rules of the 2nd sort, presence or absence of diabetes of pancreatogenic genesis was redefined in 42 patients presented for the prognosis.

As a result of computer research, 10 patients were redefined positively, i.e. the diabetes of pancreatogenic genesis was diagnosed, 28 patients were redefined negatively, i.e. they weren’t predicted to have the diabetes of pancreatogenic genesis, whereas 19 of patients belonged to the older age group. Results of extension of a

definition coincided with real medical data. Thus, the accuracy of extension of a definition in this computer research was equal to 100%, and completeness — to 90%.

The second computer study was conducted on a survey data of 123 patients (a training set of 81 patients was added by data of 42 patients who were presented for the prognosis in the first study). 33 patients were diagnosed with diabetes of pancreatogenic genesis, and 89 — the absence of diagnosis. Results of the examination of 35 patients were given for the prognosis.

For this case, using the criterion of "extension of a definition one by one", strategy and the parameters of the experiment were chosen:

1. a simple method with a prohibition of counterexamples;
2. number of parents «+» — 20;
3. number of parents «-» — 30.

The results of computer research: 8 cases — right positive extension of a definition one by one, 24 cases — right negative extension of a definition one by one, 1 case — wrong positive extension of a definition one by one. Thus, the accuracy of the present study was 95%, and completeness — 92%.

Discussion of results

The offered intelligent system is the instrument of support of medical researches with difficult structured data and number of facts being vast without use of computer technologies. Results of the JSM-method use, as means for the analysis of medical data, demonstrate usefulness of this method, which is the new tool of evidence-based medicine.

Unfortunately, common term "evidence-based medicine" isn’t translated as "medicine based on the obvious facts". Methods relating to the evidence-based medicine are based on the findings, and these facts are used as arguments to these decisions in medicine, particularly for diagnostic purposes.

But the meaning of this term is that the decisions made by physicians must be reasoned (but not proven — in fact the authors talk about the probable prognosis, i.e. hypotheses!), which means that the decisions use arguments drawn from clinical data.

Indeed, the hypotheses about the causes of the effects found in BF by identifying similarities are the arguments or counter-arguments in favor of hypotheses generated by intelligent system about the presence or absence of the studied effect.

Above-said is an argument in favor of the statement that the JSM-method is a useful instrument of evidence-based medicine [7, 8].

Therefore, this confirms the possibility of using the JSM-method for automatic generation of hypotheses in the field of medical diagnostics, where the lack of formal knowledge can be "compensated" by rich factual material. In some areas of medicine, at the stage of phenomenological description and data storage, logic and mathematical methods restore the causal relationships, serve to enhance knowledge on the basis of available evidence and support the intellectual capacities of medical experts.

References

1. Алгоритм прогнозирования рецидивирующего непрерывного течения язвенного колита у детей на основе математического моделирования / Э. Н. Федулова, Е. И. Шубанина, Е. А. Жукова [и др.] // Фарматека. — 2014. — № 6. — С. 74–79.
2. ДСМ-метод автоматического порождения гипотез: логические и эпистемологические основания / Под ред. О. Аншакова. — М. : Либроком, 2009. — 432 с.
3. Интеллектуальные системы для анализа медицинских данных. Ч. 1 / В. К. Финн, В. Г. Блинова, Е. С. Панкратова, Е. Ф. Фабрикантова // Врач и информационные технологии. — 2006. — № 5. — С. 62–70.
4. Интеллектуальные системы для анализа медицинских данных. Ч. 2 / В. К. Финн, В. Г. Блинова, Е. С. Панкратова, Е. Ф. Фабрикантова // Врач и информационные технологии. — 2006. — № 6. — С. 50–60.
5. Интеллектуальные системы для анализа медицинских данных. Ч. 3 / В. К. Финн, В. Г. Блинова, Е. С. Панкратова, Е. Ф. Фабрикантова // Врач и информационные технологии. — 2007. — № 1. — С. 51–57.
6. Механизм развития панкреатической недостаточности при хроническом панкреатите / Л. В. Винокурова, И. Е. Трубицына, Е. А. Дубцова [и др.] // Клиническая геронтология. — 2014. — Т. 20, № 1–2. — С. 46–49.
7. О применении интеллектуальной компьютерной системы для анализа клинических данных больных меланомой / И. Н. Михайлова, Е. С. Панкратова, Д. А. Добрынин [и др.] // Российский биотерапевтичский журнал. — 2010. — Т. 9, № 2. — С. 54
8. Флетчер Р. Клиническая эпидемиология. Основы доказательной медицины / Р. Флетчер, С. Флетчер, Э. Вагнер / Пер. с англ. под общ. ред. С. Е. Бащинского, С. Ю. Варшавского. — М.: Медиа Сфера, 2004. — 347 с.
9. Харнас С. С. Дооперационная и интраоперационная диагностика хирургических заболеваний поджелудочной железы / С. С. Харнас, Ю. А. Кулезнева, Д. Лачман // Хирургия. — 2003. — № 3. — С. 62–65.

10. Хронический панкреатит и рак поджелудочной железы / Л. Б. Лазебник, Л. В. Винокурова, Н. И. Яшина [и др.] // Клиническая и экспериментальная гастроэнтерология. — 2012. — № 7. — С. 3–9.
11. Frequent detection of pancreatic lesions in asymptomatic high-risk individuals / M. I. Canto, R. H. Hruban, E. K. Fishman [et al.] // Gastroenterology. — 2012. — Vol. 142, No 4. — P. 796–804.
12. Krejs G. J. Pancreatic cancer: epidemiology and risk factors / G. J. Krejs // Dig. Dis. — 2010. — Vol. 28, No 2. — P. 355–358.
13. Schneider A. The M-ANNHEIN classification of chronic pancreatitis: introduction of a unifying classification system based on review of previous classification of the disease / A. Schneider, J.-M. Löhr, M. V. Singer // J. Gastroenterol. — 2007. — Vol. 42. — P. 101–119.

The use of intelligent JSM-typed system for the analysis of clinical data

L.V. Vinokurova, M. A. Agafonov, G. G. Varvanina, V. K. Fynn, E. S. Pankratova, D. A. Dobrynin

[1]*Central research institute of gastroenterology, Moscow, Russia*

[2]*RAS Russian institute of scientific and technical information, Moscow, Russia*

Key words: chronic pancreatitis, intelligent system, clinical analysis, JSM-method, diabetes

Mathematical methods of analysis and mathematical modeling of pathological processes more often come to help the general practitioners today.

Aim of research is to apply a JSM-method of data analysis in intelligent systems to predict the development of diabetes of pancreatogenic genesis in patients with chronic pancreatitis.

Materials and methods. The intelligent JSM-typed system (IntJSM), based on the JSM-method of automatic generation of hypotheses, was created in the sector of intelligent systems of RAS VINITI. Existing IntJSM includes: IntJSM = Solver of tasks + Information environment (base of the facts (BF) and knowledge base (KB) + Intelligent interface (dialogue + representation of results + learning to work with system).

Results. At the first stage, by means of JSM-system by rules of a plausible conclusion of the 1st sort (induction), using the training sample of 81 patients, hypotheses of the 1st sort were generated: presence of a certain set of features in the patient was the reason for presence or absence of the diagnosis diabetes of pancreatogenic genesis.

At the second stage, with the use of hypotheses of the 1st sort by rules of the 2nd sort, presence or absence of diabetes of pancreatogenic genesis was predetermined in 42 patients presented for the prognosis.

As a result of computer research, 10 patients were predetermined positively, i.e. the diabetes of pancreatogenic genesis was diagnosed, 28 patients were predetermined negatively, i.e. they weren’t predicted to have the diabetes of

pancreatogenic genesis. Results of to definition coincided with real medical data. Thus, the accuracy of predetermination in this computer research was equal to 100%, and completeness — to 90%.

Conclusion. The offered intelligent system is the instrument of support of medical researches with difficult structured data and facts being vast without use of computer technologies. Results of the JSM-method use, as means for the analysis of medical data, demonstrate usefulness of this method, which is the new tool of evidence-based medicine.