UDC 61:004.651(075.8)

**Andrushchak I.E., Silchuk Yu.V., Prubush I.O., Umsha O.S., Semeniuk J.R.**

# SOFTWARE IMPLEMENTATION OF NUMERICAL METHODS OF OPTIMAL CONTROL IN THE PROBLEM DECISION SYSTEM

***The work is dedicated to software implementation of the direct method of optimal control and demonstrate it for solving tasks of preventive medicine. Software implementation is available graphical interpretation, which allows its use in practice.***

***Keywords:*** *numerical method for the optimal control model of coexistence of the two strains, the method of DCA, Java, SQL.*

**Андрущак І.Є., Сільчук Ю.В., Прибиш І.О., Умша О.С., Семенюк Я.Р.**

# ПРОГРАМНА РЕАЛІЗАЦІЯ ЧИСЕЛЬНОГО МЕТОДУ ОПТИМАЛЬНОГО КЕРУВАННЯ В ЗАДАЧАХ ПРОФІЛАКТИЧНОЇ МЕДИЦИНИ

***В роботі розглядається програмна реалізація прямого методу знаходження оптимального керування і демонстрування його для отримання розв'язку задач профілактичної медицини. Програмна реалізація має доступну графічну інтерпретацію, що робить можливим її використання на практиці.***

**Ключові слова:** чисельний метод оптимального керування, модель співіснування двох штамів вірусу, метод DCA, Java, SQL.

**Андрущак И.Е., Сильчук Ю.В., Прибыш И.О., Умша О.С., Семенюк Я.Р.**

# ПРОГРАММНАЯ РЕАЛИЗАЦИЯ ЧИСЛЕННОГО МЕТОДА ОПТИМАЛЬНОГО УПРАВЛЕНИЯ В ЗАДАЧАХ ПРИНЯТИЯ РЕШЕНИЙ

***В работе рассматривается программная реализация прямого метода нахождения оптимального управления и демонстрации его для получения решения задач принятия решений. Программная реализация имеет доступную графическую интерпретацию, что делает возможным ее использование на практике.***

***Ключевые слова:*** *численный метод оптимального управления, модель сосуществования двух штаммов вируса, метод DCA, Java, SQL.*

**Introduction.** In medicine, the term "differential diagnosis" means a systematic approach based on evidence, to determine the causes of the symptoms observed in the case when there are several alternative explanations, and to reduce the list of possible diagnoses.

Today, medical diagnostics can be performed automatically using computer systems and algorithms. Such systems are usually referred to diagnostic decision support system or medical diagnostic systems. They belong to a more general class of clinical decision support systems [Martsenyuk, 2004-2012]. The purpose of these systems is the systematic support of the doctor in the differential diagnosis. Many of these systems can provide results even when not enough data, ie under uncertainty, and most importantly – they are limited as to the amount of information that can store and process.

One approach that reflects the natural process of thinking in the differential diagnosis is a method of induction of decision trees. During the late 1970s, early 1980s J.R.Quinlan [Quinlan, 1986] developed an algorithm for constructing decision tree ID3 (Iterative dyhotomayzer). Later J.R.Quinlan presented algorithm C4.5 (successor ID3), which has become a benchmark, which is often compared to the latest algorithms in the field of computer knowledge. In 1984, a group of statisticians (L.Breiman, J.Friedman, R.Olshen, C.Stone) published work onClassification and Regression Trees (CART) [Breiman, 1984], which described the construction of binary decision trees. The algorithms ID3 and CART, despite the fact that developed independently and at about the same time, implementing this approach to learning decision trees based on the training data. This decision trees are built as a result of the recursive procedure of the "top-down". Most algorithms for induction of decision trees also fit this general approach. This study set recursively divided into smaller subsets as far as how the tree is constructed.

Mathematically, the problem is the induction of decision trees is formulated as follows. We set $D$, that contains $N$ sets of training data. In addition, each $i$ -й set $(A_1^i, A_2^i, \dots, A_p^i, C^i)$ consists of input data – attributes $A_1, \dots, A_p$ and output data - an attribute class $C$. Attributes $A_1, \dots, A_p$ can take as numerical and categorical values. Attribute class $C$ take one of $K$ discrete values: $C \in \{1, \dots, K\}$. The aim is to predict the decision tree class attribute value $C$ based on attribute values $A_1, \dots, A_p$. This should maximize the accuracy of the prediction attribute class, namely $P\{C=c\}$ at the terminal nodes for any $c \in \{1,\ldots,K\}$. Algorithms induction of decision trees automatically divided into units of numerical values of attributes $A_i$ two intervals: $A_i \le x_i$ та $A_i > x_i$, and categorical attributes $A_j$ –two subsets: $A_j \in S_j$, $A_j \notin S_j$. Breakdown based numerical attributes tend to Mirach based on entropy or Gini index [Han, 2001]. The process is repeated recursively partitioning as long as there will not improve the accuracy of forecasting. The last step involves removal of nodes to avoid overfitinhu model. As a result we get a set of rules that go from the root to each terminal node containing inequality for numerical attributes and conditions for the inclusion of categorical attributes.

The aim is to develop a method of induction of decision trees with the ability to program implementation as a clinical expert system.

**The method of induction of decision trees. The basis is taken of this recursive procedure** [Han, 2001].

*Generating a decision tree*

**Incoming data:** $D$ –set of training data sets $(A_1^i, A_2^i, \ldots, A_p^i, C^i)$.

**Output:** decision tree.

**Method:**

1. Create node $N$.

2. If all the sets in $D$ belong to a common class $C$, then return node $N$ a sheet with the name of the class $C$.

3. If the list of attributes (and therefore $D$) is empty, then return node $N$ a sheet with the name of the most common class $D$.

4. Apply algorithm attribute selection from the list of attributes for the set $D$ with the aim of finding the "best" attribute division.

5. Remove attribute division from the list of attributes.

6. For each subject division $j$ attribute to consider separation $D_j$ –a plurality of sets $D$, satisfying separation $j$.

7. If $D_j$ -empty, then connect to the site $N$ leaf under the most common class $D$, otherwise – attached to $N$ node returned recursive method call Generation of the decision tree input $D_j$ and a list of attributes.

8. End cycle step 6.

9. Return unit $N$.

Based on attribute selection algorithm $j$ –step of recursion put such information indicator:

$$Gain(A_i) = Info(D_j) - Info_{A_i}(D_j). \tag{1}$$

Here

$$Info(D_j) = -\sum_{k=1}^{K} p_k^j \log_2 (p_k^j) \tag{2}$$

- Information needed for classification set $(A_1, A_2, \ldots, A_p)$ in $D_j$,

$$Info_{A_i}(D_j) = \sum_{l=1}^{K_i} \frac{\#(D_j^l)}{\#(D_j)} Info(D_l) \tag{3}$$

- Information needed for classification $(A_1, A_2, \ldots, A_p)$ in $D_j$ after separation $D_j$ into subsets $D_j^l$ according to the attribute values $A_i$.

In the formula (2) the probability that a random set of $D_j$ belongs to the set $C_{k,D_j}$ rated as $p_k^j = \frac{\#(C_{k,D_j})}{\#(D_j)}$, where $C_{k,D_j}$ –plural sets of $D_j$, which attribute class $C = k$ .Here $\#(\bullet)$ –number of items in the set.

In the formula (3) $\frac{\#(D_j^l)}{\#(D_j)}$ –estimate the probability that a random set of $D_j$ belongs to the set $D_j^l$, де $D_j^l$ – Set sets with $D_j$, which attribute $A_i = a_i^l$. This attribute $A_i \in \{a_i^1, a_i^2, \ldots, a_i^{K_i}\}$.

So, $Gain(A_i)$ evaluates reduce the information required to classify a random data set $D_j$ by known attribute value $A_i$. Thus the available attributes for each node of the decision tree for separation conditions should be selected attribute $A_{i^*}$ with the largest value $Gain(A_{i^*})$. As a result of the selection process to finalize the classification of the data set in $D_j$ try the least information.

**Software implementation.** The method is implemented in Netbeans development environment in the programming language Java. Base training data deployed on the server MySQL. On fig.1 presents a conceptual model of information system. In class DecisionTree directly implemented method of induction of decision trees. In class DataManager received calls from DecisionTree to query the mysql database to obtain training data.

Mysql database consists of two tables – Table attribute, designed to store information about the attributes and tables categorized_data – for sets of training data. The structure of the tables in the SQL language for the examples below:

```
CREATE TABLE mysql.attribute (
        id integer not null unique,
        attribute_name varchar(25),
        attribute_field_name varchar(25),
    primary key (id)
  ) ENGINE=InnoDB;
 CREATE TABLE mysql.categorised_data (
        id integer not null unique,
        A1 varchar(12),
        A2 varchar(8),
        A3 varchar(7),
        A4 varchar(7),
        A5 varchar(7),
        class varchar(8),
    primary key (id)
 ) ENGINE=InnoDB;
```

Software Project classes are included in the package decision_tree.model. This includes beans-class Attribute, Attribute_for_list and CategorisedData to work with relevant data tables. SQL-queries to obtain relevant data, including calculations of the information provided in the class realized AttributeListPeer.

**Conclusions.** The paper considered the program implementation of decision trees induction based on the information provided.

In the example shown, this approach allows to develop a system to support clinical decisions.

It is shown that it has sufficient SQL syntax features that allow information to calculate performance-based database tables.

Through the use of Java-class implementing this method of decision tree induction is a web-integrated.

Prospects of research is to analyze the performance of the software depending on the number and volume attributes sets of training data.

**List of references:**

1. J.Han and M.Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, San Francisco, $1^{st}$ edition, 2001.
2. T.Hastie, R.Tibshirani and J.H.Friedman, The Elements of Statistical Learning, Springer, New York, $1^{st}$ edition, 2001.
3. C.Ordonez, Comparing association rules and decision trees for disease prediction, In Proc. ACM HIKM Workshop, 2006, pp. 17-24.
4. C.Ordonez, Integrating K-means clustering with a relational DBMS using SQL, IEEE Transactions on Knowledge and Data Engineering (TKDE) 18(2) (2006), 188-201.
5. J.R.Quinlan. Induction of decision trees. Machine Learning, 1: 81-106, 1986.
6. J.R.Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.
7. L.Breiman, J.Friedman, R.Olshen, and C.Stone. Classification and Regression Trees. Wadsworth International Group, 1984.
8. Марценюк В.П., Кравец Н.О. О программной среде проектирования интеллектуальных баз данных // Клиническая информатика и телемедицина – 2004. – №1. – с.47-53.
9. Марценюк В.П. Математичні моделі в системі підтримки прийняття рішень страхового забезпечення лікування онкологічних захворювань: підхід на основі динаміки Гомперца / В.П. Марценюк, І.Є. Андрущак, І.С. Гвоздецька,Н.Я. Климук// Доповіді Національної академії наук України. –2012. – №10. – С. 34-39.
10. Марценюк В.П. Підхід на основі актуарних математичних моделей до задач страхової медицини / В. П.Марценюк, І.Є. Андрущак, Н.Я Климук // Медична інформатика та інженерія. Науково-практичний журнал. – 2010. – №4. – С. 85-87.

Стаття надійшла до редакції 18.10.2016