METODY EKSPLORACJI WIEDZY W BAZIE DANYCH PACJENTÓW Z DYSFUNKCJĄ GŁOSU

KNOWLEDGE DISCOVERY METHODS IN VOICE DISORDERS DATABASE

Keywords: knowledge discovery, apriori, association rules

1. INTRODUCTION

Mining association rules is used to solve problems in many fields of science, including e.g. medicine. Search association rules is designed to improve the performance of database systems in terms of functionality allowing you to ask complex queries. The research database was created on the basis of the survey. The database refers to patients with voice disorders. It consists of 60 objects described by 68 attributes (numerical, symbolical and decision attribute). In this article, we conducted tests using Weka for association rule mining from data using Apriori algorithm.

2. ASSOCIATIONS RULES

We assume that $S = (X, A, V)$ is an information system, where $X$ is a nonempty, finite set of objects, $A$ is a nonempty, finite set of attributes, $V = \{ V_a : a \in A \}$ is a set of all attributes values.

Then, $a : X \rightarrow V_a$ is a function for any $a \in A$, that returns the value of the attribute of a given object. The attributes are divided into three different categories: set of stable attributes $A_1$, set of flexible attributes $A_2$ and set of decision attributes $D$, such that $A = A_1 \cup A_2 \cup D$ [3,4].

Mining Association is one of the most important data mining functionality, which works on the principle of association rules between items that are significant in the database. Obtained results form the basis for decision-making and forecasting, which is undoubtedly a great advantage of the described method [1,5]. In the first step, each set of items is called an itemset, if the support for the set is higher than a minimum threshold of support [2]. In the second step we generate rules. To confirm the rule, for example $X \rightarrow Y$, where $X$ and $Y$ are items, which are based on a confidence threshold, which is the ratio of the two items containing all transactions that contain $X$ also contain $Y$. The minimum support and confidence as a limitation for the value below the thresholds [1,2].

3. METODOLOGY

Based on the survey of 60 people we built a database consisting of 60 objects and 68 attributes. The data refer to issues related to fonoaudiology, speech therapy and voicediseases.
This database has been prepared in the extension .arff, which is accepted by the installation program, Weka. We use the Apriori Algorithm and we want to find the association rules in our database. For the purposes of analysis, the following values:
- Minimum support: 0.9 (54 instances),
- Minimum metric <confidence>: 0.9 are taking into consideration.

In the analysis we obtained interesting association rules as follows:

1. \( \text{REFLUX_TREAT} = \text{no} \ 57 \implies \text{ASTHMA} = \text{no} \ 57 \) conf: (1)
2. \( \text{VOICE_PER} = 2 \ \text{REFLUX_TREAT} = \text{no} \ 55 \implies \text{ASTHMA} = \text{no} \ 55 \) conf: (1)
3. \( \text{HORM_DISORDES} = \text{no} \ 54 \implies \text{ASTHMA} = \text{no} \ 54 \) conf: (1)
4. \( \text{EDU} = 2 \ \text{REFLUX_TREAT} = \text{no} \ 54 \implies \text{ASTHMA} = \text{no} \ 54 \) conf: (1)
5. \( \text{EDU} = 2 \ 57 \implies \text{ASTHMA} = \text{no} \ 56 \) conf: (0.98)
6. \( \text{VOICE_PER} = 2 \ 57 \implies \text{ASTHMA} = \text{no} \ 56 \) conf: (0.98)
7. \( \text{THYROID_SURGERY} = \text{no} \ 56 \implies \text{ASTHMA} = \text{no} \ 55 \) conf: (0.98)
8. \( \text{VOICE_PER} = 2 \ \text{ASTHMA} = \text{no} \ 56 \implies \text{REFLUX_TREAT} = \text{no} \ 55 \) conf: (0.98)
9. \( \text{GENDER} = \text{K} \ 55 \implies \text{ASTHMA} = \text{no} \ 54 \) conf: (0.98)
10. \( \text{LARYNG_SURGERY} = \text{no} \ 55 \implies \text{ASTHMA} = \text{no} \ 54 \) conf: (0.98)

These are the best rules found in our data. To analyze the results, we limited ourselves to describe the attributes that have occurred in the resulting rules. They are as follows:

\[
\begin{align*}
\text{@attribute} \quad \text{'REFLUX_TREAT'} & \{\text{'no'},\text{'yes'}\} \\
\text{@attribute} \quad \text{'ASTHMA'} & \{\text{'no'},\text{'yes'}\} \\
\text{@attribute} \quad \text{'VOICE_PER'} & \{0,1,2\} \quad 0 \ (0-2\text{y}), \ 1 \ (2-10 \text{y}), \ 2 \ (> 10 \text{y})
\end{align*}
\]

\[
\begin{align*}
\text{@attribute} \quad \text{'HORM_DISORDES'} & \{\text{'no'},\text{'yes'}\} \\
\text{@attribute} \quad \text{'EDU'} & \{0,1,2\} \quad \text{0-primary, 1-secondary, 2-higher} \\
\text{@attribute} \quad \text{'THYROID_SURGERY'} & \{\text{'no'},\text{'yes'}\} \\
\text{@attribute} \quad \text{'GENDER'} & \{\text{'K'},\text{'M'}\} \\
\text{@attribute} \quad \text{'LARYNG_SURGERY'} & \{\text{'no'},\text{'yes'}\}
\end{align*}
\]

4. CONCLUSION

Apriori is a common algorithm used in mining transaction database in order to extract frequent patterns. In this article, it has been implemented by means of the Weka. In the future, the results may be useful in the treatment of people with voice disorders.

REFERENCES