Data Mining to Build A Pattern of Knowledge From Psychological Consultation

Sri Mulyana
Computer Science Study Program
Department of Computer Sciences and Electronics
Gadjah Mada University
Bulaksumur Yogyakarta
smulyana@ugm.ac.id

Sri Hartati
Computer Science Study Program
Department of Computer Sciences and Electronics
Gadjah Mada University
Bulaksumur Yogyakarta
shartati@ugm.ac.id

Retantyo Wardoyo
Computer Science Study Program
Department of Computer Sciences and Electronics
Gadjah Mada University
Bulaksumur Yogyakarta
rw@ugm.ac.id

Edi Winarko
Computer Science Study Program
Department of Computer Sciences and Electronics
Gadjah Mada University
Bulaksumur Yogyakarta
ewinarko@ugm.ac.id

ABSTRACT

Psychological problems require a special handling that is usually carried out by a psychologist through a consultancy. Reasoning Systems or Expert Systems for psychological counseling, it is possible to developt. This paper will describe the process of data mining to build a pattern of knowledge from psychological consultation notes that will be used to develop an expert system in the next step. The process uses Rapidminer 4.0 software with decision tree and rule learner methods. The results of the process of data mining by decision tree shows that the data converge on the type of action ass (assistance). The results of the data mining method by the rule learner managed to find a pattern of knowledge of rules 15 rules.

Keywords

Data Mining, Knowledge Discovery, Case Based Reasoning, decision tree, rule learner.

1. INTRODUCTION

One of the most current approaches to artificial intelligence involves constructing programs that function as narrowly focused experts called expert systems. Expert system is a new innovation in the capture and integrates knowledge, has the ability to duplicate the expertise of an expert in a particular field.

Psychological problems require special handling that is usually done by a psychologist through a consultancy. The more complex problems of life will create a psychological problem and needed more the role of psychological experts to solve. Expert Systems for psychological counseling, it is possible to help resolve the issue.

The development of Expert Systems requires a large data management process of psychological consultation notes. The overall knowledge of the search process is called Knowledge Discovery in Databases (KDD). Data mining is one step in process of KDD using of a specific algorithm in the se pattern in database. This paper will explain the process of mining to construct a pattern of knowledge from psychological consultation notes, which will be used to dev the expert system in the next step using Case Based Reaso techniques.

In this research, data mining process has been carried o build a pattern based on knowledge of psychology counsel The process uses Rapidminer 4.0 software with decision tree rule learner methods.

2. CASE BASED REASONING

Case-Based Reasoning (CBR) has become a succes technique for knowledge-based systems in many domains short definition of case-based reasoning is that it methodology for solving problems by utilizing prev experiences (Kolodner, 1993). In case-based reasoning reasoner solves a new problem by noticing its similarity to o several previously solved problems and by adapting their kn solutions instead of working out a solution from scratch.

The problem-solving life cycle in a CBR system co essentially of the following four parts (Aamodt & Plaza, 1994

1. ***Retrieving*** similar previously experienced cases problem–solution–outcome triples) whose problem is ju to be similar
2. ***Reusing*** the cases by copying or integrating the solu from the cases retrieved
3. ***Revising*** or adapting the solution(s) retrieved in an att to solve the new problem
4. ***Retaining*** the new solution once it has been confirme validated

The relationship between these steps can be presented below:

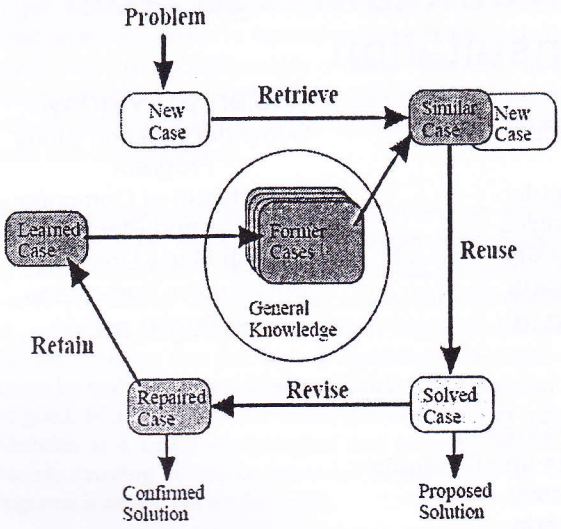

Figure 1. CBR cycle (Aamodt & Plaza, 1994)

A new problem, described as a case, is compared to the existing cases in the case base and the most similar case or case are retrieved. These cases are combined and reused to suggest a solution for the new problem. The solution proposed may be needed to be revised somewhat if it is not a valid solution. Then, the verified solution is retained by adding it as a new case to the case base or as amandements to existing cases in the case base for use in the future problem solving.

3. DATA MINING

'Data Mining' may be defined as the process of searching and analizing data in order to find implicit but potentially useful information [Frawley at all, 1991]. There are two types of data mining in its aplication (Kantardzic, 2003).

a. *Predictive data mining*
b. *Descriptive data mining*

The most popular data mining techniques and frequently used in solving the problem are (Pramudiono, 2003):

a. *Classification*
   Classification is the processes to find a model or function that distinguish data classes, in order to estimate the class of an object with unknown label. The model can be "IF-THEN" rules, *decision tree*, mathematical formula or *neural network*.
b. *Association Rule*
   An association rule is a data mining technique to find the combination of associative rule items. The importance of association rules can be identified based on two parameters y *support* and *confidence*. Support is the percentage of item combinations in the database, and confidence is the level of relations between items in association rule. The Algorithm commonly used in association rule is apriori algorithm.
c. *Clustering*
   Clustering is a process of classifying data that are not based on a specific data class. In fact, clustering can be used to provide the labels of unknown class data. Clustering is often classified as an *unsupervised learning* method.

4. METHODOLOGY

This research is part of the whole research process in the context of preparing a dissertation, entitled "Computer Reasoning System for Expert based on case" with a case study on the psychological consultation. The purpose of this research is to build a knowledge base about the psychological consultation problem. The research methods are designed as follows:

a. Data Collection
   This stage is for collecting data that contain records of clinical psychological consultation from the psychological consultant. This is a very important and strategic stage, because the computer reasoning system is developed based on the knowledge base obtained from the data.

b. Data Normalization
   Data normalization covering several processes as follows :
   - *Data Cleaning:* a process for removing noise and inconsistent data.
   - *Data integration*: the process of collecting data from various sources into a data base called data warehouse
   - *Data selection and transformation*: a process for transformting data into a form suitable for data mining and then making the selection of data.

   The purpose of this normalization is to change and choose the data that can be implemented on data mining algorithms.

c. Data Mining Process
   The purpose of this process is to build knowledge related to psychological consultation. This process will utilize Rapidminer 4.0 software.
d. Knowledge Verification
   Consultation with the psychological consultant about the knowledge that are generated from data mining process.

5. RESULTS AND DISCUSSION

The data records that have been normalized was used for data mining process. Based on the data mining process, the knowledge generated in the form of production rules (IF .. THEN) contains attribute of age, sex, class, medical diagnosis, and psychological diagnosis for antecedent, and attribute treatment for concequent. The process uses Rapidminer 4.0 software with decision tree and rule learner methods. The process for detail is as follows :

a. **Data input for Rapidminer**
   The form of file *.aml was used for Data input. The file was saved as <label></label> for the consequent and <attribute></attribute> for the antecedent. The file will call other file which form *.data acording to the sequence.
   Figure-2 is an example for the *.aml file.

```
<attributeset default_source="penelitian.data">
 <attribute
   name       = "umur"
   sourcecol  = "1"
   valuetype  = "integer"
 />
 <attribute
   name       = "jenis_kelamin"
   sourcecol  = "2"
   valuetype  = "nominal">
```

```
      <value>L</value>
      <value>p</value>
 </attribute>
 <attribute
  name       = "kelas"
  sourcecol  = "3"
  valuetype  = "nominal">
      <value>III</value>
      <value>II</value>
      <value>I</value>
  <value>VIP</value>
 </attribute>
 <attribute
  name       = "dx_medik"
  sourcecol  = "4"
  valuetype  = "nominal"
 />
 <attribute
  name       = "dx_psy"
  sourcecol  = "5"
  valuetype  = "nominal"
 />
 <label
   name     = "jenis_tindakan"
   sourcecol = "6"
   valuetype = "nominal">
 </label>
</attributeset>
```

Figure 2. Sourcecode Penelitian.aml

Figure-3 is an example for the *.data file.

```
30,L,III,conipUTHXIIFr.a,stress,ass-kons
13,L,III,Scoliosis,stress,ass
38,L,III,paraparesis,stress,ass-kons
50,L,III,RowstFractULIFr.B,Denial,ass-kons
27,L,III,SponditotitosisULI,cemas,ass
43,p,II,LBHSusp.Hnp,stress,ass
36,L,III,paraplegia,stress,ass
31,L,III,paraplegia,adjustime,kons
6,p,III,Scoliosis,normal,kons
...
```

Figure 3. Sourcecode Penelitian.data

b. Data Mining Process

1) *Decision Tree Method*

The output of data mining process by decision tree method could be shown as follow :

```
ass {ass-kons=61, ass=196, kons=16,
psiter-psitest=1, ass-psiter=19, ass-
kons-psiter=2, kons-psiter=1, psiter=2,
ass-psitest=2, Ass-Psiter=37, Psiter=40,
Ass-Kons=49, Ass-Kons-Psiter=3, Ass=49,
PsiDasar-Konseling=5, PsiDasar-
Psikoterapi=5, PsiDasar=1,
Psikoterapi=4, PlayTerapi-Konseling=1,
PsiDasar-Konseling-Psikoterapi=1, Ass-
PsiTest=2,
Kons=3}
```

The results of the process of data mining by decision tree shows that the data converge on the type of action ass (assistance), because most of the records (196 records) have the kind of ass (assistance) action although antecedents diverse.

2) *Rule Learner Method*

This method will extract the rule from penelitian.aml data input. Numerical data was divided in the range form using *FrequencyDiscretization*. The feature of age was divided into 3 ranges. Then *number of bins=3* and sample ratio = 1.0 was chosed. It means that all records were included in the determination of the rule.

This process generates the rules shown in Figure-4.

Figure-4 show that :

- There are 15 rules.
- The first rule : if dx_psy = normal then ass (5/51/1/0/1/0/0/0/0/0/0/0/0/0/0) mean that if psycological diagnosis = normal then the treatment = ass. The sequence numbers show the number of records that support and unsupport the rule. There are 51 supported data and 7 unsupported data for the rule.
- "*correct*: 277 *out of* 500 *training examples*" mean that there are 277 data from 500 data supporting all of the rules. So the significance of this rule was 55 % (277/500 * 100 %).

```
1. if dx_psy = normal then ass
(5/51/1/0/1/0/0/0/0/0/0/0/0/0/0)
2. if dx_psy = cemas then ass
(66/149/16/0/23/2/0/2/1/0/2/0/0/0/0/1)
3. if dx_psy = MoodDisorder then ass
(18/23/0/0/1/1/0/0/0/0/0/0/0/0/0)
4. if kelas = III then psiter
(14/15/1/0/21/2/1/31/1/0/0/0/2/0/0/0)
5. if jenis_kelamin = L then ass-psiter
(3/2/1/1/6/0/0/3/0/0/1/1/1/0/0/1)
6. if kelas = II then psiter (0/2/0/0/2/0/0/5/0/0/0/0/0/0/0/0)
7. if dx_psy=anxietas then psiDasar-konseling
(2/2/0/0/1/0/0/0/0/4/2/0/0/1 /0/0)
8. if dx_medik = Fr.InterochanterFemur then ass-kons
(1/0/0/0/0/0/0/0/0/0/ 0/0/0/0/0/0/0)
9. if dx_medik = MultipleFraktur then psiter
(0/0/0/0/0/0/0/1/0/0/0/0/0/0/ 0/0)
10. if umur = range1 then ass-psiter
(0/0/0/0/1/0/0/0/0/0/0/0/0/0/0/0)
11. if dx_medik = Fr.compresiVertebrae then psikoterapi
(0/0/0/0/0/0/0/0/0/0/0 /0/1/0/0/0)
12. if dx_medik = Multifractur then psiDasar-konseling-
psikoterapi (0/0/0 /0/0/ 0/0/0/0/0/0/0/0/0/1/0)
13. if dx_medik = Fr.collFemur then psiDasar-konseling
(0/0/0/0/0/0/0/0/0/0 /1/0/ 0/0/0/0/0/0)
14. if umur = range2 then ass
(0/1/0/0/0/0/0/0/0/0/0/0/0/0/0/0)
15. else ass-kons (1/0/0/0/0/0/0/0/0/0/0/0/0/0/0/0)
correct: 277 out of 500 training examples.
```

Figure 4. Generated rules

6. CONCLUSIONS

Data mining process by decision tree method can't generates pattern of knowledge. On the other hand, data mining process by the rule learner method generated 15 rules and the degree of significant was 55 % .

7. REFFERENCES

[1] Aamodt, A., 1994, *Case-based reasoning: Foundation issues. AICOM* 7 39-59

[2] Frawley W.J, Shapiro G.P., Matheus C.J., 1991, *'Knowledge Discovery in Database : An Overviw'*, in Knowledge Discovery in Database, AAAI Press, Menlo Park, CA, pp. 1-27

[3] Kartardzic, M., 2003, *Data Mining : Concepts, Models, Methods, and Algorithm.* John Wiley and Sons

[4] Kolodner, J., 1993, *Case-based reasoning.* ISBN: 1-55860-237-2, Morgan Kaufmann, San Mateo

[5] Pramudiono, I., 2003, *Pengantar Data Mining : Menambang Pertama Pengetahuan di Gunung Data,* http://www.ilmukomputer.com