Generation of Explicit Knowledge from Temporal Data

Norita Md Norwawi[a], Ku Ruhana Ku-Mahamud[b] and Safaai Deris[c]

Department of Computer Science
Faculty of Information Technology
Universiti Utara Malaysia, 06010 UUM Sintok, Kedah
Tel : 04-9284788, Fax : 04-9284753, E-mail : rita@uum.edu.my

Department of Computer Science
Faculty of Information Technology
Universiti Utara Malaysia, 06010 UUM Sintok, Kedah
Tel : 04-9284700, Fax : 04-9284753, E-mail : ruhana@uum.edu.my

Department of Software Engineering
Faculty of Computer Science and Information Science
Universitu Teknologi Malaysia, 83130 Skudai, Johor
Tel : 07-5532343, Fax : 07-5565044, E-mail : safaai@fsksm.utm.my

ABSTRACT

Expert decisions are usually based on experience knowledge regarding the decision environments. However due to the mobility or high turn-over of expert whether due to promotion, resignation or other reasons, there is a need to capture expert knowledge especially those involved critical decision. This paper will describe an approach of capturing expert knowledge through past operation data. An explicit knowledge extraction technique from temporal data representing reservoir operation will be presented. Rules regarding decision on the number of spillway gates to be open to release excess water from the dam will be discovered from the historical data. Temporal information related to a decision will be captured through sliding window method and classified into unique decision classes. This information can be transformed into decision rules for easy interpretation and understanding. These rules can be further use to guide future operation decisions.

Keywords

Knowledge Extraction, Temporal Data Mining, Knowledge Discovery

1.0 INTRODUCTION

In an organization, knowledge assets reside in many different places: database, knowledge base, text documents and human mind. Knowledge assets are the knowledge own by the organization or needs to own where targeted and timely knowledge is needed to guide business problems solution.

There are two types of knowledge: tacit and explicit knowledge. Tacit knowledge is hard to formalize and articulate such as intuitions, heuristics, best practices and special know-how. An explicit knowledge is codified, documented knowledge or rule-based knowledge (KIm, 1999) of the form situation-action pairs to guide action. It is used in designing organization routines, standard operation procedures and structure data records. Operational control and efficiency can be achieved through explicit knowledge.

Temporal data is a time-series data. The data is ordered and in sequence of time. The data is dependent to each other. An event could be an effect of a cause after some time delay. For example, floods could occur after two or three consecutive days of rainfall. Operation data usually keeps a record of situation per time index. This data can be formalized, organized into useful information where relations and patterns can be identified.

2.0 PROBLEM DEFINITION

In an organization, personnel mobility or turnover is common due to promotion, resignation or some other causes. There is a need to capture their expert knowledge especially the justification of any decisions made. This study is an effort to capture decision patterns from operation data. These decision patterns can be transformed into decision rules that can guide future decisions. The decision that is of interest is on opening or closing the spillway gates. If the gates need to be open, how many gates should be open suitable with the current situation. In this paper, technique for extracting explicit knowledge from temporal data will be described using reservoir operation as a case study.

3.0 KNOWLEDGE EXTRACTION TECHNIQUE.

A sliding window technique (Lee et al., 2001) is used to capture the knowledge from the data. The extraction process consists of the following steps:

a. Define the event characterization function.
b. Transform the data into discrete form.
c. Scan data and locate event.
d. Frame the event and situations preceding it in window slice.
e. Collect all events and store in window slices set.
f. Classify window slices into unique decision classes in the form of temporal patterns.
g. Transform the patterns into rule-based format.

Figure 1 shows the pipeline of the extraction process.

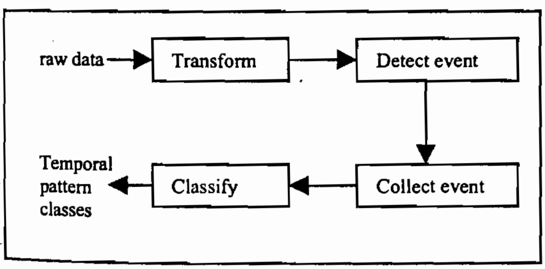

Figure 1: Pipeline of the extraction process

The general algorithm is as shown in Figure 2.

```
Transform data into discrete form

For all t in the time index
  If detect changing point t_p
  copy into window set P_w for all variables from t_{p-w+1} to t_p
  end if

For all variable vj in window Pw
    classify P_w according to similar value of v_j
```

Figure 2: Basic extraction algorithm

Figure 3 below shows an example of event detection and formation of window slices.

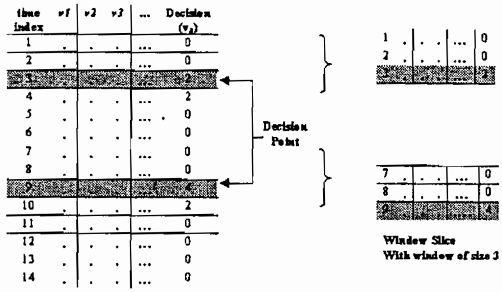

Figure 3: An example of event detection and formation of window slices

Windows are classified based on the decision value. In this case the window will be grouped according to the number of gates opened. From Figure 3 above, there are two decision classes, decision with 2 gates and 4 gates.

The performance measurement of the extraction process was adopted from Povinelli's (1999) and Gryzmala-Busse's (2000) as shown in Table 1.

Table 1: Performance measurement

Measurement	Meaning
t_p	True positive. Number of event correctly predicted
f_p	False positive. Number of predicted event but in actual non-event.
t_n	True negative. Number of non-event correctly predicted
f_n	False negative.Number of predicted non-event but in actual non-event.
Sensitivity (Sv)	The accuracy of correctly predicted event $= \dfrac{t_p}{t_p + f_n}$
Specificity (Sp)	The accuracy of correctly predicted non-event $= \dfrac{t_n}{t_n + f_p}$
$f(P)$	The ratio of total correct prediction $= \dfrac{t_p + t_n}{t_n + f_n + t_p + f_p}$

The extraction process is done n batch. All training data were read and classified. The performance will be measured by using the extracted patterns against each dataset in the training dataset.

4.0 IMPLEMENTATION AND RESULT

The algorithm was tested using C language and validated with real reservoir operation data of Timah-Tasoh dam in the state of Perlis. The data used were operation data from 1998 to 2002 since the installation of 6 rainfalls stations. Variables used are reservoir water level, rainfall measurements at 6 rainfall stations, the number of gates opened and the day of the year. It is a daily operation data. The event characterization function is as defined in equation (1) below.

$$e_t = g_{t-1} < g_t \text{ where } g_{t-1} = 0 \text{ and } g_t > 0 \qquad (1)$$

This implies that event is detected when there is a change of decision from close to opening the gate. The time point will be kept for the identifying the window frame.

The transformation of data into discrete form is based on domain expert category from Drainage and Irrigation Department (DID), Malaysia (Timah Tasoh Operation and Maintenance Manual, 1993). Table 2 below shows the categories used.

Table 2 : Domain Expert Category

Value	Water Level /m	Flood Stage
0	<29.0	Normal
1	< 29.4	Alert
2	< 29.6	Warning
3	> 29.6	Danger

Value	Rainfall /mm	Category
0	0	None
1	1-10	Light
2	11-30	Moderate
3	31-60	Heavy
4	> 60	Very heavy

The algorithm was run iteratively and the window size that gave the optimum performance is 2. This implies that the gate decision depends on the rainfall measurement of yesterday *(t-1)* and the day before yesterday *(t-2)*. Based on this window size, the temporal patterns extracted were as in Table 3 below.

Table 3: Temporal Pattern Classes

Number of Gates Open = 2			
Flood Stage/t	Avg Rain t-2	Avg Rain t-1	%Δ
0	1	1	0.1760
0	1	1	0.0701
0	3	3	0.1744
1	0	2	0.2226
1	2	1	0.0515
1	3	2	0.3607
1	4	1	0.3963
1	2	2	0.3442
1	0	3	0.4112
1	2	3	0.5659
1	2	2	0.2059
1	3	1	0.1711
1	1	1	0.1025

Number of Gates Open = 4			
Flood Stage/t	AvgRain t-2	AvgRain t-1	%Δ
0	4	4	0.7495
0	2	3	0.9609
0	2	1	0.0690
0	3	2	0.5759
0	2	2	0.3119
1	2	2	0.4480
1	2	2	0.2914
1	2	1	0.0684
1	4	2	0.3626

There are two decision classes, opening two gates and four gates. The list in each class shows the condition when the gates were first opened. This information can be transformed in rule based form as below :

Rule 1:

```
If flood stage is Normal at t and
   average rainfall at t-2 is Light and
   average rainfall at t-1 is Light
then
   open 2 spillway gates
```

Rule 2:

```
If flood stage is Alert at t and
  average rainfall at t-2 is Moderate and
  average rainfall at t-1 is Moderate
then
  open 4 spillway gates
```

The performance of this extraction process based on Table 1 are shown below in Table 4.

Table 4 : Algorithm Performance

Dataset	*tp*	*fp*	*tn*	*fn*	*Sv*	*Sp*	*f(P)*
Set1	2	12	272	0	1	0.958	0.958
Set2	9	21	334	0	1	0.941	0.942
Set3	11	18	341	0	1	0.950	0.951

From Table 4, it shows that the extraction process gave about 6% false alarm where

$$false\ alarm\ rate = Sv\text{-}Sp \quad (2)$$

This implies that the algorithm is able to detect all event ($Sv=1$) but there are false alarm cases when the gates need not to be opened were recommended otherwise. This is not really critical as compared not to be able to detect when gates should be opened.

5.0 CONCLUSION

This paper has shown a technique of extracting explicit knowledge from temporal data. The explicit knowledge can be represented in a rule-based form. This rules can be use to guide future decisions. By using a simple sliding window technique, events can be captured and classified. The algorithm is able to detect all event with some minor false alarm of gate opening.

6.0 REFERENCES

Grzymala-Busse, J.W., Zheng,X., Goodwin,L.K. and Grzymala-Busse,W.J. (2000). *Comparing Different Strategies in Using Rules Induced from Pre-term Birth Data*. In Proc. of the Information Processing and Management of Uncertainty in Knowledge Based System, Madrid, Spain, July 3-7, pp388-394.

Kim, S. (1999). *The Roles of Knowledge Professionals for Knowledge Management*. In the Proccedings of the 65th IFLA Council and General Conference, Bangkok, Thailand. Aug 20-28.

Lee, C., Lin, Ç. and Chen, M. (2001). *Sliding-Window Filtering: An efficient algorithm for incremental Mining*. In Proc. of the 10th International Conference on Information and Knowledge Management, p263-270. ISBN 1-58113-436-3

Povinnelli, R.J. (1999). *Time Series Data Mining: Identifying Temporal Patterns for Characterization and Prediction of Tim,e Series Events*. Phd Thesis, Marquette University , Winsconsin, USA.1999

Timah Tasoh Dam Operation and Maintenance Manual (1993) Drainage and Irrigation Department, Perlis, Malaysia.