ORIENTAL JOURNAL OF
COMPUTER SCIENCE & TECHNOLOGY
An International Open Free Access, Peer Reviewed Research Journal
Published By: **Oriental Scientific Publishing Co., India.**
www.computerscijournal.org

ISSN: 0974-6471
December 2014,
Vol. 7, No. (3):
Pgs. 385-389

Using Vectors of Features for Finite State Automata Dataset Reduction

TAWFIQ A. AL-ASSADI and ABBOOD KIREBUT JASSIM

[1]Collage of Information Technology, Babylon University, Iraq.

(Received: November 06, 2014; Accepted: December 01, 2014)

ABSTRACT

A finite state automata is the most important type of graphs ,which is called conceptual graphs, while the expansion of using the graphs in the process of data mining, the use of FSM is still limited because of the difficulty in processing in databases, therefore in order to find methods that make it easier to deal with large groups of machines, as a database, is encourage to use of this type of representation in this paper of graph mining . This paper present a method using vectors of features for find machines matching, which is one task of mining graph data , which are frequently found in a single environment or similar environments, thereby reducing the number of records and increase efficiency of mining tasks.

Key words: Data reduction, Essential machines, FSM**,** Graph mining, Machine matching vectors of features.

INTRODUCTION

The graph mining has been raped sudden increase in the data mining. Graph mining denotes a group of algorithms for mining the relational aspects of data represented as a graph.[1]

The graph depiction that set of nodes and links between nodes allow practice of mining algorithm. Graph-based data mining has two major approaches: frequent subgraph mining and graph relational data [2].The center of graph mining is to extract important knowledge from graph represented data by techniques from fields such as data mining, machine learning, pattern recognition, statistics and graph theorypattern recognition and graph theory. [1].

The graph is a data construction[3](which implements the algorithms on which a number of types of graph, including together direct and indirect, labeled or not and weighted or not.[4]

Finite-state machines offer a simple computational form with many applications [5]. finite-state machine (FSM) or finite-state automaton (plural: automata), or simply a state machine, is a arithmetical model of computation used to propose both computer programs and sequential logic circuits. It is conceived as an abstract machine that can be in one of a finite number of states. The machine is in only one state at a time; the state it is in at any given time is called the current state. It can transform from one state to another when initiated by a triggering occurrence or condition; this is called a transition. A particular FSM is defined by a set of its states, and the triggering condition for each transition.

fsm are important technique of graphs [6] and formulas that define the conceptual graphs and consists of states, transitions, and be in two basic formats [7].

Machines Matching Method

The selected database is composed of machines that represent data in an environment or similar other environments to enable the mining algorithms to find the best results, the following steps depict the method used as.

First, all FSM are converted into vectors of features that are extracted from environment or r environments of similar features to make the database composed of only vectors representing machines.

Second step is the process of detecting matches between vectors by comparing each one with the other vectors features in the dataset and when there is a similarity between them, the number of the original vector placement to the similar vector is given and so on and hence two types of transactions in the database.

The first type are essential records and the second type are the similar records,(see in the table 1).

Table. 1: Grequent FSM in data set

Machine	Replacement
F1	1
F2	2
F3	1
F4	1
F5	2
F6	1
F7	1
F8	1
F9	1
F10	2
F11	1
F12	2

The data mining tasks will deal with the first type, while the second type ignored.

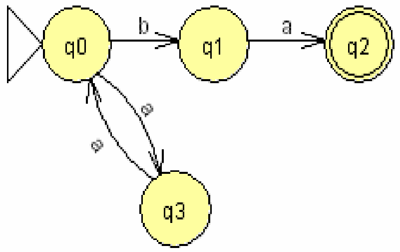
The calculation of the matching process need efficient function to get the job done accurately and therefore be choose the equality function as in the proposed algorithm ,see in fig(2).

The process continues until all vectors are allocated correctly.

The dataset are then considered to be unique machines, see table (2).

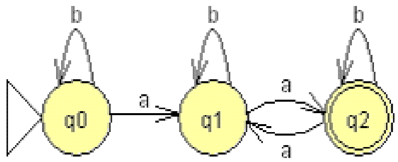
Table. 2: essential type of machine(first type)

F1
F2



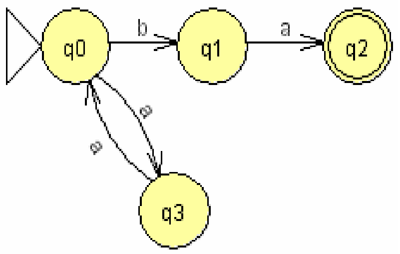
1 1 1 1 0 0 0

Fig. 1: FSM1 with their vector



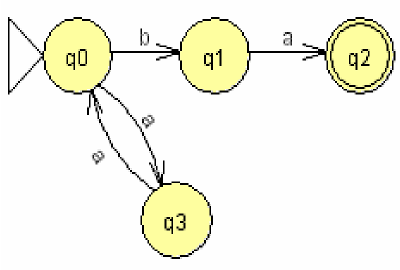
1 0 0 0 0 / 1

Fig .2: FSM2 with their vector



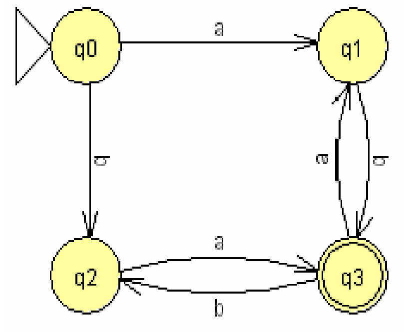
1 1 1 1 0 0 0

Fig. 3: FSM3 with their vector



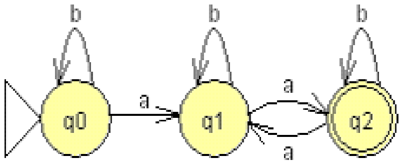
1 1 1 1 0 0 0

Fig. 4: FSM4 with their vector



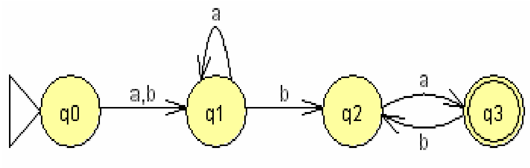
1 1 1 1 0 0 0

Fig. 5: FSM5 with their vector



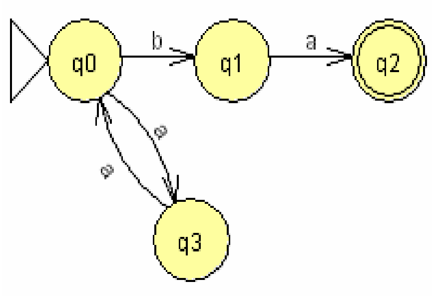
1 0 0 0 0 1 1

Fig. 6: FSM6 with their vector



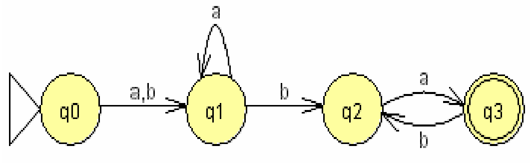
1 1 1 1 0 0 0

Fig. 7: FSM7 with their vector



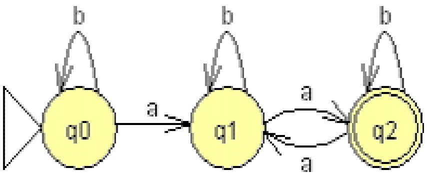
1 1 1 1 0 0 0

Fig. 8: FSM8 with their vector



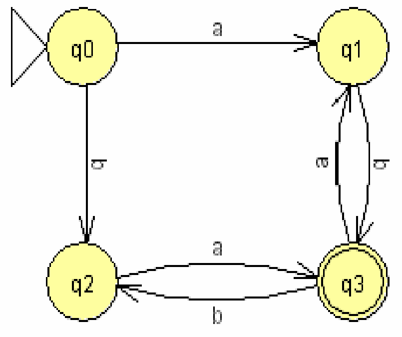
1 1 1 1 0 0 0

Fig. 9: FSM9 with their vector



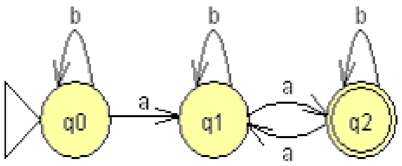
1 0 0 0 0 / 1

Fig. 10: FSM10 with their vector



1 1 1 1 0 0 0

Fig. 11: FSM11 with their vector



1 0 0 0 0 1 1

Fig. 12: FSM12 with their vector

To illustrate the method the twelve machines are used as an example to be converted in the vectors of features(1 for feature existence, 0 for not) with set of feature{a,b,ab,ba,ai,bai,abi} to be the dataset for mining task as shown in fig(1).

```
Input set of machines
Output essential of machines data set
for each machine
Extract features vector
End
Set sim[0],flag=0                    //sim is Similarity machine vector //
For k= 1 to n                        // n is a number of vectors
For i=1 to n
For j= 1 to m                        //m is a length of features vector
begin
if f(k,m)“‘ f(j,m)  flag=1
end j
if flag=0
sim(j)=k
if flag=1 then flag=0
end i
end k
for i= 1 to n
for j= i+1 to  n
if sim(i) “‘ -1
if sim(j)=sim(i)
sim(j)= -1
end for i
end for j
for I =1 to n
if sim(i) “‘ -1
ouput  f(i)
end for i
```

Figure (2) finite state machine matching algorithm

RESULTS AND DISCUSSION

After implementing the algorithm the placement for each finite state machine will be as seen in table (1).

By ignored the second type of machines the essential machines will be present, shown in table(2).

CONCLUSION

1. Data generates in one environment or in similar environments will generate many machines Similar.
2. The tasks of data mining is the most efficient and fastest method with Trimming effect.

REFERENCES

1. Daine J. Cook, Lawernce B. Holder, “mining graph data” ,willy,Washington 2007.
2. Yan, Xifeng, X. Zhou, and Jiawei Han, “frequent subgraph mining and graph relational knowledge Mining closed relational graphs with connectivity constraints.” *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. ACM, 2005.
3. Wiskott, Laurenz, et al. “Face recognition by elastic bunch graph matching.”*Pattern Analysis and Machine Intelligence”, IEEE Transactions on* 19.7 (1997): 775-779.).
4. Raymond, John W., Eleanor J. Gardiner, and Peter Willett. “Heuristics for similarity searching of chemical graphs using a maximum common edge subgraph algorithm.” *Journal of chemical information and computer sciences*42.2 (2002): 305-316.)
5. Casey, Mike. “The dynamics of discrete-time computation, with application to recurrent neural networks and finite state machine extraction.” *Neural computation* 8.6 (1996):

1135-1178.

6. Yen, Ti-Yen, and Wayne Wolf. "An efficient graph algorithm for FSM scheduling." *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* 4.1 (1996): 98-112.

7. Yang, Shih-Yao, and Von-Wun Soo, "Extract conceptual graphs from plain texts in patent claims." *Engineering Applications of Artificial Intelligence* 25.4 (2012): 874-887.