Adaptive Multi-levels Dictionaries and Singular Value Decomposition Techniques for Autonomic Problem Determination

Hoi Chan and Thomas Kwok

IBM Thomas J. Watson Research Center

19 Skyline Drive Hawthorne, NY 10532

{hychan, kwok}@us.ibm.com

Abstract

An autonomic problem determination system can adapt to changing environments, react to existing or new error condition and predict possible problems. In this report, we propose such a system using dynamic and adaptive multi-levels dictionaries and "Singular Value Decomposition techniques" (SVD). Compared to standard SVD, our system uses an iterative method that enables dynamic interaction between events and the current dictionaries with its entries being updated continuously to reflect relative importance of each event, thereby accelerating its convergence. The system captures knowledge in a hierarchical form for complex knowledge representation. It does not require a formal knowledge model or intensive training by examples. It is efficient with sufficient accuracy for autonomic problem determination.

1. Introduction

Problem determination system (PDS) is an integral part of an autonomic computer system. It plays a key role in data centers where unexpected error situations occur frequently. Traditional PDS becomes insufficient. In this paper, we propose a PDS by using Singular Value Decomposition technique (SVD) [1] together with an adaptive multi-levels dictionaries system (DD) that can react to new error situations, predict possible problems, and adapt to new environments. SVD is a classical statistical method and is widely used in latent semantic analysis for information retrieval and ink recognition [2]. Its use in autonomic system [3] has been explored recently. However, these prior studies did not consider the use of expert and learned knowledge to shorten search time and increase accuracy. Our proposed PDS enables dynamic interaction between events and the dictionaries with its entries being continuously updated to enhance correlation among the data. We further introduce the concept of "occurrence index", a user defined frequency and time based weighting factor assigned to each dictionary entry to represent its relative importance.

2. System Overview

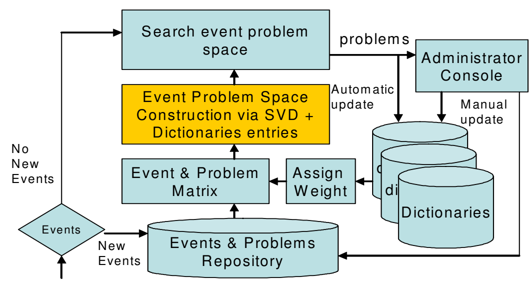

Figure 1. System Overview

Fig.1 shows the overview of the proposed PDS. It includes an event bus, an administrator console, an EP repository for storing events and problems patterns. An EP matrix module transforms an initial event-problem set obtained from the EP repository into an EP matrix, which is then decomposed by SVD transformation to form an n-dimensional EP space wherein events and closely associated problems are placed near one another. Problems which are closely associated with the incoming events in this space are selected. Administrators can accept, reject or modify the detected problems. Each episode of problem detection triggers the system to update the EP repository by putting a positive or negative weight on certain EP patterns.

A multi-levels dictionary assigns weight to each term in the EP matrix. The level of the dictionary indicates complexity of terms, for example, the first level contains only atomic terms, while higher level contains combination of terms. Each dictionary can be optionally initialized with expert defined entries and each entry is assigned an occurrence index (OI). In our PDS: **OI = (F*B)/T** where *F = occurrence of term for a fixed time interval, B = user input weight (0 to infinity, default =1), and T = elapsed time from last occurrence of term.* OI for each term is updated as events coming in and problems

0-7695-2779-5/07 $20.00 © 2007 IEEE

IEEE COMPUTER SOCIETY

detected. New terms can be added as new events occur. This dictionary system, and the EP repository, provides an adaptive knowledge base for up-to-date problem determination.

3. Experiment and Result

We used a simple example to compare the results of SVD and SVD-DD. We represent the nine event types as E1-9, and the associated problems as P1-5. For the SVD-DD, a dictionary lookup operation is performed for each term in the EP matrix, resulting in a matrix R (Fig. 2) where an OI based normalized weight is assigned to each term. E4 and E5 have new weight equals to 3 instead of 1 for an SVD based EP set. This matrix is decomposed into three matrices by standard SVD technique: **R = E S P'. E** and **P'** are the EP matrices of left and right singular vectors. **S** is the diagonal matrix of singular values. We choose the largest two singular values for a two dimensional space which represents the major correlation among events and problems where events are represented as *diamonds* and problems are represented as *squares in Fig. 3 & 4*. The dot product (cosine) between two component vectors corresponds to their estimated similarity.

	P1	P2	P3	P4	P5
E1	1	1		1	
E2	1			1	
E3					1
E4		3			
E5		3	3		
E6	1				1
E7			1		
E8	1				
E9					1

Figure 2. EP 9x5 Matrix after Dictionary lookup

With standard SVD and incoming events E4 and E5. There is no exact match in the EP patterns. A pseudo-problem (point q in Fig. 3) is constructed from E4 and E5 (centroid of E4 and E5). Two potential problems, **P2, P3** are within the dotted cone with a cosine value of 0.7 from q. With the same events E4 and E5 and SVD-DD, a pseudo-problem (point q in Fig. 4) is constructed from E4 and E5. **P2,** the most relevant problem**,** is the sole problem within the dotted cone (cosine value of 0.7 from q) while **P3** is outside. This result is more precise than standard SVD and it can be achieved with less computation time. If the administrator accepts the result, the dictionary will be updated. E4 and E5 which lead to P2 will also be recorded as new pattern in the EP repository for subsequent uses, completing the feedback loop for interactive learning.

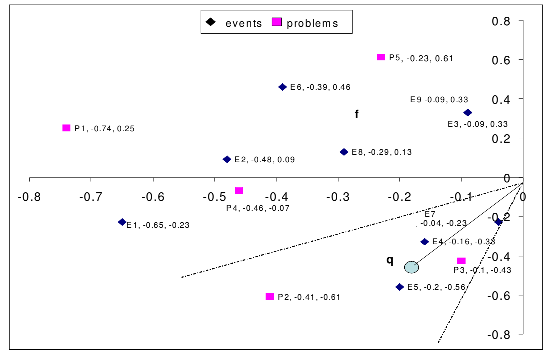

Figure 3. 2-D Plot of Es and Ps (SVD)

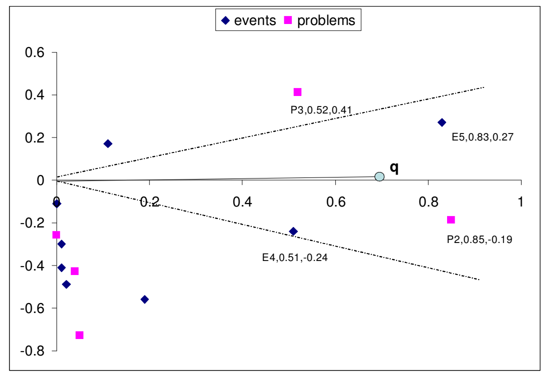

Figure 4. 2-D Plot of Es and Ps (SVD-DD)

4. Conclusion

The use of SVD for problem determination yields good results, especially in static EP sets. For ambiguous and new situations, it suffers from unpredictable accuracy and is computationally expensive. SVD-DD addresses these problems, not only it can handle ambiguous conditions and adapt to changing environments, search accuracy and computation time are greatly enhanced. The user defined index system for the dictionary allows the use of SVD-DD in different domains. Initial results are very promising. Future work includes experiments with real life large event and problem sets from a data center environment.

5. References

[1] J.E. Gentle, "Singular Value Factorization" in *Numerical Linear Algebra for Apps in Statistics,* Berlin: Springer-Verlag, (1998), pp.102-103.

[2] T. Kwok and M. Perrone, “Adaptive N-Best List Handwritten Word Recognition” in Proc. of 6th Int’l Conf. on Document Analysis and Recognition, (2001), pp. 168- 172.

[3] H. Chan and T Kwok, “Autonomic PD Agents using SVD for Ambiguous Situations” in Proc. of IEEE WIC/ACM IAT, (2006), pp. 270-275.