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

Hoi Chan and Thomas Kwok

IBM T.J.Watson Research Center
{hychan,kwok}@us.ibm.com

Motivation

- Problem determination system (PDS) is an integral part of an autonomic computer system
- Traditional approaches of symptom recognition relying on static authoring of pattern matching rules become insufficient
- The prior studies of SVD in autonomic system did not consider the use of expert and learned knowledge to shorten search time and increase accuracy.

Overview

- A PDS by using Singular Value Decomposition technique (SVD) together with an adaptive multi-levels dictionaries system (DD) that can react to new error situations, predict possible problems, and adapt to new environments

- It enables dynamic interaction between events and the dictionaries with its entries being continuously updated to enhance correlation among the data.

- The concept of “occurrence index”, a user defined frequency and time based weighting factor assigned to each dictionary entry to represent its relative importance, is also introduced in this study.

System Overview

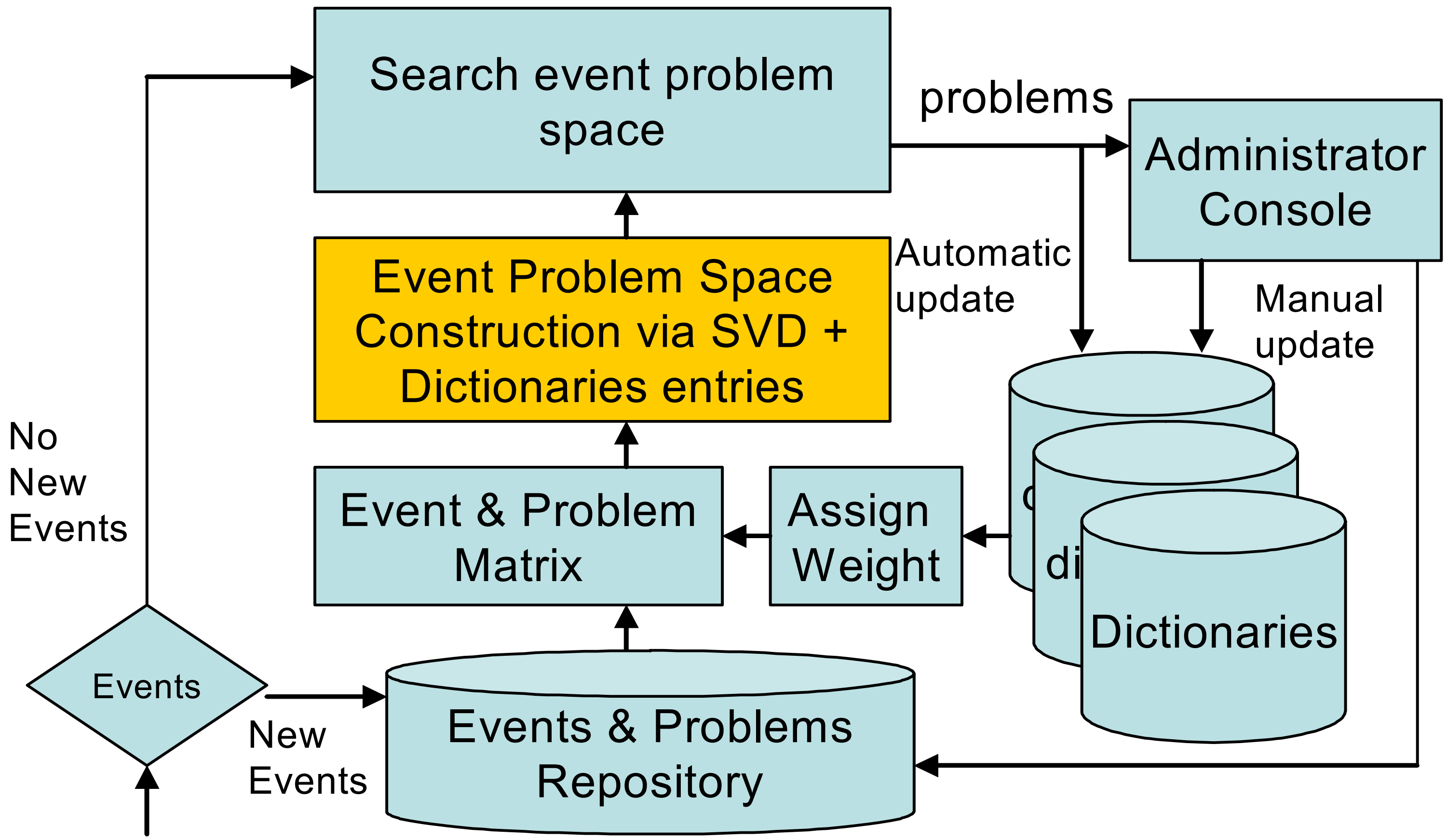

Figure 1

The proposed problem determination system

- The proposed PDS 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
- This matrix 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.

Adaptive multi-levels dictionaries system

- A multi-levels dictionary assigns weight to each term in the EP matrix
- The level of the dictionary indicates complexity 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**
- OI for each term is updated as events coming in and problems detected
- This dictionary system, and the EP repository, provides an adaptive knowledge base for up-to-date problems and events

A Statistical Approach

- A method by which problem symptoms can be recognized even in ambiguous situations
- Treats the observed event-symptom relationship represented by an event-symptom matrix as a statistical problem
- Using Singular Value Decomposition (SVD) technique, implicit higher order structure in the association of events with symptom is modeled to estimate event and symptom association

Singular Value Decomposition

- A classical mathematical technique closely related to a class of mathematical and statistical techniques, such as eigenvector decomposition, spectral and factor analysis

- Widely used in applications such as latent semantic analysis for information retrieval, ink retrieval from handwritten documents, document search

- Well-established theoretical foundation and readily available tools

Singular Value Decomposition

- Singular value decomposition takes a rectangular matrix of event and symptom data (defined as **A**, where **A** is an **m x n** matrix) in which the **m** rows represents the events, and the **n** columns represents the symptoms.

The SVD theorem states:

$$\mathbf{A}_{mxn} = \mathbf{E}_{mxm}\, \mathbf{S}_{mxn}\, \mathbf{P'}_{nxn}$$

Where

$$\mathbf{E}^{T}\mathbf{E} = \mathbf{I}_{mxm}$$

$$\mathbf{P'}^{T}\mathbf{P'} = \mathbf{I}_{nxn} \quad \text{(i.e. } \mathbf{E} \text{ and } P' \text{ are orthogonal)}$$

Where the columns of E are the left singular vectors (*gene coefficient vectors*); **S** (the same dimensions as **A**) has singular values and is diagonal (*mode amplitudes*); and **P'** has rows that are the right singular vectors (*expression level vectors*). The SVD represents an expansion of the original data in a coordinate system where the covariance matrix is diagonal

Singular Value Decomposition Calculation

Split A into E, S and P’ (visualNumerics library)

$$A = ESP'$$

Creation of Event-Symptom Matrix and Space by SVD Technique

A simplified set of events and symptoms:

E1 = request response time > 400ms
E2 = request queue length > 100
E3 = excessive logins in the entire system
E4 = excessive requests from a domain
E5 = excessive requests from individual IP
E6 = average server utilization > 90%
E7 = connection from unknown source
E8 = requests frequent timeouts
E9 = excessive unknown application terminations

Symptom 1 = saturated on demand router
Symptom 2 = unexpected peak demand
Symptom 3 = possible intruder attack
Symptom 4 = network congestion
Symptom 5 = serious security breach

Experiments

- With 9 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.
- The largest two singular values are chosen 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.

Creation of Event-Policy Matrix and Space by SVD Technique

This dataset consists of m events (**Em**) and n symptoms (**Pn**), where m=9 and n=5.

The m events are entered as rows and the n symptoms are entered as columns.

The entries in the event-symptom matrix are simply occurrences of events in different symptoms.

Figure 2

Events-Symptom Matrix - A Matrix

	S1	S2	S3	S4	S5
E1	1	1		1	
E2	1			1	
E3					1
E4		3			
E5		3	3		
E6	1				1
E7			1		
E8	1				
E9					1

2-D Plot of Es and Ps (SVD)

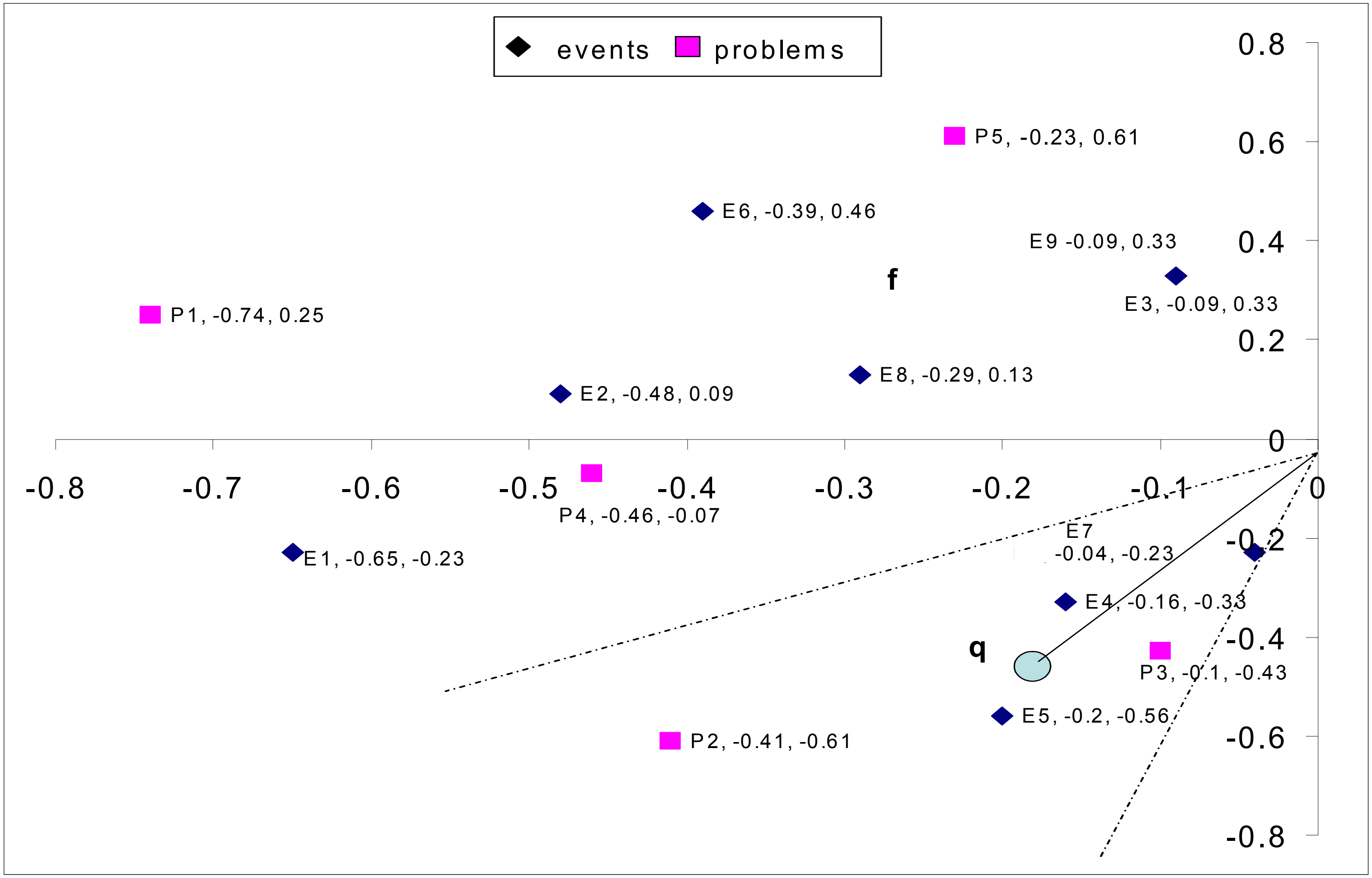

Figure 3

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

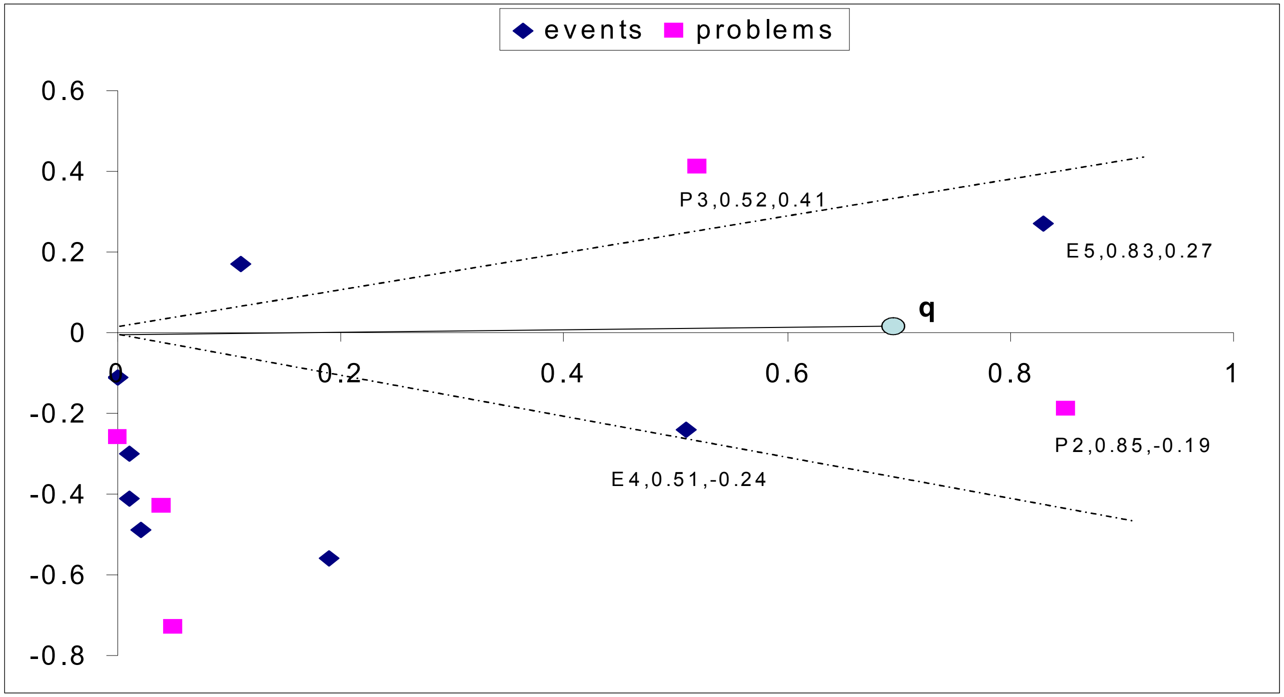

Figure 4

Results

- 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.

Conclusion / New Problems Solved

- 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
- Future work includes experiments with real life large event and problem sets from a data center environment.