US007483934B1

(12) **United States Patent**
Ide et al.

(10) **Patent No.: US 7,483,934 B1**
(45) **Date of Patent: Jan. 27, 2009**

(54) **METHODS INVOLVING COMPUTING CORRELATION ANOMALY SCORES**

(75) Inventors: **Tsuyoshi Ide**, Kanagawa (JP); **Spyridon Papadimitriou**, White Plains, NY (US)

(73) Assignee: **International Busniess Machines Corporation**, Armonk, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **11/959,073**

(22) Filed: **Dec. 18, 2007**

(51) **Int. Cl.**
G06F 17/15 (2006.01)

(52) **U.S. Cl.** **708/422**

(58) **Field of Classification Search** 708/422–426
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,621,365	A *	11/1986	Chiu	708/422
5,257,364	A *	10/1993	Melamed et al.	708/426
5,267,271	A *	11/1993	Rice	708/422
2005/0143976	A1	6/2005	Steniford	
2005/0283511	A1 *	12/2005	Fan et al.	708/306
2006/0161592	A1	7/2006	Ertoz et al.	
2007/0005256	A1 *	1/2007	Lincoln et al.	702/19
2007/0214133	A1 *	9/2007	Liberty et al.	707/5

OTHER PUBLICATIONS

Kenji et al., Dynamic Syslog Mining for Network Failure Monitoring, Aug. 21-24, 2005, ACM, pp. 499-508.*

J. Goldberger, S. Roweis, G. Hinton and R. Salakhutdinov. Neighbourhood Components Analysis, In Advances in Neural Information Processing Systems, 2005, pp. 513-520.

* cited by examiner

Primary Examiner—Chat C Do
(74) *Attorney, Agent, or Firm*—Cantor Colburn LLP; Vazken Alexanian

(57) **ABSTRACT**

An exemplary method for computing correlation anomaly scores, including, defining a first similarity matrix for a target run of data, the target run of data includes an N number of sensors, defining a second similarity matrix for a reference run of data, the target run of data includes the N number of sensors, developing a k-neighborhood graph N_i of the i-th node for the target run of data, wherein the k-neighborhood graph of the i-th node is defined as a graph comprising the i-th node and its k-nearest neighbors (NN), developing a k-neighborhood graph $\overline{N}_i$ of the i-th node for the reference run of data, defining a probability distribution p(j|i), wherein p(j|i) is the probability that the j-th node becomes one of the k-NN of the i-th node, coupling the probability between the i-th node and the neighbors of the i-th node, determining an anomaly score of the i-th node, and determining whether the target run of data has changed from the reference run of data responsive to determining the anomaly score of the i-th node.

1 Claim, 2 Drawing Sheets