



US007849124B2

(12) **United States Patent**
Ide

(10) **Patent No.:** **US 7,849,124 B2**

(45) **Date of Patent:** **Dec. 7, 2010**

(54) **METHOD AND SYSTEM FOR DETECTING
DIFFERENCE BETWEEN PLURAL
OBSERVED RESULTS**

5,574,641	A *	11/1996	Kawakami et al.	708/444
6,408,321	B1 *	6/2002	Platt	708/520
7,162,489	B2 *	1/2007	Folting et al.	707/102
7,346,593	B2	3/2008	Takeuchi et al.	
2005/0050129	A1	3/2005	Kamatani et al.	
2005/0283511	A1 *	12/2005	Fan et al.	708/306
2006/0101402	A1 *	5/2006	Miller et al.	717/124
2007/0005297	A1	1/2007	Beresniewicz et al.	

(75) Inventor: **Tsuyoshi Ide**, Kanagawa-ken (JP)

(73) Assignee: **International Business 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.: **12/264,556**

(22) Filed: **Nov. 4, 2008**

(65) **Prior Publication Data**

US 2009/0132626 A1 May 21, 2009

Related U.S. Application Data

(63) Continuation of application No. 11/768,967, filed on Jun. 27, 2007, now abandoned.

(30) **Foreign Application Priority Data**

Jul. 10, 2006 (JP) 2006-189301

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

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

(58) **Field of Classification Search** 708/400-446
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,046,179	A	9/1991	Uomori et al.
5,257,364	A	10/1993	Melamed et al.

OTHER PUBLICATIONS

Office Action from U.S. Appl. No. 11/768,967 dated Nov. 20, 2008. Keogh, et al.; Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases; 19 pages; May 16, 2000.

(Continued)

Primary Examiner—Chat C Do

(74) *Attorney, Agent, or Firm*—Schmeiser, Olsen & Watts; Shimokaji & Assoc.

(57) **ABSTRACT**

A method and system for analyzing time series data. In an embodiment, a loop is executed and terminated upon a specified maximum number of iterations of the loop being performed or upon a difference between scores in successive iterations of the loop not being greater than a specified tolerance, wherein the score in each iteration is calculated as function of an absolute value of a difference between respective cumulative probability values of first and second cumulative probability distributions which are generated from respectively first and second time series data sets. In an embodiment, time series data is processed in a sequence of time periods, wherein a combined cumulative probability distribution is generated in each time period by combining a cumulative probability distribution of new time series data with previously combined cumulative probability distribution data according to a ratio of the number of new to previous observed values.

1 Claim, 9 Drawing Sheets

