Subhash Bhalla (Ed.)

LNCS 3433

Databases in Networked Information Systems

4th International Workshop, DNIS 2005
Aizu-Wakamatsu, Japan, March 2005
Proceedings

Springer

Subhash Bhalla (Ed.)

Databases in Networked Information Systems

4th International Workshop, DNIS 2005
Aizu-Wakamatsu, Japan, March 28-30, 2005
Proceedings

Springer

Volume Editor

Subhash Bhalla
University of Aizu
Department of Computer Software, Database Systems Laboratory
Tsuruga, Ikki Machi, Aizu-Wakamatsu City, Fukushima, 965-8580, Japan
E-mail: bhalla@u-aizu.ac.jp

Library of Congress Control Number: Applied for

CR Subject Classification (1998): H.2, H.3, H.4, H.5, C.2

ISSN 0302-9743
ISBN 3-540-25361-0 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11401339 06/3142 5 4 3 2 1 0

Lecture Notes in Computer Science 3433

Lecture Notes in Computer Science

For information about Vols. 1–3332

please contact your bookseller or Springer

Vol. 3452: F. Baader, A. Voronkov (Eds.), Logic for Programming, Artificial Intelligence, and Reasoning. XI, 562 pages. 2005. (Subseries LNAI).

Vol. 3448: G.R. Raidl, J. Gottlieb (Eds.), Evolutionary Computation in Combinatorial Optimization. XI, 271 pages. 2005.

Vol. 3436: B. Bouyssounouse, J. Sifakis (Eds.), Embedded Systems Design. XV, 492 pages. 2005.

Vol. 3433: S. Bhalla (Ed.), Databases in Networked Information Systems. IX, 319 pages. 2005.

Vol. 3432: M. Beigl, P. Lukowicz (Eds.), Systems Aspects in Organic and Pervasive Computing - ARCS 2005. X, 265 pages. 2005.

Vol. 3427: G. Kotsis, O. Spaniol, Wireless Systems and Mobility in Next Generation Internet. VIII, 249 pages. 2005.

Vol. 3423: J.L. Fiadeiro, P.D. Mosses, F. Orejas (Eds.), Recent Trends in Algebraic Development Techniques. VIII, 271 pages. 2005.

Vol. 3422: R.T. Mittermeir (Ed.), From Computer Literacy to Informatics Fundamentals. X, 203 pages. 2005.

Vol. 3419: B. Faltings, A. Petcu, F. Fages, F. Rossi (Eds.), Constraint Satisfaction and Constraint Logic Programming. X, 217 pages. 2005. (Subseries LNAI).

Vol. 3418: U. Brandes, T. Erlebach (Eds.), Network Analysis. XII, 471 pages. 2005.

Vol. 3416: M. Böhlen, J. Gamper, W. Polasek, M.A. Wimmer (Eds.), E-Government: Towards Electronic Democracy. XIII, 311 pages. 2005. (Subseries LNAI).

Vol. 3415: P. Davidsson, B. Logan, K. Takadama (Eds.), Multi-Agent and Multi-Agent-Based Simulation. X, 265 pages. 2005. (Subseries LNAI).

Vol. 3414: M. Morari, L. Thiele (Eds.), Hybrid Systems: Computation and Control. XII, 684 pages. 2005.

Vol. 3412: X. Franch, D. Port (Eds.), COTS-Based Software Systems. XVI, 312 pages. 2005.

Vol. 3411: S.H. Myaeng, M. Zhou, K.-F. Wong, H.-J. Zhang (Eds.), Information Retrieval Technology. XIII, 337 pages. 2005.

Vol. 3410: C.A. Coello Coello, A. Hernández Aguirre, E. Zitzler (Eds.), Evolutionary Multi-Criterion Optimization. XVI, 912 pages. 2005.

Vol. 3409: N. Guelfi, G. Reggio, A. Romanovsky (Eds.), Scientific Engineering of Distributed Java Applications. X, 127 pages. 2005.

Vol. 3408: D.E. Losada, J.M. Fernández-Luna (Eds.), Advances in Information Retrieval. XVII, 572 pages. 2005.

Vol. 3407: Z. Liu, K. Araki (Eds.), Theoretical Aspects of Computing - ICTAC 2004. XIV, 562 pages. 2005.

Vol. 3406: A. Gelbukh (Ed.), Computational Linguistics and Intelligent Text Processing. XVII, 829 pages. 2005.

Vol. 3404: V. Diekert, B. Durand (Eds.), STACS 2005. XVI, 706 pages. 2005.

Vol. 3403: B. Ganter, R. Godin (Eds.), Formal Concept Analysis. XI, 419 pages. 2005. (Subseries LNAI).

Vol. 3401: Z. Li, L.G. Vulkov, J. Waśniewski (Eds.), Numerical Analysis and Its Applications. XIII, 630 pages. 2005.

Vol. 3398: D.-K. Baik (Ed.), Systems Modeling and Simulation: Theory and Applications. XIV, 733 pages. 2005. (Subseries LNAI).

Vol. 3397: T.G. Kim (Ed.), Artificial Intelligence and Simulation. XV, 711 pages. 2005. (Subseries LNAI).

Vol. 3396: R.M. van Eijk, M.-P. Huget, F. Dignum (Eds.), Agent Communication. X, 261 pages. 2005. (Subseries LNAI).

Vol. 3395: J. Grabowski, B. Nielsen (Eds.), Formal Approaches to Software Testing. X, 225 pages. 2005.

Vol. 3394: D. Kudenko, D. Kazakov, E. Alonso (Eds.), Adaptive Agents and Multi-Agent Systems III. VIII, 313 pages. 2005. (Subseries LNAI).

Vol. 3393: H.-J. Kreowski, U. Montanari, F. Orejas, G. Rozenberg, G. Taentzer (Eds.), Formal Methods in Software and Systems Modeling. XXVII, 413 pages. 2005.

Vol. 3391: C. Kim (Ed.), Information Networking. XVII, 936 pages. 2005.

Vol. 3390: R. Choren, A. Garcia, C. Lucena, A. Romanovsky (Eds.), Software Engineering for Multi-Agent Systems III. XII, 291 pages. 2005.

Vol. 3389: P. Van Roy (Ed.), Multiparadigm Programming in Mozart/OZ. XV, 329 pages. 2005.

Vol. 3388: J. Lagergren (Ed.), Comparative Genomics. VII, 133 pages. 2005. (Subseries LNBI).

Vol. 3387: J. Cardoso, A. Sheth (Eds.), Semantic Web Services and Web Process Composition. VIII, 147 pages. 2005.

Vol. 3386: S. Vaudenay (Ed.), Public Key Cryptography - PKC 2005. IX, 436 pages. 2005.

Vol. 3385: R. Cousot (Ed.), Verification, Model Checking, and Abstract Interpretation. XII, 483 pages. 2005.

Vol. 3383: J. Pach (Ed.), Graph Drawing. XII, 536 pages. 2005.

Vol. 3382: J. Odell, P. Giorgini, J.P. Müller (Eds.), Agent-Oriented Software Engineering V. X, 239 pages. 2005.

Vol. 3381: P. Vojtáš, M. Bieliková, B. Charron-Bost, O. Sýkora (Eds.), SOFSEM 2005: Theory and Practice of Computer Science. XV, 448 pages. 2005.

Vol. 3379: M. Hemmje, C. Niederee, T. Risse (Eds.), From Integrated Publication and Information Systems to Information and Knowledge Environments. XXIV, 321 pages. 2005.

Vol. 3378: J. Kilian (Ed.), Theory of Cryptography. XII, 621 pages. 2005.

Vol. 3377: B. Goethals, A. Siebes (Eds.), Knowledge Discovery in Inductive Databases. VII, 190 pages. 2005.

Vol. 3376: A. Menezes (Ed.), Topics in Cryptology – CT-RSA 2005. X, 385 pages. 2005.

Vol. 3375: M.A. Marsan, G. Bianchi, M. Listanti, M. Meo (Eds.), Quality of Service in Multiservice IP Networks. XIII, 656 pages. 2005.

Vol. 3374: D. Weyns, H.V.D. Parunak, F. Michel (Eds.), Environments for Multi-Agent Systems. X, 279 pages. 2005. (Subseries LNAI).

Vol. 3372: C. Bussler, V. Tannen, I. Fundulaki (Eds.), Semantic Web and Databases. X, 227 pages. 2005.

Vol. 3371: M.W. Barley, N. Kasabov (Eds.), Intelligent Agents and Multi-Agent Systems. X, 329 pages. 2005. (Subseries LNAI).

Vol. 3370: A. Konagaya, K. Satou (Eds.), Grid Computing in Life Science. X, 188 pages. 2005. (Subseries LNBI).

Vol. 3369: V.R. Benjamins, P. Casanovas, J. Breuker, A. Gangemi (Eds.), Law and the Semantic Web. XII, 249 pages. 2005. (Subseries LNAI).

Vol. 3368: L. Paletta, J.K. Tsotsos, E. Rome, G.W. Humphreys (Eds.), Attention and Performance in Computational Vision. VIII, 231 pages. 2005.

Vol. 3367: W.S. Ng, B.C. Ooi, A. Ouksel, C. Sartori (Eds.), Databases, Information Systems, and Peer-to-Peer Computing. X, 231 pages. 2005.

Vol. 3366: I. Rahwan, P. Moraitis, C. Reed (Eds.), Argumentation in Multi-Agent Systems. XII, 263 pages. 2005. (Subseries LNAI).

Vol. 3365: G. Mauri, G. Păun, M.J. Pérez-Jiménez, G. Rozenberg, A. Salomaa (Eds.), Membrane Computing. IX, 415 pages. 2005.

Vol. 3363: T. Eiter, L. Libkin (Eds.), Database Theory - ICDT 2005. XI, 413 pages. 2004.

Vol. 3362: G. Barthe, L. Burdy, M. Huisman, J.-L. Lanet, T. Muntean (Eds.), Construction and Analysis of Safe, Secure, and Interoperable Smart Devices. IX, 257 pages. 2005.

Vol. 3361: S. Bengio, H. Bourlard (Eds.), Machine Learning for Multimodal Interaction. XII, 362 pages. 2005.

Vol. 3360: S. Spaccapietra, E. Bertino, S. Jajodia, R. King, D. McLeod, M.E. Orlowska, L. Strous (Eds.), Journal on Data Semantics II. XI, 223 pages. 2005.

Vol. 3359: G. Grieser, Y. Tanaka (Eds.), Intuitive Human Interfaces for Organizing and Accessing Intellectual Assets. XIV, 257 pages. 2005. (Subseries LNAI).

Vol. 3358: J. Cao, L.T. Yang, M. Guo, F. Lau (Eds.), Parallel and Distributed Processing and Applications. XXIV, 1058 pages. 2004.

Vol. 3357: H. Handschuh, M.A. Hasan (Eds.), Selected Areas in Cryptography. XI, 354 pages. 2004.

Vol. 3356: G. Das, V.P. Gulati (Eds.), Intelligent Information Technology. XII, 428 pages. 2004.

Vol. 3355: R. Murray-Smith, R. Shorten (Eds.), Switching and Learning in Feedback Systems. X, 343 pages. 2005.

Vol. 3354: M. Margenstern (Ed.), Machines, Computations, and Universality. VIII, 329 pages. 2005.

Vol. 3353: J. Hromkovič, M. Nagl, B. Westfechtel (Eds.), Graph-Theoretic Concepts in Computer Science. XI, 404 pages. 2004.

Vol. 3352: C. Blundo, S. Cimato (Eds.), Security in Communication Networks. XI, 381 pages. 2005.

Vol. 3351: G. Persiano, R. Solis-Oba (Eds.), Approximation and Online Algorithms. VIII, 295 pages. 2005.

Vol. 3350: M. Hermenegildo, D. Cabeza (Eds.), Practical Aspects of Declarative Languages. VIII, 269 pages. 2005.

Vol. 3349: B.M. Chapman (Ed.), Shared Memory Parallel Programming with Open MP. X, 149 pages. 2005.

Vol. 3348: A. Canteaut, K. Viswanathan (Eds.), Progress in Cryptology - INDOCRYPT 2004. XIV, 431 pages. 2004.

Vol. 3347: R.K. Ghosh, H. Mohanty (Eds.), Distributed Computing and Internet Technology. XX, 472 pages. 2004.

Vol. 3346: R.H. Bordini, M. Dastani, J. Dix, A.E.F. Seghrouchni (Eds.), Programming Multi-Agent Systems. XIV, 249 pages. 2005. (Subseries LNAI).

Vol. 3345: Y. Cai (Ed.), Ambient Intelligence for Scientific Discovery. XII, 311 pages. 2005. (Subseries LNAI).

Vol. 3344: J. Malenfant, B.M. Østvold (Eds.), Object-Oriented Technology. ECOOP 2004 Workshop Reader. VIII, 215 pages. 2005.

Vol. 3343: C. Freksa, M. Knauff, B. Krieg-Brückner, B. Nebel, T. Barkowsky (Eds.), Spatial Cognition IV. Reasoning, Action, and Interaction. XIII, 519 pages. 2005. (Subseries LNAI).

Vol. 3342: E. Şahin, W.M. Spears (Eds.), Swarm Robotics. IX, 175 pages. 2005.

Vol. 3341: R. Fleischer, G. Trippen (Eds.), Algorithms and Computation. XVII, 935 pages. 2004.

Vol. 3340: C.S. Calude, E. Calude, M.J. Dinneen (Eds.), Developments in Language Theory. XI, 431 pages. 2004.

Vol. 3339: G.I. Webb, X. Yu (Eds.), AI 2004: Advances in Artificial Intelligence. XXII, 1272 pages. 2004. (Subseries LNAI).

Vol. 3338: S.Z. Li, J. Lai, T. Tan, G. Feng, Y. Wang (Eds.), Advances in Biometric Person Authentication. XVIII, 699 pages. 2004.

Vol. 3337: J.M. Barreiro, F. Martin-Sanchez, V. Maojo, F. Sanz (Eds.), Biological and Medical Data Analysis. XI, 508 pages. 2004.

Vol. 3336: D. Karagiannis, U. Reimer (Eds.), Practical Aspects of Knowledge Management. X, 523 pages. 2004. (Subseries LNAI).

Vol. 3335: M. Malek, M. Reitenspieß, J. Kaiser (Eds.), Service Availability. X, 213 pages. 2005.

Vol. 3334: Z. Chen, H. Chen, Q. Miao, Y. Fu, E. Fox, E.-p. Lim (Eds.), Digital Libraries: International Collaboration and Cross-Fertilization. XX, 690 pages. 2004.

Vol. 3333: K. Aizawa, Y. Nakamura, S. Satoh (Eds.), Advances in Multimedia Information Processing - PCM 2004, Part III. XXXV, 785 pages. 2004.

Preface

Information systems in healthcare and public utility services depend on computing infrastructures. Many research efforts are being made in related areas, such as wireless computing (e.g., Auto-ID Laboratories, and projects at MIT), Web-based computing, and information accesses by Web users. Government agencies in many countries plan to launch facilities in education, healthcare and information support as a part of e-government initiatives. In this context, information interchange management has become an active research field. A number of new opportunities have evolved in design and modeling based on new computing needs of users. Database systems play a central role in supporting networked information systems for access and storage management aspects.

The 4th International Workshop on Databases in Networked Information Systems (DNIS 2005) was held on March 28–30, 2005 at the University of Aizu in Japan. The workshop program included research contributions, and invited contributions. A view of research activity in Information Interchange Management and related research issues was provided by the session on this topic. The invited contribution was contributed by Dr. Umeshwar Dayal. The workshop session on Web Data Management Systems had invited papers by Prof. Elisa Bertino, Prof. Masahito Hirakawa, and Prof. William I. Grosky. The two sessions on Networked Information Systems included invited contributions by Prof. Sushil Jajodia, Dr. Cyrus Shahabi, Prof. Divyakant Agrawal and Dr. Harumi Kuno. I would like to thank the members of the Program Committee for their support and all authors who considered DNIS 2005 in making research contributions.

The sponsoring organizations and the Organizing Committee deserve praise for the support they provided. A number of individuals contributed to the success of the workshop. I thank Dr. Umeshwar Dayal, Prof. J. Biskup, Prof. D. Agrawal, Prof. Elisa Bertino, Dr. Cyrus Shahabi, Prof. Sushil Jajodia and Dr. Mark Sifer for providing continuous support and encouragement.

The workshop received invaluable support from the University of Aizu. In this context, I thank Prof. Ikegami, President of the University of Aizu. I thank Prof. Daming Wei, Head of the Department of Computer Software, for making the support available. I express my gratitude to the members and chairman of the International Affairs Committee, for supporting the workshop proposal. Many thanks are also due to the faculty members at the university for their cooperation and support.

March 2005 Subhash Bhalla

Organization

The DNIS 2005 international workshop was organized by the Database Systems Laboratory, University of Aizu, Aizu-Wakamatsu City, Fukushima, 965-8580, Japan.

Executive Committee

Honorary Chair	T. L. Kunii, Hosei University, Japan
Program Chair	S. Bhalla, University of Aizu, Japan
Organizing Chair	Hui Wang, University of Aizu, Japan
Executive Chair	N. Bianchi-Berthouze, University of Aizu, Japan

Program Committee

D. Agrawal, University of California, USA
F. Andres, National Center for Science Information Systems, Japan
N. Berthouze, University of Aizu, Japan
E. Bertino, Purdue University, USA
S. Bhalla, University of Aizu, Japan
P.C.P. Bhatt, Indian Institute of Information Technology, Bangalore, India
J. Biskup, University of Dortmund, Germany
L. Capretz, University of Western Ontario, Canada
M. Capretz, University of Western Ontario, Canada
U. Dayal, Hewlett-Packard Laboratories, USA
W.I. Grosky, University of Michigan-Dearborn, USA
J. Herder, University of Applied Sciences, Fachhochschule Dusseldorf, Germany
M. Hirakawa, Shimane University, Japan
Q. Jin, University of Aizu, Japan
A. Kumar, Pennsylvania State University, USA
H. Kuno, Hewlett-Packard Laboratories, USA
J. Li, University of Tsukuba, Japan
K. Myszkowski, Max-Planck-Institut fuer Informatik, Germany
T. Nishida, Kyoto University, Japan
A. Pasko, Hosei University, Tokyo, Japan
L. Pichl, University of Aizu, Japan
V. Savchenko, Hosei University, Japan
C. Shahabi, University of Southern California, USA
M. Sifer, Sydney University, Australia
H. Wang, University of Aizu, Japan

Sponsoring Institution

International Affairs Committee, University of Aizu,
Aizu-Wakamatsu City, Fukushima, 965-8580, Japan

Table of Contents

Information Interchange and Management Systems

Web Data Management Systems

Networked Information Systems: Applications

Networked Information Systems: Implementations

Data Management Systems

Predictive Business Operations Management

Malu Castellanos, Norman Salazar, Fabio Casati,
Umesh Dayal, and Ming-Chien Shan

Hewlett-Packard,
1501 Page Mill Road, MS 1142,
Palo Alto, CA, 94304, USA
firstanme.lastname@hp.com

Abstract. Having visibility into the current state of business operations doesn't seem to suffice anymore. The current competitive market forces companies to capitalize on any opportunity to become as efficient as possible. The ability to forecast metrics and performance indicators is crucial to do effective business planning, the benefits of which are obvious – more efficient operations and cost savings, among others. But achieving these benefits using traditional forecasting and reporting tools and techniques is very difficult. It typically requires forecasting experts who manually derive time series from collected data, analyze the characteristics of such series and apply appropriate techniques to create forecasting models. However, in an environment like the one for business operations management where there are thousands of time series, manual analysis is impractical, if not impossible. Fortunately, in such an environment, extreme accuracy is not required; it is usually enough to know whether a given metric is predicted to exceed a certain threshold or not, is within some specified range or not, or belongs to which one of a small number of specified classes. This gives the opportunity to automate the forecasting process at the expense of some accuracy. In this paper, we present our approach to incorporating time series forecasting functionality into our business operations management platform and show the benefits of doing this.

1 Introduction

The current competitive market situation where only the best can survive has triggered the radar for opportunities to become more efficient than the competitors. One such opportunity is given by good business planning with the goal of better meeting business objectives. Having visibility into the past and current state of business operations, as traditionally done by monitoring applications, is not enough anymore. More and more, there is a pressing need to project future performance. In particular, the ability to forecast metrics and performance indicators is crucial to do effective business planning, the benefits of which are obvious – more efficient operations and cost savings, among others. In consequence, forecasting (or *predictive analytics*, as it is now being called in the industry) is increasingly being considered by companies as a key technology to remain or become competitive. We are witnessing this trend (along with a mindset change in high-level managers to embrace prediction

S. Bhalla (Ed.): DNIS 2005, LNCS 3433, pp. 1 – 14, 2005.

technology) in virtually every domain, from telecom to supply chain, from transportation to finance. For example, transportation companies need to forecast the number of packages they will need to dispatch from a certain hub (or at least whether that number will exceed a given threshold) to plan for trucks. Correct planning allows more efficient shipments, avoids delays and saves costs. In a nutshell, forecasting provides the foundation for good planning inside and across organizations. The need for forecasting tools is so high that companies are willing to spend millions of dollars in consulting efforts to obtain these forecasts.

Traditionally, forecasting has been a human-driven task where domain forecasting experts interactively massage historical data, analyze it, try different algorithms, iterate over this cycle and finally come up with a forecasting model. This however, is absolutely impractical when there is the need of forecasting many different metrics and when the business conditions change dynamically so that prediction models need to be updated. In this situation automating the forecasting process seems to be the only practical solution. The problem is that automation can only come at the expense of accuracy and therefore in applications where extreme accuracy is required, there is no other way than to continue with the human-driven process whose output is expected to be a perfectly fit model. For example, accuracy is extremely important when providing forecasts of a company's revenue to financial analysts, because even a slight deviation of the actual revenue delivered from the forecast can have severe business consequences such as a drop in the company's stock price.

Fortunately, there are other applications, like the one discussed in this paper, where such great accuracy is not required and an approximate prediction model obtained quickly and automatically is preferred to a more accurate model obtained after long, expensive consulting efforts. In this case, automation of the forecasting process seems to be appropriate. For example, when metrics defined over business processes have thresholds that indicate undesired business operations behavior, forecasts that indicate whether future values are likely to be in the red, yellow or green zones are enough to decide the next action to take, for instance, proactively trying to prevent a metric from reaching a given value. As can be observed in situations like this, the predictions have to be made with reasonable confidence, but 100% accuracy is not required.

The goal of our work is that of developing a *completely automated* forecasting solution that could endow operation monitoring tools with the capability of providing reasonable accurate, and up-to-date predictions. Our work is characterized by three basic ingredients:

i) We have bundled and integrated into a single software application multiple algorithms that cover the entire forecasting lifecycle. Whenever a prediction is needed, our software identifies and applies all the appropriate algorithms, evaluates the confidence with which each prediction is made, and selects the most accurate one.
ii) We have developed new algorithms and use state of the art optimization technologies to enable complete automation with efficiency and performance improvement.
iii) We have architected the algorithms into a forecasting engine (G-TSFE) that can be embedded into monitoring and management software so that,

whenever users monitor metrics, as long as the monitoring system keeps a log of metrics values, they can also access forecasts for such metrics. For example, if the transportation company has a monitoring system that tracks daily shipments, then the forecasting engine can be embedded into this software to make predictions for shipments and any other metric monitored by the same system. We are planning to integrate this engine into various HP management products, starting with OpenView Business Process Insight (OV-BPI) [1] to provide forecasting on business process-related metrics.

In this paper we focus on business process monitoring applications to illustrate the concepts. In this environment, there are two kinds of prediction that are of interest:

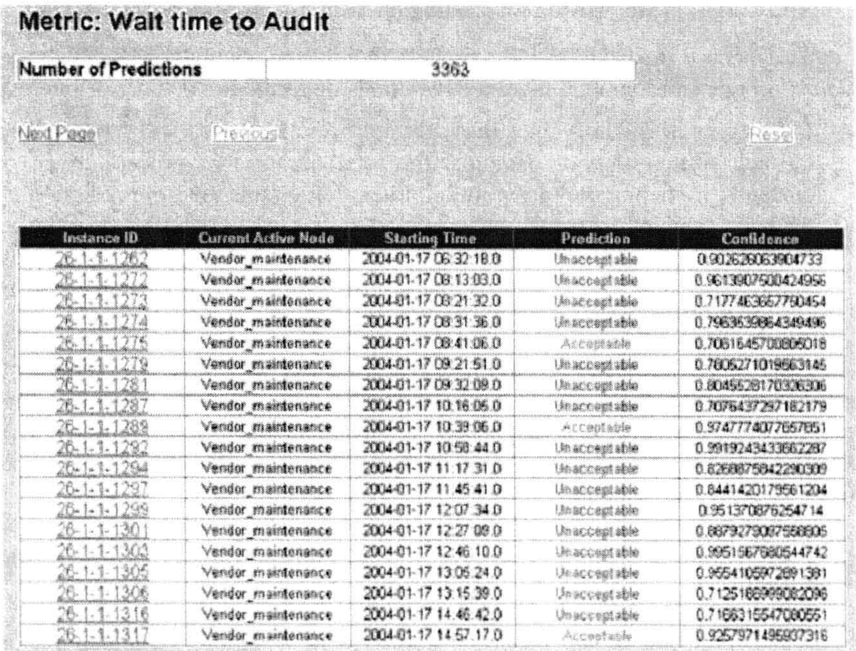
a) Predictions of metric values that active process instances will have at the end of their execution. For example, the total duration of processing an order currently in the middle of execution. Even though this kind of prediction is not the main topic of the paper, we briefly address it for completing the picture.
b) Predictions of aggregated metric values of future instances. For example, the number of orders that will be placed next Monday or even their average processing time. This is the main focus of the paper.

The remainder of the paper is organized as follows. Section 2 gives an example of the usage of an automated prediction solution in a business process monitoring environment. Section 3 gives a technical overview of our prediction engine concentrating more on its forecasting functionality. Section 4 summarizes the competitive landscape, and in Section 5 we conclude with the current state and next steps.

2 Example

Let's assume that one of the business objectives of a hypothetical manufacturing company ACME is to gain a good reputation for being punctual in making payments to its suppliers. This objective will allow ACME to be in a favorable position to negotiate with vendors the terms of the payment period in their contracts. The business operation on which this objective depends is "Pay to Vendor" and the process underlying this operation is called "Payables". Let's assume also that ACME has a 10 days payment policy from the moment it receives a vendor invoice. For ACME to keep its good standing it is important to keep monitoring the duration of the Payables process and react if the process is not meeting the payment policy. However, monitoring the current state of the process is not enough, since it only gives opportunity to "react" after the problem has occurred, which is too late. What ACME needs is the ability to be "proactive" so that problems can be prevented. Specifically, it needs to leverage its monitoring platform with predictive functionality. Two kinds of prediction are useful in this context:

a) For invoices that are currently being executed, ACME will be able to have predictions about their total duration. This will enable ACME to take action before the problem arises in order to prevent its occurrence or at least minimize the damage caused. In this case, try to speed up the processing of those invoices likely to last more than 10 days, for example, by increasing their priority. This kind of prediction is *per instance,* as illustrated in Figure 1 where those instances (each instance corresponding to the processing of an invoice) predicted to exceed the maximum duration say "unacceptable" (in red but not noticeable in this black and white paper). Notice that each prediction is accompanied by a confidence value.



Instance ID	Current Active Node	Starting Time	Prediction	Confidence
26-1-1-1262	Vendor_maintenance	2004-01-17 06:32:18.0	Unacceptable	0.902626063904733
26-1-1-1272	Vendor_maintenance	2004-01-17 08:13:03.0	Unacceptable	0.961390750042495
26-1-1-1273	Vendor_maintenance	2004-01-17 08:21:32.0	Unacceptable	0.717746366577 90454
26-1-1-1274	Vendor_maintenance	2004-01-17 08:31:36.0	Unacceptable	0.796363986434 9496
26-1-1-1275	Vendor_maintenance	2004-01-17 08:41:06.0	Acceptable	0.706164570080 6018
26-1-1-1279	Vendor_maintenance	2004-01-17 09:21:51.0	Unacceptable	0.780527101966 3145
26-1-1-1281	Vendor_maintenance	2004-01-17 09:32:09.0	Unacceptable	0.804552817032 6306
26-1-1-1287	Vendor_maintenance	2004-01-17 10:16:06.0	Unacceptable	0.707643729718 2179
26-1-1-1289	Vendor_maintenance	2004-01-17 10:39:06.0	Acceptable	0.974777407765 7851
26-1-1-1292	Vendor_maintenance	2004-01-17 10:58:44.0	Unacceptable	0.991924343366 2287
26-1-1-1294	Vendor_maintenance	2004-01-17 11:17:31.0	Unacceptable	0.826887584229 0309
26-1-1-1297	Vendor_maintenance	2004-01-17 11:45:41.0	Unacceptable	0.844142017956 1204
26-1-1-1299	Vendor_maintenance	2004-01-17 12:07:34.0	Unacceptable	0.95137087625471 4
26-1-1-1301	Vendor_maintenance	2004-01-17 12:27:09.0	Unacceptable	0.887927908755 8806
26-1-1-1303	Vendor_maintenance	2004-01-17 12:46:10.0	Unacceptable	0.995156768054 4742
26-1-1-1305	Vendor_maintenance	2004-01-17 13:06:24.0	Unacceptable	0.955410597289 1381
26-1-1-1306	Vendor_maintenance	2004-01-17 13:15:39.0	Unacceptable	0.712518699908 2096
26-1-1-1316	Vendor_maintenance	2004-01-17 14:46:42.0	Unacceptable	0.716631554708 0551
26-1-1-1317	Vendor_maintenance	2004-01-17 14:57:17.0	Acceptable	0.925797149593 7316

Fig. 1. Predictions of metric Wait Time to Audit for active process instances

b) ACME needs to know whether the average total duration of processing invoices the next day (or week, or month) is likely to be above a certain threshold in order to make plans for the resources that it will assign to this process. The idea is to provide ACME with the ability to monitor the daily (or weekly, or monthly) average total duration not only for the present and past, but also for the future. This is illustrated in Figure 2, shows a possible GUI where a user can move the slider along the time scale to set the time for the metric prediction, which is displayed on the dial chart. This kind of prediction is more in the style of Time Series Forecasting.

Instance prediction is specific to a business process monitoring environment, while time series forecasting applies to any monitoring environment.

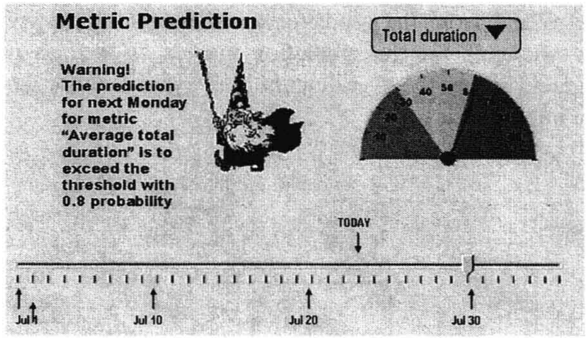


Fig. 2. Forecasting total duration metric average for the time set on the slider

It can be appreciated that these two kinds of prediction functionality empower business managers to do effective planning and decision making to better meet the goals of their business operations. They constitute a powerful tool that leverages the business process monitoring platform and gives its users the ability to see the metrics' most probable future behavior.

One of the main features of such a tool is that its design responds to the need that users have for a prediction tool that is easy and intuitive to use. The user doesn't need to write code, or be an expert in predictive technologies; rather she simply clicks a button and obtains the requested predictions.

3 Technical Overview

In this section, we explain the architecture, main concepts and principles underlying the prediction functionality of *Metric Predictor* (MP), our prediction engine to automate the prediction of metric values in a monitoring environment.

Metric Predictor has been designed with three goals in mind: ease-of-use, generality and scalability. It is

- easy to use so that business managers do not need to be experts in data mining or statistics, but can simply obtain forecasts at the click of a button;
- generic so that it works for any kind of metric in different domains; and
- scalable so that it can handle large numbers of frequent predictions.

3.1 System Architecture

Metric Predictor is an engine that can easily be integrated into a monitoring platform, enhancing it with predictive functionality. In general terms, a monitoring platform that supports metrics receives its input from external instrumentation infrastructure, allows its users to define metrics on top of data fed by the instrumentation, computes metric values from this data according to the metric definitions, and provides reports to its users. Seamlessly integrating Metric Predictor into such a platform essentially means tailoring it for the platform repository schema, creating the interfaces to let it read and write to and from the repository, and extending this repository with the necessary data structures for the prediction models as well as for the predictions themselves. Figure 3 shows the overall architecture of a monitoring platform enriched with *predictive* intelligence.

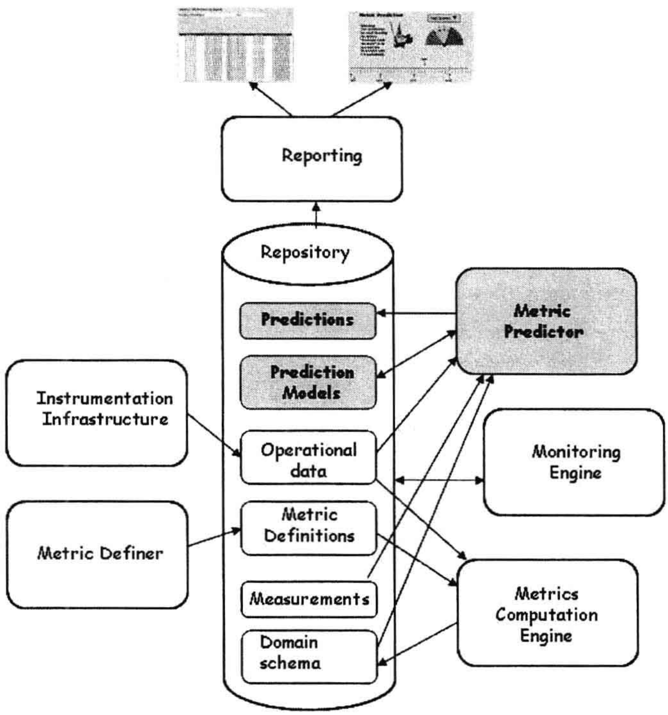


Fig. 3. General architecture of a predictive monitoring platform

In the case of a business process monitoring platform such as OV-BPI, whose task is to track the progress of process executions, the instrumentation infrastructure captures events that signal the start and end of the activities corresponding to the different process steps, as shown in Figure 4. For this purpose the platform allows its users to define abstract views of business processes as well as events and their mappings to the start and end of the steps of such abstract process views. In addition, it lets users define metrics on top of these process views. Metrics are computed on the process execution data that the monitoring engine logs after interpreting the occurrence of events according to its mappings.

In the rest of the paper, we concentrate on a Metric Predictor whose development was motivated by the need to incorporate prediction capabilities into our business operations management platform, *Business Cockpit* (built on top of OV-BPI). Details about other functionalities of this platform can be found in [2].