Dynamic Collective Work

Geetika T. Lakshmanan and Rania Khalaf
IBM T.J. Watson Research Center

Schahram Dustdar
Vienna University of Technology

The Internet has completely changed the way in which data is circulated, and has rapidly shifted us from a world of paper documents to a world of online documents, databases, and provenance systems.[1] This has also helped to increase the size and complexity of systems that support today's globally distributed, rapidly changing and agile businesses. Such businesses are becoming increasingly federated, loosely coupled, and distributed. They are also generating a huge number of events ranging from record entries representing business activities to more technical events at various granularity levels.[2]

Internet-based systems and middleware now integrate massive amounts of diverse data (email, PDF attachments, database records, system logs, phone calls, videos, photos, and so on) from distributed and heterogeneous sources and make it widely available. Furthermore, the proliferation of online collaboration tools and platforms such as YouTube, Flickr, Facebook, LinkedIn, Chatter, Second Life, Many Eyes, Pinterest, and many others have made it easier for people to collaborate.[3,4] This has enabled the growth of ad hoc, people-driven, informal processes. Industries such as healthcare, insurance, and banking have witnessed explosive growth in such processes.[5–13]

Dynamic and collective activities are characterized by their flexibility and data-driven nature. Automobile insurance claims processing, order processing of prescription drugs, patient case management in a hospital, and recovery and response assistance during natural disasters are a few examples. The execution of such informal processes isn't always controlled by a predefined model or a single execution runtime.[14] Various factors, including human judgment and document contents, determine the set of activities that humans involved in such activities must perform and in what order those activities are executed.

This special issue focuses on challenges and solutions involved in handling Internet-enabled dynamic collective work activities.

Challenges in Dynamic Collective Work

One implication of the added flexibility and data-driven nature of dynamic collective work activities is that they're more difficult to mine and analyze than traditionally rigid processes. Systems that handle such activities must address three main challenges: finding

 Published by the IEEE Computer Society 1089-7801/13/$31.00 © 2013 IEEE

patterns in ad hoc execution behavior, handling concurrent users and concurrent task execution, and providing operational support such as alerts, predictions, recommendations, knowledge sharing, and collaboration.

Finding Patterns in Ad Hoc Execution Behavior

In most instances of dynamic collective work, humans make decisions about which task to execute next based on data associated with the instance and on their independent judgment. Furthermore, people such as knowledge workers might introduce ad hoc tasks in a process instance as necessary to handle a particular situation. As a result, instances of the same process can vary dramatically in terms of which tasks were executed and in what order. This makes it challenging to mine the process to identify a meaningful set of activities and detect a coherent behavior pattern. In addition, the underlying process might change as it undergoes analysis.[15] Processes can change due to periodic or seasonal fluctuations (for example, "In August, there is more demand" or "On Monday evening, fewer people go to the movies") or to changing conditions ("Unemployment is rising"). Such changes affect processes, and the ability to detect and analyze them is vital.

Handling Concurrent Users and Task Execution

Multiple people might be involved in dynamic-collective-work-related activities, and they can initiate multiple tasks concurrently (in parallel) and repeat task execution. Collecting intelligence from such platforms and enabling knowledge sharing could be extremely useful but quite challenging. In particular, state and predictive models must be developed that capture such processes' behavior and enable prediction and recommendation.

Providing Operational Support

Because many data sources today are updated in (near) real-time, and sufficient computing power is available to analyze events when they occur, a strong demand has emerged for online operational support during the course of dynamic collective work. Such support can consist of alerting knowledge workers and supervisors to situations, prediction, and recommendation. The moment a case deviates from the predefined process, the system can detect it and generate an alert. Often, we'd like to generate such notifications immediately (so that we can still influence outcomes in a process) and not in an offline fashion. We can use historical data to build predictive models that can then help guide running process instances. For example, it's possible to predict a case's remaining processing time. Based on such predictions, we could build recommender systems that propose particular actions to reduce costs or shorten flow time. Providing operational support in such online settings creates challenges in terms of computing power and data quality.

These challenges are also echoed in IEEE's process mining manifesto,[16] which aims to increase investment, research, and development in process mining.

In This Issue

The three articles in this special issue address one or more challenges that arise from analyzing and leveraging dynamic collective work today.

A strong demand has emerged for online operational support during the course of dynamic collective work.

In "Mining Artful Processes from Knowledge Workers' Emails," Massimo Mecella and Claudio Di Ciccio propose MailOfMine. This approach and software tool aims at automatically building a set of workflow models that represent the artful processes behind knowledge workers' activities, on top of a collection of email messages. This helps to formalize unspecified agile processes that are autonomously used. Because such models aren't defined a priori by experts but are rather inferred from real-life scenarios that actually took place, they represent the actual execution of the process embedded in knowledge workers' email.

To extract these processes, the authors adopt an approach based on the concept of "speech acts" and apply it to email messages. Detecting speech acts requires that knowledge workers provide a dictionary of words from their domains.

These words are then divided into verbs and objects, each verb is concatenated with each object, and the resulting strings are kept in a collection of expressions. For each expression, the system uses an information retrieval tool to search for them within the email messages. This approach generates process models comprising a declarative workflow expressed in terms of temporal constraints on activities. These models can be shared, compared, and preserved to help knowledge workers identify bottlenecks during execution and extract and share best practices across the organization.

Arno Scharl, Alexander Hubmann-Haidvogel, Marta Sabou, Albert Weichselbraun, and Heinz-Peter Lang present "From Web Intelligence to Knowledge Co-Creation: A Platform for Analyzing and Supporting Stakeholder Communication," which introduces a Web intelligence and online collaboration platform. This platform aggregates large archives of digital content from multiple stakeholder groups and enables the co-creation and visualization of evolving knowledge archives. The authors describe the platform and a context-aware document editor as an add-on that supports concurrent authoring by multiple users. While documents are being edited, semantic methods analyze them on the fly to recommend related content. Positive or negative sentiment is computed automatically to gain a better understanding of third-party perceptions. The editor is part of an interactive dashboard that uses trend charts and map projections to show how often and where relevant information is published, and to provide a real-time account of concepts that stakeholders associate with a topic.

Analysis of dynamic collective work could provide users with insight for making better decisions and introduce process optimizations.

Finally, "Casebook: A Cloud-Based System of Engagement for Case Management," by Hamid-Reza Motahari-Nezhad, Susan Spence, Claudio Bartolini, Sven Graupner, Charlie Bess, Marianne Hickey, Parag Joshi, Roberto Mirizzi, Kivanc Ozonat, and Maher Rahmouni, presents Casebook, a system that embraces social and collaboration technology, analytics, and intelligence to provide a cloud-based system of engagement for knowledge workers. This system assists knowledge workers by capturing and codifying flexible processes as well as case-related information.

Casebook is architected as a cloud-based service to support case management in cross-organizational and cross-company settings. Workers can engage in case management by offering social collaboration around cases as a focal point for business value generation and interaction. Casebook also provides a system that integrates the information flow captured by communication and collaboration tools with information about cases, backed with process-aware components that help enact people-driven processes in a plan-as-you-do manner. Other features include roadmapping as a shared workspace for case workers, advanced analytics to recommend experts, and suggestions for next actions determined from templates based on past cases. Casebook has been implemented in research prototypes and validated for case management applications inside Hewlett Packard.

The articles in this issue make it clear that analyzing and using dynamic collective work activities is increasingly relevant today. Software professionals, founders and employees of start-up companies, researchers, and social scientists are actively participating in this space and exploring solutions to problems such as designing recommender systems,[6–8] predicting next steps for case workers,[9] analyzing and tracing healthcare processes,[10,11] adapting mined models to dynamic changes in a process,[15] and optimizing business processes in crowdsourcing platforms.[12,13] Analysis of dynamic collective work could provide this spectrum of users with insight for making better decisions and introduce process optimizations.

Acknowledgments

We thank the authors, whose submissions made this special issue possible and showed how active and exciting this field is; the reviewers, who sacrificed their time and made the effort to read and offer constructive comments to help the authors improve the quality of their articles; the *IEEE*

Internet Computing editorial staff, who helped us manage the submissions; and the magazine's editorial board for its guidance in the process.

References

1. G.T. Lakshmanan et al., "Provenance in Web Applications," *IEEE Internet Computing*, vol. 15, no. 1, 2011, pp. 17–21.
2. S. Rozsnyai et al., "Business Process Insight: An Approach and Platform for the Discovery and Analysis of End-to-End Business Processes," *Proc. SRII Global Conf.*, IEEE, 2012, pp. 80–89.
3. N. Sadat Shami, T. Erickson, and W.A. Kellogg, "Common Ground and Small Group Interaction in Large Virtual World Gatherings," *Proc. 12th European Conf. Computer Supported Cooperative Work* (ECSCW 11), Springer, 2011, pp. 393–404.
4. J.M. DiMicco et al., "People Sensemaking and Relationship Building on an Enterprise Social Network Site," *Proc. 42nd Hawaii Int'l Conf. System Sciences* (HICSS 09), IEEE, 2009, pp. 1–10.
5. S. Dustdar et al., "Principles of Elastic Processes," *IEEE Internet Computing*, vol. 15, no. 5, 2011, pp. 66–71.
6. C. Dorn et al., "Self-Learning Predictor Aggregation for the Evolution of People-Driven Ad-Hoc Processes," *Proc. 9th Int'l Conf. Business Process Management* (BPM 2011), Springer, 2011, pp. 215–230.
7. C. Dorn and S. Dustdar, "Supporting Dynamic, People-Driven Processes through Self-Learning of Message Flows," *Proc. 23rd Int'l Conf. Advanced Information Systems Eng.* (CAiSE 11), Springer, 2011, pp. 657–671.
8. H. Schonenberg et al., "Supporting Flexible Processes through Recommendations Based on History," *Proc. Int'l Conf. Business Process Management* (BPM 08), Springer, 2008, pp. 51–66.
9. G.T. Lakshmanan et al., "Predictive Analytics for Semi-Structured Case Oriented Business Processes," *Business Process Management Workshops*, Lecture Notes in Business Information Processing 66, Springer, 2010, pp. 640–651.
10. Á. Rebuge and D.R. Ferreira, "Business Process Analysis in Healthcare Environments: A Methodology Based on Process Mining," *J. Information Systems*, vol. 37, no. 2, 2012, pp. 99–116.
11. G.T. Lakshmanan, S. Rozsnyai, and F. Wang, "Investigating Clinical Care Pathways Correlated with Outcomes," *Proc. Int'l Conf. Business Process Management*, to appear, 2013 pp. 323–338.
12. R. Khazankin, B. Satzger, and S. Dustdar, "Optimized Execution of Business Processes on Crowdsourcing Platforms," *Proc. 8th IEEE Int'l Conf. Collaborative Computing: Networking, Applications, and Worksharing* (CollaborateCom 12), IEEE, 2012, pp. 443–451.
13. M. Allahbakhsh et al., "Quality Control in Crowdsourcing Systems: Issues and Directions," *IEEE Internet Computing*, vol. 17, no. 2, 2013, pp. 76–81.
14. G.T. Lakshmanan and R. Khalaf, "Leveraging Process Mining Techniques to Analyze Semi-Structured Processes," *IEEE IT Professional*, to appear; http://doi.ieeecomputersociety.org/10.1109/MITP.2012.88.
15. R.P. Jagadeesh et al. "Handling Concept Drift in Process Mining," *Proc. 23rd Int'l Conf. Advanced Information Systems Eng.* (CAiSE 11), Springer, 2011, pp. 391–405.
16. W.M.P. van der Aalst et al., "Process Mining Manifesto," *Business Process Management Workshops*, vol. 1, Springer, 2011, pp. 169–194.

Geetika T. Lakshmanan is a research staff member at IBM T.J. Watson Research Center. Her research includes the management, mining, and prediction of data-driven processes to address new problems introduced by emerging applications, including healthcare and banking. Lakshmanan has a PhD in computer science from Harvard University. She's an IEEE senior member and a member of ACM. Contact her at gtlakshm@us.ibm.com; http://researcher.watson.ibm.com/researcher/view.php?person=us-gtlakshm.

Rania Khalaf is a research staff member and manager of the Component Systems group at the IBM T.J. Watson Research Center. Her interests include service computing, business process management, and distributed computing. Khalaf has a PhD in computer science from the University of Stuttgart. She's a creator of the Bite REST-based workflow language, coeditor of the WS-BPEL standard, and developer and coarchitect of the first public BPEL implementation. Contact her at rkhalaf@us.ibm.com.

Schahram Dustdar is a full professor of computer science and head of the Distributed Systems Group, Institute of Information Systems, at the Vienna University of Technology. His research interests include service-oriented architectures and computing, cloud and elastic computing, complex and adaptive systems, and context-aware computing. Dustdar is an ACM Distinguished Scientist and IBM Faculty Award recipient. Contact him at dustdar@dsg.tuwien.ac.at.