Towards a Dynamic Declarative Service Workflow Reference Model*

Damian Clarke

Department of Computer Science, University of Miami, Miami, USA
d.clarke6@umiami.edu

Abstract. Functional, nonfunctional, just-in-time approaches to composing web services span the sub-disciplines of software engineering, data management, and artificial intelligence. Our research addresses the process that must occur once the composition has completed and stakeholders must investigate historical and online operations/data flow to reengineer the process either off-line or in real-time. This research introduces *an effective reference model to assess the message flow of long-running service workflows.* We examine Dynamic Bayesian Networks (dDBNs), a data-driven modeling technique employed in machine learning, to create service workflow reference models. Unlike other reference models, this method is not limited by static assumptions. We achieve this by including the trend and time varying variables in the model. We demonstrate this method using a flight dataset collected from various airlines.

1 Introduction

In response to today's increasingly volatile business environment, web service workflows need to be agile and dynamic. The main cause of volatility is trend and time dependent, which are secondary but influential variables implicitly within the service workflow that affect the relationship between dependent variables and other independent variables of primary interest. Therefore, reference models [1] for service workflows must consist of trend and time varying constructs that efficiently and effectively capture dynamically identifiable changes in the information-processing functionalities. One method to dynamically identify changes is to infer meaning from data the service workflow consumes and produces and then be able to recommend action based on that meaning. With the inclusion of on-demand data intensive discoveries the model can now accommodate constructive feedback and forward interventions resulting in an agile representation that can more accurately reflect trend and time varying variables.

As a motivating scenario consider an *Airline Ticket Pricing Workflow*. Such a workflow consists of a Select Airlines service that takes origin and destination cities from the user and devises a list of air carriers who fly between the two cities. Subsequently, a Collect Prices service contacts the list of airlines and develops the best ticket price. The Collect Prices service may have many attributes that are not used in

* Supervised by Prof. M. Brain Blake.

A.R. Lomuscio et al (Eds.): ICSOC 2013 Workshops, LNCS 8377, pp. 563–568, 2014.

every instance of the workflow such as number of connections, time of day of departure, and time of day of arrival. In this service workflow and in others like it, the user may be able to leverage unused message types to find the most optimal price. For this scenario, the date and time of purchase and the time-of-day for the flight also may affect the optimal price. A more sophisticated example of this workflow is used in our evaluation section later in this paper.

In our work, our research contributes to the next step after a set of functionally-adequate workflows are created, on-demand, by a third party and in use. Given historical data and the ability to strategically poll these longer-standing workflows, we believe that, by using the real operational message data generated by the workflow logic which we call *constructive feedback*, a reference model of the workflows can be dynamically created. *A **contribution of our work*** is the introduction of a model that can encapsulate declarative and predictive features that can deal with uncertainty thereby facilitating *forward intervention*.

2 Related Work

Research projects related to the optimizations of web service workflows [2][3] can be classified into three areas, service engineering and modeling projects that focus on the web service specifications, similar service engineering projects that alternatively concentrates on the operational web services and their data, also the general body of work in workflow optimization and workflow decentralization. Largely, the state-of-the-art in web service discovery, composition, and mashup operates with the specifications and not the running operational systems [4][5]. As such, these projects are not related to our approach as we look at the real data content of messages as a method to re-engineer operational web service workflow systems.

There are other service engineering projects that investigate the real data content. One such work is in the area of automated or semi-automated web service testing. To automate testing of web services, related projects must evaluate if web service outputs meet tests plans and, in other cases, predict the specific data content that requires testing. These approaches must develop models to understand the data. The most relevant approaches develop models of data that extend SOAP [6] and UDDI data models [7]. These approaches try to perturb data from specifications and then execute them in operational mode. Unlike these approaches that tend to work on just one web service, models in our work leverages models across multiple services in a web service workflow.

3 Technical Approach

We introduce a reference model for service workflows created from the underlying messages collected during the actual service workflow operation. Our approach includes the implicit secondary trend and time varying variables to more accurately reflect a volatile business environment. This section describes how we formalize the model by proposing the use of *non-homogeneous semi-flexible dynamic Bayesian networks* [8].

3.1 Bayesian Networks

Static Bayesian Networks usually referred to as simply *Bayesian Networks* (BN) are a class of *graphical models* [9]. They allow a concise representation of the probabilistic dependencies between a given set of random variables X = $\{X_1,X_2,\ldots,X_p\}$ as a directed acyclic graph (DAG) G = (V,E) where each node $v_i \in V$ corresponds to a random variable X_i and E is the set of edges between connecting nodes. A random variable denotes an attribute, feature, or hypothesis about which we may be uncertain.

An important feature of Bayesian networks is that by instantiating vertices in the directed structure independences may change to dependences, i.e. stochastic independence has specific dynamic properties. This produces the concept known as "explaining away" [10] where the confirmation of one cause of an observed or believed event reduces the need to invoke alternative causes and/or confirmation of one cause increases belief in another.

Whereas static Bayesian Networks model multiple independent "snapshots" of the process, intuitively *Dynamic Bayesian Networks* (dBN) extend the fundamentals of Bayesian networks by modeling associations from the temporal dynamics between entities of interest. We refer the reader to [11] for a comprehensive review. Each variable in a dBN is represented by several nodes across time points. In addition, temporal signatures are useful in capturing possible feedback loops that are disregarded by static Bayesian Networks. A set of sufficient conditions for a model to be represented as a dynamic network are detailed in the following works [12] [13] [8].

By combining qualitative and quantitative event data (e.g. intra-service-workflow step messages) in a coherent way, a Bayesian statistical approach allows the representation of each event with its set of mutually exclusive and collectively exhaustive values as a random variable. Each node (perhaps defining a data point from a service messages) has assigned a function that describes how the state of a node depends on the parents of the node. The topology of the graph that relates the nodes defines the probabilistic dependencies between the node variables, by means of a set of conditional distributions. In addition, we can integrate different sources of information, for example domain expert knowledge, historical and polled event data, to give a unified knowledge that allows us to manage internal and external "causal" factors such as bottlenecks. In this context, a Bayesian network is augmented with two other types of nodes, then it is possible for actions to be decided based on given evidence. These two types of nodes are utility nodes and decision nodes. Utility nodes represent the value of a particular event, while decision nodes represent the choices that might be made.

3.2 Proof of Concept and Discussion

The intent of our experimentation is to demonstrate that the behavior in web service workflow operations vary in the nuanced ways that enable our approach to predict the content-based outcomes and perform forward interventions. When using a web service workflow to manage the airline ticket purchase workflow, we believe that the businesses have encoded their purchase operations and these operations might vary from airline to airline. This variation is the basis for why our data model would be important for optimizing workflow paths. Consider the workflow path in Fig. 1. We introduce a reference model that can:

1. *Predict when it is not necessary to check the availability of an airline ticket on a particular airliner, thus the overall BPEL workflow can be truncated*
2. *Predict when the purchase date for a ticket is too close to the departure date to get an optimal price*
3. *Predict that when a passenger is restricted for a particular time-of-day certain airlines will not have optimal pricing.*

In Fig.1, we see an exemplar model for the airline ticket shopping workflow. When a service provider receives a request through a Web Service Interface Service, then this would trigger the concurrent availability (Airline N: Check Availability Service) and price checks (Airline N: Check Prices Service) from the web services of the various airlines. The final step is a decision to buy a ticket from a particular airline (Airline N. Purchase). The shaded web services represent a truncated workflow determined using our data model.

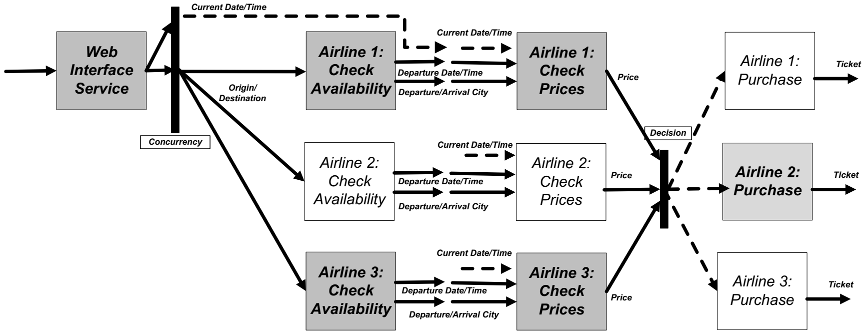

Fig. 1. An Example Airline Purchase Workflow

As a first part of our evaluation, we used 41 days of airline data from [14] where we anticipate that each airline is running a business specific workflow similar to that illustrated in Fig. 1. Consequently, when we analyze their data we saw different behaviors in how their airline prices are generated. It is evident that airlines modify their prices for the same ticket as the time to departure reduces. As we anticipated, these modifications fluctuate as the business operations for each airline differs.

Figure 2 shows a representation of a BN for the real airline pricing data for AA. Our model accurately predicts lower prices for polling dates further away from the actual flight time on 1/1/2003 and higher prices as the polling date gets closer to the flight time.

Figure 3 shows data collected from 10/12/2002 through 02/01/2003 for a United Airline's flight leaving on 02/02/2003. We demonstrate how our model behaves twenty-five time points in the future in the shaded region. The dotted lines show a gentle upward tick in the price. Figure 4, plots the data collected from 10/12/2002 to 17/12/2002. Our model accurately predicts a downward trend in the price and then a leveling off.

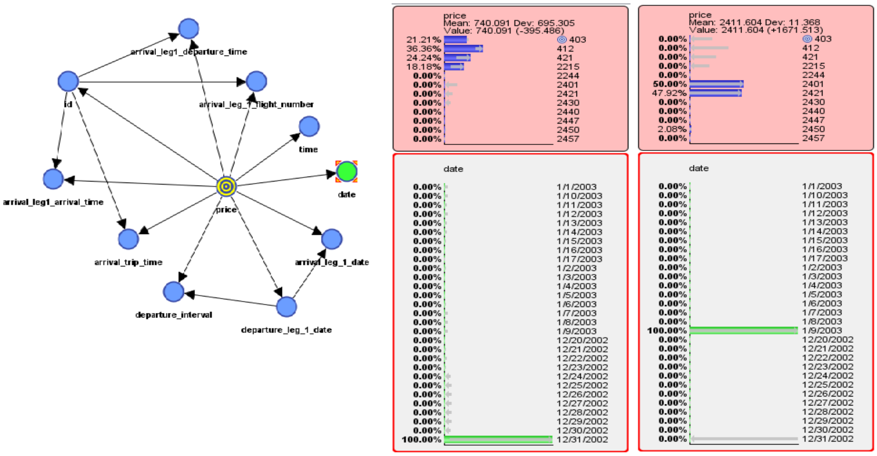

Fig. 2. An Instantiation of the BN for AA prices

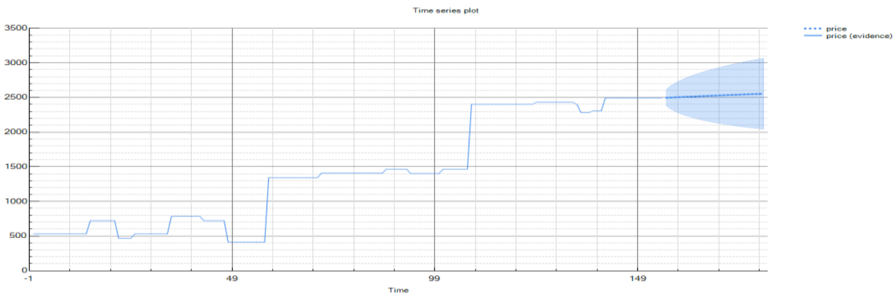

Fig. 3. UA prices (10/12/2003 to flight date on 02/01/2003) with 25 time point projection

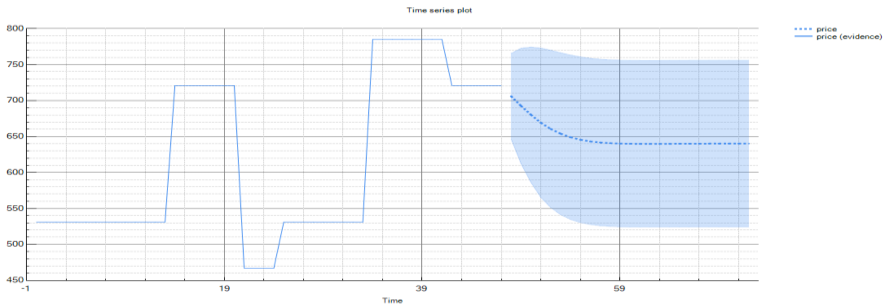

Fig. 4. UA prices taken from 10/12/2003 to 17/12/2003 with 25 time point projection

4 Conclusion and Future Work

The service workflow assessment vision that we explore in this paper is to completely automate the data centric service workflow model and to perform such assessments in a reliable, reproducible and efficient manner. The intention is that service workflow assessment can be deployed frequently and on demand. We have seen that this vision is promising with the inclusion of trend and time-varying features of a business environment. Unanswered questions such as polling rate still need to be explored.

References

[1] OASIS SOA Reference Model (SOA-RM) TC, https://www.oasis-open.org/committees/soa-rm/faq.php
[2] Blake, M.B., Huhns, M.N.: Web-Scale Workflow: Integrating Distributed Services. IEEE Internet Computing 12(1), 55–59 (2008)
[3] Blake, M.B., Gomaa, H.: Agent-oriented compositional approaches to services-based cross-organizational workflow. Decis. Support Syst. 40(1), 31–50 (2005)
[4] Liu, X., Hui, Y., Sun, W., Liang, H.: Towards Service Composition Based on Mashup. In: 2007 IEEE Congress on Services, pp. 332–339 (2007)
[5] Srivastava, B., Koehler, J.: Web Service Composition - Current Solutions and Open Problems. In: ICAPS 2003 Workshop on Planning for Web Services, pp. 28–35 (2003)
[6] Offutt, J., Xu, W.: Generating test cases for web services using data perturbation. SIGSOFT Softw. Eng. Notes 29(5), 1–10 (2004)
[7] Bai, X., Dong, W., Tsai, W.-T., Chen, Y.: WSDL-based automatic test case generation for Web services testing. In: IEEE International Workshop on Service-Oriented System Engineering, SOSE 2005, pp. 207–212 (2005)
[8] Dondelinger, F., Lèbre, S., Husmeier, D.: Non-homogeneous dynamic Bayesian networks with Bayesian regularization for inferring gene regulatory networks with gradually time-varying structure. Mach. Learn. 90(2), 191–230 (2013)
[9] Dawid, P., Lauritzen, S.L., Spiegelhalter, D.J.: Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks. Springer (2007)
[10] Wellman, P.P., Henrion, M.: Explaining 'explaining away'. IEEE Trans. Pattern Anal. Mach. Intell. 15(3), 287–292 (1993)
[11] Kim, S.Y., Imoto, S., Miyano, S.: Inferring gene networks from time series microarray data using dynamic Bayesian networks. Brief. Bioinform. 4(3), 228–235 (2003)
[12] Lèbre, S.: Inferring Dynamic Genetic Networks with Low Order Independencies. Stat. Appl. Genet. Mol. Biol. 8(1), 1–38 (2009)
[13] Lèbre, S., Becq, J., Devaux, F., Stumpf, M.P., Lelandais, G.: Statistical inference of the time-varying structure of gene-regulation networks. BMC Syst. Biol. 4(1), 130 (2010)
[14] Etzioni, O., Tuchinda, R., Knoblock, C.A., Yates, A.: To buy or not to buy: mining airfare data to minimize ticket purchase price. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, pp. 119–128 (2003)