Estimation of Parameters Sensitivity for Scientific Workflows

Fakhri Alam Khan, Yuzhang Han, Sabri Pllana, and Peter Brezany
Department of Scientific Computing, Faculty of Computer Science, University of Vienna
Nordbergstrasse 15/C/3, A-1090, Vienna Austria
{khan|han|pllana|brezany}@par.univie.ac.at

***Abstract*— Usually workflow activities in the scientific domain depend on a collection of parameters. These parameters determine the output of the activity, and consequently the output of the whole workflow. In the scientific domain, workflows have exploratory nature and are used to understand a scientific phenomenon or answer scientific questions. In the process of a scientific experiment a workflow is executed multiple times using various values of the parameters of activities. It is relevant to identify (1) which parameter strongly affects the overall result of the workflow and (2) for which combination of parameter values we obtain the expected result. Foreseeing these issues, in this paper we present our methodology to estimate the significance of all scientific workflow parameters as well as to estimate the most significant parameter to the workflow. The estimation of parameter significance will enable the scientist to fine tune, and optimize his results efficiently. Furthermore, we empirically validate our methodology on Non-Invasive Glucose Measurement workflow and discuss our results. The NIGM workflow uses the neural network model to calculate the glucose level in patient blood. The neural network model has a set of parameters, which affect the result of the workflow significantly. But, unfortunately the impact significance of these parameters is commonly unknown to the user. We present our approach for estimating and quantifying impact significance of neural network parameters.**

I. Introduction

With the emergence of the Grid infrastructure, with its heterogeneous and geographically distributed computational resources, e-Science has become a reality. E-Science [1] enables the scientists to perform complex computational experiments and share their results. Usually, the tasks that are to be executed on the Grid are specified in the form of a workflow. Workflows can be defined and represented by Directed Acyclic Graph (DAG) as $G = (N, E)$, where N represents a set of nodes and E is the set of directed edges between the nodes. The nodes indicate workflow activities (that is Grid tasks), whereas the edges indicate the control or data flow between activities. From this definition it is evident that; a) scientific workflow can be composed of any number of activities, b) secondly, the order in which these activities are executed is very important to the whole workflow output.

A scientific workflow activity represents a transformation of input data to output data, that is they take input, process them, and produce results. Activities contain one or more parameters. These parameters determine the output, which ultimately means, that the value of these parameters influence

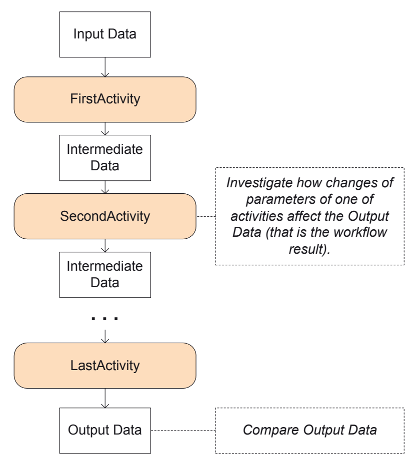

Fig. 1. Workflow parameter sensitivity

the calculation carried out by activities and accordingly affect the final result given by using these activities. Naturally, the user would want to optimize the result produced by using a model, which can be achieved through modifying the parameter values. But unfortunately, there are problems in doing so. On one hand; in some models, such as neural network, the relation between each single parameter and the result is not defined and can only be decided empirically through experiment, which is time consuming and tedious task. On the other hand, a workflow might contain tens of services, so the number of parameters in total can reach hundred or even more. Furthermore, for intermediate results of certain activities it is hard to judge the quality (typically for intermediate data we have no trusted data to compare to). In cases where we know the expected overall workflow result (for instance, obtained by real-world measurements of the system under study), we can reason about the quality of results of a certain activity by studying how this activity affects the overall workflow result (that is, which parameter values of an activity lead to the overall workflow result that is close the expected result; see Figure 1). Additionally, the user must be able to select the parameter(s), the modification on which will significantly influence the final result of the workflow, and ignore other parameter(s), the influence of which is not significant to the final workflow result. The significance of a parameter can be descriptively defined as "the effect of a parameter on the final

1530-2016/09 $26.00 © 2009 IEEE
DOI 10.1109/ICPPW.2009.9

result of a computational model".

In this paper we propose a methodology to determine numeric weight of each parameter in a scientific workflow, which reflects the influence a parameter exerts on the final result. The observation of weights can assist a user in finding the significant parameters to the result and accordingly determining which parameter to change if he or she wants to optimize the result some way. The Figure 1 depicts our methodology to calculate the impact significance of parameters:

1) *Phase 1*: In this step calculation is carried out to determine the effect of change in a single parameter of an activity to the workflow output.
2) *Phase 2*: Here the comparison for change effect of all parameters is performed to get the most significant workflow parameter.

A suitable place to record significant values for all parameters is along side provenance [2] information of each activity, so as to provide the user with information on how to adjust the model result as well as to show the most significant parameters. Basically, we may define the workflow provenance as the process of collection of information on workflow execution. Such information will enable the user to find out which parameters are candidates for modification. Modification should be done on these parameters in model initialization phase. In the later half of this paper, we empirically validate our proposed methodology via estimating the most significant parameter for the real-world workflow that addresses the Non-Invasive Glucose Measurement. The major contributions of this paper include:

- A methodology for quantification of the impact significance of workflow activity parameters.
- Estimating the most significant parameter in neural network model.

The paper is organized as follows. In Section II related work is discussed, while in Section III we provide details of proposed methodology and discuss the various steps to estimate the workflow parameter significance. In Section IV we apply our methodology to the Non-Invasive Glucose Measurement workflow and share its results as well as estimate the most significant parameter out of the neural network model. Section V concludes the paper.

II. Related Work

Recently there have been significant research efforts focused on various aspects of scientific workflows [3]: workflow planing and execution, scheduling, Quality of Service (QoS), or scientific data access and integration in the Grid. UK e-Science group [4] is actively involved in developing e-Science applications [1], which enable the scientists to perform complex computational experiments and share their results through workflows. A major aspect of their research addresses the access and integration of scientific data [5] in the Grid environments using the workflow paradigm. The TeraGrid [6] project is an open scientific discovery infrastructure that aims at providing access to high performance computers, distributed resources, and tools through high speed interconnections. The Pegasus project at ISI investigates planning and management of large-scale scientific workflows [7].

K. M. McCann et. al. [8] addresses the workflow parameter study, in a distributed environment via providing an envirnoment that allows scientists to specify their application runs at a high organizational level. X. Liu et al. [9] addresses the workflow scheduling issue via estimating the time interval of an individual activity within a workflow. X. Liu proposes time-series based forecasting strategy and conducts time-series segmentation to discover the smallest pattern set and predicts the activity duration intervals with pattern matching results. There is some related work on workflow planning and scheduling. S. Pandey and R. Buyya [10] propose an approach for the efficient selection of replica data grid resources, for execution of workflows, while I. Brandic et al. [11] focuses on Quality of Service (QoS) support for Grid workflow systems.

While most of the related work is concerned with the performance aspects of scientific workflows (such as minimizing the workflow execution time), we focus on the quality of the workflow results and its relation to the individual activities within a scientific workflow. We believe that it is very important for the users of scientific workflows to know the relation of individual activity parameters to the workflow result.

III. Methodology

In this section we provide the details of our proposed approach to estimate the parameters significance individually as well as evaluate the most significant parameter to the workflow output.

A mathematical model or algorithm can contain one or more parameters. The values of parameters influence the calculation carried out by the model and accordingly affects the final result given by using the model. Naturally, the user would want to optimize the result by modifying some of the model parameters. Unfortunately, there are raised two issues when approaching this problem. On one hand, in some models, such as the neural network, the relation between each single parameter and the result achieved by using the model is not defined and can only be decided empirically. On the other hand, a workflow might contain tens of services, and the number of parameters in total can reach hundreds. We propose a method to determine the numeric weight of each parameter in a workflow, which reflects the influence the parameter exerts on the final result. The observation of weights can assist the user in finding the significant parameters to the result and accordingly determining which parameter to change if one wants to optimize the result. The provenance record of every service should contain the weight of every parameter so as to provide the user with information on how to adjust the model result as well as to show the most significant parameters.

Our proposed methodology has the advantage that it is sufficiently generalized and can be applied to any computational model, as well as to any number of scientific workflow parameters. Our two step approach is described below:

A. Phase-I

This step mathematically quantifies the idea, "to investigate, how changes in a parameter value of one activity affects the output data (that is the workflow result)" as depicted by Figure 1. To measure significance of a parameter, the workflow output result is recorded over a periodic change in the allowed value interval of that parameter, whereas rest of the parameters values are kept constant (that is only the value of parameter whose significance is measured is changed periodically). The *Estimation*, i.e. $PC_{(i)}$, is defined as "the ratio of percentage of change in final result to the percentage of change in parameter."

$$PC_{(i)} = \frac{r_i}{p_i} \tag{1}$$

$$\begin{aligned} &where, \\ PC_{(i)} &= ratio\ of\ percentage\ change\ in\ final\ result \\ &\quad\ to\ percent\ change\ in\ parameter, \\ r_i &= percentage\ of\ change\ in\ the\ final\ result, \\ p_i &= percentage\ of\ change\ in\ parameter, \end{aligned}$$

$PC_{(i)} > PC_{(j)}$ implies that parameter i has higher influence than parameter j to the final result.

B. Phase-II

In *Phase-II* the individual values estimated from Section III-A are compared to get the most significant parameter for the workflow. In this step, first we calculate the *Normalization Factor*; that is we sum the estimated values (that is *PC*) of all the parameters and divide it by the total number of parameters (that is *n*).

$$Normalization\ Factor = \frac{\sum_{i=1}^{n} PC_{(i)}}{n} \tag{2}$$

$$\begin{aligned} &where, \\ n &= is\ the\ total\ number\ of\ parameters, \\ PC(_i) &= the\ ratio\ of\ percentage\ of\ change\ in \\ &\quad final\ result\ to\ the\ percentage\ of\ change \\ &\quad inparameter\ i,\ calculated\ in\ Step\ III - A. \end{aligned}$$

In the final step, for each parameter its weight is calculated by dividing each parameter estimated value (*PC*) on the *Normalization Factor*:

$$WP_{(i)} = PC_{(i)}/Normalization\ Factor \tag{3}$$

$$\begin{aligned} &where, \\ WP_{(i)} &= Weight\ of\ Parameter\ i, \\ PC(_i) &= Value\ of\ PC\ for\ variable\ i\ calculated \\ &\quad\ in\ Equation\ (1). \end{aligned}$$

Now, if $WP_{(i)} > 1$, then it means that the parameter *i* is more significant than the average level of all the parameters, and $WP_{(i)} < 1$ means less significant.
The definition seems not faultless as for handling nominal values, because the term *percentage* is meaningless for nominal values. To extend the usage of the formula to nominal values, we modify the meaning of p_i (percentage of change in parameter) only when coping with nominal values:

$$p_i = \frac{1}{m} \tag{4}$$

$$\begin{aligned} &where, \\ p_i &= percentage\ of\ change\ in\ parameter, \\ m &= number\ of\ values\ or\ categories. \end{aligned}$$

IV. Experimental evaluation

Before proceeding to experimental details, we give a brief introduction to the constraints and workflow domain in Subsection IV-A. In Subsections IV-B and IV-C we apply our two step proposed methodology on the selected workflow and share their results.

A. Case Study: Non-Invasive Glucose Measurement Workflow

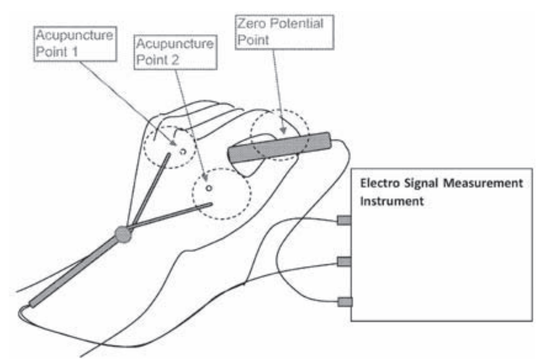

Fig. 2. Non-Invasive Glucose Measurement Technique

The non-invasive method for measuring human glucose values used in the NIGM workflow is based on the meridian-theory, which is an important part of the Traditional Chinese Medicine (TCM) [12], according to which the human body has 14 acupuncture meridians. Each of these longitudinally distributed lines on our human body has its main points, called source points, totally 24. In order to prove the meridian theory with modern methods a number of special meridian measurement instruments were developed. Analyzing these meridian measurement data with advanced data mining techniques and models can lead to important information about human illness state and other health relevant knowledge. The electro signal measurement instrument sends an electric signal (white noise) into one meridian source point and measures the corresponding signal output at another source point either on the same meridian or on another meridian. In particular a random electro signal with the maximal voltage less than 2.0 V is produced by the instrument. This process is illustrated in Figure 2. The measurements obtained in this process can, if analyzed by the meridian electro information transmission model, derive diabetic patients blood glucose values. The support with an enhanced Grid infrastructures allowing collaborative research with advanced data mining services, efficient data and workflow management services, and visualization services contribute to the progress in this domain.

We have chosen the sequential Non-Invasive Glucose Measurement (NIGM) [13] workflow for analysis and estimation of most significant parameter in neural network model. NIGM workflow as depicted by Figure 3 consists of five WS-I [14] and WSRF-compliant [15] Web services executed by WEEP[1] [16]. NIGM workflow takes data obtained by measurements (see Figure 2) on patient meridians as input and then performs several activities, like *System Identification* [17], *Kalman Filtering* [18], and *Fast Fourier Transformation* [19].

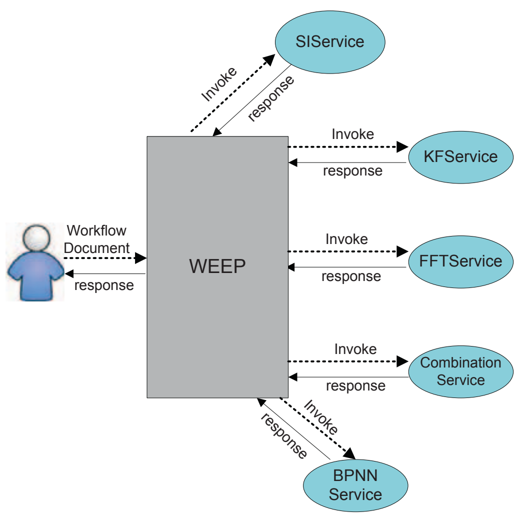

Fig. 3. NIGM Workflow execution by WEEP

NIGM efficiently computes the glucose value in patient blood by a method developed with the help of a Traditional Chinese Medicine (TCM) theory. The first three services compute eigenvalues for a given set of input value pairs (meridian measurements) as shown in Figure 4 along with all associated input(s) and output(s). The combination service combines the results given by first three services and the neural network service accurately predicts the blood glucose value. Computational models used in NIGM have several parameters. So it is very important for the user of the model to know about the significance of these parameters, if he/she

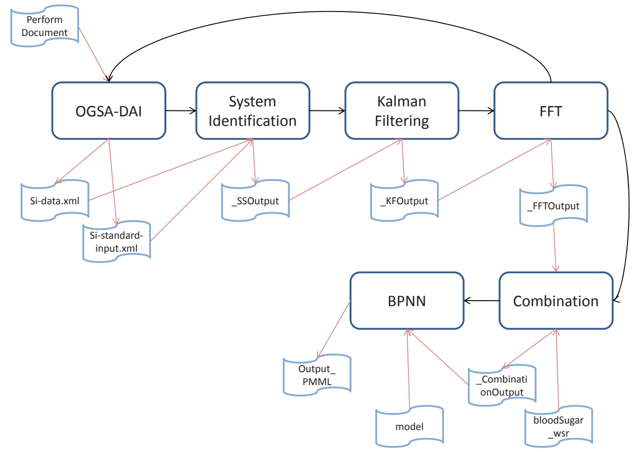

Fig. 4. NIGM Workflow

wants to fine tune the model. The neural network model, which is a core component of the NIGM workflow, has two parameters: *learning rate* and *momentum*. *Learning rate* is used in the training phase of neural network model, and for the determination of the numerical values of the weights. The *momentum* parameter allows a change to the neural network weights to persist for a number of adjustement cycles, in other words, the magnitude of the persistence is controlled by this parameter. Both these parameters have effect on the result of the model, but to the best of our knowledge, so far no work has been done to quantify the significance of these two parameters. To address this issue, we empirically validate our proposed methodology via estimating the most significant parameter to the neural network model.

The *BPNNService (Back Propagation Neural Network Service)* of the NIGM workflow has two functions:

1) It builds an individual health model for each considered person. This model is patient specific and has the form of neural net. It is stored in a PMML file (PMML[2] is the standardized XML format for representation of statistical and data mining models).
2) The patient specific model is used to predict blood glucose value for the considered particular patient.

B. Phase-I:

Both of the neural network parameters that is *learning rate* and *momentum* are of type real and have the value interval of (0, 1]. To estimate the significance of *learning rate* we set *momentum* to 0.3 and it remains unchanged throughout all the invocations. We sample *learning rate* for an interval of 0.05, starting from 0.01 to the maximum possible value of *learning rate* i.e. 1.00. Then we calculate mean absolute error for each sample. The specification of the sample input data is shown in Table I. Mean absolute error is a value that reflects the difference between prediction and supervisor values of a

[1] Workflow Enactment Engine Project (WEEP) is developed by Department of Scientific Computing at the University of Vienna. The source code of the engine is freely available at http://weep.gridminer.org for download under the terms and conditions of the Apache License, Version 2.0.

[2] www.dmg.org

TABLE I

ATTRIBUTES OF THE SAMPLE INPUT DATA

Input data for SI:- input to NIGM	
File Name	Attributes
input_SI1.xml - input_SI50.xml -50 input Files	Number of rows in each input file: 5000 Number of columns in each input file: 1 Data type: real Data format: WebRowSet
standardInput_SI.xml -1 input file	Number of rows: 5000 Number of columns: 1 Data type: real Data format: WebRowSet
Input data for NN:- Training data for NN	
Input_NN.xml -1 input file	Number of rows: 50 Number of columns: 4 Column 1,2,3: eigenvalues of FFT output curves Column 4: supervisor value of blood-sugar Data type: real Data format: WebRowSet

TABLE II

RESULTS OF CHANGING THE *learning rate* BY 0.05

Learning Rate	Mean Absolute Error
0.01	3.30
0.05	2.98
0.10	2.92
0.15	2.91
0.20	3.09
0.25	3.35
0.30	3.58
0.35	3.89
0.40	4.14
0.45	4.29
0.50	4.47
0.55	4.68
0.60	4.89
0.65	5.58
0.70	6.04
0.75	5.87
0.80	6.93
0.85	7.35
0.90	7.12
0.95	7.54
1.00	7.67

neural network model. Predicted and supervisor values are both human blood sugar values of the data type real. The predicted values are calculated and displayed by the neural network model, whereas the supervisor values are measured from the patient. The supervisor values are used in the training process of a model as target values for the training. Here we take mean absolute error as a measurement of the model result, namely the prediction values. Change in the result is reflected as a change in the mean absolute error. The results are shown in Table II.

In the next step, the differences of mean absolute error values between every two samples, namely the differences of every two adjacent rows, are calculated, which are recorded in Table III.

The mean of these absolute differences, according to the calculation, is 0.30, which is then 6.2% of the possible interval (max. mean absolute error - min. mean absolute error). The sample interval 0.05 is 5.0% of the possible interval (0, 1]. This implies that the change of 5.0% in *learning rate* results in 6.2% change in result in average. Therefore the percent

TABLE III

DIFFERENCE OF ERROR BETWEEN ADJACENT ROWS

Learning Rate	Absolute Diff: in MAE
0.01 - 0.05	0.32
0.05 - 0.10	0.06
0.10 - 0.15	0.01
0.15 - 0.20	0.18
0.20 - 0.25	0.26
0.25 - 0.30	0.23
0.30 - 0.35	0.31
0.35 - 0.40	0.25
0.40 - 0.45	0.15
0.45 - 0.50	0.18
0.50 - 0.55	0.21
0.55 - 0.60	0.21
0.60 - 0.65	0.69
0.65 - 0.70	0.46
0.70 - 0.75	0.17
0.75 - 0.80	1.06
0.80 - 0.85	0.42
0.85 - 0.90	0.23
0.90 - 0.95	0.42
0.95 - 1.00	0.13

change in parameter *learning rate* is:

$$\begin{aligned} PC_{(learningRate)} &= 6.2/5.0 \\ PC_{(learningRate)} &= 1.24 \end{aligned} \tag{5}$$

Now, we calculate the percent change for *momentum* in the same manner. We set *learning rate* to 0.3 and sample *momentum* for every 0.05 in the meaningful interval (0,1] and record once again mean absolute error values. We calculate

$$\begin{aligned} PC_{(momentum)} &= 3.1/5.0 \\ PC_{(momentum)} &= 0.62 \end{aligned} \tag{6}$$

C. Phase-II:

To be able to compare and get the most significant parameter, we first calculate the normalization factor i.e. $\frac{\sum_{i=1}^{n} PC_{(i)}}{n}$, which is the mean of Equation (5) and Equation (6) i.e. $PC_{(learningRate)}$ and $PC_{(momentum)}$:

$$\begin{aligned} Normalization\ Factor &= \frac{\sum_{i=1}^{n} PC_{(i)}}{n} \\ &= \frac{PC_{(learningRate)} + PC_{(momentum)}}{2} \\ &= \frac{1.24 + 0.62}{2} \\ Normalization\ Factor &= 0.93 \end{aligned} \tag{7}$$

So the Weight of *learning rate* parameter to the BPNN according to the Equation (3) is:

$$
\begin{aligned}
WP_{(learningRate)} &= \frac{PC_{(learningRate)}}{Normalization\ Factor} \\
&= \frac{1.24}{0.93} \\
WP_{(learningRate)} &= 1.333
\end{aligned}
\tag{8}
$$

Similarly, the weight of *momentum* parameter to the BPNN is:

$$
\begin{aligned}
WP_{(momentum)} &= \frac{PC_{(momentum)}}{Normalization\ Factor} \\
&= \frac{0.62}{0.93} \\
WP_{(momentum)} &= 0.667
\end{aligned}
\tag{9}
$$

From Equation (8) and (9) it is clear that the *learning rate* posses a higher weight than *momentum*, which means that the modification on *learning rate* should exert higher influence to the result than that of *momentum*.

We repeat the experiment on NIGM services workflow with *momentum* set to 0.7 and respectively calculate the mean absolute errors with *learning rate* 0.2 and 0.8, the result i.e. mean absolute errors, turns out to be 3.24 and 6.28, the difference of which is 3.04. Then we set *learning rate* to 0.7 and calculate the mean absolute errors with *momentum* 0.2 and 0.8, the result is 4.30 and 5.97, with the difference 1.67. The experiment implies that for the same modification to the parameters, the change in workflow result related to *learning rate* is higher than the one related to *momentum*. This complies with our first experiment that the *learning rate* poses higher significance than *momentum*.

V. Conclusions and Discussion

We have proposed a novel approach to estimate the significance of all workflow parameters and estimate the most sensitive workflow parameter as well. From the implementation point of view, the provenance record of a service should contain weights for every or at least significant parameters which are thought to be useful in influencing the model result. The parameter significance method asses the relative importance of parameters, and thus will assist scientists to optimize their experiments, by enabling the scientists to focus only on the most significant parameters. Furthermore, we have empirically validated our methodology using real world case study that is Non-Invasive Glucose Measurement workflow. We experimentally demonstrated that in cases where we know the expected overall workflow result (such as NIGM workflow in our case study), we can reason about the quality of results of a certain activity by studying how this activity affects the overall workflow result (that is, which parameter values of an activity leads to the overall workflow result that is close the expected result). We have observed that *learning rate* is the most significant parameter to the neural network model and small changes to the learning rate can greatly influence the BPNN result. In future, we will attempt to study the effect of this work in relation to provenance.

Acknowledgment

This research is done in the context of GridMiner [20] and ADMIRE [21] projects. As a starting point it has been implemented and tested on the NIGM workflow [13] of CADGrid Project [22].

References

[1] e Science, "UK e-Science," http://www.rcuk.ac.uk/escience.
[2] F. A. Khan, Y. Han, S. Pllana, and P. Brezany, "Provenance Support for Grid-Enabled Scientific Workflows," in *Proceedings of Fourth International Conference on Semantics, Knowledge and Grid*. Beijing, China: IEEE, December 3-5 2008, pp. 173–180.
[3] I. J. Taylor, E. Deelman, D. B. Gannon, and M. S. (Eds.), *Workflows for e-Science: Scientific Workflows for Grid*. Springer, 2006.
[4] NeSC, "National e-Science Centre," http://www.nesc.ac.uk/.
[5] OGSA-DAI, http://www.ogsadai.org.uk/.
[6] TeraGrid, "TeraGrid US," http://www.teragrid.org/.
[7] E. Deelman, D. Gannon, M. Shields, and I. Taylor, "Workflows and e-Science: An Overview of Workflow System Features and Capabilities," *Future Generation Computer System*, July 2008.
[8] K. M. McCann, M. Yarrow, A. DeVivo, and P. Mehrotra, "ScyFlow: an Environment for the Visual Specification and Execution of Scientific Workflows," *Concurrency and Computation: Practice and Experience*, vol. 18, no. 10, pp. 1155 – 1167, August 2006.
[9] X. Liu, J. Chen, K. Liu, and Y. Yang, "Forecasting Duration Intervals of Scientific Workflow Activities Based on Time-Series Patterns," in *Proceedings of IEEE Fourth International Conference on eScience, 2008. eScience '08.* IEEE, December 7-12 2008, pp. 23–30.
[10] S. Pandey and R. Buyya, "Scheduling of Scientific Workflows on Data Grids," in *Proceedings of the 2008 Eighth IEEE International Symposium on Cluster Computing and the Grid*. IEEE, 2008, pp. 548–553.
[11] I. Brandic, S. Pllana, and S. Benkner, "Specification, Planning, and Execution of QoS-aware Grid Workflows within the Amadeus Environment," *Concurrency and Computation: Practice and Experience*, vol. 20, no. 4, pp. 331–345, March 2008.
[12] G. Maciocia, *The Foundations of Chinese Medicine: A Comprehensive Text for Acupuncturists and Herbalists*. Elsevier Churchill: Livingstone, 2005.
[13] I. Elsayed, J. Han, T. Liu, A. Wohrer, F. A. Khan, and P. Brezany, "Grid-Enabled Non-Invasive Blood Glucose Measurement," in *Computational Science - ICCS 2008*. Krakow, Poland: LNCS 5101, June 2008, pp. 76–85.
[14] K. Ballinger, D. Ehnebuske, C. Ferris, M. Gudgin, C. K., Liu, M. Nottingham, and P. Yendluri, "WS-I Basic Profile Version 1.1," http://www.ws-i.org/Profiles/BasicProfile-1.1.html, 2006.
[15] I. Foster, J. Frey, S. Graham, S. Tuecke, K. Czajkowski, D. Ferguson, F. Leymann, M. Nally, T. Storey, and S. Weerawaranna, "Modeling Stateful Resources with Web Services," v1.1. Technical report, Globus Alliance, 2004.
[16] I. Janciak and P. Brezany, "Workflow Enactment Engine for WSRF-compliant services orchestration," in *The 9th IEEE/ACM International Conference on Grid Computing (Grid 2008)*. Tsukuba, Japan: IEE/ACM, September 29 - October 1 2008.
[17] L. Ljung, *System Identification - Theory For the Use*. Upper Saddle River, N.J.: PTR Prentice Hall, 1999.
[18] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," ASME - Journal of Basic Engineering, 1960.
[19] S. Bochner and K. Chandrasekharan, *Fourier Transforms*. Princeton Book Comp., 2001.
[20] GridMiner, http://www.gridminer.org/.
[21] ADMIRE, "Advanced Data Mining and Integration Research for Europe," http://www.admire-project.eu/.
[22] CADGrid, "China Austria Data Grid Project," http://www.par.univie.ac.at/project/cadgrid/.