Computer Industry Laboratory

Shigeru Kanemoto
Professor

Incheon Paik
Senior Associate Professor

Kenta Ofuji
Associate Professor

Computer Industry Laboratory

Research by faculties and students in the Computer Industry Laboratory has been carried out in the fields of both basic computer science and industrial applications, such as Semantic Web Services, Web Data Mining and Security, Semantic e-Business, Situation Awareness, Signal processing, Simulation engineering, and Functional Safety, and environmental impact analysis of energy industry. Several articles were published as book chapters, journal-contributions, proceedings-contributions in conferences, and technical reports.

The research activities of the Computer Industry Laboratory include the following topics:

[Shigeru KANEMOTO] - Signal processing for plant monitoring and control - Dynamical system identification and analysis - Image processing and 3D shape reconstruction - Human-Computer hybrid simulation - Functional safety of embedded system [Incheon PAIK] - Semantic Web Service and Composition - Big Data Analytics and Infrastructure - Web Data Mining - Semantic e-Business Application and Workflow - Situation Awareness

[Kenta OFUJI] - Socio-economic study of energy efficiency and renewable energy

Refereed Journal Papers

[o-fu-01:2014] Kenta Ofuji. Recent improvements and challenges in Fukushima: prefecture's economy and living environment. *Wiley Interdisciplinary Reviews: Energy and Environment*, page doi: 10.1002/wene.139, 2014.

Today as we approach the 3-year mark since the earthquake disaster and nuclear accident, Fukushima Prefecture 's economy is gradually improving, albeit not yet completely, and the living environment in urban areas is showing signs of recovery. Yet much remains to be done to reassure people, who have still not returned home since evacuating, about their fears concerning radiation, and to dispel harmful rumors about the tourism and the farming and fishing industries. In this article, I raise three issues: (1) decontamination and target dose, (2) contaminatedwater and revival of the fishing industry, and (3) a vision for the industrialization of renewable energy; and present examples of the progress made in recent efforts in these three issues to close the gap between the ideal and reality.

[paikic-01:2014] Incheon Paik Keunho Ryu Chenghao Jin, Goutchol Pok. Short-Term Electricity Load and Price Forecasting Based on Clustering and Next Symbol Prediction. *IEEJ Transactions ON Electrical and Electronic Engineeringg*, 10(2):175–180, 2015.

Short-term electricity load and price forecasting is an important issue in competitive electricity markets. In this paper, we propose a new direct time series forecasting method based on clustering and next symbol prediction. First, the cluster label sequence is obtained from time series clustering. Then a lossless compression algorithm of prediction by partial match version C coder (PPMC) is applied on this obtained discrete cluster label sequence to predict the next cluster label. Finally, the whole time series values of one-step-ahead can be directly forecast from the predicted cluster label. The proposed method is evaluated on electricity time series datasets, and the numerical experiments show that the proposed method can achieve promising results in day-ahead electricity load and price forecasting.

[paikic-02:2014] Koswatte R. C. Koswatte Wuhui Chen Banage T. G. S. Kumara, Incheon Paik. IMPROVING WEB SERVICE CLUSTERING THROUGH POST FILTERING TO BOOTSTRAP THE SERVICE DISCOVERY. *International Journal of Services Computing (ISSN 2330-4472)*, 2(3):1–13, 2014.

Service Oriented Architecture (SOA) has altered the way programmers develop applications. Instead of using standalone libraries, programmers today often incorporate curated web services, accessed via well-defined interfaces (APIs), as modules in their applications. Web APIs, however, evolve rapidly, making it critical for developers to be able to compare APIs for similarity and estimate the workload associated with "porting" applications to use different or new APIs (or API versions). Unfortunately, today there is no simple automated mechanism for analyzing the similarity between web APIs and reasoning about the porting effort that will be necessary when the web APIs that an application uses change. To address this limitation, we describe an automated methodology for analyzing API similarity and quantifying the porting effort associated with the use of web APIs. Our approach defines a simple type system and a language with which API developers specify the syntactic and semantic features of APIs. We also define algorithms that transform the syntactic and semantic features of APIs into similarity and porting effort information. We evaluate our approach using both randomly generated and real-world APIs and show that our metric captures the relative difficulty that developers associate with porting an application from one API to another.

[paikic-03:2014] I. Paik W. Chen.   Toward Better Quality of Service Composition Based on Global Social Service Network. *IEEE Transactions on Parallel and Distributed Systems*, 26(5):1466–1476, 2014.

Automatic service composition can create new value-added services dynamically and automatically from existing services in an envisioned service-oriented architecture. However, despite considerable progress, web-scale uptake has been significantly less than initially anticipated because of several challenging issues, such as poor scalability, exponentially expanding search time in large search spaces, and the lack of service sociability caused by the isolation of services. In this paper, we propose an innovative methodology for moving from isolated service islands to a global social service network (GSSN) by developing a network model that supports service sociability. First, we propose the construction of a GSSN based on the quality of social links. We then propose an algorithm that maps the GSSN into a service cluster network to reduce the search space, and a quality-driven composition approach that enables exploitation of the service cluster network by providing workflow as a service. Finally, experimental results show that our GSSN-based approach can solve the service composition problem well, im-

proving not only the response time but also the success rate.

[paikic-04:2014] W. Chen K. Ryu B.T.G.S. Kumara, I. Paik. Web Service Clustering using a Hybrid Term-Similarity Measure with Ontology Learning. *International Journal of Web Services Research*, 11(2):22–45, 2014.

Clustering Web services into functionally similar clusters is a very efficient approach to service discovery. A principal issue for clustering is computing the semantic similarity between services. Current approaches use similarity-distance measurement methods such as keyword, information-retrieval or ontology based methods. These approaches have problems that include discovering semantic characteristics, loss of semantic information and a shortage of high-quality ontologies. In this paper, the authors present a method that first adopts ontology learning to generate ontologies via the hidden semantic patterns existing within complex terms. If calculating similarity using the generated ontology fails, it then applies an information-retrieval-based method. Another important issue is identifying the most suitable cluster representative. This paper proposes an approach to identifying the cluster center by combining service similarity with term frequencyinverse document frequency values of service names. Experimental results show that our term-similarity approach outperforms comparable existing approaches. They also demonstrate the positive effects of our cluster-center identification approach.

Refereed Proceeding Papers

[kanemoto-01:2014] Masaya WATANABE Shigeru KANEMOTO and Noritaka YUSA. Advanced Machine learning Algorithm Application for Rotating Machine Health Monitoring. In *ISOFIC/ISSNP 2014 (International Symposium on Future I&C for Nuclear Power Plants/International Symposium on Symbiotic Nuclear Power Systems)*,, Jeju, Korea, August 2014.

The present paper tries to evaluate the applicability of conventional sound analysis techniques and modern machine learning algorithms to rotating machine health monitoring. These techniques include support vector machine, deep leaning neural network, etc. The inner ring defect and misalignment anomaly sound data measured by a rotating machine mockup test facility are used to verify the above various kinds of algorithms. Although we cannot find remarkable difference of anomaly discrimination performance, some methods give us the very interesting eigen

patterns corresponding to normal and abnormal states. These results will be useful for future more sensitive and robust anomaly monitoring technology.

[kanemoto-02:2014] Kenta Suenaga Ryota Yamada Shigeru Kanemoto, Taku Hagihara. Application of new hazard analysis model for embedded systems. In *The 2nd International Workshop on Functional Modelling for Design and Operation of Engineering Systems and Infrastructures(IWFM2015)*〞,, Okayama, Japan, March 2015.

Nancy Leveson proposed a new analysis method called STAMP/STPA(Systems-Theoretic Accident Model and Process / System-Theoretic Process Analysis) to analyze hazard of complicated system which has many interactions between human and machines. The present paper demonstrates its usefulness through a case study of simulated chemical plant accident. In this case study, we found some advantages of STAMP/STPA comparing the conventional hazard analysis methods. They are easy to use than the conventional hazard analysis and can extract more detailed hazard causal factors(HCF).

[o-fu-02:2014] Osamu Kimura and Kenta Ofuji. Estimation of the cost-effectiveness of subsidy programs for energy efficiency. In Society for Environmental Economics and Policy Studies (SEEPS), editors, *2014 Annual Conference of the Society for Environmental Economics and Policy Studies (SEEPS)*, Tokyo, Japan, September 2014. Society for Environmental Economics and Policy Studies (SEEPS), Society for Environmental Economics and Policy Studies (SEEPS).

This report analyses the effectiveness of subsidy propgrams for energy efficiency investment in Japan. Since the first Oil Crisis the government has been conducting various direct subsidies, tax incentives, and low-interest loans. The budgets for direct subsidies have been increasing since the late 1990 ' s, which now amount to about 100 billiion JPY per year. When tax incentives for energy efficiency are added the total governmental expenditure amounts to about 200 billion JPY per year. Despite such emphasis on subsidy programs, there has been scarce effort on their expost evaluation so far. Therefore, we conducted a survey to commercial and industrial facilities which participated in the three major subsidiy programs for energy efficiency investments by NEDO in order to estimate their freerider rates and, based on them, net-to-gross-ratios. Using

a silimar method to those adopted in California and New York, which are two of the most active States in the US in demand-side management, we estimated net-to-gross ratios as about 50 to 60 net-to-gross-ratios we also evaluated levelized cost per energy saved by the programs. The estimated program administrator costs (PAC) were from 10000 to 40000 JPY per kilolitters of crude oil equivalent saved and 5000 to 20000 JPY per t-CO2 saved, and the estimated total resource costs (TRC) from 20000 to 60000 JPY per kilolitters of crude oil equivalent saved and 7000 to 30000 JPY per t-CO2 saved. Those results imply that the subsidy programs were mostly cost-effective when compared to the avoided costs and carbon emission prices.

[paikic-05:2014] Hiroki Ohashi Yuichi Yaguchi Wuhui Chen Banage T. G. S. Kumara, Incheon Paik. Context Aware Filtering and Visualization of Web Service Clusters. In IEEE ICWS Committee, editor, *Proceedings of IEEE International Conference on Web Services*, pages 89 – 96, Alaska, USA, July 2014. IEEE ICWS Committee, IEEE CPS.

Web service filtering is an efficient approach to address some big challenges in service computing, such as discovery, clustering and recommendation. The key operation of the filtering process is measuring the similarity of services. Several methods are used in current similarity calculation approaches such as string-based, corpus-based, knowledge-based and hybrid methods. These approaches do not consider domain-specific contexts in measuring similarity because they have failed to capture the semantic similarity of Web services in a given domain and this has affected their filtering performance. In this paper, we propose a context-aware similarity method that uses a support vector machine and a domain dataset from a context-specific search engine query. Our filtering approach uses a spherical associated keyword space algorithm that projects filtering results from a three-dimensional sphere to a two-dimensional (2D) spherical surface for 2D visualization. Experimental results show that our filtering approach works efficiently.

[paikic-06:2014] Hiroki Ohashi Wuhui Chen Koswatte R. C. Koswatte Banage T. G. S. Kumara, Incheon Paik. Context Aware Post-Filtering for Web Service Clusters. In IEEE SCC Committee, editor, *Proceedings of IEEE International Conference on Service Computing*,

pages 440 – 447, Alaska, USA, July 2014. IEEE SCC 2014 Organizing Committee, IEEE CPS.

Web service discovery is becoming a challenging and time consuming task due to large number of Web services available on the Internet. Organizing the Web services into functionally similar clusters is one of a very efficient approach for reducing the search space. However, similarity calculation methods that are used in current approaches such as string-based, corpus-based, knowledge-based and hybrid methods have problems that include discovering semantic characteristics, loss of semantic information, encoding fine-grained information and shortage of high-quality ontologies. Because of these issues, the approaches couldn't identify the correct clusters for some services and placed them in wrong clusters. As a result of this, cluster performance is reduced. This paper proposes post-filtering approach to increase precision by rearranging services incorrectly clustered. Our approach uses context aware method that learns term similarity by machine learning under domain context. Experimental results show that our post-filtering approach works efficiently.

[paikic-07:2014] Incheon Paik K.R.C. Koswatte and B.T.G.S. Kumara. Meta-Framework for Semantic TRIZ. In IEEE SSCI Publication Committee, editor, *Proceedings of IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–17, Orlando, Florida, U.S.A., December 2014. IEEE SSCI Organizing Committee, IEEE CPS.

In the manufacturing industry, SCM (Supply Chain Management) is playing an important role which gives profit to enterprise. Extracting innovative design considerations for a product from the information infrastructure requires large knowledge base and solutions to technical and physical contradictions. Construction of information infrastructure with the integration of four designs attributes: component cost, quality, function and technology, innovation can be enhanced. Information which is useful to improve the existing products and development of a new product can be acquired from the database and from the ontology. The TRIZ (Theory of Inventive Problem Solving) supports designers for innovative product design by searching from a knowledge base. The existing TRIZ ontology support innovative design of a product. But it is considering about a specific product (flashlight) for TRIZ ontology. Our final goal is to construct meta-TRIZ ontology that can be applied to multiple products. To achieve this goal, we try to apply the semantic TRIZ to another

product; multifunction fan (Smart Fan), as an interim stage towards meta-ontology. This may open up new possibilities to innovative product designs with multifunction.

[paikic-08:2014] Incheon Paik Wuhui Chen, Banage T. G. S. Kumara and Zhenni Li. Topology-aware Heuristic Data Allocation Algorithm for Big Data Infrastructure. In Big Data Service Publication Committee, editor, *Proceedings of IEEE International Conference on Big Data Computing Service and Applications*, pages 11–13, San Francisco, U.S.A, March 2015. Big Data Service 2015 Organization Committee, IEEE CPS.

We propose a novel optimal data replacement technique considering not only the data locality but also the global data access cost to improve the performance of MapReduce in cloud data centers. We first conduct analytical and experimental study to identify the performance issues of MapReduce in data center and show that MapReduce tasks which are involved in unexpected remote data access take much more communication cost and execution time, and could significant deteriorate the all over performance. To solve optimal data replacement problem, we propose a topology-aware heuristic Algorithm by firstly constructing a replica-equalized structure for abstract tree structure, and then building replica-similarity structure for detail tree construction. The experimental results demonstrated that our optimal data replacement approach can minimize global data access cost effectively with low communication cost and less execution time, by reducing the unexpected remote data access.

[paikic-09:2014] Banage T. G. S. Kumara T.H.A.S. Siriweera Koswatte R. C. Koswatte, Incheon Paik. Meta-Ontology for Innovative Product Design with Semantic TRIZ. In ICEIT 2015 Committee, editor, *Proceedings of International Conference on Electrical and Information Technologies (ICEIT'15)*, pages 1429–1435, Moroco, Feburary 2015. ICEIT 2015 Organizing Committee, IEEE CPS.

In the manufacturing industry, SCM (Supply Chain Management) is playing an important role is providing profit to the enterprise. Market globalization and rapid advancement of technologies require that companies differentiate themselves with innovative products and services to create competitive advantage. Increasingly, manufacturers face higher pressure to shorten their time for product design to enter market. TRIZ (Theory of Inventive Problem Solving) supports designers for innovative prod-

uct design by searching from knowledge base. The existing TRIZ ontology supports innovative design of a product. But it considers about specific products (Flashlight and SmartFan) for TRIZ ontology. We applied the semantic TRIZ to products as an interim stage towards meta-ontology that can manage general products and other concepts. The research in this paper aims at developing meta-ontology for innovative product design that can be applied to multiple products in different domain areas. Modeling ontologies of real world (SmartPen and SmartMachine) is taken as an evaluation for the meta-ontology. This may open up new possibilities to innovative product designs.

[paikic-10:2014] Incheon Paik Sheng-Han Yang. Improving Trust Based Recommender System By Mining User. In FTRA, editor, *Comments on the Web World IT Congress*, pages 1–10, Jeju, Korea, Feburary 2015. FTRA, Springer.

Recommender systems help a user to provide a list of preferable items to the user. And with growth of social network service on the Web, trust based recommender systems with the collaborative filtering has been emerged, and are being studied actively. One important research issue has been to find optimal trust values by several stages of updating the trust values. The approaches usually start the updating with random or average value of trust at the initial stage. We are motivated that defining initial trust value between users contributes to better performance in trust based recommender system. In this paper, we proposed a novel method to calculate the initial value from users' comments for product items by machine learning technology. We incorporate the initial trust value from our model with the trust based propagation model, and evaluate performance of the recommendation.

[paikic-11:2014] K. R. C. Koswatte B. T. G. S. Kumara, Incheon Paik and Wuhui Chen. Ontology Learning with Complex Data Type for Web Service Clustering. In IEEE SSCI Publication Committee, editor, *Proceedings of IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 129–136, Orlando, Florida, U.S.A., December 2014. IEEE SSCI Organizing Committee, IEEE CPS.

Clustering Web services into functionally similar clusters is a very efficient approach to service discovery. A principal issue for clustering is computing the semantic similarity between services. Current approaches use

similarity-distance measurement methods such as keyword, information-retrieval or ontology based methods. These approaches have problems that include discovering semantic characteristics, loss of semantic information and a shortage of high-quality ontologies. Further, current clustering approaches are considered only have simple data types in services' input and output. However, services that published on the web have input/ output parameter of complex data type. In this research, we propose clustering approach that considers the simple type as well as complex data type in measuring the service similarity. We use hybrid term similarity method which we proposed in our previous work to measure the similarity. We capture the semantic pattern exist in complex data types and simple data types to improve the ontology learning method. Experimental results show our clustering approach which uses complex data types in measuring similarity works efficiently.

[paikic-12:2014] K. R. C. Koswatte B. T. G. S. Kumara, Incheon Paik and Wuhui Chen. Recommendation for Web Services with Domain Specific Context Awareness. In IEEE SSCI Publication Committee, editor, *Proceedings of IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 281–287, Orlando, Florida, U.S.A., December 2014. IEEE SSCI Organizing Committee, IEEE CPS.

Construction of Web service recommendation systems for users has become an important issue in service computing area. Content-based service recommendation is one category of recommendation systems. The system recommends services based on functionality of the services. Current content-based approaches use syntactic or semantic methods to calculate the similarity. However, syntactic methods are insufficient in expressing semantic concepts and semantic content-based methods only consider basic semantic level. Further, the approaches do not consider the domain specific context in measuring the similarity. Thus, they have been failed to capture the semantic similarity of Web services under a certain domain and this is affected to the performance of the recommendation. In this paper, we propose domain specific context aware recommendation approach that uses support vector machine and domain data set from search engine in similarity calculation process. Experimental results show that our approach works efficiently.

[paikic-13:2014] Keunho Ryu Hosun Shon Incheon Paik Dingkun Li, Minghao Piao. One Pass Preprocessing for Token-based Source Code

Clone Detection. In IEEE iCAST 2014 Publication Committee, editor, *Proceedings of International Conference on Awareness Science and Technology (iCAST) 2014*, pages 101–106, Paris, France, November 2014. IEEE iCAST 2014 Organizing Committee, iCAST 2014 Publication Committee.

Token-based source code clones detection provides a promising way to detect the source code duplication and re-dundancy. While preprocessing of clone detection plays an important role in KDD for further processing as the old saying goes: well begun is half done. However, processing unstructured source code files of large software systems is really challenging and time or space consuming. This paper introduces a novel way to clean, tokenize and transform the source code into the appropriate form for mining. A tool called OPP (One Pass Preprocessor) has been developed to preprocess the source code files efficiently and flexibly. The paper experimented on three large open source projects like Wildfly1.02 Linux core-3.6, VTK of different host languages, and the result showed that our tool has great power and flexibility to preprocess the source code files and products high quality output.

[paikic-14:2014] Incheon Paik. Improved Malicious Code Classification Considering Sequence by Machine Learning. In IEEE ISCE Publication Committee, editor, *Proceedings of The 18th IEEE International Symposium on Consumer Electronics*, pages 1–4, Jeju, Korea, June 2014. IEEE ISCE 2014 Organizing Committee, IEEE ISCE Publication Committee.

Classification of malicious code by machine learning gives more flexible and adaptable prediction result than by existing approaches [1]. But the approach just can identify looks-like malicious code instead of real malicious one. In this research, a novel method to reduce the vagueness in the classification by machine learning to consider code sequence.

Unrefereed Papers

[o-fu-03:2014] Kenta Ofuji and Osamu Kimura. Freerider Estimation of C&I Energy Efficiency Subsidy Programs in Japan. In *Proceedings of the 30th Conference on Energy, Economy, and Environment*, pages 27–5. Japan Society of Energy and Resources, January 2015.

Summary of Achievement

This paper presents free-rider (FR) estimation of energy efficiency subsidy programs for Japanese commercial and industrial energy users. The estimation is based on methodologies practiced in two U.S. states, New York(NY) and California(CA). To prepare data, we conducted a field questionnaire survey in late 2013 and collected 506 valid responses. The calculated FR rates were almost in good agreement between the two states' methods, validating the counter-factual, and multi-faceted question concepts. The counterfactural questions help avoid conflicting responses, while multi-faceted questions better depict diverse situations behind the energy users' investment decision making. Results between the two methodologies, however, presented dissimilarities in sensitivity to some question items, leaving room for continued study.

Academic Activities

[o-fu-04:2014] K. Ofuji, September 2013-.

   member, editorial committee

[paikic-15:2014] Incheon Paik, November 2014.

   Vice Chair

[paikic-16:2014] Incheon Paik, July 2014.

   Program Committee Member

[paikic-17:2014] Incheon Paik, July 2014.

   Program Committee Member

[paikic-18:2014] Incheon Paik, Nov. 2014.

   Publicity Chair

[paikic-19:2014] Incheon Paik, January 2014.

   Chair

Ph.D and Others Theses

[o-fu-05:2014] Mina Sasaki. Graduation Thesis: Intertemporal changes in JEPX forward contracts: a count model analysis, University of Aizu, 2014.

   Thesis Advisor: K. Ofuji

[o-fu-06:2014] Shiho Fujiwara. Graduation Thesis: Household power conservation effects: a multi-level regression analysis, University of Aizu, 2014.

Thesis Advisor: K. Ofuji

[o-fu-07:2014] Michika Nakano. Graduation Thesis: Barriers in domestic energy-efficient water heater choice using SEM and discrete choice models, University of Aizu, 2014.

Thesis Advisor: K. Ofuji

[o-fu-08:2014] Chika Go. Graduation Thesis: Intertemporal changes in JEPX spot trading: a vector error correction analysis, University of Aizu, 2014.

Thesis Advisor: K. Ofuji

[paikic-20:2014] Kenta Takahashi. Automating Bigdata Analysis using Service Composition, University of Aizu, 2014.

Graduation Thesis Advisor: I. Paik

[paikic-21:2014] Koswatte Ralalage Chathurika Koswatte. Innovative Product Design using Meta Ontology with Semantic TRIZ, Graduate School, University of Aizu, 2014.

Master Thesis Advisor: I. Paik

[paikic-22:2014] Banage Thenne Gedara Samantha Kumara. Improving Web Service Clustering through Ontology Learning and Context Awareness, Graduate School, University of Aizu, 2014.

Ph.D. Thesis Advisor: I. Paik

[paikic-23:2014] Yutaka Koshiba. Investigation of Distributed Data Centers for Performance Improvement, University of Aizu, 2014.

Graduation Thesis Advisor: I. Paik

[paikic-24:2014] Hirotoki Okamoto. Performance Observation in Distributed Data Center, University of Aizu, 2014.

Graduation Thesis Advisor: I. Paik

[paikic-25:2014] Awa Yousuke. Classification of Enterprise SNS Users using Machine Learning, University of Aizu, 2014.

Graduation Thesis Advisor: I. Paik

Summary of Achievement

[paikic-26:2014] Mizuki Watanabe. Analysis of Enterprise SNS User Characteristics on Big Data Infrastructure, University of Aizu, 2014.

    Graduation Thesis Advisor: I. Paik

Others

[o-fu-09:2014] Kenta Ofuji, 9 2014.

    Invited Speech: The most fundamental motivation for big data analysis , UoA, Joint seminar with Accenture, Sep. 19, 2014