

# Data intelligence on the Internet of Things

Zhangbing Zhou<sup>1,2</sup> · Kim-Fung Tsang<sup>3</sup> · Zhuofeng Zhao<sup>4</sup> · Walid Gaaloul<sup>2</sup>

Published online: 25 April 2016  
© Springer-Verlag London 2016

The Internet of Things, which enables the interconnection, interoperation, and collaboration between smart things, allows collecting data from various sources, including GPS data of vehicles, real-time traffic data of road cameras, weather data (e.g., temperature or air quality data) from environment sensors and user-generated contents (e.g., tweets, micro-blog, check-ins, photos) from mobile social APPs. In fact, sensory data have been widely available in large volume and variety nowadays. Sensory data exhibit specific characteristics, including multi-sources, heterogeneous, large-scale, real-time streaming, continuous, ever-expanding, and spatial-temporal. Traditional approaches or platforms are limited in processing these sensory data, which are big data actually. The engineering and intelligence on sensory data covers the theories and technologies of different disciplines to provide efficient processing and smart analysis. Intensive research is required on sensory data engineering and intelligence. This special issue, as a dedicated forum, aims for the scientific and industrial community to present their novel models, methodologies, techniques, and solutions which can address theoretical and practical issues. It is worth mentioning that the majority of submissions are the selected papers with high quality

which have been reported at the 2015 Smart World Congress. These selected papers are seriously improved and recommended to this special issue. The other submissions are from the open-call. After carefully reviewing submissions, there are 15 articles being accepted finally. A brief summary about each article is presented as follows:

Internet of Things (IoT) connects billions of devices in an Internet-like structure. Each device encapsulated as a real-world service in which provides functionality and exchanges information with other devices. This large-scale information exchange results in novel interactions between things and people. Unlike traditional Web services, Internet of Services (IoS) is highly dynamic and continuously changing due to constant degradation, vanishing, and possible reappearance of the devices, and this opens a new challenge in the process of resource discovery and selection. In response to increasing numbers of services in the discovery and selection process, there is a corresponding increase in number of service consumers and consequent diversity of quality of service (QoS) available. Increase in both sides leads to the diversity in the demand and supply of services, which would result in the partial match of the requirements and offers. In the article [1], Ahmed et al. propose an IoT service ranking and selection algorithm by considering multiple QoS requirements and allowing partially matched services to be counted as a candidate for the selection process. One of the applications of IoT sensory data that attracts many researchers is transportation especially emergency and accident services which is used as a case study in this article. Experimental results from real-world services showed that the proposed method achieved significant improvement in the accuracy and performance in the selection process.

The key idea of mobile cloud computing (MCC) is using powerful backend computing nodes to enhance capabilities

---

✉ Zhangbing Zhou  
zhangbing.zhou@gmail.com

<sup>1</sup> School of Information Engineering, China University of Geosciences, Beijing, China

<sup>2</sup> Computer Science Department, TELECOM SudParis, Paris, France

<sup>3</sup> Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong

<sup>4</sup> Research Center for Cloud Computing, North China University of Technology, Beijing, China

of small mobile devices and provide better user experience. An effective and widely used approach to realizing this vision is task migration. Decision making is an important aspect of migration which affects the feasibility and effectiveness of task migration. There have been a number of research efforts to MCC which help make decisions for task migration. In the article [2], Zhang et al. present a comprehensive survey on decision making for task migrations in MCC, including decision factors and algorithms. The authors observe that there are still some challenges such as comprehensive context-awareness, migration standards, large-scale experiments, and more involvement of latest achievement from artificial intelligence, and flexible decision-making mechanisms are needed to support task migration.

Semantic-oriented service matching is one of the challenges in automatic service discovery. A number of approaches to computing the semantic similarity between terms have been developed to compare service profiles and user requests. Most corpus-based approaches rely on the distributional statistics of context to represent a term as a single vector, which makes it difficult to handle polysemous problems. Meanwhile, even though knowledge-based approaches can benefit from the differentiated concept information provided by knowledge bases, they are limited to the lexical coverage. In this article [3], in order to address the issue of lexical ambiguity, Lu et al. propose a multiple semantic fusion model to generate sense-specific vector per term. In this model, various semantic properties of ontology WordNet are integrated to fine-tune the continuous word vectors learned from corpus in terms of diverse vector combination strategies. The retrofitted vectors are taken as semantic vectors for estimating semantic similarity of terms. Our similarity measure based on the MSF model is validated against different similarity measures on benchmark data sets. Experimental results indicate that our computational assessment is more conform to human judgment in most cases. Furthermore, by the application in a concrete service discovery task, it is indicated that the proposed measure based on semantic fusion could markedly improve the performance of WordNet-based measures and corpus-based measures. Accordingly, our findings provide a novel perspective to understand and represent lexical semantics.

Mobile Crowd Photography (MCP) is a widely used technique in crowd sensing. In MCP, a picture stream is generated when delivering intermittently to the backend server by participants. Pictures contributed later in the stream may be semantically or visually relevant to previous ones, which can result in data redundancy. To meet diverse constraints (e.g., spatiotemporal contexts, single or multiple shooting angles) on the data to be collected in MCP tasks, a data selection process is needed to eliminate data

redundancy and reduce network overhead. This issue has little been investigated in existing studies. To address this requirement, in the article [4] Guo et al. propose a generic data collection framework called PicPick. It first presents a multifaceted task model that allows for various MCP task specifications. A pyramid-tree (PTree) method is further proposed to select an optimal set of pictures from picture streams based on multi-dimensional constraints. Experimental results on two real-world data sets indicate that PTree effectively reduces data redundancy while maintaining the coverage requests, and the overall framework is flexible.

Traffic jam has grown to be a more and more difficult problem to be solved in big cities around the world, and people are getting less chance to take taxis. In order to remit this issue, in the article [5] Jing et al. propose a recommendation strategy based on taxi traces data for passenger by the roads to make it easier. Considering the scale of taxi traces data, Hadoop is employed to handle the traces data, whose tasks include filtering and cleaning of the data, mapping taxi traces, and computing the average passage time and empty taxi arrival rate on the roads. When a user uploads his position and the time, assisted by weather condition gain from the Internet, the authors get the model that corresponds to the date and the weather, and the time interval is thought to be the expected waiting time between the moment when user requests the service and the moment when the cumulative number of empty taxi is larger than or equal to 1 after adding the time the taxi spends on the road, and is pushed to the user. The experiment is conducted on the base of a real-world data set generated by over 12,000 taxis over a period of 3 months in Beijing. Experimental results demonstrate that the processing speed of Hadoop is nine times faster than serial's, which displays the feasibility of Hadoop in the application of massive traces data. In addition, the accuracy of the recommendation strategy reaches up to 88.75 %, and it meets the demand of real-time services.

In a resource-constrained wireless sensor network, energy efficiency is a principle issue for monitoring the movement of continuous objects, such as wildlife and hazardous chemical material. In the article [6], Han et al. propose a continuous object tracking scheme with two-layer grid model (TGM-COT). To address the problem of boundary distortion caused by uneven node distribution, the authors put forward a novel mechanism for boundary nodes identification. Besides, a streamlining mechanism is designed to reduce the amount of uploaded data. Simulation results demonstrate that, without sacrificing additional energy consumption, TGM-COT is able to achieve high tracking accuracy and significantly reduce the communication overhead.

Sensory data have becoming widely available in large volume and variety due to the increasing presence and adoption of the Internet of Things (IoT). Such data can be tremendously useful if they are processed properly in a timely fashion. They could play a key role in the coordination of industrial production. It is thus desirable to explore an effective and efficient scheme to support data tracking and monitoring. In the article [7], Luo et al. propose a novel automatic learning scheme to improve the tracking efficiency while maintaining or improving the data tracking accuracy. A core strategy in the proposed scheme is the design of Laguerre neural network (LaNN)-based approximate dynamic programming (ADP). As a traditional optimal learning strategy, ADP is a popular approach for data processing. The action neural network (NN) and the critic NN as two important components in ADP have big impact on the performance of ADP. In this article, a LaNN is employed as the implementation of the action NN in ADP considering Laguerre polynomials' approximation capability. In addition, this LaNN-based ADP is integrated into an online parameter-tuning framework to optimize those parameters of characteristic model (CM) that is used to trace the data in the tracking control system. Meanwhile, this article provides an associated Lyapunov convergence analysis to guarantee a uniformly ultimately boundedness (UUB) property for tracking errors in the proposed approach.

As the huge number of mobile devices (e.g., smart phones, tablets, and netbooks) increases, more and more people choose to use the internet services financed by mobile Internet service providers (MISPs). To provide better services, it is quite necessary for MISPs to analyze the information hidden in the big data stream generated by users. Therefore, processing the real-time big data stream efficiently has become increasingly important. However, traditional static data storage technology fails to meet the demands of real-time data processing. To improve processing capacity, many parallel processing structures are proposed, which brings up the problem about how the parallel devices can be scheduled to maximize their efficiency. In the article [8], Liu et al. propose a dynamic assignment scheduling algorithm for big data stream processing in mobile internet services and build a stream query graph to calculate the weight of every edge. The edge with the minimum weight is selected to send tuples. Simulation results show that the proper number of the logic devices can dramatically reduce system response time. Furthermore, system context switching is reduced by increasing the number of tuples sent each time.

Concomitant with rapid advancements in video streaming technology, concurrence video traffic has increased significantly on communication channels. Conflicts often arise among the various video streams on these

communication channels when the available bandwidth is limited because the required bitrates and transmission range often vary. In the article [9], Zhang et al. propose a server-side-based rate allocation algorithm for content delivery networks (CDNs). Instead of simply considering bitrate selection from the perspective of network and client conditions, the algorithm combines user experience with video bitrate allocation to maximize viewer engagement. First, the values of users are evaluated and a user value computation method is proposed. Second, the authors developed a profit maximization bitrate allocation approach (PMBAA) that enables content providers and CDNs to maximize profits by providing guaranteed video quality while users are watching programs. The results of experiments and analyses verify that PMBAA enables high-value clients to obtain more preferable bitrates than the HTTP live streaming algorithm developed by Apple Inc.

The wide distribution of mobile vehicles installed with various sensing devices and wireless communication interfaces has made vehicular mobile crowd sensing possible in practice. Owing to the heterogeneity of vehicles in terms of sensing interfaces and mobility, collecting comprehensive tempo-spatial sensing data with only one sensing vehicle is impossible. Moreover, the sensing data collected may expire in the future; as a result, sensing vehicles may have to continuously collect sensing data to ensure the relevance of these data. Although including more sensing vehicles can improve the quality of collected sensing data, this step requires additional cost. Thus, how to continuously collect comprehensive tempo-spatial sensing data with a limited number of heterogeneous sensing vehicles is a critical issue in vehicular mobile crowd sensing systems. In the article [10], Liu et al. propose a heterogeneous sensing vehicle selection (HVS) method for the collection of comprehensive tempo-spatial sensing data. On the basis of the spatial distribution and sensing interfaces of sensing vehicles and the tempo-spatial coverage of collected sensing data, a utility function is designed in HVS to estimate the sensing capacity of sensing vehicles. Then, according to the utilities of sensing vehicles and the restriction on the number of recruited sensing vehicles, sensing vehicle selection is modeled as a knapsack problem. A greedy optimal sensing vehicle selection algorithm is designed to solve the sensing vehicle selection problem.

The road network has gained more and more attention in the research area of databases. Existing works mainly focus on standalone queries, such as *k*-nearest neighbors queries over a single type of objects (e.g., facility like restaurant or hotel). In the article [11], Lin et al. propose a *k*-multi-preference (*k*MP) query over road networks, involving complex query predicates and multiple facilities. In

particular, given a query graph, a kMP query retrieves of the top-k groups of vertices (of k facility types) satisfying the label constraints and their aggregate distances are the smallest. A naive solution is to enumerate all combinations of vertices with k possible facility types and then select the one with the minimum sum distance. This method, however, incurs rather high computation cost due to exponential possible combinations. In addition, the existing solutions to other standalone queries are for a single type of facilities and cannot be directly used to answer kMP queries. Generally, this technique reduces the computation cost of the shortest path queries.

Target tracking is one of the important applications of wireless sensor networks (WSNs). Most of existing approaches assume that the nodes are dense enough and ignore the coverage holes which are very common in WSNs because of random deployment of the sensor nodes, block of obstacles, etc. Besides, predicting the target's location of the next time instance is unwise because of quite a lot random factors. In the article [12], Liu et al. propose a novel target tracking approach without any predicting, called k-nearest neighbors tracking (k-NNT), to tackle the problems of energy efficiency, continuity, and coverage holes. In k-NNT, only the k-nearest neighbors keep active and track the target when more than k nodes can sense the target; the  $k'$ -nearest neighbors work when there are only  $k'$  nodes ( $k' < k$ ) can sense the target. A sophisticated rotation mechanism is designed to improve the continuity of the tracking process. In the worst case, none of the nodes can sense the target, i.e., the target enters into the coverage holes, and then k-NNT recovers by the roundup mode (RU mode). The nodes on the perimeter of the coverage hole always keep active for a time threshold  $t$  and sense the around environment to find the target in time. Once a node finds the target, the RU mode is over and the irrelevant nodes turn into inactive mode. Experiment shows that k-NNT performs superiorly compared with several existing approaches in terms of tracking accuracy, continuity, and energy efficiency.

As the identification technology is developed day by day, so is the counterfeit, any accreditation system can be tricked. Therefore, a complete biometric identification system is supposed to distinguish between real and fake. Aiming at the liveness detection problems during the dorsal hand vein recognition process, in the article [13], Wang et al. propose a method which combines principal component analysis and power spectrum estimation of the AR model together, 3 kinds of fake hand vein images which are paper printed, wearing thin rubber gloves and wearing thick rubber gloves has been tested. Experimental result shows that the recognition rate of fake samples can reach 98.3 %, which proves that this method can realize in liveness detection of dorsal hand veins effectively.

Recommendation techniques greatly promote the development of online services in the interconnection environment. Personalized recommendation has attracted researchers' special attention because it is more targeted to individual tasks with the characteristics of diversification and novelty. However, the data sets that personalized recommendation process usually possess the characteristics of data sparseness and information loss, which is more likely to have problems such as cognitive deviation and interest drift. To solve these issues, in recent years people gradually notice the important role in which trust factor plays in promoting the development of personalized recommendation. Given the difference between online social trust and traditional offline social trust in facilitating personalized recommendation, in the article [14], Cheng et al. propose a novel technique of online social trust-reinforced personal recommendation to improve the recommendation performance. Compared with traditional offline social trust-based personal recommendation, this work synthesizes both factors of online social trust and offline social trust to identify private and public trusted user communities. The trusted degree or the accredited degree can be deduced by Bayesian network probabilistic inferences. In this way, the performance of personalized recommendation can be improved by avoiding excessive interest deviation. Moreover, the authors also get time sequence into personal recommendation model to effectively track how user's interest changes over time. Accordingly, the recommendation accuracy can be improved by eliminating the unfavorable effect of interest drift caused by temporal variation.

Along with the development of IoT applications, wearable devices are becoming popular for monitoring user data to provide intelligent service support. The wearable devices confront severe security issues compared with traditional short-range communications. Due to the limitations of computation capabilities and communication resources, it brings more challenges to design security solutions for the resource-constrained wearable devices in IoT applications. In the article [15], Liu et al. propose a yoking-proofs-based authentication protocol (YPAP) for cloud-assisted wearable devices. In the YPAP, a physical un-clonable function and lightweight cryptographic operators are jointly applied to realize mutual authentication between a smart phone and two wearable devices, and yoking-proofs are established for the cloud server to perform simultaneous verification. Meanwhile, Rubin logic-based security formal analysis is performed to prove that the YPAP has theoretical design correctness. This indicates that the proposed YPAP is flexible for lightweight wearable devices in IoT applications.

Finally, we would like to thank the authors to submit their high-quality manuscripts to this special issue. We

would like to acknowledge the contribution of the reviewers who have participated in the reviewing process and provided helpful comments and suggestions to the authors on improving their manuscripts. We specially thank Professor Peter Thomas, the Editor-in-Chief of the *Personal and Ubiquitous Computing*, for his advice and strong support during the process of putting together this special issue. We also hope that the readers will enjoy reading the articles included in this special issue.

## References

- Ahmed M, Liu L, Hardy J, Yuan B, Antonopoulos N (2016) An efficient algorithm for partially matched services in internet of services. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0917-9](https://doi.org/10.1007/s00779-016-0917-9)
- Zhang W, Tan S, Lu Q, Gong W, Li Z, Liu X, Yang S (2016) A survey on decision making for task migration in mobile cloud environments. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0915-y](https://doi.org/10.1007/s00779-016-0915-y)
- Lu W, Cai Y, Che X, Lu Y (2016) Joint semantic similarity assessment with raw corpus and structured ontology for semantic-oriented service discovery. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0921-0](https://doi.org/10.1007/s00779-016-0921-0)
- Guo B, Chen H, Yu Z, Xie X, Zhang D (2016) PicPick: a generic data selection framework for mobile crowd photography. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0924-x](https://doi.org/10.1007/s00779-016-0924-x)
- Jing W, Hu L, Shu L, Mukherjee M, Hara T (2016) RPR: recommendation for passengers by roads based on cloud computing and taxis traces data. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0925-9](https://doi.org/10.1007/s00779-016-0925-9)
- Han G, Shen J, Liu L, Qian A, Shu L (2016) TGM-COT: energy-efficient continuous object tracking scheme with two-layer grid model in wireless sensor networks. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0927-7](https://doi.org/10.1007/s00779-016-0927-7)
- Luo X, Lv Y, Zhou M, Wang W, Zhao W (2016) A Laguerre neural network-based ADP learning scheme with its application to tracking control in the internet of things. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0916-x](https://doi.org/10.1007/s00779-016-0916-x)
- Liu Y, Wang K, Yu Y, Qi J, Sun Y (2016) A dynamic assignment scheduling algorithm for big data stream processing in mobile internet services. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0914-z](https://doi.org/10.1007/s00779-016-0914-z)
- Zhang D, He H, Li W (2016) Bitrate allocation among multiple video streams to maximize profit in content delivery networks. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0919-7](https://doi.org/10.1007/s00779-016-0919-7)
- Liu Y, Niu J (2016) Comprehensive tempo-spatial data collection in crowd sensing using a heterogeneous sensing vehicle selection method. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0920-1](https://doi.org/10.1007/s00779-016-0920-1)
- Lin P, Yin Y, Nie P (2016) K-multi-preference query over road networks. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0913-0](https://doi.org/10.1007/s00779-016-0913-0)
- Liu Y, Fu J-S, Zhang Z (2016) k-Nearest neighbors tracking in wireless sensor networks with coverage holes. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0918-8](https://doi.org/10.1007/s00779-016-0918-8)
- Wang Y, Zhang D, Qi Q (2016) Liveness detection for dorsal hand vein recognition. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0922-z](https://doi.org/10.1007/s00779-016-0922-z)
- Cheng Y, Liu J, Yu X (2016) Online social trust reinforced personalized recommendation. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0923-y](https://doi.org/10.1007/s00779-016-0923-y)
- Liu W, Liu H, Kong H, Wan Y, Ning H (2016) The Yoking-proofs based authentication protocol for cloud assisted wearable devices. *Pers Ubiquit Comput*. doi:[10.1007/s00779-016-0926-8](https://doi.org/10.1007/s00779-016-0926-8)