Correlation Discovery in Web of Things

Lina Yao and Quan Z. Sheng
School of Computer Science, The University of Adelaide
Adelaide, SA 5005, Australia
{lina, qsheng}@cs.adelaide.edu.au

ABSTRACT

With recent advances in radio-frequency identification (RFID), wireless sensor networks, and Web services, Web of Things (WoT) is gaining a considerable momentum as an emerging paradigm where billions of physical objects will be interconnected and present over the World Wide Web. One inevitable challenge in the new era of WoT lies in how to efficiently and effectively manage things, which is critical for a number of important applications such as object search, recommendation, and composition. In this paper, we propose a novel approach to discover the correlations of things by constructing a relational network of things (RNT) where similar things are linked via virtual edges according to their latent correlations by mining three dimensional information in the things usage events in terms of user, temporality and spatiality. With RNT, many problems centered around things management such as objects classification, discovery and recommendation can be solved by exploiting graph-based algorithms. We conducted experiments using real-world data collected over a period of four months to verify and evaluate our model and the results demonstrate the feasibility of our approach.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Services; H.4.0 [**Information Systems**]: General

General Terms

Algorithms, Design, Experimentation

Keywords

Web of Things, correlation discovery, random walk with restart

1. INTRODUCTION

With the recent advances in radio-frequency identification (RFID), wireless sensors network, and Web services, "Web of Things" (WoT) is gaining a considerable momentum as an emerging paradigm where physical objects are interconnected and present over the World Wide Web [1]. With many things connected and interacted over the Web, there is an urgent need to efficiently index, organize, and manage these things for object discovery (e.g., finding a quiet restaurant), recommendation (e.g., suggesting a device that can consume a video stream), and mashup (e.g., composing device functional-

ities for a new service), and effectively reveal interesting patterns from the things.

Unfortunately, it is quite challenging to discover explicitly the relationships between heterogeneous things. Correlations among things are not obvious and explicit. Unlike social networks of people, where users have observable and direct links and connections, things often exist in isolated settings and the interconnections between them are usually limited. Targeting the challenges for deriving the implicit correlations of things, we approach this problem by exploring regularities of users' interactions with similar things. Such interactions contain enriched information including user information (who used the thing), temporal information (when the thing was used) and location information (where the thing was used). The problem targeted in this paper can be therefore formulated as discovering the *latent correlations* among things by exploiting observable *things usage events* with the goal of automatically distinguishing strong correlations of things from the weak ones. A usage event happens when a person interacts with a particular thing and can be captured explicitly, e.g., via RFID readings.

2. RELATIONAL NETWORK OF THINGS

To discover latent correlations among similar things, we develop a novel algorithm that constructs a *relational network of things* (RNT), which captures the relevance between things by considering three aspects (see Figure 1 (a)): *user*, *time*, and *location*. Three graphs, namely *thing-user* graph, *thing-time* graph, and *thing-location* graph, are developed accordingly (Figure 1 (b)). We will not give formal definitions of these graphs due to space constraint.

We need to derive the relevance among pair-wise things in order to build RNT, where edges in the network are explicitly established among things according to their relevance values. For each pair of things, t_i and t_j, we can derive the relevance values from the three graphs respectively, i.e., $r^u_{t_i,t_j}$ of the thing-user graph, $r^{ts}_{t_i,t_j}$ of the thing-time graph, and $r^l_{t_i,t_j}$ of the thing-location graph. The overall correlation value between a pair of things t_i and t_j, denoted by r_{t_i,t_j}, is calculated as: $r_{t_i,t_j} = \gamma r^u_{t_i,t_j} + \epsilon r^{ts}_{t_i,t_j} + \eta r^l_{t_i,t_j}$ where $\gamma + \epsilon + \eta = 1, \forall \gamma, \epsilon, \eta \in [0,1]$. γ, ϵ and η are parameters that can be used to control preferences over three different aspects (i.e., user, time, and location). Based on the formula, things accessed by the same user at the same time in the same location show strong correlation since all the three aspects are captured in the formula. To obtain the relevance values $r^u_{t_i,t_j}$, $r^{ts}_{t_i,t_j}$, and $r^l_{t_i,t_j}$ between a pair of things, t_i and t_j, we propose to use the Random Walk and Restart method [2] over the three graphs, respectively. A relational network of things (RNT) is constructed in a way that each thing is connected to things with top-k overall correlation values. Formally, an RNT is denoted as $\mathcal{G}_{rnt} = (\mathcal{T}, \mathcal{E})$. For each thing $t_i \in \mathcal{T}$, let $\mathcal{T}_i$ denote the set of related things to t_i. $\mathcal{E} = e(x,i)|\forall t_i \in \mathcal{T}, t_x \in \mathcal{T}_i,$
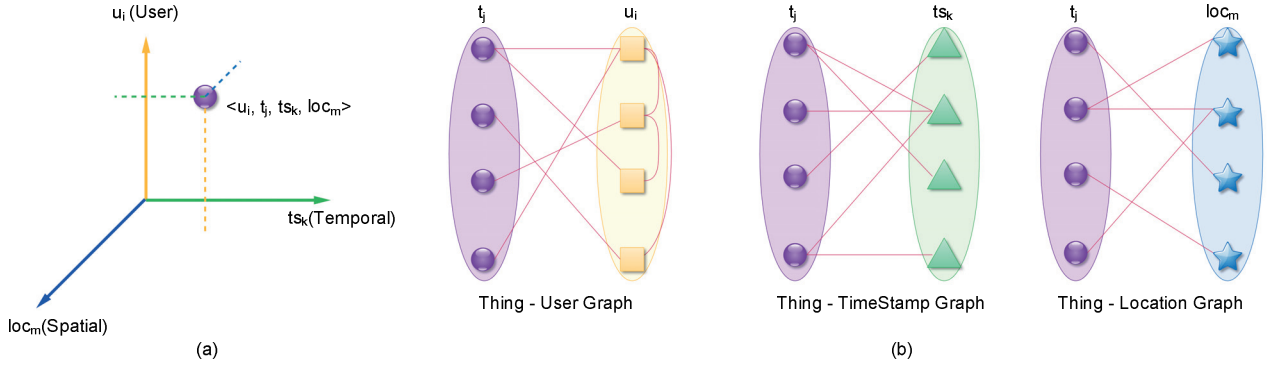
Figure 1: (a) Three dimensional information in things usage events; (b) Three graph representations: thing-user graph, thing-time graph, and thing-location graph

where $e(x, i)$ is an edge from t_x to t_i. Each edge is associated with a weight w_{t_x, t_i} with the correlation value r_{t_x, t_i}. Due to space constraint, we will not give our algorithm here.

Because the relational network of things (RNT) is essentially a graph, with the proposed RNT, many problems centered around things management (e.g., object discovery and recommendation) can be solved by tapping sizable literature of graph algorithms.

3. EXPERIMENTAL STUDY

Web of Things (WoT) is still relatively new and it is difficult to collect large number of things from the Web for the experimental study. We set up a testbed to collect data based on our previous work [3]. We tagged 127 different things of different categories using RFID tags and exposed them on the Web using RESTful Web services. We manually labeled these things with 397 different labels. It should be noted that some things belong to multiple categories, therefore having multiple labels. This dataset serves as the ground-truth dataset in our experiments for performance evaluation. Ten volunteers participated in data collection by interacting with RFID tagged things for a period of four months, generating 20,179 interactions records to be used in the experiments.

To test our approach in correlation discovery, we performed a classification task. Given a new object, the classifier automatically decides whether the object belongs to the category of the corresponding tag. There are two types of features: i) the latent features that represent the hidden correlations among things, and ii) the explicit features of things. The former can be extracted from our RNT's structure and the latter can be obtained from things description (i.e., textual features). These features are different but complementary. Both features are fed into a SVM classifier for the classification. We adopted the hamming-loss, one-error, coverage and average precision to evaluate the performance.

We randomly removed the category labels of a certain percentage, ranging from 10% to 50%, of things from each category of the ground-truth dataset. These things were used to test our approach while the rest were used as the training set. We particularly compared the classification performance by using i) the features obtained from RNT's structure (i.e., latent features), ii) the features obtained from things' descriptions (i.e., content features), and iii) the combination of the both. Each process was repeated 10 times and the average results were recorded. Similar observations were obtained for different testing percentages.

Figure 2 shows the result when we removed 30% of things from each category of the ground-truth dataset. Descriptions of things are normally short and noisy. It is therefore not surprising that the performance based on content features only is worse than the
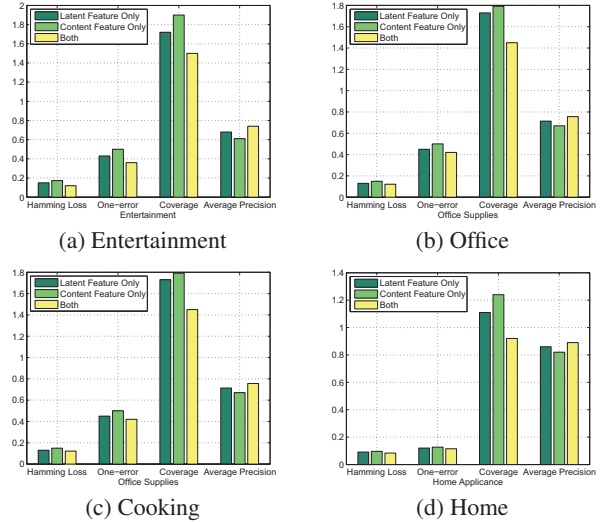

Figure 2: Performance comparison

one based on latent features in the most categories. The consistent good performance from the latent features also indicates that our proposed relational network of things is able to capture the correlations among things well. It should be noted that for the metrics of hamming-loss, one-error, and coverage, higher value means worse performance, while for the average precision metric, higher value means better performance.

4. FUTURE WORK

We view the work presented in this paper as a first step towards efficient management of things in the Web of Things era. In our current correlation model, we only consider the interactions between users and things. One important task of our future work is to extend the model by considering the interactions between things. Another challenge lies on handling the dynamic nature of things.

5. REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito. The Internet of Things: A Survey. *Computer Networks*, 54(15):2787–2805, 2010.

[2] H. Tong, C. Faloutsos, and J. Pan. Random walk with restart: fast solutions and applications. *Knowledge and Information Systems*, 14(3):327–346, 2008.

[3] Y. Wu et al. PeerTrack: A Platform for Tracking and Tracing Objects in Large-scale Traceability Networks. In *Proc. of EDBT 2012*, 2012.