

Contextual Synchronization for Online Co-browsing on Peer-to-Peer Environment

Jason J. Jung

INRIA Rhône-Alpes
ZIRST 655 avenue de l'Europe, Montbonnot
38334 Saint Ismier cedex, France
jjjung@intelligent.pe.kr

Abstract. In this paper, we propose a novel synchronization method based on contextual information elicited from a group of peers for online collaborative browsing on p2p environment. Thereby, the users are semantically tracked for modeling the context about their information searching tasks. The co-browsing system embedding our proposed method was shown to improve 52.7% and 11.5% communication performance, compared to single browsing and the asynchronous system, respectively.

1 Introduction

In communication systems, synchronization can be defined as the process of making sure that two or more entities contain the same up-to-date information for consistency. More particularly, in case of online cooperation between people, we consider that it means the process of comparing the user's current context and categorizing them into the groups of similar users. We focus on detecting the moment at which the context is switched to the others.

In this study, we propose semantics-based co-browsing system on peer-to-peer (p2p) environment with two main contributions; *i*) automatic organization of the groups of like-mined users by comparing the context represented as hierarchical topic paths, and more importantly, *ii*) recognition of temporal transition of context. Thus, when a user's context is changed over time, he has to be shifted to the more relevant groups. Thereby, the contextual transitions in information searching should be detected by analyzing the sequential patterns of user browsing actions.

2 Modeling Browsing Contexts and Detecting Contextual Transitions

The semantic factors are defined to measure the various relationships between users in a group, and between web pages. The web access patterns of a user group G_a aggregated during time interval T is given by a matrix $\mathcal{W}(G_a)$ of size $U \times T$, where U is the number of users in G_a . Therefore, we can extract two kinds of browsing context; *i*) personal context ($\mathcal{PC}(u_i)$) of user u_i from each row component in P , and *ii*) group context ($\mathcal{GC}(t_j)$) at a certain moment t_j from each column components.

After labeling two arbitrary web requests based on function WD , which is referring to the web directory, we can obtain both sets of hierarchical paths of the corresponding topics. Here, lightweight ontologies (e.g., web directory) are deployed to label (more exactly, conceptualize) the web pages accessed by the corresponding users¹. Let a web page wp^i categorized to the set $\{p_m^i | p_m^i \in WD(wp^i), m \in [1, \dots, M]\}$ where M is the number of all possible topic paths.

Semantic distance δ^\diamond is $\min_{m=1, n=1}^{M, N} \frac{\min((L_m^i - L_{m,n}^C), (L_m^j - L_{m,n}^C))}{\exp(L_{m,n}^C)}$ where L_m^i , L_m^j , and $L_{m,n}^C$ are the lengths of p_m^i , p_m^j , and common part of both, respectively. It is the minimum value among all combinatorial comparisons of two sets ($|p^i| \times |p^j|$). Semantic distance is assigned in the interval $[0, 1]$, and in case of complete matching, it is 0. Exponent function in denominator is for increasing the effect of $L_{m,n}^C$.

Semantic distance matrix Δ^\diamond The adjacent wp 's aggregated by either users' accesses at a certain time point (column components in \mathcal{W} for \mathcal{GC}) or a particular user's accesses during the time interval (row components in \mathcal{W} for \mathcal{PC}) of which size is either $U \times U$ for \mathcal{GC} or $T \times T$ for \mathcal{PC} , and the diagonal elements are all zero.

Semantic distance mean μ^\diamond is given by $\mu^\diamond = \frac{2}{T(T-1)} \sum_{i=1}^{T-1} \sum_{j=i+1}^T \Delta^\diamond(i, j)$, and means the mean value of upper triangular elements in Δ^\diamond except diagonal components. It can measure the semantic consistency of the given set of web requests.

Semantic distance deviation σ^\diamond is $\sqrt{\frac{2}{T(T-1)} \sum_{i=1}^{T-1} \sum_{j=i+1}^T (\Delta^\diamond(i, j) - \mu^\diamond)^2}$. measuring the degree of dispersion of the semantic distance values from a given set of web requests.

To identify the contextual transition from the streaming web accesses of a group of users, we have to consider to compute not only the semantic factors in a given interval but also the distributions of μ^\diamond and σ^\diamond like sliding windows method. Hence, the triggering patterns from these signals are regarded as important evidence. This process should be conducted by following the objective function $\min \left[\sum_{g=1}^G \mathcal{GC}_g(t_{j+1}) - \mathcal{GC}_g(t_j) \right]$ where G is the total number of groups. It means that the temporal differences of \mathcal{GC} of each group should be minimized. This process is organized as two steps;

1. **Alarming step.** Semantic distance deviation σ^\diamond of column components in \mathcal{P} is applied to capture the contextual transitions of a particular user u_i whose personal context \mathcal{PC} is different from the corresponding group's context \mathcal{GC} . By using the objective function, alarming at time t_j can be characterized to

$$\Delta(\sigma^\diamond(\mathcal{GC})) = |\sigma^\diamond(\mathcal{GC}(t_j)) - \sigma^\diamond(\mathcal{GC}(t_{j-1}))| \geq \lambda_A \quad (1)$$

where λ_A is a threshold for alarming of \mathcal{GC} transitions. Then, a user is detected by

$$u_i = \arg_i \min \Delta'_i(\sigma^\diamond(\mathcal{GC})) \quad (2)$$

$$= \arg_i \min |\sigma^\diamond(\mathcal{GC}'_i(t_j)) - \sigma^\diamond(\mathcal{GC}'_i(t_{j-1}))| \quad (3)$$

where $\mathcal{GC}'_i = \mathcal{GC} - \{i\text{-th component}\}$. Until $\Delta'_i(\sigma^\diamond(\mathcal{GC})) \leq \lambda_A$, alarming process has to be repeated.

¹ This web directory-based labeling process is explained in [1] in detail.

2. **Confirming step.** Semantic distance mean μ^\diamond of the i -th row component in \mathcal{P} is applied to make sure that his context \mathcal{PC} is changed or not. If so, the exact time point can be detected. Similarly to the previous step, confirming the conceptual transition of user u_i can be characterized as

$$\Delta(\mu^\diamond(\mathcal{PC})) = |\mu^\diamond(\mathcal{PC}[t_1, t_{s-1}]) - \mu^\diamond(\mathcal{PC}[t_s, t_T])| \geq \lambda_C \quad (4)$$

where λ_C is the threshold for confirming of \mathcal{PC} transitions. Then, the time point s surely is the moment when \mathcal{PC} of the corresponding user is changed.

Each time the streaming web accesses within a group is stored in \mathcal{W} , the alarming step has to be fulfilled. We employ the semantic distance deviation σ^\diamond to recognize the dispersion of members in a group, rather than the group context itself. Afterward, if some users would be detected in this step, the confirming step can justify if their transitions are validated or not, because the semantic distance mean μ^\diamond is useful to measure the semantic cohesion within a certain time interval.

3 Online Co-browsing with Contextual Transition

The users detected their context transitions should be reorganized to the most relevant group. Thus, they can get all information about the group members' browsing patterns through blackboard module. User interface is simply composed of three frames for web browser, blackboard, and lists of friends/groups. On p2p environment, all users utilize homogeneous system, except to the facilities for super-peers.

Even multiple users can be the super-peer in a group, but we assume that the only one be the super-peer to monitor and control the rest of members. Thus the super-peer of g -th group \mathcal{G}_g is selected by $SuP_g = \arg_i \min \sum_{k=1, u_k \in \mathcal{G}_g}^{|\mathcal{G}_g|} |\mathcal{PC}(u_k) - \mathcal{PC}(u_i)|$ which means that the super-peer user must be in the most middle of the group as minimizing the semantic distances between other members. This selection process should be conducted when the members are changed. We can say it is similar to the adaptation process of k -nearest neighborhood (k -NN) method.

Now, we want to explain reorganization process of the users whose contextual transition has been detected. By using the objective functions and super peer selection function, the reorganization process is given by $\arg_g \min_{g=1}^G |\mathcal{PC}(SuP_g) - \mathcal{PC}(u_i)|$. where G is the total number of groups in p2p network. It searches for the super-peer whose context is closest to user u_i , because the super-peer is surely regarded as the representative context of the corresponding group context. Hence, u_i can join the most relevant group \mathcal{G}_g .

4 Experimental Results

We conducted simulations to evaluate the performance of communications on the proposed method for co-browsing system. Three groups \mathcal{G}_A , \mathcal{G}_B , and \mathcal{G}_C are organized by 30 users (each ten users in a group), and then we collected the web logs dataset by letting these users to browse the testing bed space² in the fixed personal context.

² As extended from the dataset applied in [1], it is composed of 4610 web pages labeled from ODP (open directory project, <http://www.dmoz.org>).

Firstly, in order to evaluate the detection of contextual transitions of each user, we generated 30 testing sequences, including totally 583 contextual transitions, by randomly intermixing the fragments which are randomly segmented from the web log dataset. We examined how exactly the transitions could be detected with respect to measurement recall and precision. According to F1-value, set $\frac{2R \times P}{R+P}$, we empirically uncovered the best threshold values $\lambda_A = 0.4$ and $\lambda_C = 0.6$. The threshold level of confirming step seems slightly more critical, because it is for the personal context. Overall, the average of confirming step is about 33% higher than that of alarming step.

Secondly, we evaluated the performance of communications by group organization. This proves the efficiency of online co-browsing, rather than single browsing or basic co-browsing systems. While users in \mathcal{G}_A browsed without any collaboration, \mathcal{G}_B and \mathcal{G}_C was under co-browsing. But, users in \mathcal{G}_C was only considered the contextual transitions. As the final results of three group members' browsing for three weeks, \mathcal{G}_C in online co-browsing has shown only 53% web access with helping each other according to the context. Compared with \mathcal{G}_B , our proposed method has slightly improved by 11.5%.

5 Concluding Remarks and Future Work

In this paper, we have proposed online co-browsing system of which characteristics are spatially remote and temporal synchronous. It is capable of detecting the contextual transitions of users in a group, so that they are efficiently shifted into the relevant group communications. As main contribution of this paper, most importantly, we propose tracking the contextual dynamic of the groups while co-browsing, rather than modeling the consensual context of the groups.

However, this system has still many problems that have to be dealt with in future work. We modified *Levenshtein* edit distance [2] to measure the hierarchical path-labeled web pages. In order to support more general users, we obviously consider various semantic annotation methods [3] to compare the relationships between them. Another issue is topology in p2p network. Because, as mentioned in [4], the hyperlinked environment has various topological features such as authorities and hubs, we have to think over about selection process of super-peers.

References

1. Jung, J.J.: Semantic preprocessing of web request streams for web usage mining. *Journal of Universal Computer Science* **11**(8) (2005) 1383–1396
2. Levenshtein, I.: Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory* **10**(8) (1996) 707–710
3. Jung, J.J., Lee, K.S., Park, S.B., Jo, G.S.: Efficient web browsing with semantic annotation: A case study of product images in e-commerce sites. *IEICE Transactions on Information and Systems* **E88-D**(5) (2005) 843–850
4. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. *Journal of the ACM* **46**(5) (1999) 604–632