DBSJ Journal, Vol.11, No.2
October 2012

Looking into Socio-cognitive Relations between Urban Areas based on Crowd Movements Monitoring with Twitter

Shoko WAKAMIYA♥ Ryong LEE♦
Kazutoshi SUMIYA♠

Due to the proliferation of location-based information services, there is abundant urban information which makes us difficult to catch up with the characteristics and dynamics of our living space. However, nowadays, crowd lifelogs shared over social network sites are attracting a great deal of attention as a novel source to search for local information from the massive voices and lifelogs of crowds. In this regard, we can further look into urban images representing how we recognize a city in mind through the direct massive crowd experiences. In this work, we explore crowd-experienced local information over location-based social network sites to derive much better understandable and useful urban images. In detail, we propose a method to generate a socio-cognitive map where characteristic urban clusters are projected based on cognitive distance between urban areas. Specifically, in order to measure cognitive distances between urban clusters and examine their influential strengths, we observe crowd's movements over Twitter. Finally, we show an experimental result of generating a socio-cognitive map illustrating crowd-sourced cognitive relations between urban clusters in Kinki area, Japan.

1. Introduction

Recently, due to the widespread of location-based information services, the explosive growth of local information like maps, geographical statistics and local web pages virtually enables us to search for whatever we want for our daily decision makings. However, it is increasingly becoming difficult more and more to explore useful information from the overflowing local information. In order to support user's urban lives with local information, we need to extract the essence of the complex and dynamic nature of urban area. In many respects, we need much intuitive and succinct maps which are well-arranged to describe the overall characteristics and dynamics. Therefore, it is critical to generate a kind of cognitive map enabling us to easily take the image of our living space.

In this work, we challenge to explore crowd-experienced local information over social network sites such as Twitter[1] and Foursquare[2] to extract urban images based on crowd experiences in the real space. Indeed, so far, in order to grasp how people live in an urban space, person trip (PT) survey[3] has often been conducted. Especially, in Japan, this kind of survey has been made every 10 years or more under the jurisdiction of the Ministry of Land, Infrastructure, Transport and Tourism. Although we can obtain a snapshot of people's daily trip in an urban area by this kind of survey, it would not be enough to examine dynamic social nature of urban space because practically impossible to consecutively capture our daily lives.

In this paper, for the purpose of exploring crowd's common cognition to an urban area, we generate a socio-cognitive map using crowd's movements observed over Twitter. Especially, as for crowd's common cognition, we focus on cognitive distances between urban areas. On the basis of crowd's movements with spatio-temporally indexed tweets, we measure cognitive distances between urban clusters and reallocating the urban clusters based on the relative closeness by means of Multi-dimensional Scaling (MDS). Furthermore, we compute their influential strengths for emphasizing the urban clusters eventually by drawing a cognitive map with Weighted Voronoi Diagram. In particular, we summarize our contributions of this work as follows:

- exploiting crowd-experienced local information from massive crowd lifelogs over location-based social network sites,
- presenting cognitive distances between urban clusters for deriving common urban cognition, and
- representing crowd-experienced urban image as a socio-cognitive map.

The remainder of this paper is organized as follows. Section 2 addresses our research model for generating a socio-cognitive map with social network sites and describes related work. Section 3 details a method for socio-cognitive map generation exploiting crowd's lifelogs. Section 4 shows an experimental result with geo-tagged tweets over Twitter. Section 5 describes the conclusion and future work.

2. Exploiting Crowd-experienced Local Information

In this section, we introduce our approach to explore crowd's urban cognition by exploiting crowd-experienced local information publicly shared over social network sites. Then, we briefly describe some related work in terms of crowd-experienced local information.

2.1 Exploring Crowd's Cognition for Urban Area

In this work, we attempt to extract crowd's common

♥ Student Member. Graduate School of Human Science and Environment, University of Hyogo
ne11n002@stshse.u-hyogo.ac.jp
♦ Regular Member. Graduate School of Human Science and Environment, University of Hyogo, NICT
lee.ryong@gmail.com
♠ Regular Member. Graduate School of Human Science and Environment, University of Hyogo
sumiya@shse.u-hyogo.ac.jp

[1] Twitter: http://twitter.com/
[2] Foursquare: https://foursquare.com/
[3] Person Trip Survey (PT) Japan, C.C.: http://www.mlit.go.jp/crd/tosiko/pt/map_e.html

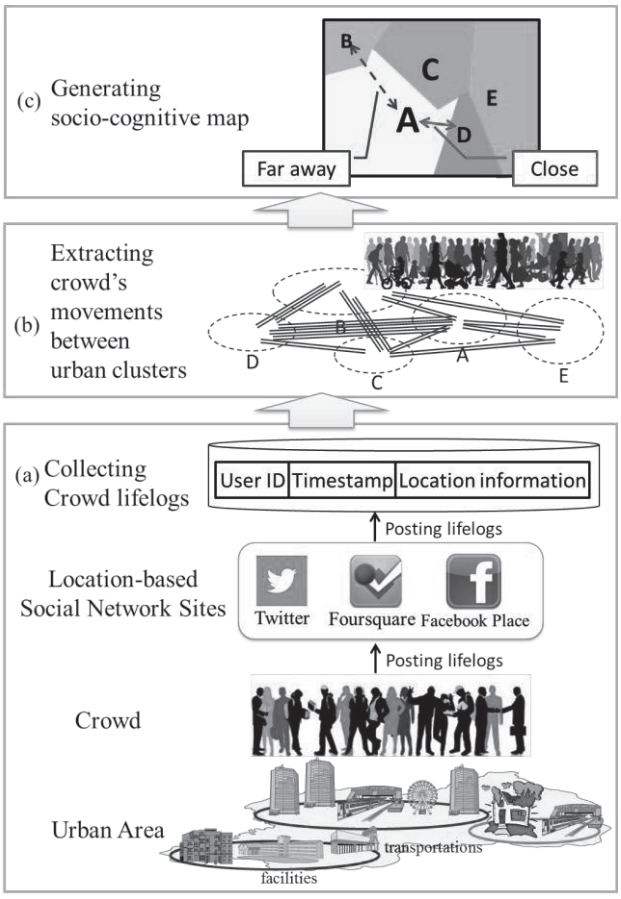


Fig. 1. Research model: Distilling crowd's cognition to urban areas through location-based social network sites

images of an urban area with crowd-sourced lifelogs over social network sites and eventually represent them as a socio-cognitive map. For this purpose, we first propose a model for generating a socio-cognitive map with location-based social networks sites as shown in Fig. 1. As illustrated in Fig. 1 (a), nowadays, thanks to the proliferation of social network sites and the widespread smartphone technology, crowd-soured lifelogs reflecting their activities, minds, etc. in urban areas are available in an unprecedented large scale. Next, we extract crowd's movements as shown in Fig. 2 (b) from the collected crowd's urban lifelogs as depicted in Fig. 2 (a) and explore distance-based relations between urban areas based on the crowd's movements as illustrated in Fig. 1 (b). Finally, in order to represent images of urban areas hidden in crowd's minds, we generate a socio-cognitive map where landmark clusters are re-mapped based on the observed relations as shown in Fig. 1 (c).

2.2 Related Work

One of noteworthy characteristics of crowd lifelogs as local information is the capability to monitor most up-to-date social situation by massive crowd's experiences in the real world. Some research work by Sakaki et al [1] and Lee et al. [2] focused on detecting geo-social events or phenomena from the crowd's messages on Twitter.

Furthermore, the other critical characteristic is that

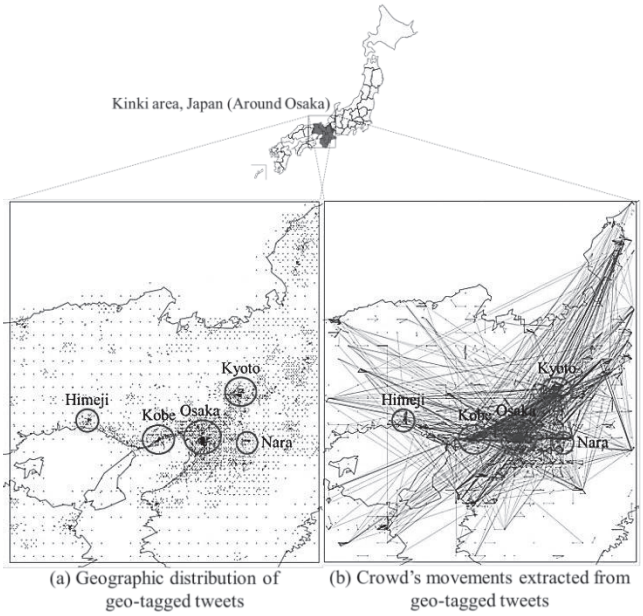


Fig. 2. Crowd's movements in an urban area observable via Twitter

crowd lifelogs are regarded as the accumulation of crowd's experiences and knowledge in urban space. Thus, exploring crowd-sourced experiences effectively is one of key topics of recent research studies of location-based services and social network sites. McArdle et al. [3] investigated the usefulness of digital footprints of individual movement for calibrating human mobility models within an urban traffic micro-simulation framework. Hsieh et al. [4] presented a method to recommend a time-sensitive trip route by squeezing a lot of knowledge from check-in data over location-based services. In our previous work [5], we proposed a method to characterize urban area by exploring latent patterns of crowd behavior using Twitter.

This paper is significantly different from these studies because we aim to measure urban relations focusing on crowd's movements between urban clusters and generate a socio-cognitive map based on the relations.

3. Integrating Crowd-experienced Local Information with a Map

In this section, we describe a detailed procedure to generate a socio-cognitive map with crowd's movements through Twitter.

3.1 Extracting Crowd's Movements from Social Network Sites

In order to extract crowd-experienced local information from crowd lifelogs over social network sites, we utilize Twitter with which we collect geo-tagged tweets as shown in Fig. 3 (1). However, it takes a considerable amount of time and efforts to acquire a significant number of geo-tagged tweets due to practical limitation of an open API[4] provided by Twitter. Therefore, in order to efficiently

[4] Twitter Open API: http://apiwiki.twitter.com/Twitter-Search-API -Method%3A-search

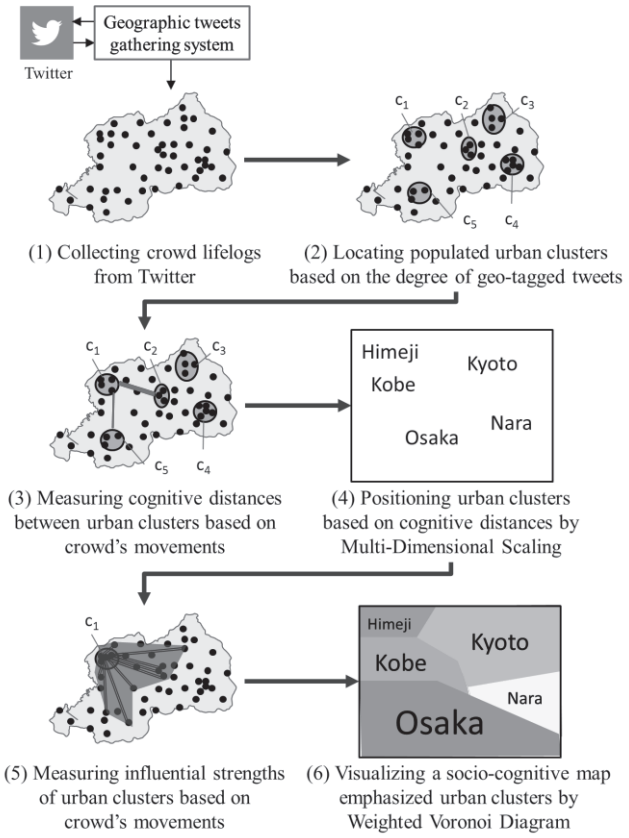


Fig. 3. Process of socio-cognitive map generation

monitor crowd behavior for a specific region of any size, we utilized a geographic tweets gathering system which was developed in our previous work [2].

Next, we are interested in crowd's movements which can be extracted with primary metadata of geo-tagged tweet; user ID, timestamp, and location information already identified as coordinates by means of geo-coding which can translate place names into the corresponding exact locations with Google Map's geo-coding service [5]. Furthermore, in order to effectively search moving segments, we arrange the tweets based on user ID and sort each user's tweets in the order of the timestamp. Here, if the successive geo-tagged tweets were written at different locational points, we consider this means a user's moving segment depicted by a blue line in Fig. 2 (b). Consequently, we could observe crowd's movements using their lifelogs on Twitter.

3.2 Looking for Crowded Urban Clusters

In general, a map consists of some important landmarks. Similarly, we should find representative geographical objects or areas. In this paper, we define a populated space as an urban cluster and locate them based on the geographical density of geo-tagged tweets as shown in Fig. 3 (2). However, because there are too many tweet occurrence locations at which one or more persons were published geo-tagged tweets, it would require unbearable

[5] Google Geocoding API: http://code.google.com/intl/ja/apis/maps/documentation/geocoding/

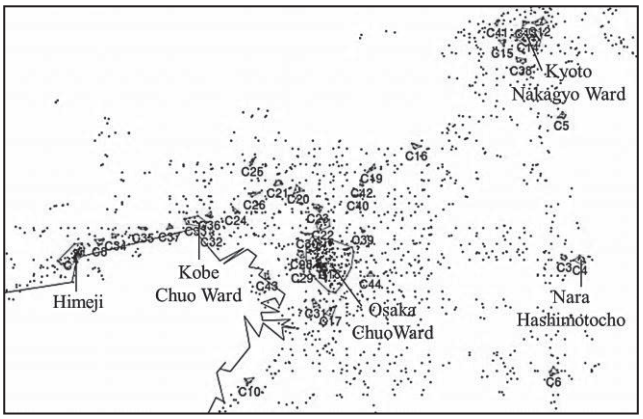


Fig. 4. Urban clusters based on geographic distribution of geo-tagged tweets

computational efforts to cluster and finally find out urban clusters. Hence, we need to reduce the data size in much smaller and computable size with less loss of essential quality of the original distribution. For this, we adopted NNClean (Nearest Neighbor Clutter Removal) algorithm [6] which can filter high-frequency locational points. Then, in order to locate urban clusters with the ideally reduced dataset, we applied DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm [7]. Here, we defined the original variant as a minimum number of points *MinPts* within a specified radius *radius*. In Fig. 4, we show urban clusters located by empirically setting *MinPts* and *radius* to 6 tweets and 650m, respectively. Eventually, we found 44 urban clusters (c_1~c_{44}) which are represented by convex-hull based boundary polygons [8]. In this method, we are easily and successfully able to achieve simple findings of significant urban clusters only according to geo-tagged tweets.

3.3 Examining Cognitive Distances between Urban Clusters

Then, we need to define and measure cognitive distances between urban clusters as shown in Fig. 3 (3). We assume that if the distance between urban clusters is somewhat long in the real space and lots of crowds are moving in and out between these clusters, the crowds may cognitively construct a map in mind, where these clusters become closer than the actual physical distance. Therefore, we consider that urban clusters which are tightly associated with each other in terms of crowd experiences should be projected close on a socio-cognitive map. On the basis of this assumption, we made a formula to calculate cognitive distances between two urban clusters as follows:

$$CogDist(c_i, c_j) = w_1 \cdot EucDist(c_i, c_j) + w_2 \cdot ExpDist(c_i, c_j) \quad (1)$$
$$(w_1 + w_2 = 1.0, \quad w_1, w_2 \geq 0)$$

$$ExpDist(c_i, c_j) = \frac{1}{\#MovSeg(c_i, c_j)+1} \quad (2)$$

where the function $CogDist(c_i, c_j)$ calculates cognitive distances between two distinct urban clusters c_i and c_j. Specifically, cognitive distance is calculated by physical distance and experiential distance between the urban clusters. The Euclidean distance calculates physical distance between the urban clusters by the function

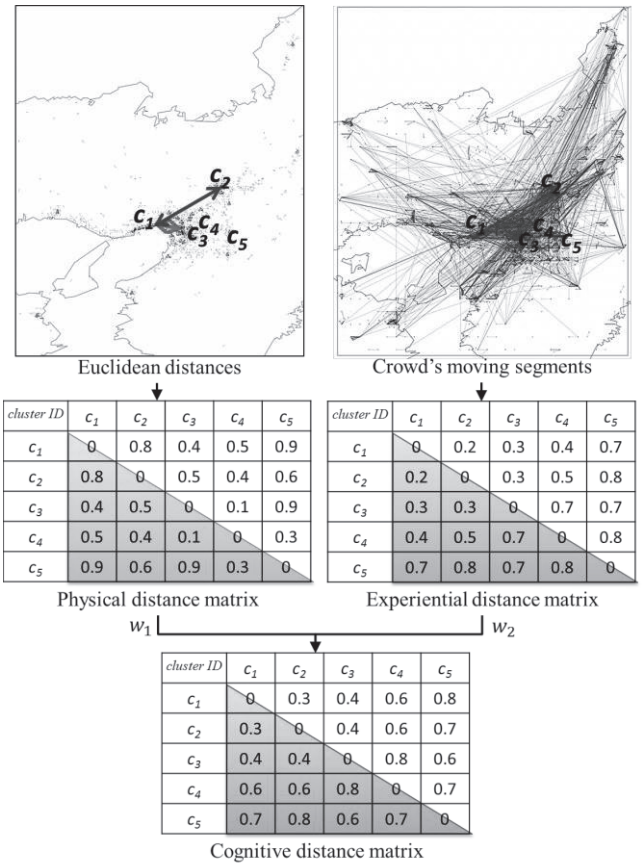


Fig. 5. Generating a cognitive distance matrix

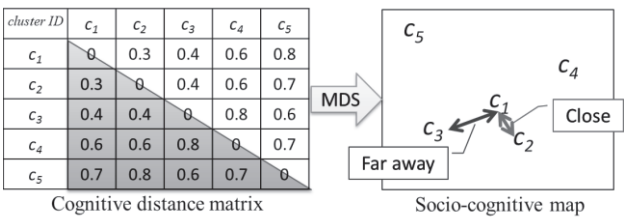


Fig. 6. Projecting urban clusters based on cognitive distances by Multi-Dimensional Scaling (MDS)

$EucDist(c_i, c_j)$. The experiential distance is calculated by the function $ExpDist(c_i, c_j)$ which is based on the total number of crowd moving segments between the urban clusters and calculate by the function $\#MovSeg(c_i, c_j)$. $EucDist$ and $ExpDist$ are normalized and weighted by two values w_1 and w_2, respectively. We calculate cognitive distances between all pairs of urban clusters located in the urban area. As a result, we construct matrixes in terms of physical distance, experiential distance and cognitive distance, respectively as illustrated in Fig. 5. These weighting values can be adjusted by user's purposes for generating a cognitive map. For example, if a user wants to generate a cognitive map by focusing on the crowd's movements, he/she can set a high weight to w_2. In the experiment in Section 4, we will show a cognitive map generated in extreme cases weighting the maximum value on physical distance and experiential distance respectively.

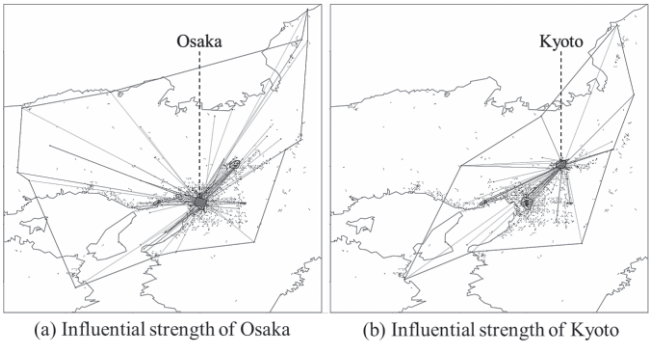


Fig. 7. Geographic influential strengths of urban clusters

3.4 Projecting Urban Clusters based on Cognitive Distances

Next, we need to project urban clusters based on the calculated cognitive distances as shown in Fig. 3 (4). In order to represent intuitive distance between urban clusters, we need to appropriately allocate the clusters on a cognitive map.

In this paper, we decided to apply Multi-Dimensional Scaling (MDS) [9] which can allocate given dataset of multi-dimensional space by considering similarities (or dissimilarities) in the dataset. Therefore, it can allocate data in the neighborhood if the similarity between the data is high, vice versa. Specifically, this algorithm starts with a matrix of similarities between multiple objects, then assigns a location to each item in N-dimensional space, where N is specified a priori as shown in Fig. 6. In the experiment described in Section 4, we mapped urban clusters in two-dimensional space by applying MDS.

We also labeled names of generated clusters by a Reverse Geocoding service[6] which translates a given location coordinate into a textual place name. For each cluster, we obtain its representative place name by using this service.

3.5 Emphasizing Urban Clusters based on Influential Strengths

We could generate a tentative socio-cognitive map by the above steps. However, its representation would be far from a map apparently, because it just represents point-based distribution of urban clusters. Therefore, in order to show a socio-cognitive map in an intuitive way, we decided to emphasize featured urban clusters on the socio-cognitive map by focusing on crowd's movements. Thus, we measure an influential strength in an urban area by examining crowd's movements involving the urban cluster as illustrated in Fig. 3 (5). We define an influential strength involving an urban cluster c_i by the formula as follows.

$$Influence(c_i) = \#MovSeg(c_i)$$

Fig. 7 shows two urban clusters in Osaka and in Kyoto in our finding. Interestingly, effective ranges of a cluster located in Osaka in Fig. 7 (a) covers the most major urban clusters in our experimental area of Kinki area, Japan.

Then, we emphasize urban clusters based on the

[6] Google Reverse Geocoding API: https://developers.google.com/maps/documentation/geocoding/#ReverseGeocoding

influential strengths of urban clusters as shown in Fig. 3 (6). For intuitive representation as a map, if an urban cluster has strong influential strength in an urban area, the cluster should be represented largely. For this, we applied a space partitioning method, Weighted Voronoi Diagram [10] to the above result. Here, Voronoi Diagram is a well-known geometric representation for partitioning a space drawing a line between two points keeping with the same distance. However, this simple method cannot show influential strengths of urban clusters. Thus, we decided to utilize Weighted Voronoi Diagram which can have a weight extending its size. In our method, we applied an influential strength as a weighting value on each partitioned area by Voronoi Diagram.

4. Experiment

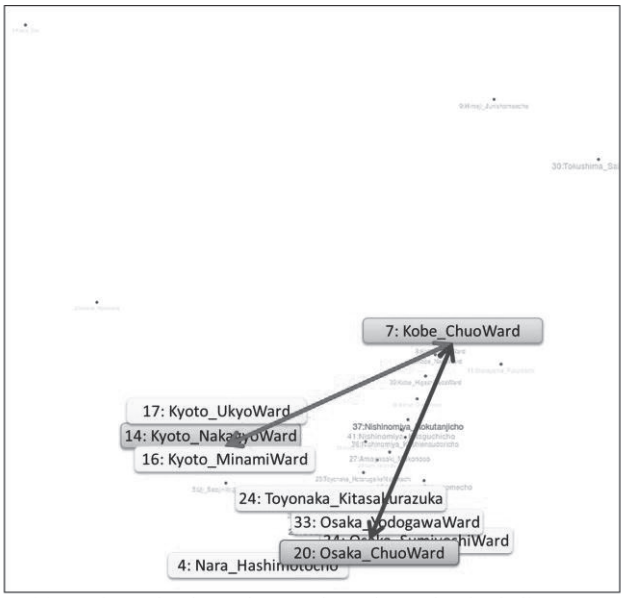
We describe the experiment for extracting crowd's cognition for an urban area and visualizing it as a socio-cognitive map as described in Section 3. We present our dataset collected from Twitter in Section 4.1 and show socio-cognitive maps generated with the dataset in Section 4.2.

4.1 Dataset

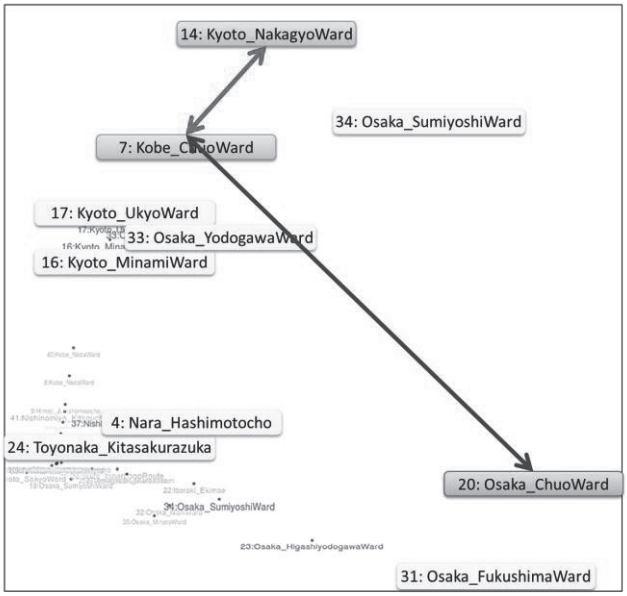
In this experiment, we were able to collect 157,097 geo-tagged tweets by 25,674 distinct users in a weekday (April 23rd, 2012) in a Kinki area including Kobe, Osaka, Kyoto and Nara in Japan (longitude range = [134.122433, 136.337186], latitude range = [33.810804, 36.785050]) from Twitter using the geographical tweet gathering system developed in our previous work [2]. By using these geo-tagged tweets distributed as shown in Fig. 2 (a), we locate urban clusters as shown in Fig. 4 and extract crowd's moving segments as illustrated in Fig. 2 (b).

4.2 Looking into Crowd-experienced Urban Relations

In order to generate a socio-cognitive map, we examined crowd-experienced proximity of urban clusters by calculating cognitive distances between urban clusters. Especially, in order to observe how geographical proximities between urban clusters are sophisticated by crowd's cognition, we defined cognitive distances by adjusting weighting values on physical distance and experiential distance. As a result, Fig. 8 show two different socio-cognitive maps which projected urban clusters based on cognitive distances by means of MDS. In these maps, texts represent labels of urban clusters. Specifically, Fig. 8 (a) present a socio-cognitive map which was generated by projecting urban clusters based on cognitive distance with the maximum weighting value (= 1.0) on physical distance. Specifically, on the figure, the distance between KobeChuoWard (c_7) and OsakaChuo Ward (c_{20}) is nearly similar with the distance between KobeChuoWard (c_7) and KyotoNakagyoWard (c_{14}). Not surprisingly, these relative distances are the same with geographical relationship. In contrast, in case of Fig. 8 (b), the socio-cognitive map was generated based on cognitive distance with the maximum weighting value (= 1.0) on experiential distance. As a result, OsakaChuoWard (c_{20}) is projected far from KobeChuoWard (c_7) than Kyoto NakagyoWard (c_{14}). Thus, we could successfully extract



(a) Projected urban clusters based on physical distance



(b) Projected urban clusters based on experiential distance

Fig. 8. Socio-cognitive maps based on cognitive distance

crowd's cognition for the urban area which is often different from conventional local information.

However, the visual representation of MDS-based socio-cognitive map is hard to be regarded as a map. Therefore, we decided to emphasize urban clusters by partitioning the generated map for emphasizing each cluster's influence more effectively and intuitively. Specifically, Fig. 9 shows a socio-cognitive map visualized top-10 urban clusters by allocating regions to them according to their influential strengths. Interestingly, from this map, we can know which urban clusters are significant in crowd's cognition. For example, Kobe ChuoWard (c_7) is closer to KyotoNakagyoWard (c_{14}) than OsakaChuoWard (c_{20}) which is different from the real space. In addition, we can grasp which urban cluster plays

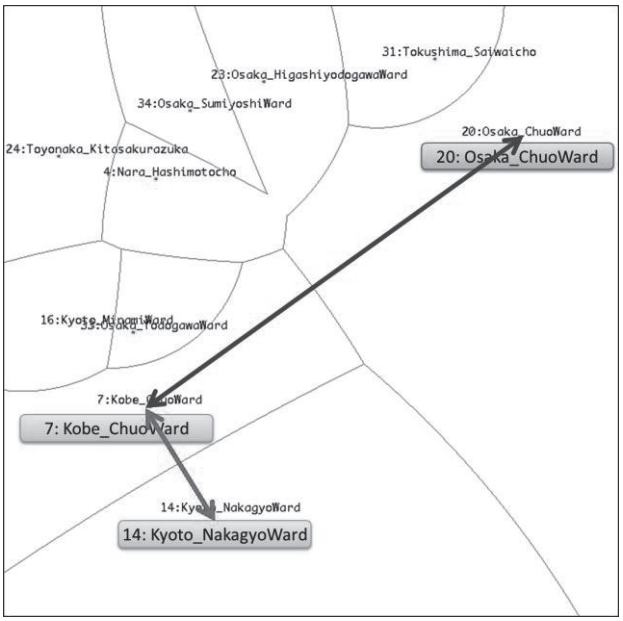


Fig.9. Weighted Voronoi-based socio-cognitive map

the role of hub in this urban area. For example, regions of clusters in OsakaChuoWard (c_{20}) covered a large space on this map. In fact, some other clusters located in Osaka such as OsakaHigashiyodogawaWard (c_{23}), Osaka SumiyoshiWard (c_{34}), OsakaYodogawaWard (c_{33}), Toyo-nakaKitasakuraduka (c_{24}) are projected close to urban clusters located in other prefectures like Nara Hashimotocho (c_4) or KyotoMinamiWard (c_{16}). Voronoi-based visualization of socio- cognitive map would be helpful to grasp significant clusters and their relationships intuitively.

5. Conclusion and Future Work

In this paper, we propose a method to draw a socio-cognitive map for exploring crowd's cognition for an urban area by exploiting crowd-experienced local information over Twitter. In the experiment, we successfully showed a definite distinction between geographical relationships in the real space and crowd's cognition for the urban area.

In future work, we will conduct on comprehensive socio-cognitive map generation by further observing various crowd's experiences utilizing crowd lifelogs over social network sites. In addition, we will study various useful metrics of distance for supporting user's decision makings such as distance based on crowd's traveling time.

[Acknowledgement]

This research was supported in part by the 7th Microsoft Research IJARC Core Project and Grant-in-Aid for JSPS Fellows 24.9154 from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

[References]

[1] Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, In Proc. 18th International World Wide Web Conference (2010)

[2] Lee, R., Wakamiya, S., Sumiya, K.: Discovery of unusual regional social activities using geo-tagged microblogs. World Wide Web Special Issue on Mobile Services on the Web 14(4), 321-349 (2011)

[3] McArdle, G., Lawlor, A., Furey, E., Pozdnoukhov, A., City-scale traffic simulation from digital footprints, In Proc. of the ACM SIGKDD International Workshop on Urban Computing, pp. 47-54 (2012)

[4] Hsieh, H.-P., Li, C.-T., Lin, S.-D., Exploiting large-scale check-in data to recommend time-sensitive routes, In Proc. of the ACM SIGKDD International Workshop on Urban Computing, pp. 55-62 (2012)

[5] Lee, R., Wakamiya, S., Sumiya, K.: Urban area characterization based on crowd behavioral lifelogs over twitter. Personal and Ubiquitous Computing pp. 1-16 (2012)

[6] Byers, S., Raftery, A.: Nearest-neighbour clutter removal for estimating features in spatial point processes. Journal of the American Statistical Association 93, 577-584 (1998)

[7] Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proc. of the Second International Conference on Knowledge Discovery and Data Mining, pp. 226-231 (1996)

[8] Andrew, A.M.: Another efficient algorithm for convex hulls in two dimensions. Information Processing Letters 9(5), 216-219 (1979)

[9] Kruskal, J., Wish, M.: Multidimensional scaling. In: Sage University Papers on Quantitative Applications in the Social Sciences, pp. 07-011 (1978)

[10] Aurenhammer, F.: Voronoi diagrams-a survey of a fundamental geometric data structure. ACM Comput. Surv. 23(3), 345-405 (1991)

Shoko WAKAMIYA

She is a student of Graduate School of Human Science and Environment, University of Hyogo. Her research interests include socio-geographic mining and user-generated contents. She is a student member of Information Processing Society of Japan and the Database Society of Japan.

Ryong LEE

He is a researcher in NICT and a visiting researcher of School of Human Science and Environment, University of Hyogo. He received a Ph.D. and a M.S. from Graduate School of Social Informatics, Kyoto University in 2003. His research interests include geographic information systems, social network analysis and web information retrieval systems.

Kazutoshi SUMIYA

He is a Professor at School of Human Science and Environment, University of Hyogo. He received the Ph.D. degree in engineering in 1998 from the Kobe University Graduate School of Science and Technology. His research interests include web information systems, spatio-temporal database, broadcast computing, mobile multimedia, and multimedia databases.