FlierMeet: Cross-Space Public Information Reposting with Mobile Crowd Sensing

Bin Guo
Northwestern Polytechnical Univ.
No. 128, Youyi-West Road
Xi'an, China
guob@nwpu.edu.cn

Xing Xie
Microsoft Research Asia
No. 5, Danling Raod
Beijing, China
xing.xie@microsoft.com

Huihui Chen
Northwestern Polytechnical Univ.
No. 128, Youyi-West Road
Xi'an, China
ddchh@163.com

Shenlong Huangfu
Northwestern Polytechnical Univ.
No. 128, Youyi-West Road
Xi'an, China
huangfusl@163.com

Zhiwen Yu
Northwestern Polytechnical Univ.
No. 128, Youyi-West Road
Xi'an, China
zhiwenyu@nwpu.edu.cn

Zhu Wang
Northwestern Polytechnical Univ.
No. 128, Youyi-West Road
Xi'an, China
wangzhu@nwpu.edu.cn

Abstract
Bulletin boards serve an important function for public information sharing. Posted fliers advertise services, events and other announcements. However, fliers posted offline suffer from problems such as limited spatial-temporal coverage and inefficient search aid. In recent years, with the development of sensor-enhanced mobile devices, mobile crowd sensing has been used in a variety of application areas. In this paper we present FlierMeet, a crowd-powered sensing system for cross-space public information reposting, tagging and sharing. The tags are auto-labeled based on a set of visual and crowd-object interaction features. Initial deployments and experiments prove the effectiveness of our system.

Author Keywords
Mobile crowd sensing, cross-space, urban sensing, crowd-object interaction, tagging, recommendation.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
Bulletin boards serve an important communication function within communities. Posted fliers advertise public information (sales, recruitments, notice, etc.) and invite community members to interact. Though the

bulletin board has proved useful and significant in our daily lives, it suffers from issues such as limited spatial-temporal coverage, lacking order, low search speed, etc. Therefore, there would be benefits to transfer fliers from the physical space to the cyber space, facilitating remote manipulation and information sharing. We define such a transfer as the "repost" activity.

With the recent surge of sensor-enhanced mobile devices, mobile crowd sensing (MCS) [1] has become an emerging paradigm for large-scale sensing. It has demonstrated its usefulness in a variety of application areas [2, 3]. However, no existing method focuses on distributed public flier information collection and cross-space reposting.

In this paper, we propose FlierMeet, a system that attempts to digitize the 'silent' paper fliers and widen their propagation. We leverage crowdsourcing to repost fliers from the physical space to the cyber space, and identify both the category and semantic tags based on crowdsourced data. The system can be applied to various application areas, such as public information collection, target advertising, mobile socializing, etc. Specifically, our work makes the following contributions:

- Develops a mobile platform for participatory public information reposting, automatic grouping, intelligent tagging and sharing.
- Introduces a novel set of crowdsensing-specific features to characterize the reposted fliers, and provides a hybrid inference model to predict varied category and semantic tags using these features.

We evaluated FlierMeet with an eight-week, 38 person deployment using commercially-available smartphones.

FlierMeet: The System Overview

The system architecture of FlierMeet is shown in Fig. 1. It has the following major components.

The *cross-space reposting* component builds the connection between mobile clients and the backend server. Using the application running at the mobile clients, users can capture interesting fliers from bulletin boards and transmit them to the backend server. The *flier grouping* module clusters fliers with duplicate reposts from different reposters to a flier group. The *data selection* module chooses the best view of a flier in a flier group, which can be used for result display. *Intelligent tagging* auto-assigns tags to reposts.

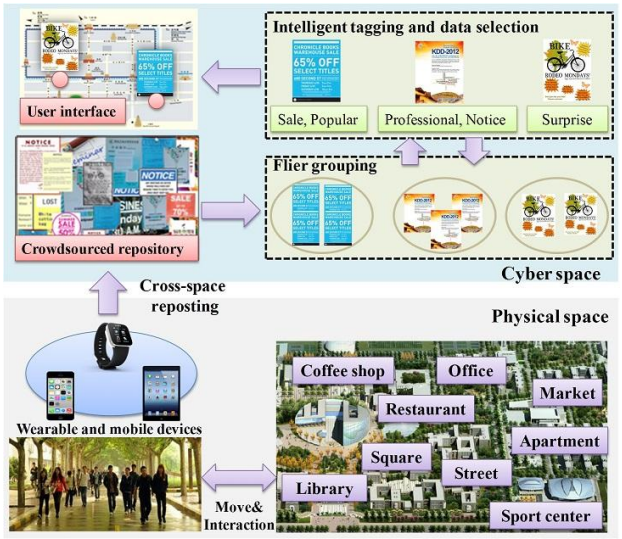

Figure 1. The FlierMeet system architecture.

Considering general user demands, we characterize fliers at two different levels using two types of tags: category tags and semantic tags. The following

Figure 3. Map & Tag view of reposts.

Figure 4. Detailed info of a repost.

category tags are considered: (1) ads (e.g., special deals, sales), (2) academic fliers (e.g., seminar posters), (3) notices, and (4) recruitments.

We also define the following *semantic tags*.

- ***Popular***. A 'popular' flier refers to a flier whose content will be enjoyed by a variety of people (e.g., with different age, occupation, etc.).
- ***Professional***. In contrast to 'popular' fliers, a professional flier shows its influence to a specific community of users who share some commons.
- ***Social.*** People from *existing groups* usually show high similarity, which motivates us to characterize a flier at the social structure level.
- ***Surprise***. We believe that the type of fliers that a person commonly sees is not novel to the user. So fliers that do not often appear in a user's daily life have more potential to be a surprise to the user.

Reposting, Grouping, and Selection

Similarity-based Flier Grouping. The aim of flier grouping is to cluster duplicate reposts from different users. It can be framed as a near-duplicate image detection problem. The SIFT approach is used here for near duplicate image retrieval [4].

Quality-based Repost Selection. Selection is important because reposts are in the format of images, which are often too blurry or dark to be useful. We design heuristics based on the "reposting action" context learned from the associated sensory readings to converge on the "best repost" selection. Three contexts are used, including *light intensity*, *motion blur*, and the *shooting angle*. They can be learned from the built-in sensors (accelerometer, light, magnetometer) of smartphones.

Crowd-Powered Flier Tagging

Assigning tags to fliers is important for selective viewing and intelligent recommendation. We have two types of tags and use different approaches for tagging.

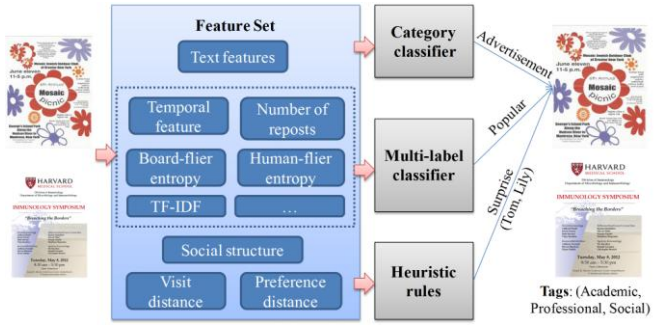

Figure 2. Hybrid classifiers of FlierMeet.

Table 1. Crowd-object interaction features for flier tagging.

Feature	Description
Flier group	The number of reposts of the same flier, the number of source boards of a flier group
Social structure	The link density of social connections among the reposters of a flier group
Human-flier entropy	To measure the diversity of fliers on a board, motivated by the use of entropy in biodiversity
Frequency	To quantify the diversity of reposts of a person
Temporal	The intervals among reposts to the same flier
Visit distance	To tell the distance between the location of a flier and a person's daily visiting area
Preference distance	To tell the distance between the type of a flier and a person's flier preference.

Category tagging. There are two steps. We first use a commercial-grade optical character recognition (OCR)

Initial Experiment Results

Table 2. Performance for flier grouping.

TP	FP	FN	Precise	Recall
1491	12	45	99.4%	97.1%

As shown in Table 2, there were 45 reposts that should have been clustered in a group but were not; 12 reposts were wrongly clustered. The precision and recall for flier grouping are high.

Table 3. Category tagging results.

Category	Ad	Recruit	Acade-mic
Ad	68.87	10.30	11.96
Recruit	1.05	75.79	6.32
Academic	3.67	6.42	81.80

We used the Naïve Bayes algorithm provided by MALLET for document classification. 270 (90 for each) reposts were used as the test set. The results shown in Table 3 indicate that our method for category tagging is effective.

Table 4. Multi-label classification results.

Algorithm	RAkEL	HMC	HOMER
Precision	0.73	0.62	0.65
Recall	0.74	0.64	0.69
F Measure	0.72	0.62	0.66

Three multi-label classification algorithms in Mulan were used, including RAkEL, HMC, and HOMER. The classification results are shown in Table 4. It is clear that *RAkEL* achieved the best performance in almost all measures.

tool to recognize texts from reposts. MALLET (http://mallet.cs umass.edu/) is then used for natural language processing and document classification.

Multi-label classifier. Since there can be overlaps among different semantic tags, we propose a multi-label classifier. A variety of features extracted from crowd-object interaction are used for semantic tagging, as listed in Table 1. We use Mulan [5], a java library for multi-labeling.

Heuristic rules. We use the social structure of a group to determine whether a given flier should be labeled as 'social'. The fliers whose visit and preference distances (see Table 1) to a person are all above the predefined thresholds are labeled as 'surprise' to that person.

Implementation and Evaluation

FlierMeet is developed on Android and users can choose among different tags and browse by 'tag' on the map (see Fig. 3). If the user clicks a repost on the map, detailed information about that repost, such as its reposters and user comments, will be listed (see Fig. 4).

In the initial deployment, we recruited 38 student volunteers to repost fliers on our university campus. They varied in their demographics such as age, major, interests, etc. During the eight-week period, we collected 2,035 reposts, grouped into 921 flier groups.

Based on the deployment, we have made initial experiments, as shown in Table 2 to 4. The results indicate that the methods we developed are effective.

Work in Progress

Incentives. To make FlierMeet a success, we should have numerous participants to contribute. There are four stakeholders: the publisher, reposters, service provider, and crowd. Many flier publishers (e.g., store ads, recruitments) want to disseminate their info to a wide range of people, and they are willing to give payment to well-performed reposters. We intend to build economic models to enhance user participation.

Activity lifecycle management. This work focuses on the preparation phase of the activity lifecycle. The other important phase is activity running. The people who attend an activity want to have 'summarized' memories about the activity, while someone who has interest but not available to attend the activity may want to query 'important' info from offline activities. We want to develop crowd-powered approaches to broadcast offline activities and bridge online-offline activities.

Acknowledgements

This work was partially supported by the National Basic Research Program of China (No.2012CB316400), and NSF of China (No. 61332005, 61373119, 61222209).

References

[1] Guo, B., et al. From participatory sensing to mobile crowd sensing. *IEEE PerCom'14 Workshops* (2014).

[2] Chon, Y., et al. Automatically characterizing places with opportunistic crowdsensing using smartphones. *UbiComp'12*, ACM Press (2012), 481-490.

[3] Koukoumidis, E., et al. SignalGuru: leveraging mobile phones for collaborative traffic signal schedule advisory. MobiSys'11, ACM Press (2011), 127-140.

[4] Lowe, D. G. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91-110.

[5] Tsoumakas, G., Ioannis K., and Ioannis V. Mining multi-label data. *Data mining and knowledge discovery handbook*, Springer, (2010), 667-685.