Discovering the City by Mining Diverse and Multimodal Data Streams

Yin-Hsi Kuo*, Yan-Ying Chen, Bor-Chun Chen[1], Wen-Yu Lee, Chun-Che Wu, Chia-Hung Lin, Yu-Lin Hou, Wen-Feng Cheng, Yi-Chih Tsai, Chung-Yen Hung, Liang-Chi Hsieh, Winston Hsu
National Taiwan University and [1]Academia Sinica, Taipei, Taiwan

ABSTRACT

This work attempts to tackle the IBM grand challenge – seeing the daily life of New York City (NYC) in various perspectives by exploring rich and diverse social media content. Most existing works address this problem relying on single media source and covering limited life aspects. Because different social media are usually chosen for specific purposes, multiple social media mining and integration are essential to understand a city comprehensively. In this work, we first discover the similar and unique natures (e.g., attractions, topics) across social media in terms of visual and semantic perceptions. For example, Instagram users share more food and travel photos while Twitter users discuss more about sports and news. Based on these characteristics, we analyze a broad spectrum of life aspects – trends, events, food, wearing and transportation in NYC by mining a huge amount of diverse and freely available media (e.g., 1.6M Instagram photos, 5.3M Twitter posts). Because transportation logs are hardly available in social media, the NYC Open Data (e.g., 6.5B subway station transactions) is leveraged to visualize temporal traffic patterns. Furthermore, the experiments demonstrate that our approaches can effectively overview urban life with considerable technical improvement, e.g., having 16% relative gains in food recognition accuracy by a hierarchy cross-media learning strategy, reducing the feature dimensions of sentiment analysis by 10 times without sacrificing precision.

Categories and Subject Descriptors

H.4.0 [**Information Systems Applications**]: General

Keywords

Multiple media sources; cross-media mining; visualization

1. INTRODUCTION

Due to the explosion of social media, people are used to sharing their daily life on the Internet. Over the decade, there have been many social websites developed for different user groups. Based on the characteristics within, people will choose different media to share their thoughts, photos, personal preferences, etc. For example, people may utilize Foursquare for recording places where they go, Instagram for sharing their visual experiences via photos or videos, and Twitter for updating things that interest them. Therefore, these social media capture different moments in the daily life. The popularity of social media also brings an emerging need of social marketing campaigns for brands [6]; hence, it is essential to discover and mine the user behaviors on diverse social media.

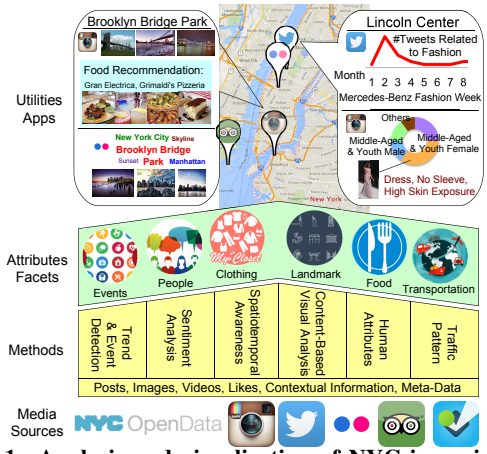

Figure 1: Analysis and visualization of NYC in various perspectives. We collect multiple media sources and adopt different methods for discovering the latent facets. Observing the complementary properties across sources, we enable a systematic framework to leverage the cross-domain and multimodal data for numerous applications for NYC analytics.

Table 1: The percentage of overlapped attractions among Twitter, Instagram, TripAdvisor, and subway traffics (open data).

	Twitter	Instagram	TripAdvisor	subway
Twitter	1	0.35	0.08	0.22
Instagram	0.35	1	0.18	0.43
Tripadvisor	0.08	0.18	1	0.32
subway	0.22	0.43	0.32	1

It is expected that each social website has unique information and features. As a case study, we conducted an experiment for three major social media, including Twitter, Instagram, TripAdvisor, and one NYC open data. The goal is to preliminarily understand the synergies and differences between the data streams, sampled and contributed for human activities. We collected data from the four websites to generate the datasets, and then extracted the top-100 popular places respectively. Table 1 shows the overlaps of each pair of the datasets. We observed that the top-100 places from Twitter and Instagram mainly focus on office buildings, residential places, and chain stores (such as Starbucks and Subway), while those from TripAdvisor focus on the places of famous landmarks, museums, and busy restaurants. The locations of the places obtained from Twitter are denser than those in Instagram. As has been mentioned, using single source is not enough. To obtain richer information, integrating multiple media sources is in dire need. In addition, based on the four datasets, we further observed that many of the popular places are close to subway stations in NYC.

Prior works usually rely on single media source (e.g., Flickr) with different modalities (e.g, image, tag, user) to address traditional applications such as image annotation, travel recommendation, tag ranking, and user preference mining [2][5]. Instead of relying on single source, for this challenge, we aim to leverage multiple data streams including social media and open data for mining

*All authors contributed equally to this work. The work was supported in part by Intel-NTU Connected Context Computing Center.

MM'14, November 3–7, 2014, Orlando, Florida, USA.

http://dx.doi.org/10.1145/2647868.2656406.

Table 2: Sentiment analysis results over noisy tweets. The letter trigram is robust to textual noises, compact, yet informative.

Methods	Error	Feature Dimensions
letter trigram	0.1447	17,030
BoW	0.1611	178,717
letter trigram + BoW (early fusion)	**0.1289**	195,747
letter trigram + BoW (late fusion)	0.1295	195,747

urban activities. We argue a systematic framework (Fig. 1) which detects rich and multi-angle semantics and attributes to facilitate further applications. Moreover, we devise advanced methods to leverage complementary information across sources and deal with the noises in such user-contributed data streams.

Our contributions include, 1) leveraging the complementary information for data analytics, 2) discovering the latent behaviors, synergies and differences among media sources, 3) mining effective factors parameterizing emerging applications from heterogeneous and multi-granularity signals, 4) devising detection and learning methods robust to noisy user-contributed data, and 5) identifying the opportunities and challenges for mining diverse data streams.

2. MEDIA SOURCE COLLECTION

Social media presents rich aspects to overview life in a city. To gain from the diverse and complementary resources, we collect the visual/text content along with metadata in NYC from four social media: 1) *Instagram* – almost ten thousands check-in locations from Instagram which contains 1.6 million photos from October 2010 to May 2014; 2) *Twitter* – 5,337,958 geotag tweets from January 2013 to September 2013; 3) *Flickr* – 261,907 geotag photos from 2012 to 2014; 4) *TripAdvisor* – 842 attractions (e.g., interesting points, landmarks) and 11,566 restaurants in 2014.

Open data is gaining more popularity because it is freely available without copyright concerns. We exploit two sorts of open transportation logs which are hardly available from social media. The first includes 5.5 million trip logs of public bike sharing system (*CitiBike*) from July 2013 to February 2014 and the second comprises 6.5 billion accumulated entries/exits of the turnstiles at NYC subway stations (*MTA*) from January 2012 to May 2014.

3. MINING FROM MULTIPLE MEDIA

3.1 Trend and Event Detection

Detecting what is buzzing on multiple media sources is a good way to understand the urban life, while using Twitter as text-oriented media and Instagram as image-oriented media. When dealing with social media, we use both bottom-up and top-down methods. The bottom-up means detecting trends from social media with multiple granularities. In *spatial aspect*, we detect the trends not only of the whole NYC, but also of some popular regions. As for *temporal aspect*, some big news or large scale events like "World Cup" may lead to monthly trends, but some, such as the news "Suarez Bite Italian Defender" which may survive for a day. On the other hand, the top-down means analyzing a known event. It is easy to show an event's volume change by analyzing its temporal distribution. While considering both spatial and temporal information, we can even see how an event moves.

We brief our trend and event detection method, mainly consisting four parts. 1) *Preprocessing.* We apply stemming and stop word removal on text messages (i.e., Instagram's hash tags, and both Twitter's surrounding text and hash tags). 2) *Trending term detection.* We aim to extract the term which is buzzing at current time, and calculate the buzz score which is similar to the method [3][8]. 3) *Trending topic generation.* After the trending terms are collected, the next step is to group them into topics. To achieve the goal, we aggregate the trending terms with frequent concurrences. 4) *Event analysis.* While we have known an event, we can distinct what is related to the event, and see how it distributes on spatial and temporal domains. A result is illustrated in Fig. 3(b).

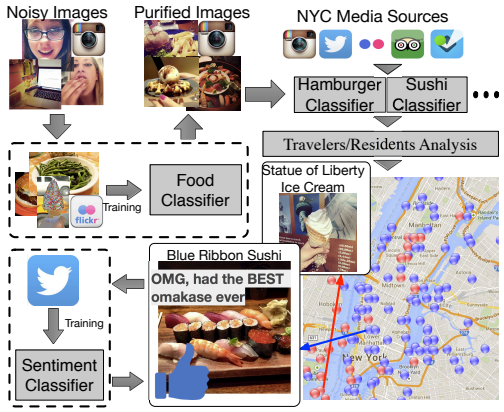

Figure 2: Food recommendation by leveraging food recognition (images), user analysis (logs), and sentiment analysis (comments) over multiple media sources. Red spots on the map are restaurants popular among travelers, while blue spots show the favorites for the (local) residents.

3.2 Sentiment Analysis

Sentiment is another important information contained in social media. For training data, we collect 210,000 (noisy) tweets with emoticons. Half of these tweets with ":)" means containing positive sentiment; others with ":(" means containing negative sentiment. We use word unigram (BoW model) and letter trigram as features. However, there are many typos and variant tags of the same targets. We are the first to adopt letter trigram [4] to represent nosiy textual data in social media. We add starting and ending marks to a word (e.g., #goal# and #goooooal#), break the word into letter trigrams (e.g., **#go**, goa, **oal**, **al#** and **#go**, goo, ooo, ooo, ooa, **oal**, **al#**), and use logistic regression to train data. As shown in Table 2, we can predict sentiment on user comments by only 12.89% error rate.

3.3 Human Attributes

Human attributes such as gender, clothing are important dimensions to understand a city. Because user profiles are very sparse, we detect human attributes in visual content from user contributed photos to observe this city. These photos are like million social sensors which can profile not only social media users but also their communities in real life.

Clothing Attributes Detection. Motivated by [1], we aim to learn semantic attributes to mine fashion trends in urban life. We first use a face detector to trim false positive examples incurred by the pose estimator. We then use Log Gabor, SURF, Skin, and LAB-color to extract low-level features from the five parts (i.e., torso, upper and lower arms). We train classifiers using SVM with totally 26 attributes. We use the dataset in [1] as the training data and test the media sources mentioned in Sec. 2. We focus on those significant attributes across time, e.g., clothing categories for season, patterns for fashion and trend.

Facial Attributes Detection. We construct 11 facial attribute detectors including gender, age, race and glasses by using a generic framework for learning facial attributes adaptively [7]. We collect training data with 10,000 images from Flickr. Given an image, we detect face regions and locate facial parts (eyes, nose and mouth) to extract low-level features (edge, color, texture); mid-level features are then learned by SVM with low-level features from facial parts. Each attribute detector is an ensemble with mid-level features weighted by Adaboost. We apply this framework to each test photo and predict the presence of each attribute.

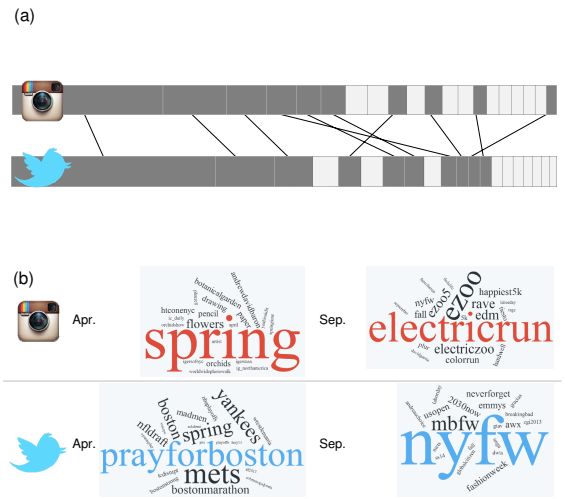

Figure 3: The upper is the top 20 hashtags in Instagram and Twitter. The overlapped tags in grey bricks are linked, and the unique tags are in colors. The bottom shows tag clouds of different trends discovered in April and September. They tend to cover different topics in different contexts so that we argue to integrate emerging and diverse media sources.

3.4 Food Recognition

A large proportion of the images in social media contains food (e.g., more than 100 million images contains hashtag "food" in Instagram). To mine information from these images, it is insufficient to simply use tags to decide what is in the image since tags are really noisy (in our preliminary study, only about 30% of images tagged food on Instagram are correct). Therefore, we adopt state-of-the-art recognition method, convolution neural network (CNN), to recognize food in images. By analyzing different social media sources, we found that Instagram contains more images that can represent users' daily life, but images in Instagram also contain more noises (our experiment shows that about 26% of images tagged food are selfies instead of food images, while the same number is 7.5% in Flickr). Hence, we propose a hierarchy cross-media learning (Fig. 2), where the less noisy Flickr images is used to train a general food classifier and is used to clean the images in Instagram, and purified Instagram images are then further used to train 9 food sub-categories such as hamburger. Using the proposed method, we can achieve 65% accuracy on testing compared to 56% when Instagram data is directly used for training and testing.

3.5 Traffic Patterns

City-scale traffic patterns reveal rich people activities, e.g., commuting, economics and leisure, which are essential for urban computing, advertising and trip planning. To comprehensively analyze NYC traffics, we address the differences between travelers and (local) residents, and the temporal impacts by using Instagram data and open transportation data (CitiBike and subway data in Sec. 2).

Travelers and Residents may care about different matters, and thus have divergent trip patterns. To differentiate their patterns, we calculate the proportion of NYC photos (p_i) in a given user's Instagram album i. If $p_i > 0.5$, the user is classified as a resident and otherwise a traveler. Meanwhile, we treat the CitiBike users registered in "Customer" (one-day/week pass) as travelers and those of "Subscriber" (annual use) as residents. By the trip logs and registrations, we can figure out their visiting patterns for validation.

Daily and Hourly Patterns can roughly depict how residents move for the purpose of working or leisure. Such trends reflect commuting behavior and economic activities that help traffic flow and transportation management. Because subway is a major transportation services in NYC, we use 6.5B entries and exits of each turnstile to discover moving patterns. The accumulated counts of entries and exits indicate how many people leave and enter somewhere near this station, respectively. Since subway logs are not sampled by hour, we leverage the accumulated trips leaving and arriving CitiBike stations, to discover the hourly patterns.

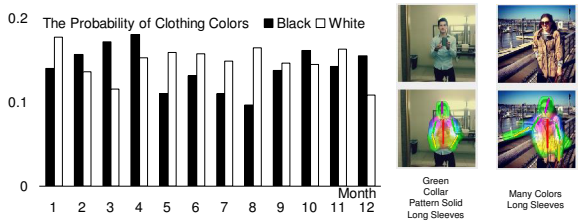

Figure 4: Varying clothing colors across months and two clothing attribute examples. It shows that people tend to wear clothing with dark color in winter whereas light color in summer.

4. VISUALIZATION AND APPLICATIONS

We believe that the data are of the users, by the users, and for the users. In order for the public to easily get across NYC through our rich mining aspects, we design various visualizations and apps to draw those technical data closer to the users. With the interactive interfaces, we can add priceless values in these data and help users to explore by themselves.

4.1 Social Media Association and Comparison

To integrate multiple media sources, we investigate the differences of hashtags between Instagram and Twitter by the proposed method (Sec. 3.1). For overall comparisons, Fig. 3(a) shows the stacked bar charts of the top 20 hashtags. The overlapped hashtags in grey bricks are linked and the unique tags are colored in red/blue. In Fig. 3(b), we show the monthly trends by tag clouds with the top 20 trending hashtags in April and September 2013. Most hot terms (grey bricks) are always popular and ranked higher in (a). Moreover, each media source has their own special tags, e.g., "instagood," "tweetmyjobs." We find that Twitter could provide detailed information of events or institutions, e.g., jobs or the fashion week. In addition, Twitter hashtags are more about the issues which highly concerned by but might not directly happen to users. For example, Boston marathon explosion and MLB season opening in April or New York fashion week in September are public events or news for the New Yorkers. On the other hand, we see fewer sports posts in Instagram (like "yankees," "mets" in Twitter) because it is hard to take photos in person. So, in Instagram, there are more check-in locations and posts regarding to the events which users are directly engaged. For example, many participants of "electricrun" posted relevant photos in April. The different user behaviors suggest that Instagram is a visually-oriented media for users to share their daily life. Conversely, Twitter is a platform for users to publish the opinions of trending events. Due to various purposes, the trending hashtags are quite different and provide complementary information indicating diverse aspects of urban life.

4.2 Human Activities and Fashion Styles

The photos for analyzing in the challenge are much noisier than the training data commonly used in vision community; for example, selfies usually contain only faces without bodies, and people have various poses. For clothing analysis, only about 20% data have the poses detected correctly. These phenomenon lead to more uncertainty in traditional clothing detection methods, but we can still find some significant information. As shown in Fig. 4, we can see that people wear clothing with black color more in winter and

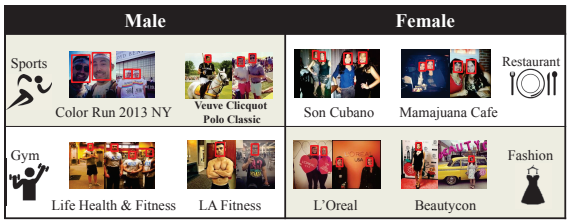

Figure 5: Facial attributes demonstrate that Instagram users have different behaviors when sharing photos. Sports are favored by men, and restaurant visitings or fashion activities are favored by women.

less in summer, while white color has the opposite tendency. Besides, we might observe the fashion and trend as the seasons change and some special fashion styles for each area. As shown in Fig. 1, we can analyze the fashion styles for the New York Fashion Week (NYFW) which is automatically discovered from Twitter.

Facial attributes is one of important information to analyze the user behaviors when sharing photos. We choose top 500 locations based on the number of images to detect 11 facial attributes and classify each location by the proportion of each attribute in the same type which is more than 50%, for example, gender, race, and glasses. As shown in Fig. 5, men prefer sport activities and women commonly share the fun in restaurants or during shopping. The attribute "glasses" dominated locations are always at open spaces like beaches and parks because many people will wear sunglasses in these spaces. Besides, we might combine other applications such as food recommendation or event detection to personalize the results with the detected attributes. For example, the female favourite restaurants in New York or what kind of people will attend the New York Fashion Week (Fig. 1). Such visual information is very different and complementary to traditional methods by text only.

4.3 Food Recommendation

Eating is one of the most important things in our live, therefore we provide a food recommendation system to recommend popular food in New York City. We combine food detection results (Sec. 3.4) with similar geo-locations together to automatically find popular restaurant locations. Comparing with traditional food review websites like Yelp, our system can provide more real-time information since people will upload their images and comments to social media websites (e.g., Instagram) while they are eating, but they might post the restaurant review on Yelp days after they ate when they have free time. We can also provide more user-centered information since all of our data are directly come from users while some of the reviews on Yelp might be paid advertisements. Comparing solely using image tags to find restaurants, our system can provide more informative food images since directly using image tags might found many noisy images like selfies. In order to provide more accurate recommendation, our system further combined two different techniques. First, we combined travelers and residents information (cf. Sec. 3.5) and found restaurants that are preferred by travelers and residents respectively. As shown in Fig. 2, the restaurants around Liberty Island and Central Park mainly serve for travelers. Second, we use sentiment analysis (cf. Sec. 3.2) to find positive and negative sentiments in the image comments and use it to find whether the foods are recommended by the user or not. For example, "Quite possibly my most favourite food..Looks Devine!!" is predicted as a positive comments whereas "Waited an hour for just the 4 bowls of not so great noodle" is negative one.

4.4 Traffic Flow

As Fig. 2 shows, the popular locations for travelers (red spots) and residents (blue spots) mined from Instagram data and CitiBike

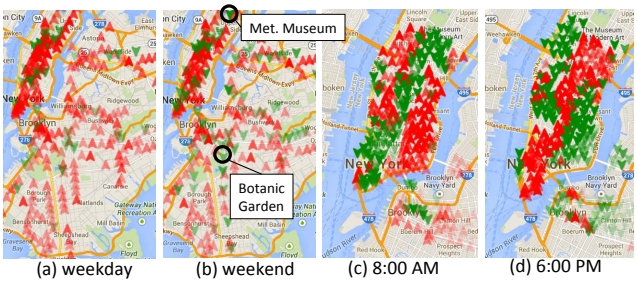

Figure 6: Traffic flow in NYC by open data. Green/red arrows indicate in/out the pointed spots. Higher color saturation means more visitors. Overall, more traffics enter Manhattan on weekdays (a) but partially turn to its neighborhood on weekends (b). The traffics during business hours (c)(d) shows the commercial areas are densely located in midtown/downtown.

logs have similar distribution in NYC. For example, both the data show that travelers prefer southern area of Central Park while residents are gathered around financial centres probably near their workplace. The phenomenon matches their moving patterns on weekdays – most heavy traffics appear in transportation centers (e.g., Penn Station, Grand Central Station) and the Financial District in NYC (Fig. 6 (a)). Overall, more traffic flow enters centre of Manhattan but partial flow on weekends (Fig. 6 (b)) turns to enter the suburban in its neighborhood (e.g., Botanical Garden in Brooklyn) or the travel landmarks (e.g., Metropolitan Museum). That suggests people tend to escape from city centers for their leisure outings or cultural activities on weekends. Meanwhile, the traffic flow around commuting time (Fig. 6 (c) and (d)) again confirms that the major commercial transactions are densely located at Manhattan midtown and downtown (green area in (c)) while the traffic flow after business hours spreads to the periphery of Manhattan (green area in (d)).

5. CONCLUSIONS AND FUTURE WORK

In this work, we show that (1) different social media do behave differently according to our experiments, such that cross-media mining is helpful to overview urban life in comprehensive perspectives, (2) the daily life has been broadly discovered in plentiful aspects by our proposed methods, (3) open data is a complementary resource beyond social media, especially good for transportation. Based on the preliminary results, a comprehensive system can be constructed outstandingly.

6. REFERENCES

[1] H. Chen et al. Describing clothing by semantic attributes. In *ECCV*, 2012.

[2] T.-S. Chua et al. Next: Nus-tsinghua center for extreme search of user-generated content. *IEEE MultiMedia*, 2012.

[3] A. Dong et al. Towards recency ranking in web search. In *WSDM*, 2010.

[4] P.-S. Huang et al. Learning deep structured semantic models for web search using clickthrough data. In *CIKM*, 2013.

[5] L. Kennedy et al. How flickr helps us make sense of the world: Context and content in community-contributed media collections. In *ACM MM*, 2007.

[6] G. Kim and E. P. Xing. Visualizing Brand Associations from Web Community Photos. In *ACM WSDM*, 2014.

[7] N. Kumar et al. FaceTracer: A Search Engine for Large Collections of Images with Faces. In *ECCV*, 2008.

[8] C.-C. Wu et al. Learning to personalize trending image search suggestion. In *ACM SIGIR*, pages 727–736, 2014.