What Makes New York So Noisy? Reasoning Noise Pollution by Mining Multimodal Geo-Social Big Data

Hsun-Ping Hsieh  
National Taiwan University  
Taipei, Taiwan  
d98944006@csie.ntu.edu.tw

Tzu-Chi Yen  
Sensoro Technology Co., Ltd.  
Beijing, China  
junipertcy@gmail.com

Cheng-Te Li  
Academia Sinica  
Taipei, Taiwan  
cylltli@citi.sinica.edu.tw

ABSTRACT

Noise pollution in modern cities is getting worse and sound sensors are sparse and costly, but it is highly demanded to have a system that can help reason and present the noise pollution at any region in urban areas. In this work, we leverage multimodal geo-social media data on Foursquare, Twitter, Flickr, and Gowalla in New York City, to infer and visualize the volume and the composition of noise pollution for every region in NYC. Using NYC 311 noise complaint records as the approximation of noise pollution for validation, we develop a joint inference and visualization system that integrates multimodal features, including geographical, mobility, visual, and social, with a graph-based learning model to infer the noise compositions of regions. Experimental results show that our model can achieve promising results with substantially few training data, compared to state-of-the-art methods. A NYC Urban Noise Diagnoser system is developed and allowed users to understand the noise composition of any region of NYC in an interactive manner.

Categories and Subject Descriptors

H.4.0 [Information Systems Applications]: General

Keywords

urban noises; multimodal data; geo-social media; big data

1. INTRODUCTION

With the prosperity of geo-social media, like Flickr, Twitter, Facebook and Foursquare, the volume of data that sensors and records the urban dynamics has increased dramatically. Geo-social media data is essentially multimodal since it is a mixture of geographical information (venue and location attributes), mobility footprints (check-in data), visual snapshots (consumer-contributed photos), and social interactions (social network). Data collected from geo-social media can be extensively considered as a kind of sensors that monitor urban human activities, and can be applied to urban-driven applications, such as detecting natural hazards events [7], tracking disease activities [8], and inferring air quality [10]. In current big cities like New York City, noise pollution had become a major problem [9] that affects the work efficiency and the sleep quality, and damages the mental health of people [1]. To tackle noise pollution of a certain urban region, the first step is to understand the elements and causes that produce the noise. If the composition of noise of any place in a city can be unveiled, such as “in afternoon rush hour, 60% noise come from vehicle traffic, 30% from loud music, and 10% from loud talking” on weekdays at Time Square, the government agencies can make the policies to cope with the problems, and the residents can plan their daily routines to prevent themselves from noise disturbance.

In this paper, we aim to leverage multimodal geo-social media data to reason urban noises of regions in NYC. It is a challenging problem because both the magnitude and the composition of noise pollution vary by places and change over time. One might think installing sound sensors everywhere in the city is an approach. However, sound sensors are costly, limited to very small areas, and hard to help diagnose the reasons that produce noises. Since it is infeasible to access the noise information of all the urban regions, we resort to use the noise complaint records of the NYC 311 service, which is a platform allowing people to report the locations, time, and categories of their near-by annoyed events via phone calls, as an approximate of urban noises at some regions and time. Then we are allowed to analyze the noise
composition of any region in a city. Nevertheless, the NYC 311 noise complaint data is sparse and gathered in a few number of regions. A model is needed to infer the noise composition of any region in the city.

Given the regions in NYC, in which a limited number of regions have sufficient noise composition information derived from 311 noise data, and multimodal geo-social media data, including check-in records in Twitter, geographical locations with categories in Foursquare, geo-tagged images in Flickr, and location-based social network in Gowalla, as illustrated in Figure 1, our goal is to infer the noise compositions of the remaining regions. The NYC 311 noise data is used as labeled data for model training and inference validation. Then based on the inferred results, we develop a visualization system that allows users to understand the noise compositions of regions in an interactive manner. To deal with the inference task, we devise a novel model, Urban Noise Diagnator (UND). In UND, for each of the noise-labeled regions (marked ones) and the unlabeled regions, we use the geo-social sensor data to extract four heterogeneous features: geographical features from Foursquare, mobility features from Gowalla check-in data, visual features from Flickr geo-tagged photos, and social features from Gowalla location-based social network. Then a three-stage inference method is devised: constructing region association map (RAM), modeling the similarity values between regions, and a noise propagation algorithm to infer the noise composition of a region.

2. URBAN NOISE DIAGNATOR

2.1 Geo-Social & NYC 311 Noise Datasets

We extract the noise complaint records from the open data of the NYC 311 service\(^1\). From Jan 2010 to Mar 2015, there are 196,631 complaint records of urban noise. The noise composition consists of 14 categories, in which the top-4 categories with the largest proportions are Loud Music/Party, Construction, Loud Talking, and Vehicle. On the other hand, we collect check-in records from Twitter\(^2\), venue data with location categories from Foursquare\(^3\) and the location-based social network (LBSN) data from Gowalla\(^4\). There are totally 47,581 venues/locations, 196,591 users, 950,327 social connections between users, and 6,442,890 check-in records. For geo-tagged photos, we use Flickr tourist photos whose GPS coordinates are in NYC, collected by Chen and Grauman [3]. Totally there are 57,200 images. According to a well-known study of urban noise analysis [4], we partition the geographical area of NYC into a number of regions, in which the size of each region is 1 km × 1 km. Totally there are 560 regions in NYC.

2.2 Multimodal Features

Geographical Features (G) are devised to describe the spatial and category distribution and the geographical interactions between venues because the type of noise can be related to the urban business environment. Regions with more/less venues or more/less venue categories can have different effects on the composition of noise. We extract two kinds of features for a region \(r\).

- **Volume-based features (G-V)** use the region density and the venue closeness as the feature values. Region density \(F_d(r)\) is the number of venues in a region. Venue closeness \(F_{vc}(r)\) is the average geographical distance between venues in region \(r\).

- **Category-based features (G-C)** consider the categories of venues, such as food, shops, outdoors, and nightlife, to characterize the activities of a region. We propose two category-based features. The first is category entropy measuring the diversity of a region, given by \(F_e(r) = -\sum_{a \in A} f_a \cdot \log(f_a)\), where \(a\) is a venue category, \(A\) is the set of all categories, and \(f_a\) is the fraction of venues with category \(a\) in region \(r\). The second is competitiveness estimating the extent that venues with the same category gather in a region, given by \(F_c(r) = \left| \{l_i, l_j \in r : a(l_i) = a(l_j)\} \right| / F_d(r)\), where \(l_i\) is a check-in venue, and \(a(l_i)\) is the category of \(l_i\).

Mobility Features (M) are developed to model the human moving dynamics among venues in a region. Regions with larger volume of visits, higher density of movement, and more incoming movements can produce more noises. Hence we use check-in data to propose the following features to characterize the human mobility.

- **Region Popularity (M-P)**, denoted by \(F_p(r)\), is the number of check-ins in region \(r\) at a certain time.

- **Movement Density (M-D)** is the density of transitions between venues in region \(r\). By denoting the set of consecutive check-ins from venue \(l_i\) and \(l_j\) in region \(r\) as \(TS_r, \{(l_i, l_j) \in TS_r\}\), the movement density \(F_{md}(r)\) is defined as the size of \(TS_r\).

- **Incoming Flow (M-F)** estimates the transitions from venues outside \(r\) to venues in \(r\). The incoming flow value of \(r\) is given by: \(F_{if}(r) = \left| \{(l_i, l_j) \in TS : l_i \notin L(r) \land l_j \in L(r)\} \right|\), where \(TS\) is the set of all consecutive check-ins, \(L(r)\) is the set of all venues in \(r\).

Visual Features (V). We combine the global features of colors and texture and local geometric features to model the visual contents of photos. The global features capture the photographic visual composition while the local ones identify the actual structural elements of objects. Such mixed global and local features have been shown to provide significant complementary information in recognition tasks [2].

- **Global (V-G)**. We extract color histograms \(F_{ch}(r)\) and color moments \(F_{cm}(r)\) as global features of photos to capture theirs visual and spatial color distributions. We also utilize the Gabor textures \(F_{gt}(r)\) to model the textural information. Feature vectors are produced for each photo by concatenating these features. The similarity between two photos is calculated according to normalized Euclidean distance.

- **Local (V-L)**. We represent photos by local interest point descriptors, given by the scale-invariant feature transform (SIFT) with a Difference of Gaussian (DoG) process. Given two photos \(p_i\) and \(p_j\) with the descriptor vectors, \(D_i = (d_{i1}, d_{i2}, ..., d_{in})\) and \(D_j = (d_{j1}, d_{j2}, ..., d_{jn})\),
we define the similarity between two photos as the number interest points shared between two photos divided by their average number of interest points.

**Social Features (S)** are designed to characterize the social interactions for people who had ever been in the region. Some public regions are crowded while people who tend to visit friends or family gathering together. Different social interactions can lead to various levels of noises with distinct compositions of noise. Therefore, we compute features from the location-based social network.

- **Centrality Metrics (S-C).** We employ network centrality metrics $F_\text{cen}(r)$, including degree, closeness, betweenness, PageRank, and SimRank, and geo-social influence measures [6] $F_\text{infl}(r)$, including spatial degree, spatial closeness, node novelty, and geographic clustering coefficient, over users who had ever visit region $r$ as feature values. The maximum, minimum, and average values for these nine features are computed.

- **Group Measures (S-G).** We characterize the extent of cohesion or separation for people who visit venues in region $r$, by modeling the structure connectivity of the graph $G(r)$ induced by users who had ever visited venues in $r$. We compute the values of density, clustering coefficient, the number of communities, and the number of components in $G(r)$ as group features $F_g(r)$.

### 2.3 Region Association Map

We construct the Region Association Map (RAM) to model the correlation between regions. We first introduce the noise composition and RAM, then describe how to exploit the extracted features to derive region similarities as edge weights.

The noise composition of a region $r$, denoted by $n(r)$, is a vector, in which each element $n_v(r)$ is the percentage of a certain noise category $c$ happening to region $r$. The dimension of $n(r)$ is the number of noise categories, and $\sum_{c \in C} n_v(r) = 1$, where $C$ is the set of all noise categories.

A RAM is a multi-layer weighted connected graph $G = (G^{t_1}, G^{t_2}, ..., G^{t_n})$, in which $t$ is the total number of layers for time periods $t_1, t_2, ..., t_n$, and $G^{t_i} = (V, E, W)$ is the layer graph in at the $t_i$-th time period, where $V$ is the set of regions, $E$ is the set of edges between regions, and $W = W^{t_i} + W^{t_i, t_j}$ is the matrix representing edge weights, where $W^{t_i}$ and $W^{t_i, t_j}$ are edge weights learned from nodes within time period $t_i$ and across time periods $t_i$ and $t_j (i \neq j)$ respectively. The node set $V$ consists of (a) regions whose noise compositions have known from NYC 311 noise complaint records, termed by labeled nodes $V_L$, and (b) regions whose noise compositions are unknown, termed by unlabeled nodes $V_U$, where $V = V_L \cup V_U$. Each labeled node $u_i$ is associated with the corresponding noise composition $n(u_i)$, and the noise composition $n(v_o)$ of each unlabeled node $v_o$ is initialized as a zero vector. The edge set $E$ also consists of two parts: the set of edges $E_c$ connecting nodes within each layer graph $G^{t_i}$, and the set of edges $E_n$ connecting the same nodes across different layer graphs $G^{t_i}$ and $G^{t_j} (i \neq j)$, where $E = E_c \cup E_n$.

The construction of RAM consists of three parts. First, since we aim to infer the noise composition of unlabeled regions from labeled ones, we connect each $v_o \in V_U$ to all $u_i \in V_L$ in each layer graph. Second, owing to the fact that the noise composition of a region can be highly correlated to its history, we connect each $v_o$ within time period $t_i$ to the corresponding unlabeled node at time period $t_i (i < j)$ whose noise composition had ever observed. Third, since the noise composition of near-by regions tends to be similar (due to sharing similar characteristics), each $v_o \in V_U$ is connected to the direct near-by $u_o \in V_L$.

### 2.4 Region Similarity

Learning edge weights in RAM from features plays a key role in the inference of noise compositions for unlabeled nodes. We model the similarity between regions as edge weights. The idea is that for a certain time period, if two regions with higher similarity in terms of features, they tend to have similar noise composition. While various features can contribute different degree of effect on the similarity for noise composition, the importance of each feature should be considered separately.

Given a certain feature $F_k$, we compute the corresponding individual similarity $rs_{F_k}(u(t_i), v(t_j))$ between nodes $u$ and $v$ in time periods $t_i$ and $t_j$ respectively from their feature difference $\Delta F_k = ||f_k(u(t_i)) - f_k(v(t_j))||$, and $f_k(u(t_i))$ is the feature vector of feature $F_k$. Then, given a set of features $F = \{F_1, F_2, ..., F_m\}$, we can derive the combined region similarity $rs(u(t_i), v(t_j))$ via the weighted sum of $rs_{F_k}$, given by $rs(u(t_i), v(t_j)) = \exp(-\sum_{k=1}^{m} \pi_k \times rs_{F_k}(u(t_i), v(t_j)))$, where $\pi_k$ is the weight of feature $F_k$.

The combined region similarity is considered as the edge weight $M(u,v; t_i, t_j) = rs(u(t_i), v(t_j))$ between $u(t_i)$ and $v(t_j)$ ($\{u(t_i), v(t_j)\} \in E$). We will empirically learn the feature weight $\pi_k$ from the validation set.

### 2.5 Noise Random Walk

We infer the noise composition of each unlabeled region using RAM. The idea is to apply the random walk mechanism to iteratively update the noise composition $n(v_o)$ of each unlabeled node $v_o$ until the change of its noise composition converges. The noise composition $n(v_o)$ of $v_o$ is computed from its neighboring labeled or unlabeled nodes in RAM, given by: $n_v(v_o) = \sum_{c \in C} \sum_{u_i \in N(u_o)} w_{v_o,u_i} \times n_v(u_i)$, $c \in C$, where $\Gamma(v_o)$ is the set of neighboring nodes in RAM. That said, we aggregate the noise composition of $v_o$'s neighbors via edge weights. Note that the noise composition for each labeled node $v_o$ will not be updated (i.e., they are fixed). Since RAM provides the benefit on modeling the region similarities between nodes, a near-by region $u$ with higher similarity with unlabeled node $v_o$ can contribute more weights on the propagation of noise to $v_o$.

### 3. EXPERIMENTS

The experiments consists of two parts: show the effectiveness of our UND model, understand the performance under time intervals, and report how various features affects the results. First, we control the percentage of regions used as the training set from 10% to 90%, fix 5% as the validation set, and use the remaining as the test set. Second, we present the performance of each feature set. As for evaluation metrics, we use Normalization Root Mean Square Error (NRMSE) that measure the difference between ground-truth and inferred noise compositions, and Average Accuracy that is defined by whether the inferred noise category with the highest percentage of an unlabeled region is its correspond-
ing ground-truth category in average. We compare UND with several typical supervised learning methods, including Support Vector Machine (SVM), Random Forest (RF), $L_1$ logistic regression (LR), and $k$-nearest neighbor (kNN) (averaging the noise composition from the top-$k$ most similar labeled regions using features, $k = 5$).

The main results are shown in Figure 2. We can find that our UND model outperforms the competitors in every case, especially when the training percentage is higher than 50%. The best accuracy (above 0.90) exhibits that UND can accurately identify the major noise category of an unlabeled region. In addition, we use UND to examine the accuracy of each feature set in Figure 3, and find that the geographical features are more influential. Though it may result from the semantics of locations can reflect human activities in geography, we think this deserve more effort to investigate the correlation between urban noise and fine-grained geographical and mobility factors.

4. SYSTEM DEMONSTRATION

We develop the proposed model as an Urban Noise Diagnosor system\(^3\). A snapshot is shown in Figure 4. The essential function is to visualize the degree of noise pollution (highlighted by different colors) for each grid in New York City. The noise pollution is derived from the volume of noise complaints, which are either obtained from both the NYC 311 noise data (for grids having adequate noise complaints) and the inferred noise composition by our UND model (for grids without or having rare complaints). When users select a certain grid, the corresponding noise composition will be displayed. The corresponding multimodal geo-social media information will be presented as well, including the category distribution of locations from Foursquare, the number of check-ins from Twitter, and a few representative venue names with photos from Flickr. Moreover, users are allowed to input the addresses to understand the noise pollution of their nearby areas, and choose the time period of interest (i.e., seasons, weekday/weekend, and hours of day) to observe how noise pollution varies in the selected grid.

5. CONCLUSIONS

The contribution of this paper is four-fold. First, we extract a series of features from multimodal geo-social media data to characterize urban noise dynamics. Second, we develop the UND model that is validated to accurately infer the noise compositions of unlabeled regions. Third, we show that the visual feature of geo-tagged photos is the most important factor on inferring urban noises, and such finding can encourage future advanced analysis on this direction. Fourth, an Urban Noise Diagnosor system is developed for users to understand what make NYC so noisy.

6. ACKNOWLEDGEMENTS

This work was sponsored by Ministry of Science and Technology of Taiwan under grant 104-2221-E-001-027-MY2. This work is also supported by Multidisciplinary Health Cloud Research Program: Technology Development and Application of Big Health Data, Academia Sinica, Taipei, Taiwan under grant MP10212-0318. In addition, we especially thank Qiao Lin, Xiaodong Liu, and Shijing Wu for their fruitful discussions as well as the Sensoro team for the open research atmosphere.

7. REFERENCES

[8] A. Signorini, A. M. Segre, and P. M. Polgreen. The use of twitter to track levels of disease activity and public concern in the u.s. during the influenza a h1n1 pandemic. PloS ONE, 2011.

\(^3\)http://uwak.elasticbeanstalk.com/noise