

# Learning Functional Compositions of Urban Spaces with Crowd-Augmented Travel Survey Data

Zack Zhu\*, Jing Yang\*, Chen Zhong†, Julia Seiter\*, Gerhard Tröster\*

\*Wearable Computing Lab  
ETH Zurich  
Zurich, Switzerland

†Centre for Advanced Spatial Analysis  
University College London  
London, United Kingdom

{zack.zhu, julia.seiter, troester}@ife.ee.ethz.ch, yangj@ethz.ch, c.zhong@ucl.ac.uk

## ABSTRACT

Regions in urban environments often afford a mixture of different utilities. Their identification allows urban planners to leverage important insights on the emerging functional dynamics of cities. With the increasing availability of human mobility data and other forms of online digital breadcrumbs, we can now characterize urban regions with multi-source features. In this work, we form a comprehensive view of urban regions by fusing features depicting their temporal, spatial, and demographic aspects. Aggregating 47K explicitly stated trip purposes into their respective destination regions, we obtain multi-dimensional ground-truths on the functionalities of urban spaces. Given fused features and training labels, we can perform supervised learning, via multi-output regression, to estimate the functional composition of urban spaces. With 14 functional dimensions, our approach using crowd-augmented travel survey predictors delivers a mean absolute error of 3.9, approximately half of the error resulting from a mean-based straw man approach (mean absolute error of 7.9). Clustering estimated regional functionalities, we find highly coherent cluster assignments (adjusted Rand Index of 0.81) compared to clustering directly on regional functionality labels. Finally, we provide an illustrative case-study where clustering of estimated region functionalities can be used to intuitively differentiate prototypical spatial neighbourhoods of a large metropolitan.

## Categories and Subject Descriptors

H.2.8 [Database Applications]: [Spatial Databases and GIS, Data Mining]

## Keywords

Urban Analytics, Travel Survey, Multi-Output Regression

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org. SIGSPATIAL'15, November 03 - 06, 2015, Bellevue, WA, USA. Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-3967-4/15/11\$15.00

DOI: <http://dx.doi.org/10.1145/2820783.2820832>

## 1. INTRODUCTION

Affordances of urban space dictate how residents conduct their day-to-day lives. To a great extent, much of people's travelling is motivated by a misalignment of their immediate needs and the affordances of their current space. From home, employees travel to places affording work while students travel to spaces affording education. Residents travel back to residential spaces to end the day. Understanding the inherent functionality of urban regions, or urban affordances, forms the basis for many urban planning applications, such as designing more accessible spaces or suitably distributing urban resources and amenities.

Recently, an increasing amount of research is beginning to assess urban functionality through implicit observations of human mobility. For example, the proliferation of GPS-enabled ubiquitous computing devices [10], transit records [23], or geo-tagged social media data [13], are all providing informative signals that hint at the functionality associated with fine-grained spatial regions. Compared to understanding functionalities through traditional methods such as urban infrastructure analysis or purpose zoning, these implicit observations yield a more natural representation as they directly measures how the city is utilized. In addition, they offer a dynamic view, allowing for longitudinal analysis. While a large quantity of data can indeed be obtained via implicit observations, ground-truth labels for why people travel (consequently, the affordance of a space) is often unavailable and difficult to infer. Therefore, objective and quantitative evaluations are difficult to conduct and results often require subjective interpretation.

In this paper, we leverage large-scale travel surveys to obtain proxy labels for regional functionalities. Often containing tens of thousands of trips with explicitly stated travel purposes (e.g. Shopping, Education, Go Home), substantial area within municipal regions are often covered by these trips. As a result, the explicitly stated purposes form a multi-dimensional functional vector for their destination region. These labels allow us to utilize supervised learning methodology to model functionality distributions from observable variables, such as temporal and demographic aspects of trips. To complement the *who* and *when* composition of an urban space, we also fuse in crowd-generated, semantic venue labels to complement with *what*.

Our contributions beyond state-of-the-art work are:

- A supervised approach using multi-output regression to capture the distribution and intensity of fine-grained

regional functionalities from fused travel survey and crowd-generated data.

- Objective evaluations and analysis of our approach in both estimating multi-dimensional functionalities and clustering of spatial regions.
- An illustrative case-study with prototypical neighbourhoods of Seattle to demonstrate our pipeline’s ability to grasp the dynamics of a city.

## 2. RELATED WORK

The use of human mobility to capture dynamics of urban spaces is currently an active area of research. Various approaches have been investigated, including mining of taxi-cab patterns [25, 15, 14], bike rental pick up/drop off locations [4], call detail records [17, 21], and web/social media streams [13, 5, 22, 11, 3].

Of these studies, many employ unsupervised approaches to extract significant mobility patterns, from which land usage patterns is subsequently interpreted by a handful of local-knowledge experts. For example, a recent work by Yuan et. al. [23] discovers urban functional zones by using a novel topic model-based approach to fuse together points of interests and mobility patterns extracted from taxi-cab and public transit data. However, without a substantial quantity of explicitly labelled regional purposes, objective evaluation and comparison of methodologies yield to subjective interpretations. Some studies have explored the use of municipal land zoning as ground-truth labels to enable supervised learning. For example, earlier work by Toole et. al. [21] classify patterns in mobile call detail records to determine land usage type as one of Residential, Commercial, Industrial, Parks, or Others. More recent work, such as [24], perform similar experiments with social media data as input signals. Compared to municipal zoning or other types of static single-purpose definition of urban spaces, our labels are derived from individual travel purposes. As such, we learn actual land usage as opposed to intended. In addition, single-purpose space usage is not assumed as our approach aggregates travel purposes to represent functionality as a vector of different purposes. To the best of our knowledge, this work is the first to employ a supervised approach to simultaneously learn multi-dimensional functional intensity of urban spaces by considering both dynamic human mobility patterns as well as static regional attributes. We are able to conduct supervised learning due to the availability of regional purpose labelling from over 10K residents.

Other researchers have also leveraged this rich data source. For example, Jiang et. al. [9] assesses the demographic profile of individuals based on their travelling patterns. Krumm and Rouhana [12] uses the 2006 Puget Sound Travel Survey and American Time Use Study datasets to predict the semantics of particular venues from people’s travel and demographic patterns. Recently, Kim et. al. [10] demonstrated the collection and utilization of travel survey data to predict people’s activities based on their travel location. Our work leverages travel survey data for a different purpose of urban space characterization. By aggregating trip instances into segmented spatial instances, we effectively model urban spaces.

## 3. DATA AND FEATURES

### 3.1 Data Source

In this work, our primary source of data comes from the Puget Sound Regional Council’s 2014 travel survey data (PSRC)<sup>1</sup>. Available for public download, this large-scale travel survey was conducted between April and mid-June 2014, where 6,000 households composed of ~10K people reported ~47K trips as part of their weekday travel patterns. The area surveyed composes of multiple counties in the Puget Sound region, including King, Kitsap, Pierce, and Snohomish counties. In addition to capturing detailed temporal and demographic aspects of trips and the participants, this dataset also provides the *purpose* of trips, as explicitly indicated by participants. Categorized as one of 16 purpose categories (e.g. Go Home, Shopping, Work, etc.) covering typical aspects of daily life, these labels provide a valuable perspective into the functionality of the destination region. We slightly reduce the number of categories to 14 by combining the categories of “Go to workplace” and “Go to other work-related place (e.g. meeting, delivery)” as well as “Go grocery shopping” and “Go to other shopping (e.g. mall, pet store)”.

We also collect crowd-generated data from the social media platforms Foursquare and Twitter to augment our characterization of the Puget Sound region. As the leading location-based social media platform in the United States, Foursquare allows users to construct and categorize points of interests as well as check into existing ones. Using the Venues API service from the platform<sup>2</sup>, we populate corresponding regions surveyed by the PSRC with 111,725 venues. The venues captured are categorized into 580 semantically meaningful categories, as defined by the Foursquare venue hierarchy<sup>3</sup>. From Twitter, we use the Streaming API<sup>4</sup> to capture geo-tagged tweets generated in the Puget Sound region. We stream for approximately three weeks between May and June of 2015. To boost the tweets we collect from the Streaming API, we use the author IDs of the collected tweets to back-search other geo-tagged tweets generated by these authors. In total, we leverage 383,140 tweets from 12,228 authors in the Puget Sound region.

### 3.2 Spatial Segmentation

Given the entire region, a necessary initial step is to partition for atomic units of analysis. A standard method is to employ grid-based partitioning to divide an area into equal-sized instances. However, such partitioning ignores the natural composition of the urban space for factors such as population density, geographical terrain, or urban artefacts with divisive properties (e.g. roads).

To address these issues, previous work of Yuan et. al. [23] proposes the use of road networks for a more natural partitioning of the urban space. In this work, we employ census cells to form atomic regions for analysis. Typically developed by national census departments, these cells represent the basis for most geographic reporting. In their construction, both physical and non-physical factors are considered. Considerations include streets, rail tracks, streams as well as property lines, municipal limits, etc<sup>5</sup>. Within a densely

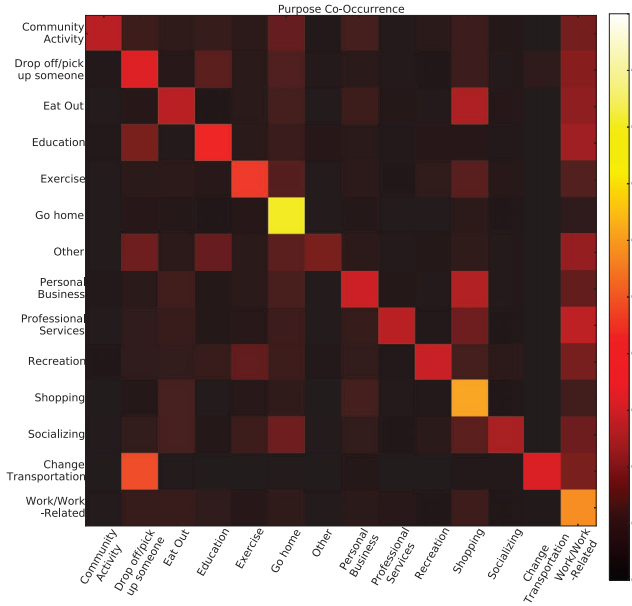
<sup>1</sup><http://www.psrc.org/data/transportation/travel-surveys/2014-household>

<sup>2</sup><https://developer.foursquare.com/overview/venues.html>

<sup>3</sup><https://developer.foursquare.com/categorytree>

<sup>4</sup><https://dev.twitter.com/streaming/overview>

<sup>5</sup><http://www.psrc.org/data/gis/shapefiles>



**Figure 1: Normalized co-occurrences of trip purposes within individual destination cells. The colour intensity of the off-diagonals represent the proportion of co-occurrence with another functionality within the same cell.**

populated metropolitan, these cells often represent one city-block. However, they may span larger areas to cover, for example, an entire conservation area. In addition to the wide range of factors considered in the construction of census cells, we also utilize census cells as atomic regions of analysis due to the obfuscation of trip coordinates in the PSRC dataset, where origin and destination coordinates are aligned to the centroids of census cells.

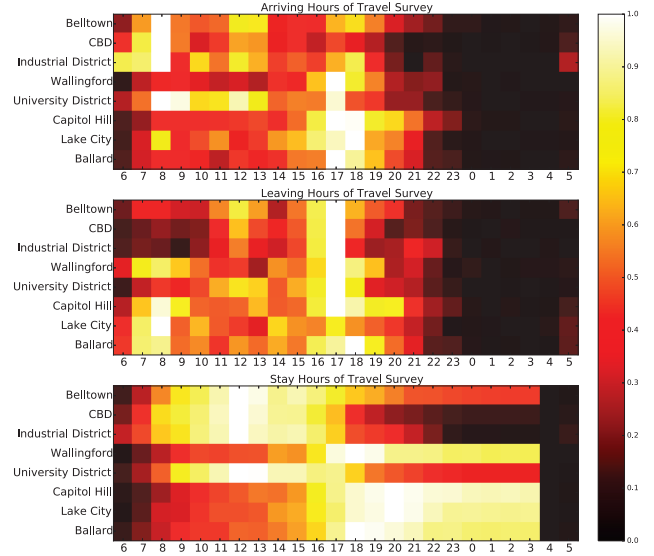
### 3.3 Co-occurrences of Functionalities

In Figure 1, we plot the co-occurrence rate of PSRC trip purposes aggregated into their respective destination cells. To better highlight the co-occurrence distribution, we conduct row-wise,  $l1$ -normalization. Of the 10610 cells containing at least one purpose, 64% contain one purpose while 36% contain multiple purposes.

Given the 14 categories of purposes, we see reasonable co-occurrences of functionalities. For example, Education has a relatively high co-occurrence rate with Drop Off/Pick Up Someone, likely arising from cells school children travel to with their parents. Similarly, it is reasonable to encounter co-occurrences of Eat Out and Shopping, for example, in shopping centres. Given the existence of numerous off-diagonal entries, we substantiate the need for multi-purpose learning to better reflect observed urban space usage.

### 3.4 Signals for Characterizing Regional Functionalities

In this section, we visualize the signals we collect and provide an intuitive basis for our feature extraction process. We select 8 characteristic regions from the city of Seattle (part of King county) as case-studies for illustrative purposes. These regions are briefly described as follows:



**Figure 2: Temporal distribution of trips from the PSRC dataset for 8 prototypical neighbourhoods of Seattle.**

***Belltown:*** The most densely populated area in Seattle, it is a city-core neighbourhood containing various residential and office high-rises. A variety of restaurants, nightclubs, and art galleries are also splattered in this neighbourhood.

***Central Business District:*** Typical CBD containing various office buildings, complemented with restaurants and other auxiliary amenities for day-time workers.

***Industrial District:*** The principal industrial area in Seattle, with heavy industry and railroad construction companies.

***University District:*** The main campus of University of Washington is here. This district also includes venues for cultural life and recreation.

***Wallingford:*** A classic residential suburb in Seattle, with shops, restaurants, bars and movie theatres to complement residential homes of middle and upper class families.

***Capitol Hill:*** A densely populated residential district in Seattle with nightlife and entertainment.

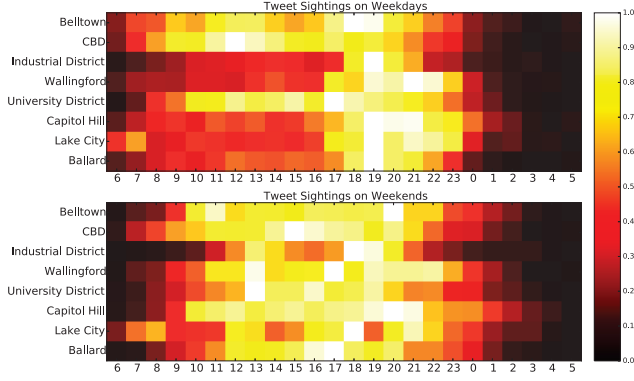
***Lake City:*** A relatively outlying suburban residential area containing some retail commerce.

***Ballard:*** A northwestern neighbourhood by the sea. It is a residential region with various restaurants, retails, markets, bars, etc.

#### 3.4.1 Temporal Characterization

From the PSRC travel survey, we can derive the arriving and leaving hours of trips aggregated for different census cells. In Figure 2, we plot a heatmap based on the counts of trips leaving, arriving, and staying in the different neighbourhoods over the different hours of the day. We conduct row-wise min-max normalization to scale data counts between  $[0, 1]$  for better visualization.

By chunking the data into hourly bins, we can spot intuitive temporal patterns. For example, Belltown and CBD areas have similar arriving and leaving time profiles: people tend to come for work around 8 in the morning, leave and arrive back around lunch time, and finally leave after 5pm.



**Figure 3: Temporal distribution of tweets. Counts of tweets sightings are scaled row-wise between  $[0, 1]$  for visualization purposes.**

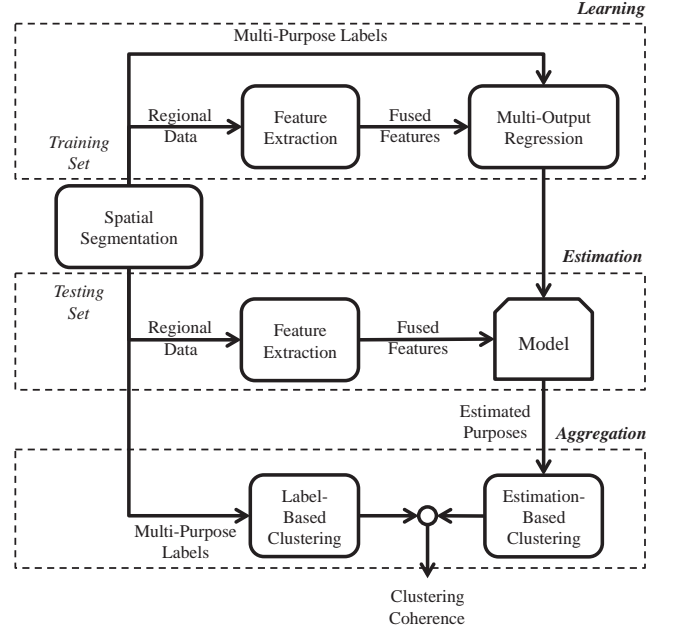
Comparing Belltown with CBD, people tend to stay later during the evening in Belltown, likely due to the restaurant and nightlife amenities available. These two neighbourhoods are markedly different than the suburban residential regions of Wallingford, Capitol Hill, Lake City, and Ballard. Examining the stay hours, we see they are typically shifted towards the evening hours, when people tend to be at home. Finally, the temporal profile of the University District is also quite intuitive, where the main hours of stay are during the day (likely for working and going to school), while some stay during the evening as there also exists a residential component in this neighbourhood.

Therefore, for the temporal profiling of a census block cell,  $C_i$ , we bin the hours of ingress and egress as  $T_{in}(i) = [t_0, t_1, \dots, t_h, \dots, t_{23}]$  and  $T_{out}(i) = [t_0, t_1, \dots, t_h, \dots, t_{23}]$ , respectively, where each element  $t_h$  represents the count of trips ingress-ing or egressing during hour  $h$ . From the time of entry and exit, we also derive the hours in which residents stayed in the region:  $T_{stay}(i) = [t_0, t_1, \dots, t_h, \dots, t_{23}]$  where, again,  $t_h$  is a count of the number of residents staying during the hour  $h$  in region  $i$ .

As an additional source of temporal characterization, our geo-tagged Twitter dataset allows us to extract the number of tweets generated for a given spatial-temporal context. In Figure 3, we plot the scaled quantity of hourly tweet sightings for the various neighbourhoods, during weekdays and weekends. For weekdays, we observe a similar pattern to that of stay hours captured from the travel survey, with the exception of the Industrial District. We observe higher variation between hours in the Twitter data compared to the features extracted from the travel survey. The weekend patterns are distinctively different, which is expected as the travel survey was conducted for weekdays only. Similar to the feature extraction method for travel survey data, we construct two temporal vectors for each cell  $i$  to reflect the raw quantity of tweet sightings:  $T_{weekend}(i) = [t_0, t_1, \dots, t_{23}]$  and  $T_{weekdays}(i) = [t_0, t_1, \dots, t_{23}]$ .

### 3.4.2 Spatial Attributes

We query the Foursquare platform to obtain the venues contained within the Puget Sound region. Using crowd-generated venue categories, we form a count distribution of venue types for each region:  $V(i) = [v_0, v_1, \dots, v_{579}]$ . These



**Figure 4: Conceptual framework for functional learning and aggregation of urban spaces. The methodological flow and associated evaluations are outlined in the three main phases of learning, estimation, and aggregation. Their details are discussed in Section 4.**

venues contain a wide range of illustrative venues, such as Residential Buildings, Office, Opera House, etc.

### 3.4.3 Demographic Signatures

Similar to temporal characterization, we aggregate the demographic profiles of users travelling to each  $C_i$  based on their age group and occupation. Following the categorization used in the PSRC dataset, the age dimensions composing the count vector  $D_{age}(i)$  are: *Under 5 years old*; *5-11*; *12-15*; *16-17*; *18-24*; *25-34*; *35-44*; *45-54*; *55-64*; *65-74*; *75-84*; and *85 or older*. The occupation count vector  $D_{occupation}(i)$  contains the dimensions: *Employed full-time (paid)*; *Employed part-time (paid)*; *Self-employed*; *Unpaid volunteer or intern*; *Homemaker*; *Retired*; and *Not currently employed*. For users without an employment specification, we count their presence using their reported school categories instead: *Daycare*; *Pre-school*; *K-12 public or private school*; *K-12 home school (full-time or part-time)*; *College, graduate, or professional school*; *Vocational/technical school*; *Other (school)*; and *None (school)*.

Using the aforementioned feature sets, we can concatenate the spatial, temporal, and demographic aspects to form a comprehensive view of atomic regions defined by census block-based segmentation of the Puget Sound region. In effect, we capture both *static* attributes of the physical region as well as *dynamic* resident-space interaction patterns.

## 4. MODELLING METHODOLOGY

In this section, we first describe our methodology for modelling the mixture of functions afforded by various atomic regions. Then, we describe a subsequent aggregation of these



regions to form spatial clusterings of similar functionalities. A conceptual overview of our methodology is presented in Figure 4. Section 4.1 corresponds to the learning and estimation sections of Figure 4 while Section 4.2 details the aggregation component.

## 4.1 Functionality Learning from Trip Purposes

Aggregating the trip purposes of a spatial region,  $C_i$ , we obtain the ground-truth functionality vector  $Y_i$ . From the spatial, temporal, and demographic features we extract (see Section 3), our aim is to learn a statistical model capable of accurately estimating an  $\hat{Y}_i$  given a fused feature vector  $X_i$  for  $C_i$ . In other words, we aim to learn the mapping from a multi-view characterization of a region to its multi-dimensional functional mixture. Furthermore, we also estimate the intensity of its functionalities to provide a more fine-grained understanding of usage dynamics.

Without the loss of generality, for a dataset containing  $N$  rows,  $L$  feature dimensions and  $K$  dimensions corresponding with different regional functionalities, we require the mapping  $f(X)$  such that the error term,  $\epsilon$  is minimized for unseen testing instances:

$$Y = f(X) + \epsilon \quad (1)$$

where  $Y$  and  $\epsilon$  are  $N \times K$  matrices and  $X$  is an  $N \times L$  feature matrix.

For such regression problems, a common method is to construct  $K$ , individual  $f_k(X)$  mappings to estimate functionalities independently. However, as urban functionalities may be highly correlated (e.g. Eat Out and Shopping), better testing accuracies and computational efficiency can be achieved by estimating a single model capable of simultaneously estimating the target dimensions [8]. One class of methods capable of naturally conducting the necessary multi-output regression is tree-based regression algorithms. For regression purposes, this family of nonparametric models recursively make partitions in the feature space to minimize sum of squares with respect to the training labels. With considerations for overfitting, the estimation at the final leaf nodes inherently contain distributions in the multi-dimensional space of the response variable. In our implementation, we leverage the ensemble method Extremely Randomized Trees (ERTs) by Geurts et. al. [6] as implemented in [16].

## 4.2 Functional Group Aggregation

Having estimated the functionality vectors of spatial instances, an aggregation step would be helpful for two purposes: to reduce the dimensionality for ease of interpretation and provide robustness when comparing noisy functionality estimations. In this aggregation step, we apply standard K-Means clustering algorithm to  $\hat{Y}$  and obtain  $D$  clusters of functionally similar regions. Later in Section 5.2, we show an illustrative application of differentiating spatially contiguous based on clustered regional functionalities.

One way to internally validate the results of cluster assignments is to use the Silhouette Index (SI) [19]. Ranging between -1 and 1, the SI measures how compact clusters are by computing the average intra-cluster and inter-cluster distances. As such, positive SI indicates compact clustering while loose clustering is indicated by negative SI. However, a more objective validation of the clustering results could be done with “ground-truth” cluster-assignment memberships.

As illustrated in Figure 4, the outcome of a clustering algorithm applied to *estimated* regional purposes should be as similar as possible to the clustering outcome of the same algorithm applied to the *labelled* regional purposes. Therefore, we can objectively validate our clustering results by calculating a clustering coherence measure between estimation-based clustering and label-based clustering. One commonly used metric to compare clustering results with an external clustering is the Rand Index [18]. Acting as a type of accuracy measure, the Rand Index penalizes assignments where a pair of data is assigned to different clusters by the model while sharing the same cluster assignments in ground-truth clustering. Similarly, pairs assigned to the same cluster by the model, but exist in different clusters in the ground-truth, are also marked as incorrect. In our experimentation, we employ the Adjusted Rand Index (ARI) measure to also discount for expected clustering coherence due to chance.

## 5. EXPERIMENTAL RESULTS

Given the methodology described above, we evaluate the performance of our approach on two fronts: first, the modelling of trip purposes afforded by atomic regions. Second, we examine the clustering coherence of functional zones derived from estimated purpose mixtures. We report results on the 7218 out of 10610 census cells containing features from all three data sources: 2014 PSRC travel survey, Foursquare and Twitter. Consequently, we are leveraging approximately 80% of the trip purpose labels from the travel survey.

### 5.1 Regional Purpose Estimation

#### 5.1.1 Model Fit and Interpretation

Having aggregated the various feature sets from the travel survey, Foursquare and Twitter, we fit a multi-output regression model to estimate  $y_i$  using our approach based on ERTs described in Section 4.1. To gauge the explanatory power of the fitted model for unseen data instances, we conduct 10-fold cross-validation. As such, the mean performance over the 10 testing folds are reported below.

To quantify the ability of our model to explain unseen instances, we calculate the coefficient of determination on unseen instances[7]. Intuitively, the coefficient of determination calculates the fraction of sum of squares of residuals over the total sum of squares. To express the best model as 1 and worst as 0, the coefficient of determination ( $R^2$ ) is 1 subtract this fraction. Adapting this concept to our multi-output case, we calculate  $R^2$  as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N_{test}} \left( \sum_{k=1}^K (\hat{y}_k^i - y_k^i)^2 \right)}{\sum_{i=1}^{N_{test}} \left( \sum_{k=1}^K (y_k^i - \bar{y}_k)^2 \right)} \quad (2)$$

where  $N_{test}$  is the number of testing instances unseen in model training,  $K$  is the number of purpose dimensions for each instance,  $\hat{y}_k^i$  is the model estimation of purpose  $k$  for  $C_i$ . Corresponding to the PSRC travel purpose labels,  $y_k^i$  and  $\bar{y}_k$  are count of travel purposes in category  $k$  for region  $i$  and mean of purpose  $k$  over all testing instances, respectively.

Clearly, a straw man model where the estimations are simply the mean of each purpose dimension  $k$ , the  $R^2$  is to be interpreted as 0. In Table 1, we tabulate the testing  $R^2$  scores achieved with different combinations of the feature sets: travel survey demographic features (TrD), travel

	MAE	aRMSE	R <sup>2</sup> Coefficient
Mean Model	7.900	1.204	0.000
TrT	4.378	0.827	0.620
TrT + TrD	4.137	0.788	0.660
TrT + FsS	4.095	0.800	0.648
TrT + TwT	4.426	0.829	0.614
TrT + TrD + FsS	<b>3.937</b>	0.774	<b>0.675</b>
TrT + TrD + TwT	4.165	0.796	0.655
TrT + TrD + FsS + TwT	3.984	<b>0.773</b>	0.672

Table 1: Testing errors obtained with different feature sets: travel survey demographic features (TrD), travel survey temporal features (TrT), Foursquare spatial features (FsF), and Twitter temporal features (TwT). A straw man approach of using the mean of the dataset is used for comparison. All approaches significantly outperform the straw man approach while the combination of travel survey and Foursquare features delivers the best performance.

survey temporal features (TrT), Foursquare spatial features (FsF), and Twitter temporal features (TwT). The multi-feature set models are trained by concatenating the individual feature sets. Using the travel survey data alone, with features describing the temporal ingress/egress and stay times of people along with their demographic profiles, our model achieves a mean  $R^2$  of 0.6598. Coupling these features with the Foursquare venue features describing the spatial aspect of regions, we obtain an increased mean  $R^2$  value of 0.6750. However, with the further addition of tweet sighting features, the mean  $R^2$  value remains similar at 0.6721.

Aside from gauging the testing  $R^2$  of our model, our supervised approach also allows us to gauge the importances of features from the different feature sets in correctly estimating regional purposes. We conduct feature importance analysis by fitting ERTs on the entire dataset with all feature sets and rank according to their Gini importances [2]. In Figure 5, we plot the distribution of the ranks from most important (rank of 0) to least important (rank of 869) on the x-axis. On the y-axis, we plot the distribution of ranks obtained for each grouping. From the distributions, we can clearly see the dominating importance of travel survey features. However, contributions from Foursquare features, capturing the points of interests in regions, are also seen in the top 100 most important features. On the other hand, the most informative tweet sighting features contribute only in the 100-200 ranks.

From the quantification via the  $R^2$  and the feature importance characterization, we see evidence that the travel survey data can be successfully complemented by Foursquare venues in estimating purposes of regions. In addition to the dynamic patterns of people-space interaction, static spatial characteristics are useful in explaining the variance of unseen data instances.

### 5.1.2 Quantification of Modelling Error

We use standard mean absolute error (MAE) to quantify the deviation of model estimates and labelled regional

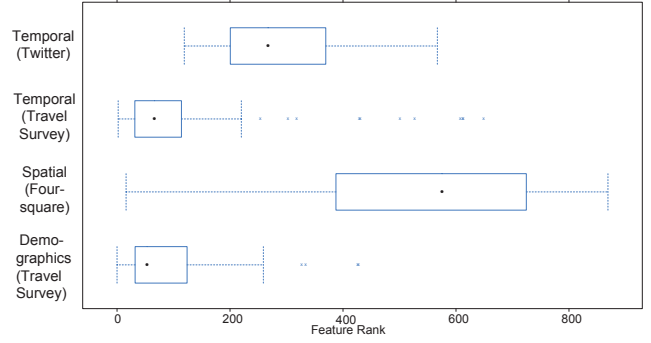


Figure 5: The importance of each feature is calculated according to its Gini importance [2] during the fitting of the model. We rank and plot the distribution of the ranks for each feature category (rank 0 is most important). The medians of the distributions are marked with the black dot while the boxes and whiskers indicate the interquartile range and 1.5 times the interquartile range of the distribution, respectively.

functionalities given by the PSRC dataset. Furthermore, we evaluate the average root mean squared error (aRMSE) to demonstrate the sample standard deviation of how predicted functional intensities differ with travel survey labels. Given our multi-output regression task, we adapt these two error metrics, similar to [1], as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{N_{test}} \left( \sum_{k=1}^K |\hat{y}_k^i - y_k^i| \right) \quad (3)$$

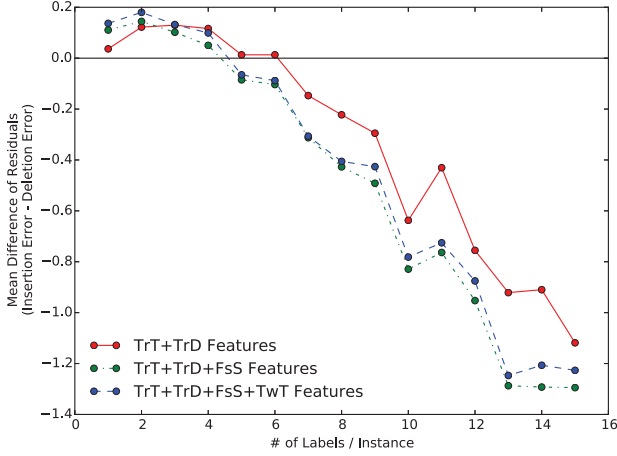
$$aRMSE = \frac{1}{K} \sum_{k=1}^K \sqrt{\frac{\sum_i^{N_{test}} (\hat{y}_k^i - y_k^i)^2}{N_{test}}} \quad (4)$$

From Table 1, we note the best performing combination of features utilizes the travel survey demographic and temporal features, coupled with Foursquare spatial features. On the other hand, the straw man approach of using the mean as an estimate results in almost double MAE. To assess the significance of the different approaches, we conduct the two-sample Kolmogorov-Smirnov test[20] on the distribution of residuals. Compared to utilizing travel survey data without augmentation, the residual distribution of the other approaches are all statistically significant ( $p < 0.01$ ). We also note that  $TrT+TrD+FsS$  results in significantly different residuals than  $TrT+TrD+TwT$  ( $p < 0.01$ ). However, the residuals of  $TrT+TrD+FsS+TwT$  and  $TrT+TrD+FsS$  are not significantly different ( $p = 0.036$ ).

These results imply two things: first, we obtain superior testing performance when augmenting travel survey with Foursquare spatial features. Second, the addition of temporal features extracted from Twitter delivers inferior performance and the addition of these features do not make a significant difference when Foursquare spatial features are already added to the travel survey features.

### 5.1.3 Regression Error Analysis

Although the previous two subsections indicate that features resulting from tweet sighting times are not particularly



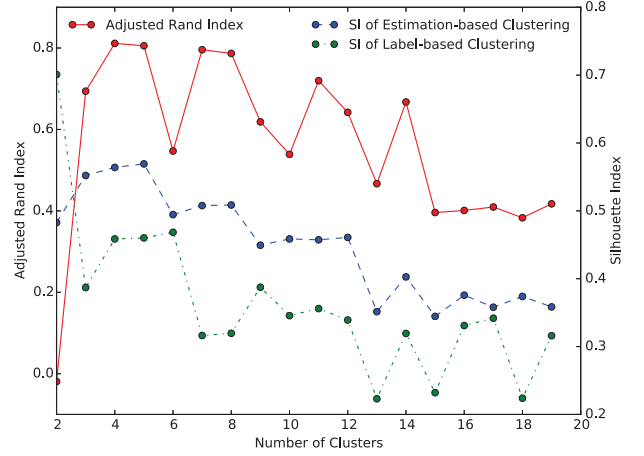
**Figure 6: Grouping instances by the number of labels per instance, we plot the average difference of deletion errors from insertion errors on the y-axis. As the number of labels increases, insertion errors are less dominant for regions with more labelling.**

useful for explaining variances in regional purposes or estimating them, we must point out that the labels are merely a sampling of the entire population of purposes. While approximately 47K trips are recorded in the PSRC dataset, the associated trip purposes are only a subset of all trip purposes afforded by each region. Here, we conduct a closer examination of the testing residuals to better understand the estimation outputs.

Two forms of residuals exist when comparing the model output to the trip purposes labelled in the PSRC dataset: insertion errors and deletion errors. We refer to insertion errors as cases where the model estimates existence of purposes not contained in the PSRC labels. Contrarily, deletion errors are where our model underestimates the existence of purposes stated in the PSRC labels.

We can assess the contribution of the two types of residuals by calculating the residual mass difference arising from insertion errors and deletion errors. To do so, we first determine the instance-specific sum of  $r_i^{pos}$  and  $r_i^{neg}$  from  $r_i = \hat{y}_i - y_i$ , where  $r_i^{pos}$  indicates total residual for insertion errors and  $r_i^{neg}$  indicates total residual arising from deletion errors. Then their difference can be calculated as  $r_i^\Delta = r_i^{pos} - r_i^{neg}$ . Therefore, an  $r_i^\Delta \approx 0$  indicates similar error contribution from insertion errors and deletion errors. A positive  $r_i^\Delta$  implies stronger insertion error while it is negative when a stronger deletion error component is present.

In Figure 6, we plot the mean of error mass differences grouped by the number of PSRC trip labels per instance. For all three approaches leveraging different feature sets, we notice a common decline in the error mass difference as instances contain an increasing number of PSRC trip labels. In other words, insertion errors are less prevalent than deletion errors when a higher number of labels are provided per instance. One possible explanation for this phenomenon is that our model is capable of identifying distinct patterns for multiple regional purposes. However, as the PSRC travel survey dataset does not capture all the activities afforded within a region, our model may be inappropriately penalized



**Figure 7: Comparing K-Means clustering on estimated regional purposes and labelled regional purposes, we quantify the performance of our approach with the Adjusted Rand Index (ARI). Plotting the resultant ARI on the left y-axis, a peak is found at  $D = 4$ , where the ARI is 0.81. On the right y-axis, we plot the Silhouette Index (SI) [19] of the two clusterings, quantifying the compactness of the resultant cluster shapes.**

for correctly overestimating beyond the labels provided. We discuss this point further in the discussion section.

## 5.2 Aggregating Estimated Functionalities

To leverage functionalities as input for understanding dynamics of urban neighbourhoods, the estimated functionalities need to be summarized for better interpretation and to provide robustness against estimation errors. Below, we show results for this aggregation step using  $TrT + TrD + FsS + TwT$  regression estimates as input.

### 5.2.1 Clustering Output Coherence

In Figure 7, we plot the ARI obtained for different numbers of clusters, indicated by  $D$ . In addition to the ARI measured by the left y-axis, we also plot internal validation for the compactness of clusters on the right y-axis. Our results indicate that the ARI peaks at 0.81 for  $D = 4$ . In Figure 8, we compare the clustering memberships of census cells in three characteristic regions in Seattle. Comparing the estimation-based clustering and label-based clustering, we notice the majority of cells match in their cluster assignments.

Also at  $D = 4$ , the SI achieved for estimation-based clustering is also near its peak at 0.56, which indicates a reasonably compact clustering. Interestingly, the SI of estimation-based clustering is almost always higher than that of label-based clustering, even with high ARI values indicating similar groupings. This shows that, while cluster assignment stays relatively similar, more compact clusters are achieved when grouping based on estimated regional purposes.

### 5.2.2 Semantic Interpretation of Clusters

By clustering from a semantically meaningful basis of functionalities, we can provide concrete interpretations of the





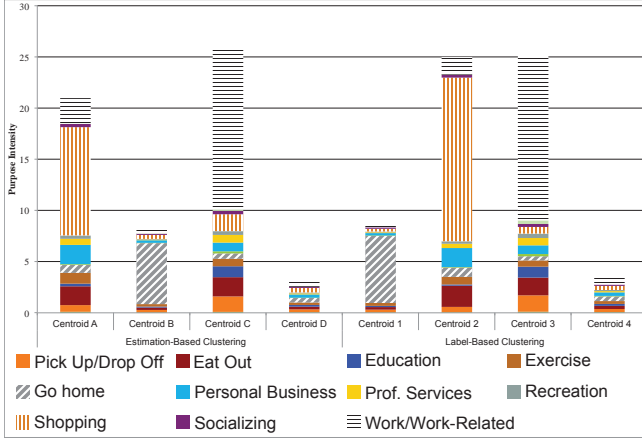
**Figure 8: Cluster membership are highly similar between estimation-based and label-based clusterings. Three prototypical neighbourhoods of Seattle are illustrated. The colour scheme corresponds to the cluster centroids depicted in Figure 9 as follows: Blue - mixture of various purposes with weak indications (Centroid D), Red - predominantly residential areas (Centroid B), Purple - predominantly working (Centroid C), and Green - Shopping and Going Out (Centroid A).**

clusters obtained. In Figure 9, we plot the composition of the inferred cluster centroids, with the 14 regional functionalities as basis. For each centroid, we plot the raw intensity of regional purposes as stacked bars. The legend selectively plots 11 of the 14 purposes due to space constraints. As can be expected from the high ARI, the centroids found between the estimation-based clustering and label-based clustering have a one-to-one correspondence and demonstrate high similarities between corresponding centroids (e.g. Centroid C and Centroid 3).

Examining the centroids from estimation-based clustering, there exists three centroids with predominant functionalities (Centroids A, B, and C) while Centroid D is low in intensity and contain a relatively uniform distribution of functionalities. The corresponding cluster, Cluster D (blue in Figure 8), captures regions where the model makes an unclear estimation of purposes, which is likely due to weakly

indicative data collected for those cells. In Cluster A (green in Figure 8), a strong functional component is Shopping, coupled with the presence of Working and Personal Business with less intensity. From Figure 8, these cells are found more frequently in the downtown area. Upon examining the individual cells, they contain landmark sites such as Pacific Place, an upscale shopping centre in downtown Seattle, Broadway Market in Capitol Hill, or University Village Shopping Center in the University District. In Capitol Hill and the surrounding areas of the University of Washington campus, we notice a prevalent number of cells belong to Cluster B (Figure 8, red). From the centroid composition and the spatial distribution of Cluster B cells, we can easily identify these as residential areas with suitable amenities, such as Shopping. Finally, Centroid C captures many office buildings in downtown Seattle. Dominated by Working, there also exists elements of Eat Out and Shopping, charac-





**Figure 9: The composition of the cluster centroids are plotted as stacked bar graphs. A one-to-one correspondence between the centroids of estimation-based clustering and label-based clustering can be seen.**

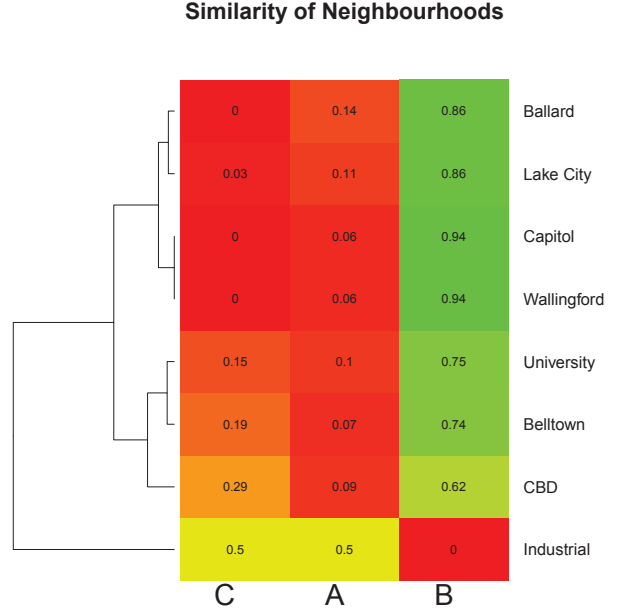
teristic of cells in multi-functional downtown regions.

In Figure 10, we plot the distribution of cell memberships of all 8 prototypical neighbourhoods we described initially at the beginning of this paper. Using the three indicative clusters (Clusters C, A, and B) described above, we see an intuitive semantic grouping of these neighbourhoods via the dendrogram calculated from the normalized cluster distributions. The four predominantly residential neighbourhoods (Capitol, Wallingford, Lake City, and Ballard) are highly similar to each other with a large proportion of Cluster B cells. Although Belltown and University District also contain significant Cluster B elements, the mixed-in Cluster C and A cells give these neighbourhoods stronger work and shopping/entertainment aspects. Belltown and University District are further grouped with the CBD district, which trades off residential elements for more working regions. Finally, a distinctively separate district from these previous two groupings is the Industrial District, without any Cluster B elements.

## 6. DISCUSSION & FUTURE WORK

From our results, we illustrate that trip purposes from travel surveys can be used as an important source of ground-truth labels to define functionalities afforded by specific urban spaces. Moreover, the availability of this ground-truth opens up the possibility to conduct *objective* quantitative model evaluation and interpretation of feature importances. As the dataset used in our work covers multiple cities in the Puget Sound area, we believe our model is capable of generalization to other cities as long as similar travel survey samples and social media data exist. One aspect of our future work will explore generalization in two aspects: first, how well do models generalize between cities (e.g. train in Seattle and test in Chicago). Second, we will investigate the use of semi-supervised methods to propagate accurate but sparse labels from travel surveys to unlabelled but abundant instances of social media data.

Our analysis also demonstrates that travel surveys only represent a sampled representation of all functionalities. As shown in Figure 6, we see evidence that the model does not overestimate as much when more labels are present per in-



**Figure 10: The distribution of (row-wise) normalized counts of cluster members for each neighbourhood. From the dendrograms, an intuitive semantic grouping of the neighbourhoods emerge: residential areas (Capitol, Wallingford, Lake City, and Ballard), working districts (Belltown, University District, and CBD), and the Industrial zone.**

stance. Partially due to the large regional coverage of travel surveys, comprehensive coverage of regional functionality labelling is not always possible. Therefore, another aspect of our future work will investigate amending this limitation in two ways: first, modifications of the current modelling approach and corresponding metrics to handle the partial-label problem. One possibility to mitigate the lack of labels is to conduct spatial smoothing to gain additional labelling from spatially near cells. Second, we will also further examine the possibility of obtaining additional labels from geo-tagged, crowd-generated data. One approach is to conduct textual content analysis (e.g. Foursquare venue tips) to extract topics to infer implicit urban functionalities.

## 7. CONCLUSION

In this work, we demonstrate the usefulness of travel survey data in providing both *signals*, indicating human-space interactions, as well as *labels*, indicating the functionality of a region. With this functional ground-truth, supervised methods can then be applied to learn the mapping from temporal, spatial, and demographic aspects of a region to the mixture of functionalities that it affords. Utilizing multi-output regression, we demonstrate that our approach strongly outperforms a mean-based straw man approach in terms of estimation error while the augmentation of travel survey features with crowd-generated Foursquare data delivers a statistically significant boost in performance. For more interpretability and robustness to estimation errors, we cluster spatial regions based on model-estimated functionalities to better grasp urban dynamics. Comparing estimation-based

clustering and clustering directly on ground-truth functionalities, we see a high level of clustering coherence (ARI of 0.81). Based on these findings, we argue our work provides an effective and novel approach for urban planners to gain insight on the functional composition of their cities.

## 8. ACKNOWLEDGEMENTS

We thank Ulf Blanke, Alberto Calatroni, and Yu Jiang for their insightful and on-going discussions. We also thank Neil Kilgren from the Puget Sound Regional Council for providing us with the travel survey data. This work is partially supported by the Hasler Foundation (SmartDAYs project).

## 9. REFERENCES

- [1] H. Borchani, G. Varando, C. Bielza, and P. Larrañaga. A survey on multi-output regression. *WIREs Data Mining and Knowledge Discovery*, 2015.
- [2] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [3] P. Brindley, J. Goulding, and M. L. Wilson. A data driven approach to mapping urban neighbourhoods. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '14, pages 437–440, New York, NY, USA, 2014. ACM.
- [4] C. Coffey and A. Pozdnoukhov. Temporal decomposition and semantic enrichment of mobility flows. In *Proceedings of the 6th ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, LBSN '13, pages 34–43, New York, NY, USA, 2013. ACM.
- [5] V. Frias-Martinez, V. Soto, H. Hohwald, and E. Frias-Martinez. Characterizing urban landscapes using geolocated tweets. In *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust*, SOCIALCOM-PASSAT '12, pages 239–248, Washington, DC, USA, 2012. IEEE Computer Society.
- [6] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.
- [7] S. Glantz and B. Slinker. *Primer of Applied Regression & Analysis of Variance*. McGraw-Hill Education, 2000.
- [8] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.
- [9] S. Jiang, J. Ferreira, and M. González. Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery*, 25(3):478–510, 2012.
- [10] Y. Kim, F. C. Pereira, F. Zhao, A. Ghorpade, P. C. Zegras, and M. E. Ben-Akiva. Activity recognition for a smartphone based travel survey based on cross-user history data. In *22nd International Conference on Pattern Recognition, ICPR 2014, Stockholm, Sweden, August 24–28, 2014*, pages 432–437, 2014.
- [11] F. Kling and A. Pozdnoukhov. When a city tells a story: Urban topic analysis. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '12, pages 482–485, New York, NY, USA, 2012. ACM.
- [12] J. Krumm and D. Rouhana. Placer: Semantic place labels from diary data. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 163–172, New York, NY, USA, 2013. ACM.
- [13] R. Lee, S. Wakamiya, and K. Sumiya. Urban area characterization based on crowd behavioral lifelogs over twitter. *Personal and ubiquitous computing*, 17(4):605–620, 2013.
- [14] X. Liu, L. Gong, Y. Gong, and Y. Liu. Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography*, 43(0):78 – 90, 2015.
- [15] G. Pan, G. Qi, Z. Wu, D. Zhang, and S. Li. Land-use classification using taxi gps traces. *Intelligent Transportation Systems, IEEE Transactions on*, 14(1):113–123, 2013.
- [16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [17] T. Pei, S. Sobolevsky, C. Ratti, S.-L. Shaw, T. Li, and C. Zhou. A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science*, 28(9):1988–2007, 2014.
- [18] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.
- [19] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- [20] N. Smirnov. Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Statist.*, 19(2):279–281, 06 1948.
- [21] J. L. Toole, M. Ulm, M. C. González, and D. Bauer. Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, UrbComp '12, pages 1–8, New York, NY, USA, 2012. ACM.
- [22] S. Wakamiya, R. Lee, and K. Sumiya. Urban area characterization based on semantics of crowd activities in twitter. In *GeoSpatial Semantics*, pages 108–123. Springer, 2011.
- [23] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong. Discovering urban functional zones using latent activity trajectories. *Knowledge and Data Engineering, IEEE Transactions on*, 27(3):712–725, 2015.
- [24] X. Zhan, S. Ukkusuri, and F. Zhu. Inferring urban land use using large-scale social media check-in data. *Networks and Spatial Economics*, 14(3-4):647–667, 2014.
- [25] W. Zhang, S. Li, and G. Pan. Mining the semantics of origin-destination flows using taxi traces. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 943–949, New York, NY, USA, 2012. ACM.