Extracting Latent Semantic Relation from Multi-modality Sensor Data in Different Cities[*]

Mingcheng Chen[†]
515030910568
Shanghai Jiao Tong University
800 Dongchuan Road,
Minhang District, Shanghai
3505251997
1211001x@sjtu.edu.cn

Chongyao Xia[‡]
5140309546
Shanghai Jiao Tong University
800 Dongchuan Road,
Minhang District, Shanghai
5140309546xcy
996@sjtu.edu.cn

Xintian Yu[§]
515030910583
Shanghai Jiao Tong University
800 Dongchuan Road,
Minhang District, Shanghai
yxt2012123@sjtu.edu.cn

ABSTRACT

The rapid development of sensor technology brought us a surge of multi-modality data, however the label scarcity and data insufficiency problem become more and more serious as the booming of big raw data. In urban computing, how to make use of multi-modality sensor data in one city to infer and predict the data in another city which suffered from lack of label and data insufficiency is always a hot topic. We proposed a novel method based on previous research using methods of transfer learning, aim to find out the relation between dataset, that is extracting latent semantic relation from multi-modality sensor data in different cities. We learns semantically related dictionaries for multiple modalities from a source domain, and simultaneously transfers the dictionaries and labelled instances from the source into a target domain. We evaluate the proposed method with a case study of air quality prediction.

Keywords

Relation Extraction; Urban computing; Sensor Data; Multi-modality; Transfer Learning

1. INTRODUCTION

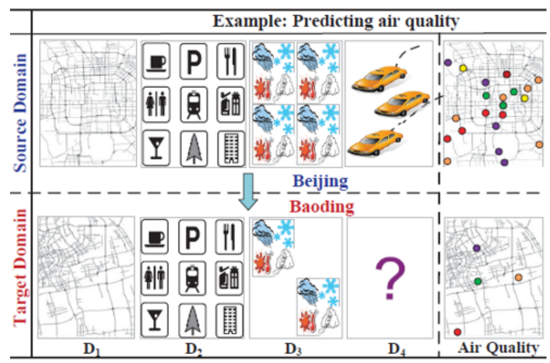


Figure 1: Multi-modality sensor data in prediction of air quality.

In the era of big data, a varieties of sensors have become an important way for people to access data and real-time information. Thanks to these sensor data, emerging branch of computer science like urban computing have a significant development. However, the explosive growth of multi-modality sensor information also brought many new problems, among those problems, the following twos are the most challenging:

1) The contraction between *Big Data* and *Label Scarcity*.

2) *Data Insufficiency* when using multi-modality sensor data solving specific issues.

The problem 1) of contraction between big data and label scarcity reflects the booming of sensor data increases the demand for labels. In the field of urban computing, although we can get a surge of sensor data of kinds of modalities (such as GPS, Acceleromete, Gyroscope, Magnetometer, etc), those raw data are usually primitive and few of the sensor data can be correctly labeled. What's more, labeling data is a time-consuming and expensive process, until now there is no effective method to label the multi-modality sensor data in urban computing. The other problem 2) of data insufficiency is usually caused by sensor services are not ready or fundamental infrastructures are just built.

In this paper, we will take air quality prediction as background to discuss the two main problems of multi-modality sensor data mentioned above. As shown in figure 1, meteorological sensor data and taxi trajectories data in Baoding suffers from both label scarcity and data insufficiency. It

[*]Multi-modal data refers to data collected through different methods or perspectives for a thing to be described. We call each method or perspective that collects a set of data a *Modality*

[†]Mingcheng Chen completed main part of surveys and determined the topic of project as team leader. He participated in algorithm design and coding.

[‡]Chongyao Xia participated in algorithm design and implemented the algorithm. He drew charts for experiment results.

[§]Xintian Yu is responsible for testing the code and optimizing parameters for models. He did most of debugging work.

is critical to make use of known multi-modality sensor data in Beijing to predict unknown label in Baoding and complement insufficiency data. To achieve this goal, the key is extracting latent semantic relation from multi-modality sensor data in different cities, so that we can transfer relation we learned from one city to another and use such relation dealing with both the label scarcity and data insufficiency problem.

We analyzed existing work on data mining and transfer learning which using multi-task, multi-model and multi-view [3] then chose one of the feasible method, improved it to adjust our problem and application background with theoretical support of transfer learning. Furthermore, experiment was carried on over the datasets download from Microsoft Urban Computing website [1].

We evaluate our method on air quality prediction in three cities, with performances outperforming existing methods.

2. RELATED WORK

In this section, we firstly introduce what multi-modality data is, then describes the basic principles of urban computing and transfer learning. At the end briefly review the related work of representative research on multimodal data fusion, and state-of-the-art transfer learning methods.

2.1 Multi-modality Data Mining

Multi-modal data refers to data collected through different methods or perspectives for a thing to be described. We call each method or perspective that collects a set of data a *Modality*.

For example, in multi-modality face recognition, multi-modal data may be composed of two modalities of a human face 2D image and a 3D shape model. In multi-modality video processing, video may be decomposed into modalities such as audio, video, and images. Texts and pictures on web pages can also be viewed as different modalities that they describe the information of web pages from different perspectives.

Here we focus on urban multi-modality data in this paper. The multi-modality sensor data for air quality prediction mainly composed of four parts:

1) Road Network Data

2) Point-of-Interests(POI) Data

3) Meteorological Data

4) Taxi Trajectories Data

Detailed structures of these four modalities of sensor data will be further described in the section 4 of **EXPERIMENT**.

2.2 Urban Computing

Urban computing is a process of acquisition, integration, and analysis of big and heterogeneous data generated by a diversity of sources in urban spaces, such as sensors, devices, vehicles, buildings, and human, to tackle the major issues that cities face, e.g. air pollution, increased energy consumption and traffic congestion.[10] Urban computing

computing.png

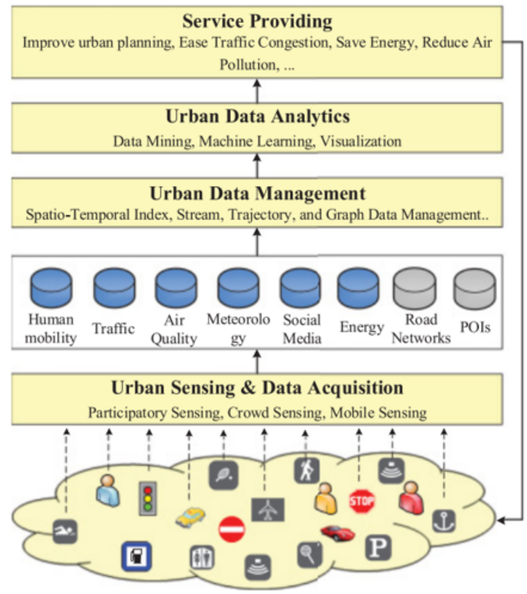


Figure 2: Urban computing framework.

connects unobtrusive and ubiquitous sensing technologies, advanced data management and analytics models, and novel visualization methods, to create win-win-win solutions that improve urban environment, human life quality, and city operation systems. Urban computing also helps us understand the nature of urban phenomena and even predict the future of cities.

Figure 2 shows the basic framework of urban computing, which consist of four layers. Our work in this paper is focus on the second layer *Urban Data Analytics*. And the *Urban Data* we used have the character of Multi-source, label scarcity and data insufficiency.

2.3 Transfer Learning

Transfer learning refers to taking use of data, tasks, or similarities between models aim to apply a model learned in the old field to a learning process in a new field.[4]

Transfer learning can help dealing with the contraction between big data and label scarcity. We are in an era of big data, every day, social networks, intelligent transportation, video surveillance, and industry logistics generate huge amounts of data such as images, texts, and voices. With the increase of data, machine learning and new deep learning models more and more rely on such a large amount of data. Continuously training and updating the corresponding models makes the performance of the model better and better, and more and more suitable for the application of specific scenes. However, these big data have brought serious problems: there is always a lack of perfect data annotation.

As mentioned above, the training and updating of machine learning models depend on the annotation of data. However, despite the fact that we have access to vast amounts of data, these data are often very primitive in their original form, and few of the data have been properly labeled manually. Labeling data is a time-consuming and expensive operation. So far, there is no effective way to solve this problem, which presents challenges for machine learning and

[1]Download from Urban Computing on Microsoft Research homepage https://www.microsoft.com/en-us/research/project/urban-computing/

learning.png

Figure 3: Sample of Transfer Learning.

deep learning model training and updates. Conversely, data insufficiency in specific areas also result in the development in these field slow down.

When it comes to the urban multi-modality sensor datasets for air quality prediction in this paper, problems of *Label Scarcity* and *Data Insufficiency* become key challenges. We use the method of transfer learning framework FLORA

12345

dealing with this two challenges and improve it to fit our requirement.

2.4 Review of Existing Work

We did a long term survey and read some outstanding papers in related fields. The following Table is an overview of related work on representative research on multimodal data fusion.

Co-training model has the characteristics of generating models for each data modality and unifying these models' outputs as the final result. However the shortcoming is it introduces overfitting and ignores non-linear interactions between modalities.

MAMUDA and HiMLS are good models, but cannot fully handle the data insufficiency problem; Aligning embrace translation method, Canonical correlation analysis and Matrix factorization method devote to extract a semantic latent subspace or build a translator to align different modalities, but they require sufficient data in each modality and abundant correspondence between instances across modalities; popular $IteM^2$ model can only tackle non-negative feature values; FLORAL model is the only one can handle both label scarcity and data insufficiency problems well, while FLORAL also exists problems like semantically related dictionaries generated are not optimized.

3. OUR PROPOSAL

Based on the characteristics of our dataset and motivation for extracting latent semantic relation from multi-modality sensor data, we finally choose the existing model named FLORA[], carry out our work on the basis of it. We improve the theoretical model to suit for specific issues of multi-modality sensor data air prediction, inherit its clustering method and the *Adaboos* transfer learning model for learning relation between datasets in different cities. We redefined the feature extraction process of the source domain data, and optimized the process of enriched representation.

3.1 Overview

The total process for extracting latent semantic relation from multi-modality sensr data in different cities by transfer learning method could be broke down into three main parts:

1) Data preprocess to regularly define data and labels.

2) Learn semantically related dictionaries from source domain.

3) Transfer dictionaries and instances for target domain prediction.

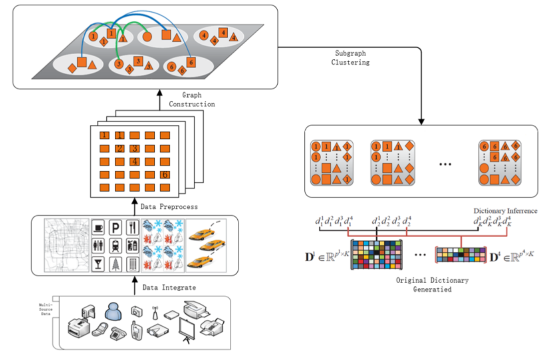


Figure 4: Learning Semantically Related Dictionaries from Source Domain.

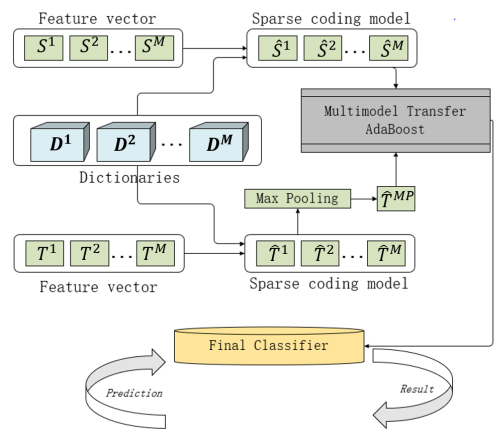


Figure 5: Transfer Dictionaries and Instances for Target Domain Prediction.

Sparse coding is a technique widely used in machine learning, represents data vectors as sparse linear combinations of basis elements. The set of basis elements is called dictionary. Sparse coding provides an effective way to homogenize representation structures of multi-modalities, by enforcing all modalities dictionaries semantically related and learning linear combination coefficients over the corresponding dictionary for each modality as new representations. There are three main categories of techniques to learn dictionaries: 1) probabilistic learning, 2) reconstruction error minimization, and 3) clustering. Here we prefer clustering because of its advantage in extracting semantically related dictionaries. [9]

When we learning semantically related dictionaries from source domain, we use method of graph clustering. After data preprocess, we build a weighted graph to model pairwise similarities between vertices across different modalities and within each modality, run graph clustering algorithm on it and get K subgraphs. Then take represent elements from each subgraph so that we can derive K dimension of dictionaries with each dimension containing 4 modalities of vectors.

As shown in figure 5, after learned the dictinoaries which contain the latent relation between sensor data in different cities, we transfer dictionaries and instances to generate the prediction model. We learn M semantically related dictionaries from the source domain to unlock the power of sparse coding in homogenizing different modalities as stated before. More importantly, we transfer the M semantically related dictionaries to the target domain, and apply them to learn enriched representations of target instances, in order to address the data insufficiency problem.

3.2 Data Preprocess

Data preprocess is important for raw sensor data, firstly we regularize the raw data in time and space.

Figure 6: Partitioning each city into grid regions.

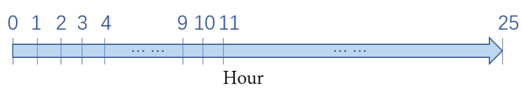


Figure 7: Partitioning each day by hours instance.

Spatially, we partitioned each city into grid regions in the size of $1.5km * 1.5km$ as shown in the figure 6. We take such a method to tackle the raw data for the reason that image dictionary learning usually processed images in the same way, e.g. 16*16 pixel patch [8].

On the term of time, we characterize a grid region in an hour of a day as an instance, as shown in figure 7, which is consistent with another air-quality predict project named U-air [11].

The key idea is the chosen of data label, which is to be predicted. For this project, we select Air Quality Index (AQI) as our data label, we labeled all valid data in grids mentioned above into 6 levels 1.Good, 2.Moderate, 3.Unhealthy for sensitive groups, 4.Unhealthy , 5.Very unhealthy and 6.Hazardous by AQI according to rules in table 1.

3.3 Instance Clustering

For the purpose of graph clustering, we must first define edges, which describes the relation of different instances. Figure 8 shows the process of instance clustering. In the figure, different shapes represent different modalities, while the eclipses enclosing shapes denote instances. The eclipses with numbered shapes are labelled instances.

There are two kinds of edges were added into the set of data instances 1) Intra-edge: edges between instances of the same modality and 2) Inter-edge: edges between instances of different modality. The weight of added edges are defined as follows. 1) The i_{th} and j_{th} vertices of m_{th} modality are connected with an intra-edge weighted by *Gaussian Kernel*: top Knn edges would be added to the graph. The weight is given by

$$w_{ij}^m = \exp - \frac{||s_i^m - s_j^m||^2}{2\delta^2} \tag{1}$$

2) As for the inter-edges, a pair of vertices s_i^m and s_j^n in the m_{th} and n_{th} modality, respectively, we connect them with an inter-edge whose weight equals to 1, i.e.,

$$w_{ij}^{m,n} = 1 \tag{2}$$

if the i_{th} and j_{th} instances are known to be correlated.

After the construction of graph, a random walk, starting at a vertex and then randomly travelling to a connected vertex, is more likely to stay within a cluster than travelling between. Therefore conducting random walks on the graph can discover clusters where the flow tends to gather. The transition probability from a vertex v_i to a vertex v_j is defined as a set function $P_{ij}(A) : 2^E \rightarrow R$ for the graph $G(V, A)$

$$P_{ij}(A) = \begin{cases} 1 - \frac{\sum_{j:e_{ij} \in A} w_{ij}}{w_i} & \text{if } i = j, \\ \frac{w_{ij}}{w_i} & \text{if } i \neq j, e_{ij} \in A, \\ 0 & \text{if } i \neq j, e_{ij} \notin A, \end{cases} \tag{3}$$

which encourages random walks within clusters ($e_{ij} \in A$) and eliminates those between clusters ($e_{ij} \notin A$). $\sum_{j:e_{ij} \in A} w_{ij}$ is the total weights incident to v_i. We add a self loop transition ($i = j$) to maintain the total transition probability out of v_i to be 1. [6][5]

3.4 Dictionary Inference

We recurrent the original dictionary inference process in FLORA [7], soon we found that the process in FLORA is not so good. In the original dictionary inference process, only an average of the representative data is taken, and the dictionary fixed once it be used to generate enriched representation.

Enriched representation of original extracted featured is discussed as the process of combining instances in the both

Table 1: Rule for labeling data.

AQI	Values Levels of Health Concern	Label
0-50	Good	1
51-100	Moderate	2
101-150	Unhealthy for sensitive groups	3
151-200	Unhealthy	4
201-300	Very unhealthy	5
301-500	Hazardous	6

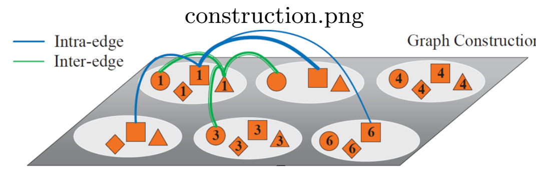
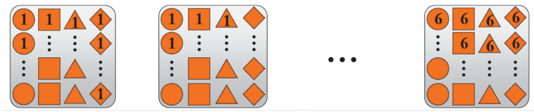


Figure 8: Graph Construction.



$$\mathbf{D}^m = \mathbf{D}^m \cup \{(1/|V_k^m|) \sum_{j:\mathbf{s}_j^m \in G_k} \mathbf{s}_j^m\}$$

Figure 9: Original Dictionary Inference Process.

Table 2: Definition of Notations

Notations	Definition	Dimension
$\mathbf{D}^m$	The dictionary of m_{th} modality	$p^m * K$
$\mathbf{t}_{li}^m$	The i_{th} labeled instance of m_{th} modality	$p^m * 1$
$\hat{\mathbf{t}}_{li}^m$	The enriched representation of t_{li}^m	$K * 1$

source and target domain with the dictionary we learned, given by [9]

$$\min_{\hat{\mathbf{t}}_{li}^m} ||\mathbf{t}_{li}^m - \mathbf{D}^m \hat{\mathbf{t}}_{li}^m||_F^2 + \alpha ||\hat{\mathbf{t}}_{li}^m||_1 \tag{4}$$

Notice that $\mathbf{D}^m$ fixed in the proposed process, it means to require a reasonable $\hat{\mathbf{t}}_{li}^m$ used for further transfer learning. Differ from that, we referred the *Traditional Method of Sparse Coding*. After a dictionary is created for the image analysis, it doesn't just use the dictionary for further analysis in the traditional way. [1]

$$\begin{aligned} \Phi^* &= \arg\min_{\Phi,\alpha} E(y, \alpha|\Phi) \\ &= \arg\min_{\Phi,\alpha} [||y - \Phi\alpha||_2^2 + \lambda||\alpha||_1] \end{aligned} \tag{5}$$

Similar to this method, our novel dictionary generated process is described as following steps:

1) The first step, dictionary is kept constant and the loss function is minimized with respect to train a set of α.

2) The second step, is the learning step. It keeps the α constant, while performing the gradient descent on dictionary to minimize the average loss.

The new dictionary inference process is shown as figure 10.

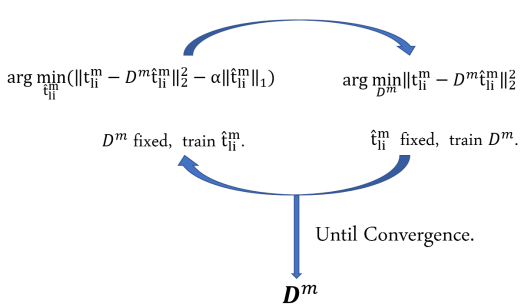


Figure 10: Improved Dictionary Inference Process.

In this way, we training Dictionary until the loss didn't change and derive the final dictionary. To accelerate this process, we apply technique of K-SVD [2] and by this way the convergence speed of the model has been improved.

3.5 Main Contribution

- It enforces multi-modalities to share knowledge and representation structures by learning semantically related dictionaries.
- It settles the data insufficiency problem, by transferring semantically related dictionaries learned from a source to enrich feature representations of a target domain.
- We evaluate our method on air quality prediction in three cities, with performances outperforming baseline.

4. EXPERIMENT

The experiment was carried out on the datasets download from Microsoft Urban Computing website, our baseline is recurrent theory model of FLORA. The result used our proposal was compared with baseline.

4.1 Dataset

Data containing 4 modality of multi-modality sensor data in Beijing, Shanghai, Tianjin, Baoding. The detail of data is shown in table 3

Table 3: Detail of Dataset.

Road Networks	Road segments (represented by end points), Length, Level of capacity
POI	Name, Address, Coordinates, Category of a venue
Meteorological	Weather, Temperature, Humidity, Barometer pressure, Wind strength , and etc.
Taxi Trajectories	Generated by over 32,000 taxicals from February 2^{nd} to May 26^{th}, 20 14

The statistics of four modalities for all cities shown as follows:

Modalities		Cities			
		Beijing	Shanghai	Tianjin	Baoding
Road	#. Segments	249,080	313,736	97,258	69,383
	Highways	994km	2,016km	1,681km	795km
	Roads	24,643km	40,944km	18,595km	17,884km
POI	#. of POIs	379,022	433,016	152,797	88,698
Meteorology	Time span(2014)	2/1-5/31	8/1-9/10	9/10-11/30	8/1-11/30

4.2 Comparison Result

The accuracy of prediction results are drawn into chart and listed.

We report the average accuracy of several times of experiments. The prediction accuracy was influenced by several factors, our model works best in data set of Shanghai and works worst in Baoding. Prediction accuracy of our novel method is 3-7% higher than original method in FLORA. Experiment is valid and successful, we extract latent semantic relation from multi-modality sensor data in differnet cities by generated corresponding dictionaries, and use transfer learning to apply those dictionaries in a strange city for air quality prediction successfully.

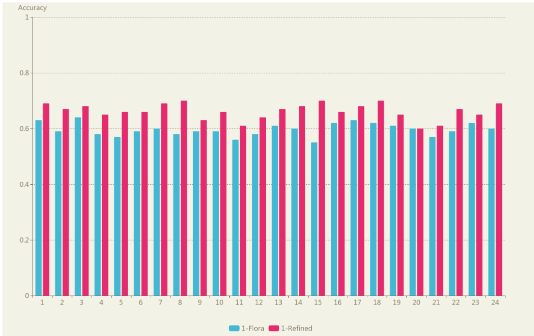


Figure 11: Hourly air quality prediction accuracies in Shanghai.

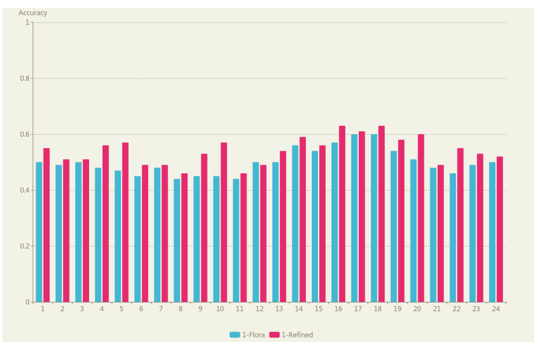


Figure 12: Hourly air quality prediction accuracies in Tianjin.

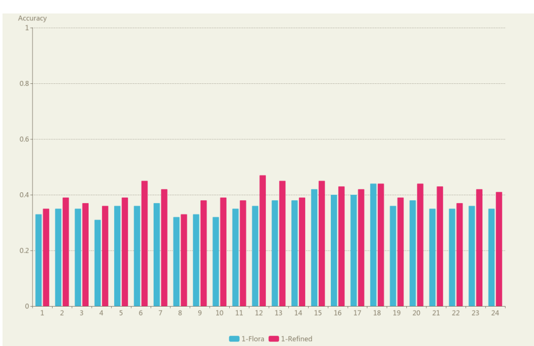


Figure 13: Hourly air quality prediction accuracies in Baoding.

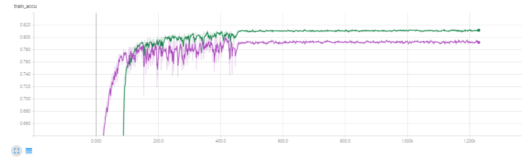


Figure 14: Improvement of Convergence Speed and Accuracy.

5. CONCLUSION

In this paper, we propose a improved method based on previous research to extract latent semantic relation from multi-modality sensor data in different cities. Particularly, enriches feature representations in a target domain with semantically related dictionaries learnt from a source, and transfers labelled instances from the source. Extensive experimental results in the case study of air quality prediction demonstrate the superiority of improvement. In the future, we would like to extend this method in more fields.

6. ADDITIONAL AUTHORS

7. REFERENCES

[1] Dictionary learning. *IEEE Signal Processing Magazine*, 2011.

[2] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 2006.

[3] W. Dai, Q. Yang, G.-R, Xue, and Y. Yu. Boosting for transfer learning. *ICML*, 2007.

[4] P. S. J. and Y. Q. A survey on transfer learning. *IEEE TKDE*, 2010.

[5] Z. Jiang, G. Zhang, and L. S. Davis. Submodular dictionary learning for sparse coding. *CVPR*, 2012.

[6] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa. Entropy-rate clustering: Cluster analysis via maximizing a submodular function subject to a matroid constraint. *PAMI*, 2014.

[7] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 2014.

[8] B. A. Olshausen, C. F. Cadieu, and D. K. Warland. Learning real and complex overcomplete representations from the statistics of natural images. *Proc. SPIE*, 2009.

[9] Y. Wei, Y. Zheng, and Q. Yang. Transfer knowledge between cities. *SIGKDD*, 2016.

[10] Y. Zheng, L. Capra, O. Wolfson, and H. Yang. Urban computing: Concepts, methodologies, and applications. *ACM TIST*, 2014.

[11] Y. Zheng, F. Liu, and H.-P. Hsieh. U-air: When urban air quality inference meets big data. *SIGKDD*, 2013.