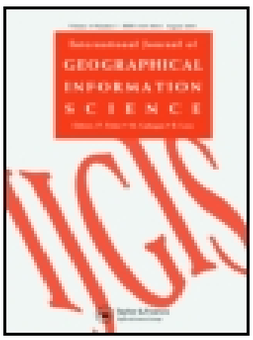

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: http://www.tandfonline.com/loi/tgis20

Classifying urban land use by integrating remote sensing and social media data

Xiaoping Liu, Jialv He, Yao Yao, Jinbao Zhang, Haolin Liang, Huan Wang & Ye Hong

To cite this article: Xiaoping Liu, Jialv He, Yao Yao, Jinbao Zhang, Haolin Liang, Huan Wang & Ye Hong (2017): Classifying urban land use by integrating remote sensing and social media data, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2017.1324976

To link to this article: http://dx.doi.org/10.1080/13658816.2017.1324976

Published online: 10 May 2017.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE, 2017
https://doi.org/10.1080/13658816.2017.1324976

Taylor & Francis
Taylor & Francis Group

ARTICLE

Classifying urban land use by integrating remote sensing and social media data

Xiaoping Liu [a], Jialv He [a], Yao Yao [a], Jinbao Zhang [a], Haolin Liang [b], Huan Wang [b] and Ye Hong [b]

[a]School of Geography and Planning, Guangdong Key Laboratory for Urbanization and Geo-simulation, Sun Yat-sen University, Guangzhou, China; [b]School of Geography and Planning, Sun Yat-sen University, Guangzhou, China

ABSTRACT

Urban land use information plays an important role in urban management, government policy-making, and population activity monitoring. However, the accurate classification of urban functional zones is challenging due to the complexity of urban systems. Many studies have focused on urban land use classification by considering features that are extracted from either high spatial resolution (HSR) remote sensing images or social media data, but few studies consider both features due to the lack of available models. In our study, we propose a novel scene classification framework to identify dominant urban land use type at the level of traffic analysis zone by integrating probabilistic topic models and support vector machine. A land use word dictionary inside the framework was built by fusing natural–physical features from HSR images and socioeconomic semantic features from multisource social media data. In addition to comparing with manual interpretation data, we designed several experiments to test the land use classification accuracy of our proposed model with different combinations of previously acquired semantic features. The classification results (overall accuracy = 0.865, Kappa = 0.828) demonstrate the effectiveness of our strategy that blends features extracted from multisource geospatial data as semantic features to train the classification model. This method can be applied to help urban planners analyze fine urban structures and monitor urban land use changes, and additional data from multiple sources will be blended into this proposed framework in the future.

ARTICLE HISTORY
Received 22 December 2016
Accepted 21 April 2017

KEYWORDS
Land use classification; probabilistic topic models; social media data; remote sensing; social sensing

1. Introduction

Land use and land cover (LULC) information comprises essential geographical spatial features for many fields, such as urban planning, government management, and sustainable development (Herold *et al*. 2003, Ellis and Pontius 2007, Arsanjani *et al*. 2013, Liu *et al*. in press). China's rapid economic and urban developments have generated diverse and sophisticated urban functional zones which reflect in urban land use patterns (Huang *et al*. 2013, Liu *et al*. 2014a, Long and Shen 2015, Zhang and Du 2015).

CONTACT Yao Yao whuyao@foxmail.com, yaoy33@mail2.sysu.edu.cn

However, urban land use patterns are affected not only by government policies but also indoor lifestyles, which are continuously changing with urban development (Yuan *et al*. 2012). Therefore, the effective detection of urban land use patterns, which are significant for formulating effective urban planning policies, has been a controversial issue in recent studies.

High spatial resolution (HSR) remote sensing images enable computation-based urban land use detection, where HSR image classification models have been extensively applied to extract and analyze LULC in recent studies (Hu and Wang 2013, Huang *et al*. 2014, 2015, Wu *et al*. 2015, Zhang and Du 2015, Wen *et al*. 2016). Analyses of urban LULC are primarily conducted with three types of spatial units; units of pixels and objects are usually employed to evaluate land cover, whereas scenes are commonly used to identify urban functional zones and accurate urban land use patterns (Liu *et al*. 2015a, Zhang and Du 2015, Zhang *et al*. 2015b). Many studies applied object-oriented classification (OOC) models to extract urban land use patterns using physical features (such as spectral, shape, and texture features) of ground components (Blaschke 2010, Blaschke *et al*. 2014, Dupuy *et al*. 2012, Hu and Wang 2013). However, OOC models often overlook the spatial distribution and semantic features of ground components because they were only designed to mine the low-level semantic land cover information of ground components. Because of the difficulties in mining sufficient information, the above traditional classification models are also difficult to identify land use classification with typical thematic features using traditional remote sensing classification models. The difficulty is due to the issue of crossing the 'Semantic Gap' (Benz *et al*. 2004, Durand *et al*. 2007, Tokarczyk *et al*. 2015).

In simple terms, low-level semantic features indicate 'information' that comes with the data directly, while high-level semantic features refer to 'knowledge' specific for each user and application (Bratasanu *et al*. 2011). The semantic gap refers to the

Table 1. Abbreviation list.

Abbreviation	Full name
BoW	Bag-of-words
CBD	Central business district
GLCM	Grey-level co-occurrence matrix
GPS	Global positioning system
GSA	Global sensitivity analysis
HSR	High spatial resolution
LDA	Latent Dirichlet Allocation
LDMM	Linear Dirichlet Mixture Model
LULC	Land use and land cover
NLP	Natural language processing
OA	Overall accuracy
OOC	Object-oriented classification
OSM	OpenStreetMap.org
pLSA	Probabilistic latent semantic analysis
POI	Point-of-interest
PTM	Probabilistic topic model
RBF	Radial basis function
RTUD	Real-time Tencent user density
SAL	Semantic allocation level
SIFT	Scale invariant feature transform
STD	Standard deviation
SVM	Support vector machine
TAZ	Traffic analysis zone

disparity of features identified between these two levels. In the field of image understanding, low-level features extracted directly from the image data (Bratasanu *et al*. 2011), such as color and texture, only express the physical properties. As disparate objects may have the same physical properties, and identical objects may possess different attributes, image classification only with low-level features is most likely inaccurate. Nevertheless, introducing the high-level semantic features (Bratasanu *et al*. 2011, Zhao *et al*. 2013, Zhong *et al*. 2015), which is the diverse attribute attached to the objects given by the human operator in accordance with the usage and other information, into image classification is likely to be more explicit classifications at higher accuracy. For example, given a set of HSR images containing different scenes, the land cover objects can be recognized based on the low-level feature description, e.g. buildings. However, attempts to capture high-level latent semantic concepts aim to seek different functional types, such as residential, commercial, and industrial areas (Zhao *et al*. 2013, Zhong *et al*. 2015).

To bridge the 'Semantic Gap' between LULC, recent studies have introduced the concept of 'scene classification' into HSR image classification to label a scene with a single category (Zhang and Du 2015). The majority of current studies apply the bag-of-words (BoW) modeling approach and fuse physical features of ground scenes via probabilistic topic models (PTMs) to improve the detection accuracy of urban land use types with high-level semantic information (Yang and Newsam 2010, Sun *et al*. 2012, Chen *et al*. 2013, Zhao *et al*. 2013, Tokarczyk *et al*. 2015, Zhang and Du 2015, Zhong *et al*. 2015, Wen *et al*. 2016). Zhang *et al*. (2015b) introduced the Linear Dirichlet Mixture Model (LDMM), a strategy to fuse HSR images and road network data to detect the percentage of land use in each land parcel (Liu *et al*. 2015a, Zhang and Du 2015, Zhang *et al*. 2015b). However, extracting features from remote sensing images can only represent the external natural–physical properties of ground components, whereas regional land use types often have a strong correlation with indoor human socioeconomic activities, which are difficult to extract from HSR images.

To solve this problem, recent studies have proposed the concepts of 'social sensing' and 'urban computing' (Zheng *et al*. 2014, Liu *et al*. 2015b). Multisource social media data, such as global positioning system (GPS) trajectories of floating cars, mobile phone signals, check-in data of social media, and point of interests (POIs), have been introduced to monitor residential activities and urban land use dynamics. Many in-depth discussions suggest that multi-social media data have great potential to reveal urban land use patterns (Yuan *et al*. 2012, Jiang *et al*. 2015, Yao *et al*. 2016). Yuan *et al*. (2012) proposed a POI-based semantic analysis model named DRoF to map urban functional zones (Yuan *et al*. 2012). Yuan and Zheng (2015) introduced the Latent Dirichlet Allocation (LDA) model, which combines GPS trajectories of floating cars (De Fabritiis *et al*. 2008) and POI frequencies, to mine urban land use types with senior semantic information, which can improve compared with HSR image-based methods (Yuan *et al*. 2015).

However, these methods utilize only one type of data rather than fusing geospatial information from HSR images and social media data into the detection of land use types. Regions with similar types of urban land use tend to have similar external natural–physical properties and indoor human socioeconomic activity patterns (Yao *et al*. 2016). For example, central business districts (CBDs) and residential zones with high-rise towers

are difficult to distinguish from remote sensing images alone but have distinct indoor human activities. On the other hand, bare fields and farmland, both with less human activities, can be differentiated by identifying natural–physical properties from remote sensing images. As previously summarized, our study aims to build a dominant urban land use sensing framework by combining several machine learning and natural language processing (NLP) models to fuse the geospatial latent semantic information extracted from HSR images (remote sensing information) and multisource social media data (social sensing information) as patterns to classify the urban land use and evaluate the accuracies and reliabilities of classification models by manual interpretation. Our model is applied to detect land use patterns in the Haizhu district in Guangzhou, which is one of most developed cities in Southern China with a mixture of land use types. By combining different kinds of features and comparing the corresponding classification results, we obtained the optimal combination of features and land use classification result.

2. Study area and data

The study area is located in the Haizhu district (Figure 1(a)) of Guangzhou in Guangdong Province with a total area of 102 km^2 and a permanent population of approximately 1,010,500. Guangzhou has been considered to be a political, cultural, and economic center in South China. As one of the four central downtown districts in Guangzhou, the urban structures of Haizhu district are very complex and contain a mixture of several land use types, such as residential communities, shopping malls, clinical facilities, and educational buildings.

Figure 1(b) displays a high spatial resolution (HSR) Worldview-2 image of Haizhu district in 2014 with a grid size of 34,263 × 14,382 and a spatial resolution of 0.5 m. Based on the road net data provided by OpenStreetMap, HSR image and official urban planning data, we divided the image into 593 land patches, similar to traffic analysis zones (TAZs) (Long and Thill 2015). Figure 1(a) displays the classification results of the dominant land use types in the study area via manual interpretation, which contains public management services land (M), industrial land (I), green land (G), commercial land (C), residential land (R), park land, (P) and urban villages (U).

Social media data, including OpenStreetMap (OSM) road networks (http://www.openstreepmap.org), Gaode POIs, and real-time Tencent user density (RTUD) (http://heat.qq.com), are used to complement HSR-image extracted features and enrich additional information for land use identification in the study area. POIs in our study are provided by Gaode Map Services (http://lbs.amap.com/), which is one of the most popular and largest web map service providers in China. We obtained POIs from approximately 123,915 records with 432 categories in the study area via Gaode Map APIs (Figure 1(c)), including business establishments, commercial sites, educational facilities (kindergartens, primary schools, and middle schools), residential communities, clinical facilities, and scenic locations. RTUD, which is a new dataset applicable to t semantic classification (Li *et al*. 2015), contains the hourly numbers of smartphone users who use Tencent applications, such as Tencent Mobile App QQ (a messenger-like software), WeChat (a mobile chat software), Soso Maps (a web mapping service and navigation software), and other mobile applications that provide LBS services. Figure 2 shows the RTUD time-series

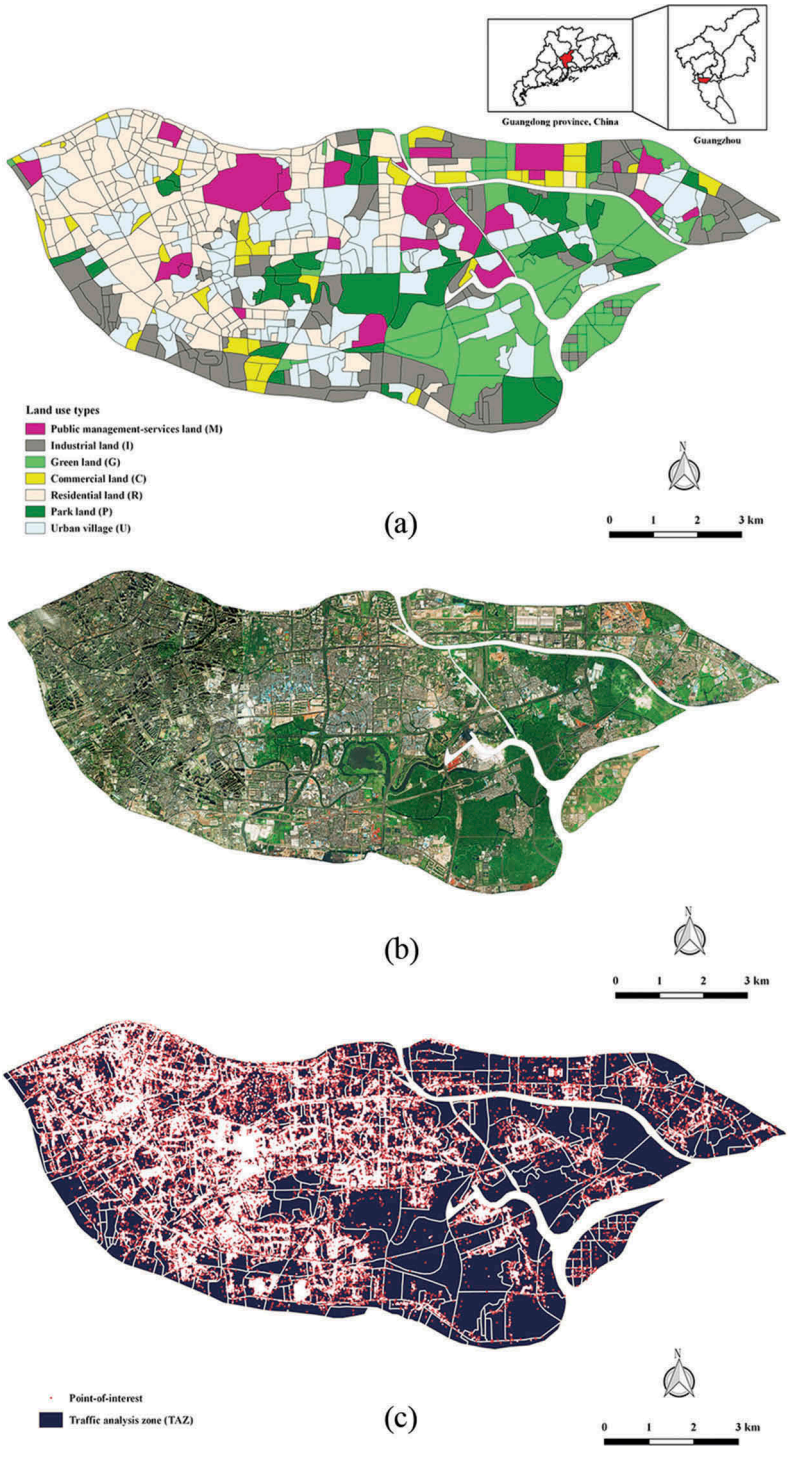

Figure 1. Case Study area: Haizhu district, Guangzhou, Guangdong Province. (a) Urban land use data obtained from manual interpretation within the unit of the traffic analysis zone (TAZ) level; (b) High spatial resolution (HSR) remote sensing image provided by Worldview-2 satellite in the study area; the black lines in the front represent roads downloaded from OpenStreetMap (OSM); (c) Spatial distribution density of Gaode point of interest (POIs).

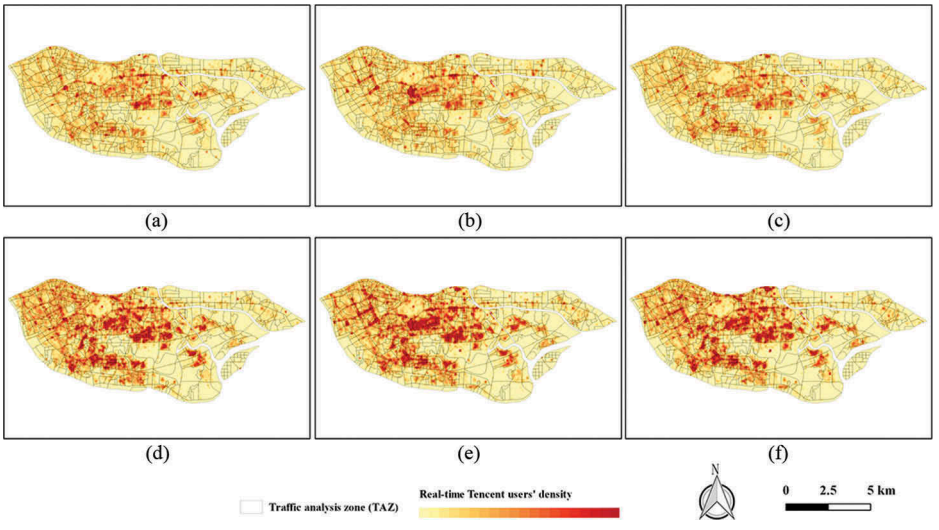

Figure 2. Time-series dataset of real-time Tencent user density (RTUD) in study area. (a) 9:00, (b) 17:00 and (c) 22:00 on a workday, (d) 9:00, (e) 17:00, and (f) 22:00 on a rest day.

data by calculating the average of workday and rest day data, respectively, at a spatial resolution of 25 m. Previous studies indicated that mean filtering is an effective social media data preprocessing method to reduce the data size and computational demands without much information loss (Liu *et al*. 2015b, Chen *et al*. 2017).

3. Methodology

The flowchart of the proposed model is illustrated in Figure 3. The purpose of our study is to classify dominant urban land use types by fusing multisource features

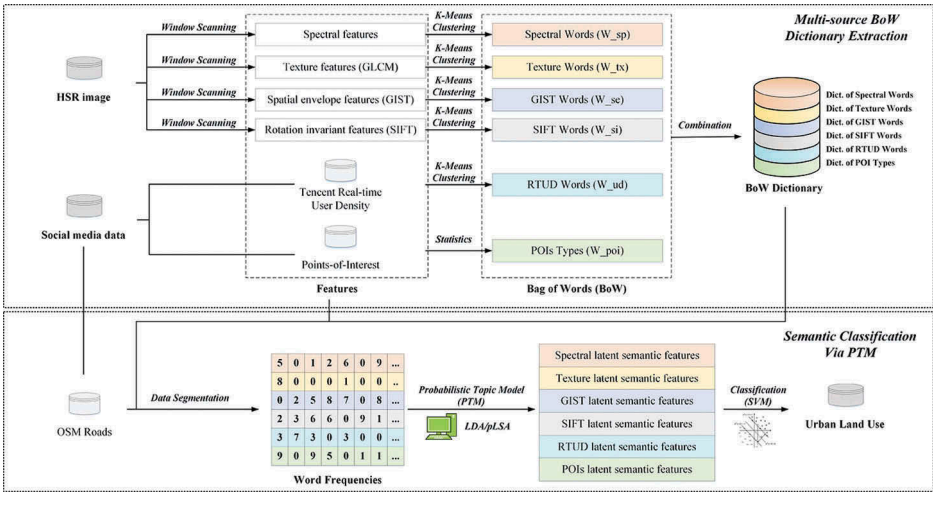

Figure 3. Flowchart of the proposed model for classifying urban land use by fusing multisource geospatial data (including HSR image and multisource social media data) via sematic models.

from HSR remote sensing images and social media data. In this study, we applied the following four steps to identify the urban land use type in each traffic analysis zone (TAZ). First, we extracted features from the remote sensing images using window scanning. The extracted features were characterized by spectral, texture, and spatial envelope characters and simultaneously extract rotation invariant features using the scale invariant feature transform (SIFT). Second, we applied the k-means clustering method to classify the features extracted in the previous step and RTUD data into several classes, and subjectively define the types of POIs. We obtained a large amount of visual words, which were labels clustered by k-means algorithm and considered as mid-level features in order to distinguish them from the low-level raw features and high-level semantic vocabulary features (Liu *et al*. 2009, Vedaldi and Fulkerson 2010), and constructed a multisource dictionary of BoW. Third, we delineated the TAZs in the study area based on the open-sourced OSM road network data and counted the feature words extracted from the HSR images and social media data in each TAZ. By using PTMs, we mined the latent semantic features based on the frequencies of feature words into high-dimensional semantic vectors. Finally, we applied a multi-class support vector machine (SVM) model. We trained the SVM model with selected land use data verified on the ground to classify urban land use types, and estimated the performance of classification across different combinations of semantic features.

The models described below were implemented by our research team using C++ on Windows 8.1(×64), linking with several open-source libraries, including CGAL (http://www.cgal.org), GDAL (http://www.gdal.org/), OpenCV (http://opencv.org/), and LIBSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). The source codes of the LDA-based topic model are available at Princeton University (http://www.cs.princeton.edu/~blei/topicmodeling.html).

3.1. *Spatial feature extraction*

HSR images contain abundant spectral and spatial information. Among all feature descriptors, the spectral and texture features of HSR images are able to reflect the inner components and tonal variation of ground components. The SIFT feature descriptors can handle the stretch, rotation, and changes in visual angle of pattern recognition for ground components, which has been extensively applied in image analysis (Kupfer *et al*. 2013, 2015, Yan *et al*. 2016). The patterns that are extracted from HSR images in this paper are similar to the patterns in the semantic allocation level (SAL)-based PTM model of Zhong *et al*. (2015), as follows.

To reduce the computational complexity during the extraction of spectral features from HSR images, we adopt a window and gap with a certain size and extract the mean and standard deviation (STD) of each band for each HSR image. Thus, the spectral feature of the center of the *i*-th window $loc_i(x_i, y_i)$, can be expressed as $spe_i = \{mean_1, std_1, mean_2, std_2, \ldots, mean_B, std_B\}_i$, where B denotes the band count. We can continue with some evenly spaced spectral feature vectors.

The grey-level co-occurrence matrix (GLCM) effectively describes the patterns of images and textures (Hua *et al*. 2006, Mohanaiah *et al*. 2013). Similar to spectral features, we compress the grey level of the images to eight images and extract four GLCM-based

Haralick's feature statistics, including correlation, ASM, energy, contrast, and homogeneity within each window with a certain size (Sebastian *et al*. 2012, Zhong *et al*. 2015). Assuming B as the band count, the texture feature of the i-th window $loc_i(x_i, y_i)$ can be described as:

$$tex_i = \{cor_1, asm_1, ene_1, con_1, hom_1, \ldots, cor_B, asm_B, ene_B, con_B, hom_B\}_i.$$

In this study, we introduce two methods (SIFT and GIST) to describe the images' local features. The first method is to calculate the SIFT feature in each window. A previous study has indicated that when a 128-dimensional vector is adopted to represent the SIFT feature, it can achieve the best optimized registration performance (Zhong *et al*. 2015). To reduce the computation cost, we obtain the first component of the HSR image and then adopt a window-scanning method to extract the SIFT feature vectors, where the SIFT features of the center of the i-th window are $sif_i = \{sif_1, sif_2, \ldots, sif_{128}\}_i$. When describing an entire scene, we usually employ partial patterns instead of global patterns. Due to the complexity and uncertainty of HSR images, this method will not only be costly in computation and storage but also cause misclassification in the cases in which two scenes are identical but have different spatial distributions of inner ground components. To solve this problem, we bring in the spatial envelope features called GIST introduced by Olivia *et al*. (2001), whose effectiveness for describing scenes of images at the macro level has been proven in recent studies (Oliva and Torralba 2001, 2006, Tung and Little 2015, Acharya *et al*. 2016).

The SIFT descriptor was originally designed for recognizing the same object appearing under different conditions, and has a strong discriminative power. The 'GIST' is an abstract representation of the scene that spontaneously activates memory representations of scene categories, and has achieved high accuracy in recognizing natural scene categories, e.g. mountain and coast (Oliva and Torralba 2001, 2006). GIST has been considered as a common spatial envelope feature descriptor that can adequately describe five different spatial envelope scenes, including degree of naturalness, openness, roughness, expansion, and ruggedness (Oliva and Torralba 2001, 2006). In our study, we segment each window into 4×4 blocks and calculate the GIST features of each band. Similar to the SIFT, the central GIST features of the i-th window are:

$$gis_i = \left\{gis_{1,1}, gis_{1,2}, \ldots, gis_{1,18}, \ldots, gis_{B,1}, gis_{B,2}, \ldots, gis_{B,18}\right\}_i,$$

where B represents the band count.

Although remote sensing data can adequately represent the physical attributes of ground components, it cannot illustrate the socioeconomic properties caused by human activities. Social media data may supplement information on human activities. Previous studies indicate that the distribution of POIs can be effectively applied to illustrate the functions of land parcels (Rodrigues *et al*. 2012, Bakillah *et al*. 2014, Jiang *et al*. 2015, Long and Liu 2015, Yao *et al*. 2016). Here, we introduce POI categories as one type of virtual words that reflect the socioeconomic properties. The patterns of human activities can then be obtained by filtered-RTUD time series (Chen *et al*. 2017, Yao *et al*. 2017), whose value and variation are important. The characteristics of urban residents' activities are closely related to the surrounding environments and urban functional areas. Thus, the time series of RTUD can represent the functional patterns of certain areas. Therefore, we describe the RTUD pattern of each window as

$rtu_i = \{rtu_{w,0}, \ldots, rtu_{w,23}, \ldots, rtu_{r,0}, \ldots, rtu_{r,23}\}_i$, where j is the j-th window in the study area, w and r represent the timing curves during workdays and rest days, respectively.

3.2. *Building multisource BoW dictionary*

Assuming the existence of a certain area R in a city, whose multisource features can be described as $f(i,j,k \in R) = \{spe_{i\in R}, tex_{i\in R}, sif_{i\in R}, gis_{i\in R}, POI_{j\in R}, rtu_{k\in R}\}$, where i, j, and k indicate the window center, the POIs, and the raster center of RTUD data in the R area, respectively. Note that $f(i,j,k \in R)$ is a multidimensional vector. We used the k-means method to cluster and converted each feature into a certain virtual word in the BoW dictionary. Thus, feature words in area R can be described as a document $Doc_R = \{word_{spe}, word_{tex}, word_{sif}, word_{gis}, word_{POI}, word_{rtu}\}_R$, where $word_{POI}$ is the set of inner POI categories, on which basis we can apply topic models to recognize and classify multisource documents.

A large amount of feature data is needed to be extracted in a large study area, which causes low efficiency during the clustering processes. When the number of features exceeds 500,000, we selected a random subset of 500,000 data points to conduct a preliminary clustering via k-means clustering and iterated to optimize the result by silhouette estimation (Rousseeuw 1987, Yuan *et al*. 2012, Da Cruz Nassif and Hruschka 2013). Based on the cluster centers obtained via the preliminary clustering process, the Euclidean distance was calculated to estimate the similarity between each center and other unlabeled semantic feature vectors; the unlabeled features were classified into the closest feature.

3.3. *Semantic classification via PTMs and SVM*

As illustrated in Figure 4, PTMs, including the probabilistic latent semantic analysis (pLSA) and LDA models, are designed to evaluate generated virtual words and mine

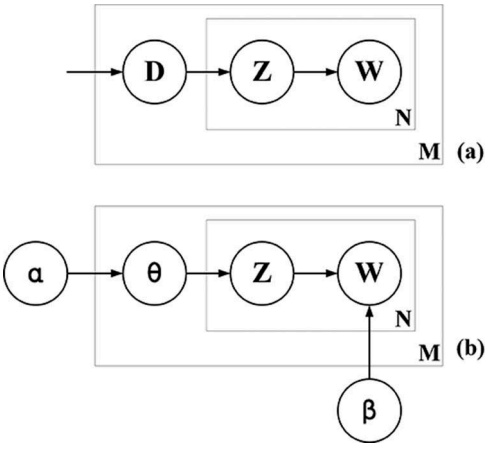

Figure 4. Probabilistic graphical models of (a) pLSA and (b) LDA. The nodes W, Z, D represent virtual words, topics, and document (or image), respectively, whereas α indicates the Dirichlet Allocation of topics in the LDA model.

latent semantic features of documents (Blei 2012). PTMs have been extensively applied in the fields of NLP; moreover, it has achieved acceptable results in the scene classification of HSR images in recent years (Huang *et al*. 2015, Liu *et al*. 2015b, Tokarczyk *et al*. 2015, Zhong *et al*. 2015, Wen *et al*. 2016).

pLSA takes advantage of the relationships among document, topics and words, and decomposes the probability $p(w_j, d_j)$ of the word w_j, which appears in a paper with the combination of BoW and the probability and total probability formulas (Bosch *et al*. 2006).

$$p(w_j, d_j) = \sum_{K}^{k=1} p(w_j|z_k)p(z_j|d_k). \tag{1}$$

In Equation (1), $\{p(w_j|z_1), \ldots, p(w_j|z_k), \ldots, p(w_j|z_K), w_j \in W\}$ demonstrates the base vectors in the latent semantic space, and $p(z_k|d_i)$ represents the topic distribution, which can be considered as the semantic features of given documents. Thus, we apply the vector set $\{p(z_1|d_i), \ldots, p(z_k|d_i), \ldots, p(z_K|d_i), d_i \in D\}$ to denote a document. The pLSA model has the problem of overfitting because each document that it represents is only the numerical form of the discrete probability of a certain topic. It is not able to mine the semantic feature outside the training datasets. To solve these problems, a new LDA-based pLSA model assumes that the semantic perplexity parameters are subject to the Dirichlet Allocation (Ramage *et al*. 2009). For certain documents with K given topics, each vector $\theta_i = \{\theta_{i,1}, \ldots, \theta_{i,k}, \ldots, \theta_{i,K}\}$ within the vector group $\theta = \{\theta_1, \ldots, \theta_i, \ldots, \theta_M\}$ obeys the Dirichlet Allocation with the parameters of $\alpha = \{\alpha_1, \ldots, \alpha_i, \ldots, \alpha_M\}$. The definition of a probability function for original latent semantic distribution by the LDA is the key to fix the disadvantage of the pLSA (Lu *et al*. 2011).

Based on the OSM road network data, we segmented the study area into several TAZs. Considering each TAZ as a land patch, we counted the distribution frequency of visual words from all feature classes and input the results into the PTMs model to calculate the high-dimensional latent semantic features. Then, the SVM, which has been proven to have high efficiency in classifying the high-dimensional feature in previous studies (Lilleberg *et al*. 2015, Zhang *et al*. 2015b), was applied to our proposed model to identify the urban land use types in the TAZs. Because the SVM is a binary classifier, we adopt a multi-classifiers-combined method to train and classify the latent semantic features in each TAZ. The final classification results were given by the class that most frequently appeared in each TAZ.

In this study, we select 50% of the training samples in each class, whose features are randomly combined and input into a multi-class SVM classifier. The other 50% of the data is used as testing data. The SVM classifier is implemented by the LIBSVM package (Chang and Lin 2012). In the training process, we use 25% of the training dataset as the validation dataset and use Kappa to assess the model calibration. Two sensitive parameters of the SVM with the radial basis function (RBF) kernel, penalty C factor, and the kernel parameters NU, need to be tuned. We set $C \in [2^{-10}, 2^{10}]$ and $NU \in [0, 1]$ and search the optimum parameters by a grid-search method, in which the optimization objective is to maximize Kappa (Li and Guo 2014).

4. Results

4.1. *Scene classification via different feature combinations*

Table 2 shows different feature combination methods and their average accuracies, and Figure 6 shows the confusion matrix of the classification results (Figure 5) of each combination that is closest to the average accuracy. To ensure the reliability and stability of the classification results, we repeated the classification process for each group 100 times and calculated the average classification accuracy.

As demonstrated in Table 2 and Figure 5, PTM-based semantic features from HSR images or social media data can be used to distinguish the functional type of land parcels. The semantic features generated by the LDA model are slightly higher than the semantic

Table 2. Scene classification results via different combinations of semantic features.

	Semantic features						pLSA		LDA	
Exp.	Spectral	Texture	SIFT	GIST	POI	RTUD	OA	Kappa	OA	Kappa
A	√	√	√				0.622	0.531	0.663	0.565
B	√	√	√	√			0.662	0.575	0.685	0.591
C					√		0.715	0.640	0.721	0.650
D						√	0.708	0.628	0.729	0.657
E					√	√	0.754	0.688	0.759	0.697
F	√	√	√	√	√	√	0.843	0.800	0.865	0.828

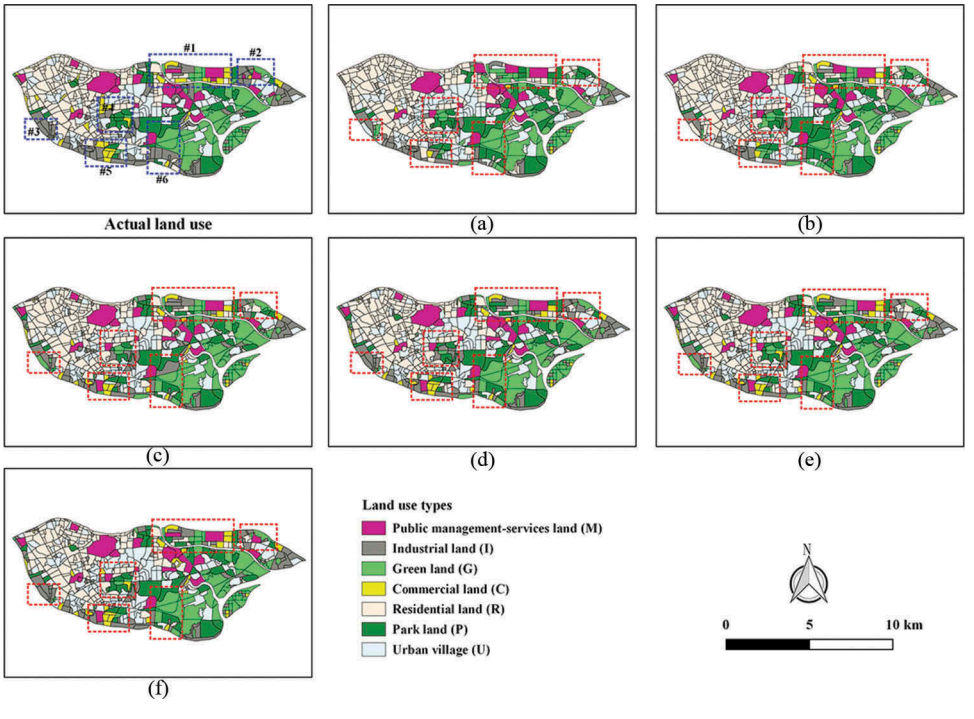

Figure 5. LDA-based land use classification results via different combinations of semantic features. (a) Spectral, texture, and SIFT, (b) Spectral, texture, SIFT, and GIST, (c) POI, (d) RTUD, (e) POI and RTUD, and (f) Spectral, texture, SIFT, GIST, POI, and RTUD.

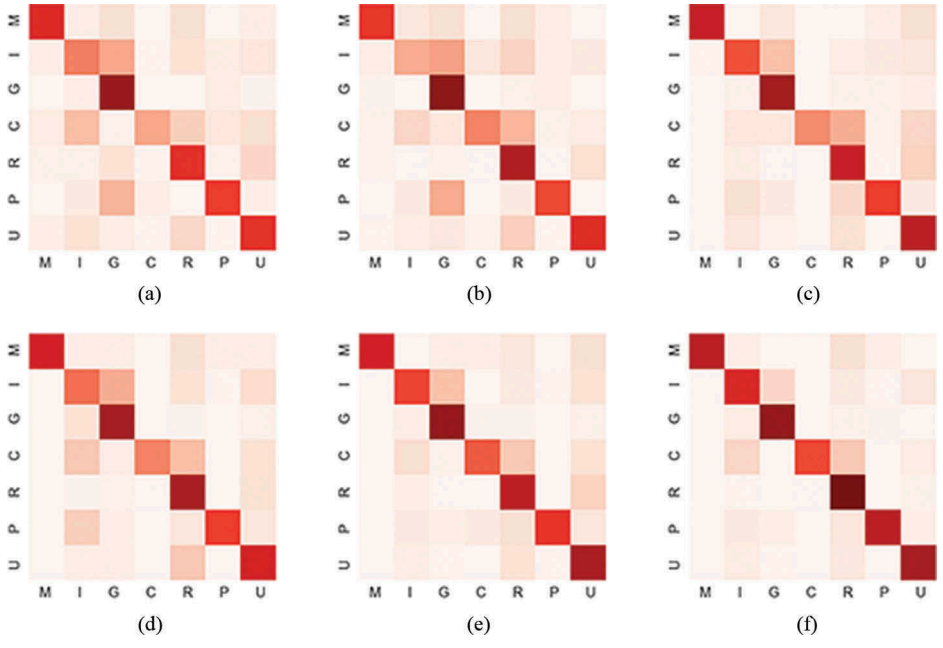

Figure 6. Confusion matrix of scene classification results via pLSA. Feature combination: (a) Spectral, texture, and SIFT, (b) Spectral, texture, SIFT. and GIST, (c) POI, (d) RTUD, (e) POI and RTUD, and (f) Spectral, texture, SIFT, GIST, POI, and RTUD.

features generated by the pLSA model regarding the accuracy of the SVM classification. Existing studies have indicated that the LDA works better than the pLSA when measuring the perplexity of predicting new documents (Lu *et al*. 2011). Our study area is located in downtown of Guangzhou, where highly mixed land use patches serve as theme-mixed documents in NLP, which would cause the LDA model to yield better classification result. This study thus applies an extra binary divisive procedure to optimize the sensitive hyper-parameter α of the LDA model, and the effect of the relevant parameters on the accuracy of the classification results will be discussed in the following section.

Conventional HSR-images-based scene classification methods (Huang and Zhang 2013, Zhang *et al*. 2015b, Zhong *et al*. 2015), which only consider spectral, texture, and SIFT features, can only attain poor classification accuracies in complicated urban land use classification. Applying GIST features to describe scenes at the macro level cannot improve the precision of classification because the use of natural–physical semantic features extracted from remote sensing images is challenging to differentiate highly mixed land use patches. As shown in Figures 6(a,b), 7(a,b), and Table 3, when only applying texture features, commercial land (#1 and #5 in Figure 5(a,b)) is easily mixed with residential land, industrial land and urban village since commercial land is often distributed extensively in remote sensing images and exhibits a complex spatial pattern. Therefore, natural–physical features that are extracted from remote sensing images are incapable of reflecting the inner properties and structures of urban functional areas.

Figures 6(c,d), 7(c,d) show semantic features extracted from social media data, such as POIs and RTUD, which are highly correlated urban land use types with human activities,

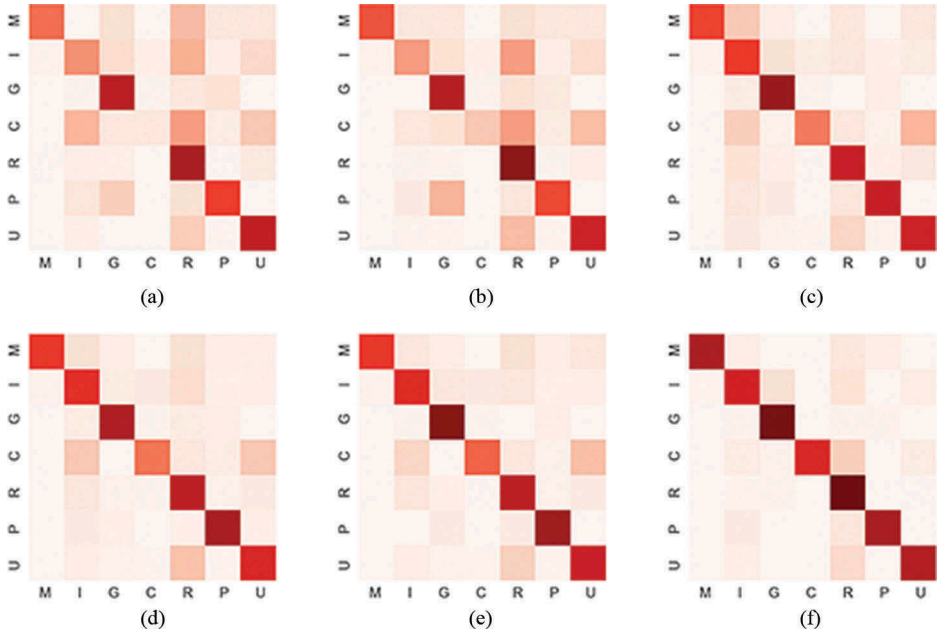

Figure 7. Confusion matrix of scene classification results via LDA. Feature combination: (a) Spectral, texture, and SIFT, (b) Spectral, texture, SIFT, and GIST, (c) POI, (d) RTUD, (e) POI and RTUD, and (f) Spectral, texture, SIFT, GIST, POI, and RTUD.

including commercial land and residential land. The classification accuracy of these two types of data is visibly improved; the total accuracy and Kappa increase 9.95% and 16.58%, respectively. Compared with the POIs, the RTUD can effectively distinguish between the residential area and an urban village, which can explain that time-series population density is more capable of reflecting the ground truth land use type from within a city area. The POI-based semantic features distinguished the urban village better than the RTUD, which explained that the distribution of POIs have a greater advantage in urban village identification by comparing with the indoor routines of people. As demonstrated in Figure 5(c,d), urban areas with sparse human activities, such as green land and park land, can be adequately identified by RTUD-based semantic features. Therefore, POIs and RTUD-based semantic features are combined for classification. This model obtains better results compared to the independent application of each type of features (Group C and Group D), where the overall accuracy (OA) and Kappa increases 0.03–0.05 and 0.04–0.05, respectively.

In Group F of the feature combination test, we input all semantic features into an SVM to classify the land use type and obtain the best classification results, where OA and Kappa exceed 0.80. From the experiment results, we discover that areas with sparse human activities (such as green land and park land) and areas with complex land use types (such as commercial land and residential land) can be distinguished. For example, as shown in Figure 5 #2, a public management services land parcel has been inaccurately classified as land for park (Figure 5(b) #2) and land for residential use (Figure 5(e) #2) when only remote sensing or social media-based semantic features are considered.

Table 3. Scene classification accuracy of each land use type via different feature combination of semantic features.

		Feature combinations (exp.)											
		A		B		C		D		E		F	
Land use type	Avg. accuracy	pLSA	LDA	pLSA	LDA	pLSA	LDA	pLSA	LDA	pLSA	LDA	pLSA	LDA
Public-management land	Commission error	0.320	0.520	0.360	0.440	0.240	0.400	0.280	0.360	0.280	0.360	0.200	0.160
	Omission error	0.469	0.200	0.407	0.125	0.050	0.118	0.000	0.111	0.000	0.111	0.000	0.045
	Mapping accuracy	0.531	0.800	0.593	0.875	0.950	0.882	1.000	0.889	1.000	0.889	1.000	0.955
	User accuracy	0.680	0.480	0.640	0.560	0.760	0.600	0.720	0.640	0.720	0.640	0.800	0.840
Industrial land	Commission error	0.575	0.632	0.708	0.660	0.434	0.368	0.528	0.340	0.406	0.321	0.311	0.274
	Omission error	0.416	0.400	0.354	0.234	0.333	0.407	0.375	0.358	0.259	0.308	0.232	0.135
	Mapping accuracy	0.584	0.600	0.646	0.766	0.667	0.593	0.625	0.642	0.741	0.692	0.768	0.865
	User accuracy	0.425	0.368	0.292	0.340	0.566	0.632	0.472	0.660	0.594	0.679	0.689	0.726
Green land	Commission error	0.108	0.203	0.081	0.189	0.122	0.108	0.135	0.162	0.095	0.068	0.095	0.041
	Omission error	0.515	0.352	0.481	0.333	0.343	0.275	0.402	0.195	0.302	0.225	0.212	0.193
	Mapping accuracy	0.485	0.648	0.519	0.667	0.657	0.725	0.598	0.805	0.698	0.775	0.788	0.807
	User accuracy	0.892	0.797	0.919	0.811	0.878	0.892	0.865	0.838	0.905	0.932	0.905	0.959
Commercial land	Commission error	0.692	0.923	0.590	0.795	0.615	0.564	0.590	0.538	0.462	0.487	0.410	0.308
	Omission error	0.294	0.571	0.429	0.333	0.000	0.346	0.059	0.400	0.160	0.333	0.042	0.000
	Mapping accuracy	0.706	0.429	0.571	0.667	1.000	0.654	0.941	0.600	0.840	0.667	0.958	1.000
	User accuracy	0.663	0.077	0.410	0.205	0.385	0.436	0.410	0.462	0.538	0.513	0.590	0.692
Residential land	Commission error	0.337	0.150	0.171	0.078	0.238	0.238	0.150	0.212	0.207	0.212	0.041	0.031
	Omission error	0.229	0.328	0.256	0.338	0.197	0.197	0.241	0.255	0.182	0.200	0.147	0.176
	Mapping accuracy	0.771	0.672	0.744	0.662	0.803	0.803	0.759	0.745	0.818	0.800	0.853	0.824
	User accuracy	0.663	0.850	0.829	0.922	0.762	0.762	0.850	0.788	0.793	0.788	0.959	0.969
Park land	Commission error	0.382	0.382	0.412	0.412	0.382	0.235	0.382	0.147	0.353	0.118	0.206	0.147
	Omission error	0.475	0.447	0.355	0.459	0.382	0.366	0.087	0.341	0.083	0.318	0.100	0.065
	Mapping accuracy	0.525	0.553	0.645	0.541	0.618	0.634	0.913	0.659	0.917	0.682	0.900	0.935
	User accuracy	0.618	0.618	0.588	0.588	0.618	0.765	0.618	0.853	0.647	0.882	0.794	0.853
Urban village	Commission error	0.344	0.221	0.328	0.262	0.205	0.262	0.287	0.303	0.156	0.254	0.139	0.172
	Omission error	0.360	0.286	0.274	0.262	0.362	0.262	0.341	0.234	0.348	0.229	0.139	0.073
	Mapping accuracy	0.640	0.714	0.726	0.738	0.638	0.738	0.659	0.766	0.652	0.771	0.861	0.927
	User accuracy	0.656	0.779	0.672	0.738	0.795	0.738	0.713	0.697	0.844	0.746	0.861	0.828

However, the land parcel can be correctly identified (Figure 5(f) #2) by fusing two proposed main features. The optimum classification results can be obtained by fusing the natural–physical properties of the HSR images and the socioeconomic properties of the social media data in each TAZ in the proposed model.

4.2. *Parameters sensitivity analysis*

In this section, we evaluate the correlation between the classification accuracies and three key parameters of our proposed models, including the size of the sample windows, the number of clustering categories that form the BoW, and the number of topics used in PTMs. The size factor of the sample window that is used to extract the natural–physical properties of the ground components on HSR images has been adequately discussed in the previous studies (Zhao *et al*. 2013, Zhang and Du 2015, Zhong *et al*. 2015). According to the conclusion of Zhong (2015), we segmented the HSR images into a set of overlapping image patches of 25 × 25 pixels to determine the spectral, texture, SIFT, and GIST features. Each pair of adjacent patches was set to overlap by 15 pixels to preserve a sufficient amount of spatial information (Zhong *et al*. 2015).

Several previous studies have indicated that the number of visual words and initial topics have a substantial impact on the classification accuracies of PTM-based scene classification; however, the way to obtain the optimum topics remains an unsolved problem (Ramage *et al*. 2009, Lu *et al*. 2011, Zhong *et al*. 2015). Figure 8 and Figure 9 display the relationship between the number of k-means clusters, the number of PTM topics, and the classification accuracy. Both the pLSA and LDA models are able to achieve relatively preferable performances within a certain window area. Figure 8 demonstrates that a decrease in the number of k-means clusters causes the accuracy of the classification results of PTM-based scene classification to stabilize and hardly fluctuate. With a fixed K value, therefore, the levels of accuracy of pLSA and LDA exhibit an initial distinct deviation and finally cause an increase in the number of initial topics (Figure 9). The accuracy of the LDA-based scene classification results exhibits strong fluctuation because the optimal results of the Dirichlet Topic Allocation Parameter α are uncertain (Figure 10). The method for obtaining the optimal number of initial topics and α remains unresolved in the field of

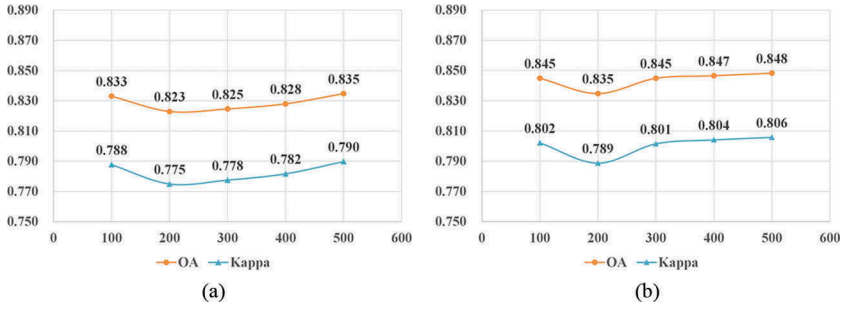

Figure 8. Accuracy assessment (y-axis) of scene classification results using different clustering numbers of k-means (x-axis), whereas topic numbers of PTMs are set to 100 and Dirichlet Topic Allocation Parameter α of LDA is set to 0.8. (a) pLSA and (b) LDA.

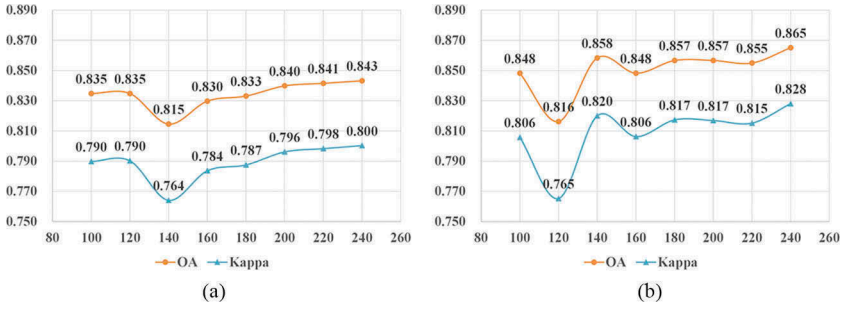

Figure 9. Accuracy assessment (y-axis) of scene classification results using different initial topic numbers of PTMs (x-axis), whereas clustering numbers of k-means are set to 500 and Dirichlet Topic Allocation Parameter α of LDA are set to 0.8. (a) pLSA and (b) LDA.

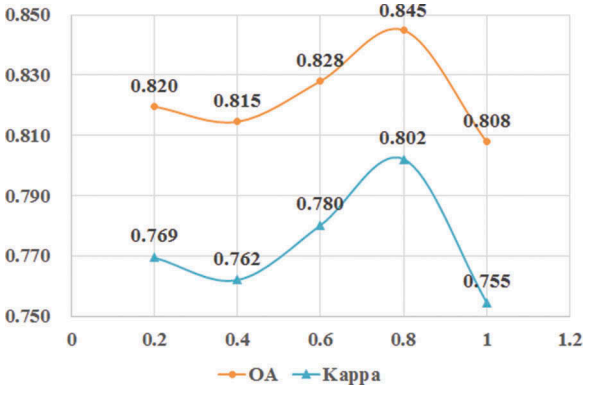

Figure 10. Accuracy assessment (y-axis) of scene classification results using different Dirichlet Topic Allocation Parameter α of LDA (x-axis), whereas topic numbers of PTMs are set to 100 and clustering numbers of k-means are set to 100.

multisource-spatial data-fused scene classification. By the above sensitivity analysis of proposed model parameters, we have chosen the clustering number of k-means, and the initial topic number of PTMs of 500 and 240, respectively.

5. Discussion

Improving urban LULC classification accuracy has been an important issue in recent literature about HSR remote sensing and social media data analysis. However, few studies have effectively fused various features extracted from multisource geospatial data (Hu *et al*. 2016) on the semantic level. This study proposed an effective framework for classifying urban land use by fusing semantic features extracted from HSR images and social media data, such as Gaode point of interest (POIs) and RTUD.

This study combines several different semantic features obtained from PTMs and input them into SVM classifiers; the results indicated that both HSR images and social media data can classify the urban land use type with high accuracy. Our finding agrees

with previous studies (Wu *et al*. 2009, Toole *et al*. 2012, Hu *et al*. 2016) that the HSR images are advantageous for identifying natural components over a rural area, whereas the use of social media data is better for a metropolitan area with a high population density. After combining all features into the SVM classifier, the results achieved the highest accuracy (OA = 0.865, Kappa = 0.828), which indicates that our model can effectively fuse natural–physical and socioeconomic information from HSR images and social media data for a high-resolution classification of urban land use.

In the future study, on the one hand, we expect to introduce more open social media sources (such as mobile data and floating car trajectories); on the other hand, based on global sensitivity analysis (GSA), the suitability of the classification of different types of land use with data from different data sources will be evaluated(Gao *et al*. 2016). In addition, we used a large amount of remote sensing image samples and multiple types of open social media data to build training datasets and perform a sensitivity analysis of different categories and accuracies. Such work will help us to establish a framework for sensing urban land use patterns at different levels.

TAZ was adopted as the basic unit in numerous studies involving urban land use division, scene classification, and urban functional zone classification (Li *et al*. 2009, Long *et al*. 2012, Yuan *et al*. 2012, Zhang and Du 2015, Zhang *et al*. 2015b, Yao *et al*. 2016), where studies in urban area of China's large cities were not uncommon. Their results indicated that using TAZ to identify urban land use patterns is reasonable and effective. However, mixed land use in urban areas does exist in reality, and even a single building contains different functions (Chen *et al*. 2017). Facing this issue, multisource geospatial data, including social media data and HSR images, may provide a new tool for quantifying the mixture of land use and distinguish between real land use situations and urban planning in the future. Hence, more fine-grained land use pattern should be identified to substitute TAZ in future studies.

The purpose of this study is to explore the integration of multisource spatial data in the framework of semantic model, thus effectively analyzing the dominant land use type of each study unit (TAZ). However, urban land use patterns are complicated and heterogeneous, especially in China's megacities. For example, many multifunctional land uses are mixed with living and commercial functions in our study area, which increases the difficulty of land use identification by manual interpretation or training sample selection. Although we have already obtained a relatively high accuracy for the hard classification of land use, the results are based on the accurate and manual interpretation of data. Therefore, the problem of mixed land use should be considered in future studies based on open social media data in the future work.

6. Conclusions

Rapid urban development leads to diversification and complication of land use types within cities (Ellis and Pontius 2007, Arsanjani *et al*. 2013, Hayashi and Roy 2013). As it is necessary for urban planners and government decision-makers to take into account the status of land use, timely and sufficient urban land use information will undoubtedly promote the sustainable development of cities (Liu *et al*. 2014b). However, the complexities and fusion of urban land use patterns create great challenges to accurate and effective mapping of urban land use (Li *et al*. 2014, Pei *et al*. 2014). This study established

a framework to classify dominant urban land use types at the level of the TAZ unit by fusing multisource semantic features that were extracted from HSR remote sensing images and multisource social media data. First, from HSR images and open social media data (including POIs and RTUD), we extracted various features and established k-means-based BoW and a dictionary of all feature categories. In second next step, PTMs that contain pLSA and LDA were introduced to extract multisource semantic information. Last, we fused different types of semantic features and input into a multi-class SVM classifier over Haizhu district, Guangzhou. The results indicated that the proposed model can effectively fuse natural–physical and socioeconomic semantic features that are extracted from HSR images and multisource social media data, respectively, to obtain the highest urban land use classification accuracy (OA = 0.865, Kappa = 0.828). We may continue our study to develop a fusing model using the following three aspects: first, discover the potential of various open social media data for detecting urban land use; second, improve the accuracy of detecting patterns of fusing urban land use with the proposed model; and last, discuss the feasibility of deep learning in which spatial information extracted from urban land use mapping is applied.

Acknowledgment

We really appreciate Prof. May Yuan, Assoc Prof. Shawn Laffan and the two anonymous reviewers for their useful comments and suggestions. This study was supported by the National Natural Science Foundation of China (Grant No. 41671398), the Key National Natural Science Foundation of China (Grant No. 41531176) and the National Natural Science Foundation of China (Grant No. 41601420).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China: [Grant Numbers 41671398, 41531176, 41601420].

ORCID

Xiaoping Liu http://orcid.org/0000-0003-4242-5392
Jialv He http://orcid.org/0000-0002-0997-558X
Yao Yao http://orcid.org/0000-0002-2830-0377
Jinbao Zhang http://orcid.org/0000-0001-8510-149X
Haolin Liang http://orcid.org/0000-0001-6865-5650
Huan Wang http://orcid.org/0000-0002-5291-2146
Ye Hong http://orcid.org/0000-0002-8996-3748

References

Acharya, U.R., *et al*., 2016. Decision support system for fatty liver disease using GIST descriptors extracted from ultrasound images. *Information Fusion*, 29, 32–39. doi:10.1016/j.inffus.2015.09.006

Arsanjani, J.J., *et al*., 2013. Toward mapping land-use patterns from volunteered geographic information. *International Journal of Geographical Information Science*, 27 (12), 2264–2278. doi:10.1080/13658816.2013.800871

Bakillah, M., *et al*., 2014. Fine-resolution population mapping using OpenStreetMap points-of-interest. *International Journal of Geographical Information Science*, 28 (9), 1940–1963. doi:10.1080/13658816.2014.909045

Benz, U.C., *et al*., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58 (3), 239–258. doi:10.1016/j.isprsjprs.2003.10.002

Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65 (1), 2–16. doi:10.1016/j.isprsjprs.2009.06.004

Blaschke, T., *et al*., 2014. Geographic object-based image analysis – towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 180–191. doi:10.1016/j.isprsjprs.2013.09.014

Blei, D.M., 2012. Probabilistic topic models. *Communications of the ACM*, 55 (4), 77–84. doi:10.1145/2133806

Bosch, A., *et al*., 2006. *Scene classification via pLSA*. Berlin Heidelberg: Springer, 517–530.

Bratasanu, D., Nedelcu, I., and Datcu, M., 2011. Bridging the semantic gap for satellite image annotation and automatic mapping applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4 (1), 193–204. doi:10.1109/JSTARS.2010.2081349

Chang, C. and Lin, C., 2012. LIBSVM: a library for support vector machine, 2001. *Software*. available from: http://www. csie. ntu. edu. tw/\~{} cjlin/libsvm.

Chen, K., *et al*., 2013. Semantic annotation of high-resolution remote sensing images via Gaussian process multi-instance multilabel learning. *IEEE Geoscience and Remote Sensing Letters*, 10 (6), 1285–1289. doi:10.1109/LGRS.2012.2237502

Chen, Y., *et al*., 2017. Delineating urban functional areas with building-level social media data: a dynamic time warping (DTW) distance based k-medoids method. *Landscape and Urban Planning*, 160, 48–60. doi:10.1016/j.landurbplan.2016.12.001

Da Cruz Nassif, L.I.S.F. and Hruschka, E.R., 2013. Document clustering for forensic analysis: an approach for improving computer inspection. *IEEE Transactions on Information Forensics and Security*, 8 (1), 46–54. doi:10.1109/TIFS.2012.2223679

De Fabritiis, C., Ragona, R., and Valenti, G., 2008. Traffic estimation and prediction based on real time floating car data. *In*: *International IEEE conference on Intelligent Transportation Systems IEEE*. 197–203.

Dupuy, S., *et al*., 2012. Land-cover dynamics in Southeast Asia: contribution of object-oriented techniques for change detection. *In*: *4th international conference on GEographic Object-Based Image Analysis (GEOBIA)*. Rio de Janeiro, Brazil, 217–222.

Durand, N., *et al*., 2007. ontology-based object recognition for remote sensing image interpretation. *In*: *International conference on TOOLS with Artificial Intelligence IEEE*. IEEE, 472–479.

Ellis, E. and Pontius, R., 2007. Land-use and land-cover change. *In*: C.J. Cleveland, ed., *Encyclopaedia of earth environmental information*. Washington, DC: Coalition.

Gao, L., *et al*., 2016. Robust global sensitivity analysis under deep uncertainty via scenario analysis. *Environmental Modelling & Software*, 76, 154–166. doi:10.1016/j.envsoft.2015.11.001

Hayashi, Y. and Roy, J., 2013. *Transport, land-use and the environment*. US: Springer Science & Business Media.

Herold, M., Liu, X., and Clarke, K.C., 2003. Spatial metrics and image texture for mapping urban land use. *Photogrammetric Engineering & Remote Sensing*, 69 (9), 991–1001. doi:10.14358/PERS.69.9.991

Hu, S. and Wang, L., 2013. Automated urban land-use classification with remote sensing. *International Journal of Remote Sensing*, 34 (3), 790–803. doi:10.1080/01431161.2012.714510

Hu, T., *et al*., 2016. Mapping urban land use by using landsat images and open social data. *Remote Sensing*, 8 (2), 151. doi:10.3390/rs8020151

Hua, B.O., Fu-Long, M., and Li-Cheng, J., 2006. Research on computation of GLCM of image texture [J]. *Acta Electronica Sinica*, 1 (1), 155–158.

Huang, K., *et al*., 2013. An improved artificial immune system for seeking the Pareto front of land-use allocation problem in large areas. *International Journal of Geographical Information Science*, 27 (5), 922–946. doi:10.1080/13658816.2012.730147

Huang, X., Liu, H., and Zhang, L., 2015. Spatiotemporal detection and analysis of urban villages in mega city regions of China using high-resolution remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 53 (7), 3639–3657. doi:10.1109/TGRS.2014.2380779

Huang, X., Lu, Q., and Zhang, L., 2014. A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 90, 36–48. doi:10.1016/j.isprsjprs.2014.01.008

Huang, X. and Zhang, L., 2013. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 51 (1), 257–272. doi:10.1109/TGRS.2012.2202912

Jiang, S., *et al*., 2015. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, 53, 36–46. doi:10.1016/j.compenvurbsys.2014.12.001

Kupfer, B., Netanyahu, N.S., and Shimshoni, I., 2013. A sift-based mode-seeking procedure for efficient, accurate registration of remotely sensed images. *In*: *Geoscience and remote sensing symposium*. IEEE, 4142–4145.

Kupfer, B., Netanyahu, N.S., and Shimshoni, I., 2015. An efficient SIFT-based mode-seeking algorithm for sub-pixel registration of remotely sensed images. *IEEE Geoscience and Remote Sensing Letters*, 12 (2), 379–383. doi:10.1109/LGRS.2014.2343471

Li, C., *et al*., 2014. Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. *Remote Sensing*, 6 (2), 964–983. doi:10.3390/rs6020964

Li, W. and Guo, Q., 2014. A new accuracy assessment method for one-class remote sensing classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52 (8), 4621–4632. doi:10.1109/TGRS.2013.2283082

Li, X.D., Yang, X.G., and Chen, H.J., 2009, Study on traffic zone division based on spatial clustering analysis. *Computer Engineering and Applications*, 45 (5), 19–22.

Li, Y., *et al*., 2015. *System and method for processing location data of target user*. U.S. Patent Application No. 14/699,073.

Lilleberg, J., Zhu, Y., and Zhang, Y., 2015. Support vector machines and Word2vec for text classification with semantic features. *In*: *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2015 IEEE 14th International Conference*. IEEE. 136–140.

Liu, J., *et al*., 2015a. Semantic classification for hyperspectral image by integrating distance measurement and relevance vector machine. *Multimedia Systems*. 1–10.

Liu, J., Yang, Y., and Shah, M., 2009. Learning semantic visual vocabularies using diffusion distance *In*: *Computer vision and pattern recognition, CVPR 2009*. IEEE. 461–468.

Liu, X., *et al*., 2014a. Simulating urban growth by integrating landscape expansion index (LEI) and cellular automata. *International Journal of Geographical Information Science*, 28 (1), 148–163. doi:10.1080/13658816.2013.831097

Liu, X., *et al*., in press. An integrated model for simulating multiple land use scenarios by coupling human and natural effects. *Landscape and Urban Planning*.

Liu, Y., *et al*., 2015b. Social sensing: a new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105 (3), 512–530. doi:10.1080/00045608.2015.1018773

Liu, Y., Fang, F., and Li, Y., 2014b. Key issues of land use in China and implications for policy making. *Land Use Policy*, 40, 6–12. doi:10.1016/j.landusepol.2013.03.013

Long, Y. and Liu, X., 2015. Automated identification and characterization of parcels (AICP) with openstreetmap and points of interest. *Environment and Planning B: Planning and Design*, 43 (2), 341–360.

Long, Y. and Shen, Z., 2015. V-BUDEM: a vector-based Beijing urban development model for simulating urban growth. *In*: *Geospatial analysis to support urban planning in Beijing*. Springer. 91–112.

Long, Y. and Thill, J., 2015. Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing. *Computers, Environment and Urban Systems*, 53, 19–35. doi:10.1016/j.compenvurbsys.2015.02.005

Long, Y., Zhang, Y., and Cui, C. Y., 2012. Identifying commuting pattern of beijing using bus smart card data. *Acta Geographica Sinica*, 67 (10), 1339–1352.

Lu, Y., Mei, Q., and Zhai, C., 2011. Investigating task performance of probabilistic topic models: an empirical study of PLSA and LDA. *Information Retrieval*, 14 (2), 178–203. doi:10.1007/s10791-010-9141-9

Mohanaiah, P., Sathyanarayana, P., and GuruKumar, L., 2013. Image texture feature extraction using GLCM approach. *International Journal of Scientific and Research Publications*, 3 (5), 1.

Oliva, A. and Torralba, A., 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42 (3), 145–175. doi:10.1023/A:1011139631724

Oliva, A. and Torralba, A., 2006. Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research*, 155, 23–36.

Pei, T., *et al*., 2014. A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science*, 28 (9), 1988–2007. doi:10.1080/13658816.2014.913794

Ramage, D., *et al*., 2009. Labeled LDA: a supervised topic model for credit attribution in multi-labeled corpora. *Association for Computational Linguistics*, 248–256.

Rodrigues, F., *et al*., 2012. Automatic classification of points-of-interest for land-use analysis. *In*: *GEOProcessing*: 2012 *The fourth international conference on advanced geographic Information systems, applications, and services*. 41–49.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. doi:10.1016/0377-0427(87)90125-7

Sebastian, V., *et al*., 2012. Gray level co-occurrence matrices: generalisation and some new features. *International Journal of Computer Science Engineering & Informa*, 2 (2), 151–157.

Sun, H., *et al*., 2012. Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model. *IEEE Geoscience and Remote Sensing Letters*, 9 (1), 109–113. doi:10.1109/LGRS.2011.2161569

Tokarczyk, P., *et al*., 2015. Features, color spaces, and boosting: new insights on semantic classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 53 (1), 280–295. doi:10.1109/TGRS.2014.2321423

Toole, J.L., *et al*., 2012. Inferring land use from mobile phone activity. *In*: *Proceedings of the Acm Sigkdd international workshop on urban computing*. 1–8.

Tung, F. and Little, J.J., 2015. Improving scene attribute recognition using web-scale object detectors. *Computer Vision and Image Understanding*, 138, 86–91. doi:10.1016/j.cviu.2015.05.012

Vedaldi, A. and Fulkerson, B., 2010. Vlfeat:anopen and portable library of computer vision algorithms. *In: International conference on multimedea*, Firenze, Italy, October DBLP. 1469–1472.

Wen, D., *et al*., 2016. A novel automatic change detection method for urban high-resolution remotely sensed imagery based on multiindex scene representation. *IEEE Transactions on Geoscience and Remote Sensing*, 54 (1), 609–625. doi:10.1109/TGRS.2015.2463075

Wu, C., Zhang, L., and Zhang, L., 2015. A scene change detection framework for multi-temporal very high resolution remote sensing images. *Signal Processing,* 124 (C), 184–197.

Wu, S., *et al*., 2009. Using geometrical, textural, and contextual information of land parcels for classification of detailed urban land use. *Annals of the Association of American Geographers*, 99 (1), 76–98. doi:10.1080/00045600802459028

Yan, K., *et al*., 2016. *CNN vs. SIFT for image retrieval: alternative or complementary?* ACM, 407–411.

Yang, Y. and Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification. *In*: *ACM sigspatial international symposium on advances in geographic information systems*, *ACM-Gis,* 3–5 November 2010, San Jose, CA, Proceedings DBLP. 270–279.

Yao, Y., *et al*., 2016. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science*, 31 (4), 825–848. doi:10.1080/13658816.2016.1244608

Yao, Y., *et al*., 2017. Mapping fine-scale population distributions at the building level by integrating multi-source geospatial big data. *International Journal of Geographical Information Science*, 31 (6), 1220–1244.

Yuan, J., Zheng, Y., and Xie, X., 2012. Discovering regions of different functions in a city using human mobility and POIs. *In: ACM SIGKDD international conference on knowledge discovery and data mining ACM*. 186–194.

Yuan, N.J., *et al*., 2015. Discovering urban functional zones using latent activity trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 27 (3), 712–725. doi:10.1109/TKDE.2014.2345405

Zhang, D., *et al*., 2015a. Chinese comments sentiment classification based on word2vec and SVM perf. *Expert Systems with Applications*, 42 (4), 1857–1863. doi:10.1016/j.eswa.2014.09.011

Zhang, X. and Du, S., 2015. A linear dirichlet mixture model for decomposing scenes: application to analyzing urban functional zonings. *Remote Sensing of Environment*, 169, 37–49. doi:10.1016/j.rse.2015.07.017

Zhang, X., Du, S., and Wang, Y., 2015b. Semantic classification of heterogeneous urban scenes using intrascene feature similarity and interscene semantic dependency. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8 (5), 2005–2014. doi:10.1109/JSTARS.2015.2414178

Zhao, B., Zhong, Y., and Zhang, L., 2013. Hybrid generative/discriminative scene classification strategy based on latent Dirichlet allocation for high spatial resolution remote sensing imagery. *In*: *Geoscience and remote sensing symposium*. IEEE, 196–199.

Zheng, Y., *et al*., 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5 (3), 38.

Zhong, Y., Zhu, Q., and Zhang, L., 2015. Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 53 (11), 6207–6222. doi:10.1109/TGRS.2015.2435801