

MACHINE LEARNING-BASED ECONOMIC DEVELOPMENT MAPPING FROM MULTI-SOURCE OPEN GEOSPATIAL DATA

Rui Cao¹, Wei Tu², Jixuan Cai^{3,*}, Tianhong Zhao², Jie Xiao², Jinzhou Cao², Qili Gao², Hanjing Su³

¹ Department of Land Surveying and Geo-Informatics & Smart Cities Research Institute,
The Hong Kong Polytechnic University, Hong Kong SAR, China - rucao@polyu.edu.hk

² Guangdong Key Laboratory of Urban Informatics & School of Architecture and Urban Planning, Shenzhen University,
Shenzhen, China - (tuwei,zhaotianhong2016,xiaojie2021,caojz,qlgao)@szu.edu.cn

³ Tencent Inc., Shenzhen, China - (codyjxcai.justinsu)@tencent.com

Commission IV, WG IV/10

KEY WORDS: Economic Development, Remote Sensing, Geospatial Big Data, Data Fusion, Machine learning.

ABSTRACT:

Timely and accurate socioeconomic indicators are the prerequisite for smart social governance. For example, the level of economic development and the structure of population are important statistics for regional or national policy-making. However, the collection of these characteristics usually depends on demographic and social surveys, which are time- and labor-intensive. To address these issues, we propose a machine learning-based approach to estimate and map the economic development from multi-source open available geospatial data, including remote sensing imagery and OpenStreetMap road networks. Specifically, we first extract knowledge-based features from different data sources; then the multi-view graphs are constructed through different perspectives of spatial adjacency and feature similarity; and a multi-view graph neural network (MVGNN) model is built on them and trained in a self-supervised learning manner. Then, the handcrafted features and the learned graph representations are combined to estimate the regional economic development indicators via random forest models. Taking China's county-level gross domestic product (GDP) as an example, extensive experiments have been conducted and the results demonstrate the effectiveness of the proposed method, and the combination of the knowledge-based and learning-based features can significantly outperform baseline methods. Our proposed approach can advance the goal of acquiring timely and accurate socioeconomic variables through widely accessible geospatial data, which has the potential to extend to more social indicators and other geographic regions to support smart governance and policy-making in the future.

1. INTRODUCTION

Monitoring national- or regional- economic development is an essential task for the policy-making of the government and the management of business (Jean et al., 2016). Traditional economic development evaluation usually depends on yearly or five-yearly economic statistics, which requires a large amount of labor and time; therefore, the collection, processing and correction of the economic development data are usually time-consuming and last for a long period of time. In addition, in developing and less developed countries, the low proportion of economic activities and weak infrastructure make it difficult to compile timely and accurate economic data, thus leading to incorrect estimation of economic development or even missing such data (Yeh et al., 2020, Steele et al., 2017).

For this reason, some studies have tried to use easily accessible remote sensing images to monitor the economic development. For example, nighttime light (NTL) images are widely used for estimating gross domestic product (GDP) as they are closely correlated with various economic parameters such as urbanization, population, and economic activities (Elvidge et al., 2021, Huang et al., 2021). For developing or less developed regions that lack reliable economic statistics, these NTL images are of great value. However, NTL images also have insurmountable limitations. For example, it is difficult to represent the economic level of agriculture due to the presence of unlit areas

such as rural areas; in addition, thermal power generation and incineration generate strong lighting, which can lead to serious overestimation of the economy (Elvidge et al., 2021).

Meanwhile, in order to predict the economic development more accurately, some studies try to exploit emerging geospatial big data. For example, points-of-interests (POI) data are recently used because they can to some extent characterize the level of infrastructure construction in the region and can also indirectly reflect the economic development level (Chen et al., 2020). However, these data sources are not very easily accessible in a complete manner, and suffer from high data bias and uneven distribution. On the other hand, some recent studies also leverage mobile phone-based positioning data to estimate the economic activities due to the directly close relationship between human activities and the regional economic development (Huang et al., 2021). Although these novel geospatial data provide alternative approaches to predicting the economic development, these data sources are not easily accessible in the real world scenarios, especially in a national or regional scale.

In this paper, we would like to focus on open available geospatial data and explore two fundamental questions: 1) What kinds of open geospatial data sources are effective for economic development estimation, and how effective are they? 2) How to make full use of the available open data to improve the prediction accuracy?

To explore these questions, in this paper, we propose a machine learning-based method to estimate and map economic devel-

* Corresponding author

opment through both knowledge-based and learning-based features extracted from multi-source open geospatial data, including NTL imagery, multispectral remote sensing imagery (MSI), and OpenStreetMap (OSM) road networks (RDN). Extensive experiments have been conducted to estimate GDP in China at county level, and the results demonstrate the effectiveness of the proposed method and also confirm the contributions of these open data in accurately estimating GDP values.

The rest of the paper is organized as follows. Section 2 reviews the related works. Section 3 describes the study area and data. Section 4 elaborates the proposed method. Section 5 presents the experimental results and analysis. Finally, Section 6 concludes the paper with discussion.

2. RELATED WORK

2.1 Socioeconomic Attributes Estimation

Smart governance and business intelligence have strong demand for up-to-date socioeconomic information (e.g. information of population and economic development) to support timely, accurate, and customized location-based services and policy-making (Dong et al., 2019, Tomor et al., 2019). However, these kinds of information are conventionally acquired through labor-intensive surveys, which have disadvantages of long-time intervals, coarse spatial granularity, and limited population coverage. Therefore, the traditional survey data are insufficient to meet the urgent needs.

To deal with the situation, studies try to leverage wide covering remote sensing imagery, such as nighttime light (NTL) and multispectral imagery (MSI), to map economic development indicators since they can either directly or implicitly reflect socioeconomic status and thus can serve as data sources to infer socioeconomic attributes (Abitbol and Karsai, 2020, Chen et al., 2020, Jean et al., 2016, Yeh et al., 2020). Alternatively, benefiting from the development of information and communication technologies (ICT), emerging geospatial big data, such as mobile phone data, vehicle trajectories, points-of-interest (POI), street view images, etc., have also been exploited to estimate socioeconomic attributes. They are usually by-products of daily socioeconomic activities and thus can be collected timely and economically (Cao et al., 2021, Dong et al., 2019, Gebru et al., 2017, Liu et al., 2015, Tu et al., 2018, Tu et al., 2020).

Both remote sensing images and geospatial big data have inevitable limitations; the former can only reflect limited human activities-related information, while the latter usually suffer from severe data sparsity and privacy issues. Therefore, there is a recent trend in combining them to complement each other to enhance the prediction performance of socioeconomic attributes (Cao et al., 2018, Cao et al., 2020, Chen et al., 2022, Steele et al., 2017). However, there are still major challenges in fusing these data for socioeconomic attributes estimation due to their inherent heterogeneity and different spatial forms. Besides, the effectiveness of different data sources has seldom examined. It is of great value to make full use of the available data sources and understand how they contribute to the estimation tasks.

2.2 Graph Representation Learning

Representation learning, a.k.a. feature learning, is a method of learning features from data that facilitates classification and

prediction without the need for handcrafted features (Bengio et al., 2013). Currently, with the development of machine learning, especially the breakthrough of deep learning in various fields, representation learning based on neural networks has become an effective data-driven feature extraction approach and is widely used in various fields (LeCun et al., 2015).

Graph representation learning is an important direction of representation learning, which learns the features (usually called embeddings) from graph-based data (Zhang et al., 2020). The learned embeddings can then be further used for various downstream classification and prediction tasks. Due to the capacity for non-Euclidean structural data modelling and feature learning, the graph neural network (GNN) models have attracted increasing attentions and have been developed rapidly in recent years, especially for graph convolutional neural networks, such as GCN (Kipf and Welling, 2017), GraphSAGE (Hamilton et al., 2017), and GAT (Velickovic et al., 2018), which have been demonstrated to be effective on graph-based data.

With the powerful ability in graph data modeling, there is great potential in exploiting GNN models to model the correlations between different geographic regions, which will be beneficial for economic development estimation, since socioeconomic activities and attributes usually present noticeable dependence between different geographic regions, as implied by the First Law of Geography.

3. STUDY AREA AND DATA

The mainland China is selected as our study area, as indicated by the light blue areas in Figure 1. Multi-source open available geospatial data, including nighttime light (NTL), multispectral imagery (MSI), and road networks are leveraged at the administrative level to estimate its GDP.

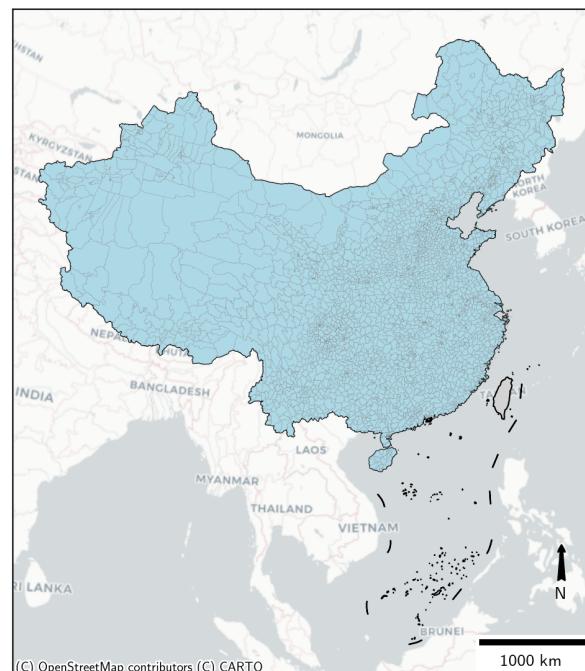


Figure 1. Study area of the mainland China.

GDP Data. The GDP data are collected and organized from

the China Statistical Yearbook and the local Bureau of Statistics, which provide county-level GDP data in 2018. It should be noted that Hong Kong, Macau, and Taiwan have different economic statistic systems, thus only the GDP data from the mainland China are used. Besides, the Sansha City, the southernmost and least populated prefecture in China, with the smallest land area but the largest maritime territory, is also excluded from the study due to its special spatial form of multiple small islands. In total, there are 2,852 county-level administrative regions used in the experiments. Due to data availability issues, among them, 2,673 counties are with GDP values.

Nighttime Light Imagery (NTL). We use the VIIRS Nighttime Light (VNL) V2 products of year 2018 provided by EOG (Elvidge et al., 2021) as the source for nighttime light imagery. The products are annual global nighttime lights time series that are consistently processed. They are produced from monthly cloud-free radiance averages which are made from low light imaging day/night band (DNB) data collected by the NASA/NOAA VIIRS nighttime light imagery. Specifically, 9 statistics (including count, area, min, max, range, mean, standard deviation, sum, and median) are calculated and exploited as the spatial attributes for each county-level region. In total, there are 9-dimensional NTL features.

Multispectral Remote Sensing Imagery (MSI). The Landsat-8 multispectral imagery is used to characterize the physical features of the landscape. Specifically, we use the Google Earth Engine (GEE) cloud computing platform (Gorelick et al., 2017) to access the level-2 Landsat-8 images and choose the surface reflection product with cloud coverage less than 10% within 2018, which contains 7 bands (i.e., ultra blue, blue, green, red, near infrared, shortwave infrared 1 and shortwave infrared 2) with 30-meter spatial resolution. For each band, we use GEE to calculate the zonal statistics of the counties, including min, max, median, mean, variance, and standard deviation of each county. Therefore, in total, there are 42-dimensional features for MSI.

OSM Road Networks (RDN). The road network data are collected from the OSM crowdsourcing platform in 2019. The OSM roads can be categorized into different types according to their importance, which are indicated by the attribute of *highway*¹, including *motorway*, *trunk*, *primary*, *secondary*, *tertiary*, etc. Considering that different types of roads will reflect different economic activities and intensities within an area, we extract the length of roads of different types, and also include the total length of all the roads. In total, we derive 26-dimensional RDN features.

4. METHODOLOGY

4.1 Preliminaries and Overview

To facilitate the description of the method, several definitions have been defined as follows.

Definition 1. *Region:* A region refers to a geographic area divided either by administrative boundaries or regular grids without overlapping. The regions of study area can be represented as $R = \{r_1, r_2, \dots, r_n\}$.

Definition 2. *Graph:* A graph is an abstract mathematical model that can be defined as $G = (V, E)$, where V is the node

¹ <https://wiki.openstreetmap.org/wiki/Key:highway>

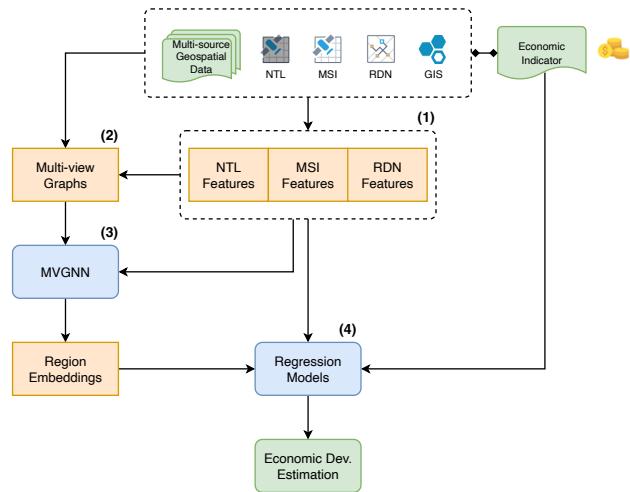


Figure 2. Overview of the workflow for economic development estimation from multi-source open geospatial data: (1) intra-region feature extraction, (2) inter-region correlation modeling via multi-view graphs, (3) region embeddings learning from MVGNN, (4) economic indicator regression.

set and E is the edge set. The geographic regions with the connections among them can be represented as the graphs, where the nodes are represented by the inner attributes of the regions and the edges are the connection between the regions such as spatial adjacency and attribute similarity.

Definition 3. *Region Embedding:* A region embedding is the distributed numerical vector used as the representation for a region. The region embeddings can be denoted as $H = \{h_1, h_2, \dots, h_n\}$ corresponding to regions R .

To effectively integrate the multi-source heterogeneous geospatial data for economic development estimation, we propose a machine learning-based solution to fully exploit both the intra-region attributes and inter-region correlations. The overview of the proposed method is presented in Figure 2.

As can be seen, the proposed method mainly consists of four steps. Firstly, handcrafted features of intra-region attributes are extracted from multi-source geospatial data, including NTL, MSI, and RDN. Secondly, inter-region correlations are modeled as multi-view graphs based on the spatial relationship and extracted intra-region features. Thirdly, the acquired features and built graphs are taken as inputs to the proposed multi-view graph neural network (MVGNN) to learn region embeddings, which include the implicit correlations between regions. Finally, the handcrafted features and learned region embeddings are combined as final region representations, together with corresponding economic development indicators such as GDP, they can be leveraged to train regression models to make the prediction. The details of the method will be elaborated in the following subsections.

4.2 Intra-region Feature Extraction from Multi-source Geospatial Data

For the multi-source geospatial data, knowledge-based handcrafted features can firstly be extracted. Specifically, for remote sensing imagery, the zonal statistics (such as min, max, mean, variance, etc.) of different bands within the regions can be extracted as corresponding features. For road networks, the statistics of road length of different road types within regions can

be leveraged as features. Details of the specific feature extraction can be found in Section 3.

4.3 Inter-region Correlation Modeling via Multi-view Graphs

Based on different perspectives from different data, multi-view graphs G are constructed to capture the correlations across regions, including region adjacency graph G^{adj} and region similarity graphs $\{G_i^{sim}\}_{i=1}^n$. For each data source, there is a region similarity graph constructed, thus $G = \{G^{adj}, G_{ntl}^{sim}, G_{msi}^{sim}, G_{rdn}^{sim}\}$ here.

4.3.1 Spatial-aware Region Adjacency Graph To model the spatial dependence of the regions, we propose to construct a spatial adjacency graph $G^{adj} = (V, E^{adj})$ based on region adjacency, where $V = \{v_i\}_{i=1}^n$ is the node set and $E^{adj} = \{e_{i,j}\}_{i,j=1}^n$ is the edge set. In the graph, each node v_i denotes a region and each edge $e_{i,j}$ represents the adjacency between node v_i and v_j . The adjacency matrix A^{adj} can be used to indicate the topological adjacency between the regions, where $A^{adj} \in R^{n \times n}$, with the entry values $A_{i,j}^{adj} \in \{0, 1\}$, and 1 indicates direct adjacency.

4.3.2 Attribute-aware Region Similarity Graph Other than spatial adjacency, some regions share similar attributes which make them closer. To capture this kind of relationship, we propose to construct region similarity graph $G^{sim} = (V, E^{sim})$ based on regional feature similarity. Similarly, $V = \{v_i\}_{i=1}^n$ is the node set and $E^{sim} = \{e_{i,j}\}_{i,j=1}^n$ is the edge set. In the graph, each node v_i denotes a region and each edge $e_{i,j}$ represents the connection between node v_i and v_j , which is determined by the similarity between their representative features. Specifically, the cosine similarity is used in our settings.

4.4 Context-aware Region Embedding Learning via Self-supervised MVGNN

To obtain the features of inter-region correlations, we propose a self-supervised multi-view GNN model to learn region embeddings which can explicitly model the inter-region relationships based on the built multi-view graphs. Specifically, the node feature matrix X , together with the corresponding multi-view graph structures G , can be leveraged to learn the region embeddings H to represent regions R , where $H = \{h_v \in \mathbb{R}^d, \forall v \in G\}$.

4.4.1 Model Architecture The architecture of the proposed MVGNN is shown in Figure 3. The input intra-region features are fed into the multi-view graph neural networks respectively to produce view-specific embeddings, which are further fused via the multi-view fusion module. The MVGNN is trained by the learning objectives, including spatial adjacency and attribute similarity reconstruction losses.

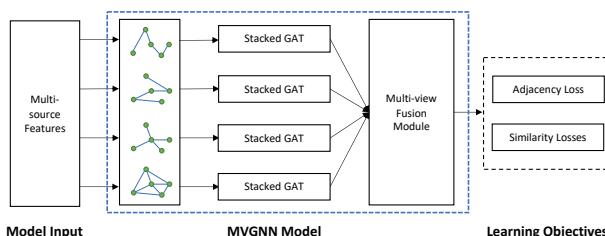


Figure 3. Multi-view graph neural network (MVGNN).

Base GNN Module: After the construction of multi-view graphs $G = \{G^{adj}, G_{ntl}^{sim}, G_{msi}^{sim}, G_{rdn}^{sim}\}$, a three-layer GAT (Velickovic et al., 2018) module is used as our backbone model to explicitly model the inter-node correlations as built in G , and further learn the representations of graph nodes, i.e. region embeddings. The GAT layers can aggregate the embeddings of neighbor nodes by weighted average with weights learned by attention mechanism, and then apply non-linear transformation (ReLU is used here).

Multi-view Fusion Module: To fuse the embeddings from the different views, adapted weighted multi-view fusion module is proposed. For each view, the fusion weights are learned from the embeddings themselves via a single layer multi-layer perceptron (MLP).

4.4.2 Learning Objectives To acquire effective region embeddings, self-supervised losses are designed as learning objectives to train the proposed MVGNN model. The overall loss function to train the model can be formulated as follows:

$$\mathcal{L} = \mathcal{L}_{adj} + \lambda \mathcal{L}_{sim} \quad (1)$$

where \mathcal{L}_{adj} and \mathcal{L}_{sim} are the spatial adjacency reconstruction loss and the attribute similarity reconstruction loss, respectively. λ is the weight to control the importance between the two losses. Specifically, \mathcal{L}_{sim} can be further formulated as follows:

$$\mathcal{L}_{sim} = \mathcal{L}_{ntl} + \alpha \mathcal{L}_{msi} + \beta \mathcal{L}_{rdn} \quad (2)$$

where \mathcal{L}_{ntl} , \mathcal{L}_{msi} , \mathcal{L}_{rdn} are the attribute similarity reconstruction losses from nighttime light, multispectral imagery, and road networks, respectively. α, β are the weights to control the importance between the losses.

Spatial Adjacency Reconstruction Loss: To reserve the properties of spatial adjacency between regions after region representation learning, the learned region embeddings should be able to reconstruct the spatial adjacency relationships. Specifically, the reconstruction loss \mathcal{L}_{adj} is designed to force the model to learn embeddings that can reconstruct the spatial adjacency matrix from the dot product of learned region embeddings by mean square error:

$$\mathcal{L}_{adj} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (A_{i,j}^{adj} - h_i^{adj T} h_j^{adj})^2 \quad (3)$$

where h_i^{adj} is the embedding vector of region r_i , which can be learned from the combination of the NTL, MSI, and RDN features.

Attribute Similarity Reconstruction Loss: To capture the inner connections of feature similarity across regions, the learned region embeddings are expected to reserve the ability to reconstruct the degree of similarity between different regions. Specifically, the similarity reconstruction losses \mathcal{L}_{sim} are proposed to enable the model to learn embeddings that can reconstruct the pairwise similarity matrix from the dot product of learned region embeddings by mean square error:

$$\mathcal{L}_{sim}(m) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (A_{i,j}^{sim}(m) - h_i^{sim}(m)^T h_j^{sim}(m))^2 \quad (4)$$

where $h_i^{sim}(m)$ is the embedding vector of region r_i for data type m , where $m \in \{\text{NTL, MSI, RDN}\}$.

4.5 Regression-based Economic Indicator Estimation

The learned region embeddings H can be further used to map to the economic development indicator y , the key is to find the mapping $f_{reg} : H \rightarrow y$, where f_{reg} is the regression model to learn from labelled data. In our experiments, Random Forest regression model is used. It should be noted that arbitrary regressors can be used to build the relation between region embeddings and corresponding economic indicators, such as linear regression (LR), support vector machine (SVR), multi-layer perceptron (MLP), etc.

5. EXPERIMENTS AND ANALYSIS

5.1 Experimental Setup

5.1.1 Evaluation Metrics The root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are used to evaluate the prediction errors. In addition, the coefficient of determination (R^2) is used to measure the goodness-of-fit of the regression models:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

where $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. n is the sample number for evaluation, y_i is the ground truth GDP value for region r_i , while \hat{y}_i is the estimated value. The best possible R^2 value is 1 and the larger values indicate better fit of the models. In our experiments, 5-fold cross-validation is used to evaluate the results.

5.1.2 Model Training Settings PyTorch (Paszke et al., 2019) and DGL (Wang et al., 2020) are used to implement the proposed graph neural networks. Adam is used as the optimizer with learning rate of 0.0001. The training epochs are empirically set to be 50. λ, α, β are set to be 1. The embedding size of the MVGNN is set to be 64. Scikit-learn (Pedregosa et al., 2011) is used to implement the regression models, including random forest, linear regression, SVR, and MLP. For targeted labels, log transformation is applied to GDP values to reduce the variance and further improve estimation performance as per normal preprocessing of economic variables (Lütkepohl and Xu, 2012).

5.2 Overall Results

The overall results with different input features are presented in Table 1. It can be seen that, in general, all the data sources, including the features from NTL, MSI, RDN, and the learned embeddings from the proposed MVGNN model, can significantly contribute to the estimation of GDP in our experiments, with almost each source over 60% goodness-of-fit in terms of R^2 measure. Besides, the results of all the errors (i.e., RMSE, MAE, and MAPE) are consistent, and are correlated with R^2 negatively.

Despite with the smallest feature dimensions (only 9-d) among the three sources of data, NTL features achieve the lowest errors in terms of both RMSE, MAE, and MAPE, while with the highest R^2 of more than 72%, significantly outperform that of MSI and RDN. These results are consistent with previous research of NTL-based GDP estimation (Chen et al., 2020, Huang et al., 2021) and further demonstrate the effectiveness of NTL data in reflecting the patterns of socioeconomic activities. RDN features have significantly higher R^2 and lower errors than MSI,

Table 1. Results of different input features using RF regressor.
 (ALL=NTL+MSI+RDN, MVGNN refers to the embeddings learned from the MVGNN model with input features of ALL)

	RMSE ↓	MAE ↓	MAPE ↓	$R^2 \uparrow$
NTL	0.2679	0.2070	3.3799	0.7240
MSI	0.3250	0.2513	4.1025	0.5937
RDN	0.2971	0.2299	3.7686	0.6604
ALL	0.2211	0.1653	2.6951	0.8120
MVGNN	0.2445	0.1859	3.0449	0.7700
MVGNN+ALL	0.2121	0.1593	2.5982	0.8269

which shows the effectiveness of road networks in economic development estimation and is quite reasonable since the lengths of road networks can well indicate the infrastructure construction situations within an area. For MSI, though it obtains significantly worse results than the other two sources, it can still achieve a reasonable prediction with R^2 slightly lower than 60%, which is inspiring since Landsat MSI data are widely available with high spatial and temporal coverage.

Furthermore, the combination of all the three data sources can achieve dramatically higher performance than single sources, with R^2 of over 81%. This shows the complementary of different data sources. Besides, the fusion of both the three data sources and the MVGNN embeddings can further boost the prediction results with an increase of about 1.5% in R^2 value, achieving 82.69%. The scatter plot of the predicted and true values of the log of GDP taking MVGNN+ALL as input is shown in Figure 4. The results show the effectiveness of including the inter-region correlations of different counties, and demonstrate the feasibility of using graphs to model inter-region correlations.

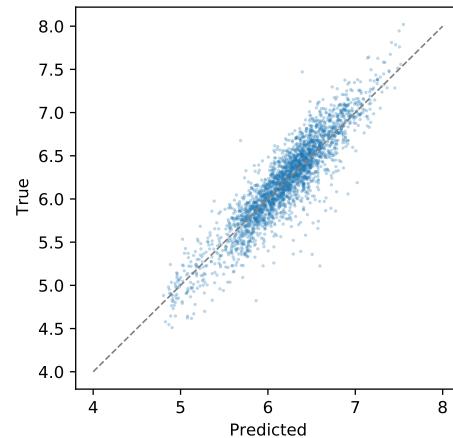


Figure 4. Scatter plot of the estimated and true $\log_{10}(GDP)$ values taken MVGNN+ALL as input features. (The GDP unit is 10^4 RMB)

The spatial distribution of the predicted GDP values are shown in Figure 5, which are the average of all the predictions made by the trained models of five folds. As can be seen, the GDP values are unevenly distributed in space, with significantly higher GDP to the east, and along the coastlines. City clusters such as Yangtze River Delta Megalopolis, Guangdong-Hong Kong-Macau Greater Bay Area, Jing-Jin-Ji Megalopolis, Yangtze River Midstream Megalopolis, Chengyu Megalopolis,

etc. can be noticeably recognized from the clustered high GDP distributions. In general, the spatial patterns of the predictions are consistent with the real situations in China, which further demonstrates the effectiveness of the proposed method in GDP estimation and mapping.

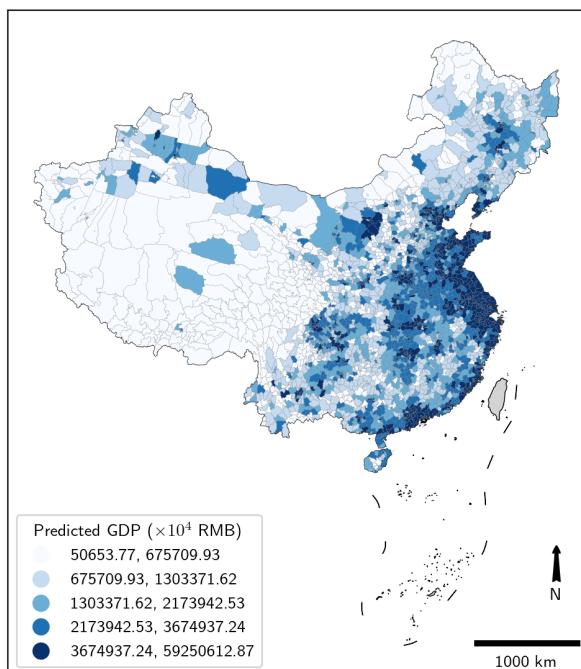


Figure 5. Spatial distribution of the predicted GDP values.

5.3 Analysis of Feature Importance

When using the random forest regressor, the importance of different features can be calculated, i.e., the impurity-based feature importances. The higher the score, the more important the feature. Specifically, the importance of a specific feature is computed as the total reduction of the selected criterion (Gini index here) contributed by that feature. The importance scores of the top 10 features, taken *ALL* and *MVGNN+ALL* as input features, are shown in Figure 6.

When taking *ALL* as the input features, the top 3 important features are both from NTL (sum, mean, and standard deviation), with the importance scores of 2.79, 0.59, and 0.07, respectively. The top 2 features contribute significantly higher than other features. The features of the lengths of the *primary* and *footway* roads are ranked within top 10. The statistics (including variance, standard deviation, median, and mean) of the red band of MSI are also ranked within top 10.

When taking *MVGNN+ALL* as the input features, the situation is similar for NTL, it can be seen that the features of the *sum* and *mean* of NTL remain to contribute to the most to the prediction, with the top-1 importance score over 2.5, which is dramatically higher than any other features. The top-1 feature (*sum* of NTL) is almost 7 times the second-important feature (about 0.37), i.e., *mean* of NTL. The third-important feature of NTL is the *standard deviation* of NTL (0.049). The other features are all from the learned embeddings of MVGNN.

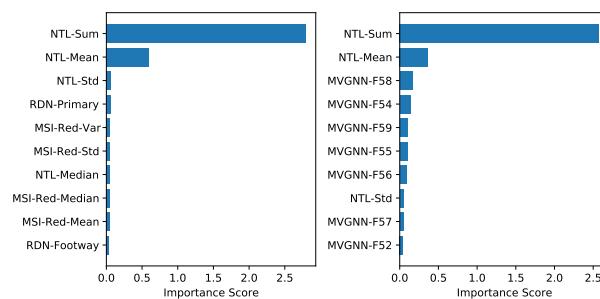


Figure 6. Bar chart of the importance scores of the top 10 contributing features with different input features. Left: *ALL*, Right: *MVGNN+ALL*.

5.4 Evaluation of Regression Models

To further test the effects of using different regressors, the results of using popular regression machine learning models are presented in Table 2. It can be seen that the RF model outperforms the other regressors by a large margin. While SVR and MLP show similar performance, which however significantly better than LR model.

Table 2. Results of using different regression models with input of *ALL* features.

	RMSE ↓	MAE ↓	MAPE ↓	$R^2 \uparrow$
LR	0.2763	0.2070	3.3877	0.7064
SVR	0.2455	0.1850	3.0405	0.7681
RF	0.2211	0.1653	2.6951	0.8120
MLP	0.2482	0.1837	3.0176	0.7630

5.5 Evaluation of GNN Backbones

To further examine the effects of different GNN backbones, we compare the results of GAT (Veličković et al., 2018) with other popular graph convolutional neural networks, including GCN (Kipf and Welling, 2017) and GraphSAGE (Hamilton et al., 2017), and the results are presented in Table 3. The results show that the GAT backbone slightly outperforms GCN and GraphSAGE backbones in our experiments, and all the results improve the performance compared with not using any GNN embedding features. These results imply that even simpler GNN models can still learn useful inter-region correlation features and add value to the GDP prediction task, which further highlights the importance of utilizing inter-region correlations.

Table 3. Results of using different GNN backbones.

	RMSE ↓	MAE ↓	MAPE ↓	$R^2 \uparrow$
GCN	0.2139	0.1606	2.6184	0.8240
GraphSAGE	0.2171	0.1632	2.6631	0.8187
GAT	0.2121	0.1593	2.5982	0.8269

5.6 Evaluation of Training Epochs

To evaluate the impacts of training MVGNN with different epochs, we present the experiment results in Table 4. As can be seen, the MVGNN features achieve the best performance when trained with 50 epochs. However, more training epochs do not always bring improvement, the results get worse when training too many epochs. Besides, it is surprising that the features extracted from MVGNN without any training can still be

effective, and even lead to better results than models trained with hundreds of epochs. This phenomenon suggests that the graph neural network (based on built multi-view graphs) with randomized model weights can already embed the inter-region relationships into the learned embeddings, and thus add extra useful information for GDP estimation.

Table 4. Results of using region embeddings learned from MVGNN of different training epochs.

Epoch	RMSE ↓	MAE ↓	MAPE ↓	$R^2 \uparrow$
0	0.2142	0.1616	2.6358	0.8235
1	0.2141	0.1616	2.6348	0.8236
10	0.2147	0.1618	2.6383	0.8227
50	0.2121	0.1593	2.5982	0.8269
100	0.2132	0.1615	2.6346	0.8251
200	0.2158	0.1622	2.6453	0.8208
500	0.2159	0.1624	2.6508	0.8207

6. DISCUSSION AND CONCLUSIONS

In this paper, we propose a machine learning-based method to estimate the economic development using multi-source open geospatial data, including nighttime light, multispectral remote sensing imagery, and OSM road networks. We firstly extract knowledge-based handcrafted features from the open data to characterize regions. Then, to take advantages of the inter-region correlations, we propose a multi-view graph neural network model, which is built on the spatial adjacency graph and attribute similarity graphs. The MVGNN model is trained in a self-supervised fashion so that it can effectively learn from different sources of data and further fuse them to derive region embeddings. After that, random forest regression models are leveraged to estimate economic development from the combination of handcrafted features and these learned embeddings. Extensive experiments have been conducted in estimating the GDP of the mainland China, and the results demonstrate the effectiveness of the proposed method which achieves an overall R^2 of 82.69% in cross-validation from the open-source geospatial data. We also analyze the importance of different features, and show the dominant contribution of NTL, and also confirm the indispensable contributions from RDN and MSI. The proposed method advances the goal of acquiring timely and accurate socioeconomic variables through open accessible geospatial data, and is promising to extend to more variables and other geographic regions to support smart governance.

The proposed method has both merits and demerits. For the merits, 1) the research is totally based on open available geospatial data, which ensures the generalization ability of the solution for economic development evaluation; 2) the proposed MVGNN model is trained in a self-supervised manner which alleviates the need for extra labeled data, and it can make full use of the inner relations between regions and their corresponding features to boost the results of using only intra-region features. For the demerits, 1) it is difficult to interpret the learned embeddings, so that it counters the demand from socioeconomic analysis, which usually require high interpretability of models and results; 2) the design of the MVGNN model architecture and learning objectives are crucial for effective region embedding learning, but also dependent on specific applications and data. In summary, though with some challenges, the proposed machine learning-based economic development estimation method shows a great potential for real-world applications and the remained issues deserve further studies.

ACKNOWLEDGEMENTS

This work was supported in part by the Tencent WeChat Rhino Bird under Grant JR-WXG-2021131, the National Natural Science Foundation of China under Grant 42101472, 42071360, and 42001393, the Hong Kong Polytechnic University Start-Up under Grant BD41, and the China Postdoctoral Science Foundation under Grant 2021M692163.

REFERENCES

- Abitbol, J. L., Karsai, M., 2020. Interpretable socioeconomic status inference from aerial imagery through urban patterns. *Nature Machine Intelligence*, 2(11), 684–692.
- Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1798–1828.
- Cao, J., Li, Q., Tu, W., Gao, Q., Cao, R., Zhong, C., 2021. Resolving Urban Mobility Networks from Individual Travel Graphs Using Massive-Scale Mobile Phone Tracking Data. *Cities*, 110, 103077.
- Cao, R., Tu, W., Yang, C., Li, Q., Liu, J., Zhu, J., Zhang, Q., Li, Q., Qiu, G., 2020. Deep learning-based remote and social sensing data fusion for urban region function recognition. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163, 82–97.
- Cao, R., Zhu, J., Tu, W., Li, Q., Cao, J., Liu, B., Zhang, Q., Qiu, G., 2018. Integrating Aerial and Street View Images for Urban Land Use Classification. *Remote Sensing*, 10(10), 1553.
- Chen, D., Tu, W., Cao, R., Zhang, Y., He, B., Wang, C., Shi, T., Li, Q., 2022. A Hierarchical Approach for Fine-Grained Urban Villages Recognition Fusing Remote and Social Sensing Data. *International Journal of Applied Earth Observation and Geoinformation*, 106, 102661.
- Chen, Q., Ye, T., Zhao, N., Ding, M., Ouyang, Z., Jia, P., Yue, W., Yang, X., 2020. Mapping China’s regional economic activity by integrating points-of-interest and remote sensing data with random forest. *Environment and Planning B: Urban Analytics and City Science*, 2399808320951580.
- Dong, L., Ratti, C., Zheng, S., 2019. Predicting neighborhoods’ socioeconomic attributes using restaurant data. *Proceedings of the National Academy of Sciences*, 116(31), 15447–15452.
- Elvidge, C. D., Zhizhin, M., Ghosh, T., Hsu, F.-C., Taneja, J., 2021. Annual Time Series of Global VIIRS Nighttime Lights Derived from Monthly Averages: 2012 to 2019. *Remote Sensing*, 13(5), 922.
- Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., Fei-Fei, L., 2017. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences*, 201700035.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale Geospatial Analysis for Everyone. *Remote Sensing of Environment*, 202, 18–27.

- Hamilton, W. L., Ying, Z., Leskovec, J., 2017. Inductive Representation Learning on Large Graphs. *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, 1024–1034.
- Huang, Z., Li, S., Gao, F., Wang, F., Lin, J., Tan, Z., 2021. Evaluating the Performance of LBSM Data to Estimate the Gross Domestic Product of China at Multiple Scales: A Comparison with NPP-VIIRS Nighttime Light Data. *Journal of Cleaner Production*, 129558.
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790–794.
- Kipf, T. N., Welling, M., 2017. Semi-Supervised Classification with Graph Convolutional Networks. *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *nature*, 521(7553), 436–444.
- Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., Shi, L., 2015. Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. *Annals of the Association of American Geographers*, 105(3), 512–530.
- Lütkepohl, H., Xu, F., 2012. The role of the log transformation in forecasting economic variables. *Empirical Economics*, 42(3), 619–638.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems 32*, 8024–8035.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Steele, J. E., Sundsøy, P. R., Pezzulo, C., Alegana, V. A., Bird, T. J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., de Montjoye, Y.-A., Iqbal, A. M., Hadiuzzaman, K. N., Lu, X., Wetter, E., Tatem, A. J., Bengtsson, L., 2017. Mapping poverty using mobile phone and satellite data. *Journal of The Royal Society Interface*, 14(127), 20160690.
- Tomor, Z., Meijer, A., Michels, A., Geertman, S., 2019. Smart governance for sustainable cities: findings from a systematic literature review. *Journal of urban technology*, 26(4), 3–27.
- Tu, W., Cao, J., Gao, Q., Cao, R., Fang, Z., Yue, Y., Li, Q., 2020. Sensing Urban Dynamics by Fusing Multi-sourced Spatiotemporal Big Data. *Geomatics and Information Science of Wuhan University*, 45(12), 1875–1883.
- Tu, W., Cao, R., Yue, Y., Zhou, B., Li, Q., Li, Q., 2018. Spatial variations in urban public ridership derived from GPS trajectories and smart card data. *Journal of Transport Geography*, 69, 45–57.
- Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y., 2018. Graph Attention Networks. *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*.
- Wang, M., Zheng, D., Ye, Z., Gan, Q., Li, M., Song, X., Zhou, J., Ma, C., Yu, L., Gai, Y., Xiao, T., He, T., Karypis, G., Li, J., Zhang, Z., 2020. Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks. *arXiv:1909.01315 [cs, stat]*.
- Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., Burke, M., 2020. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications*, 11(1), 2583.
- Zhang, Z., Cui, P., Zhu, W., 2020. Deep learning on graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*.