Location-Centered House Price Prediction: A Multi-Task Learning Approach

GUANGLIANG GAO, Jiangsu Police Institute
ZHIFENG BAO, RMIT University
JIE CAO, Nanjing University of Finance and Economics
A. K. QIN and TIMOS SELLIS, Swinburne University of Technology

Accurate house prediction is of great significance to various real estate stakeholders such as house owners, buyers, and investors. We propose a location-centered prediction framework that differs from existing work in terms of data profiling and prediction model. Regarding data profiling, we make an important observation as follows – besides the in-house features such as floor area, the location plays a critical role in house price prediction. Unfortunately, existing work either overlooked it or had a coarse grained measurement of locations. Thereby, we define and capture a fine-grained location profile powered by a diverse range of location data sources, including transportation profile, education profile, suburb profile based on census data, and facility profile. Regarding the choice of prediction model, we observe that a variety of approaches either consider the entire data for modeling, or split the entire house data and model each partition independently. However, such modeling ignores the relatedness among partitions, and for all prediction scenarios, there may not be sufficient training samples per partition for the latter approach. We address this problem by conducting a careful study of exploiting the **Multi-Task Learning (MTL)** model. Specifically, we map the strategies for splitting the entire house data to the ways the tasks are defined in MTL, and select specific MTL-based methods with different regularization terms to capture and exploit the relatedness among tasks. Based on real-world house transaction data collected in Melbourne, Australia, we design extensive experimental evaluations, and the results indicate a significant superiority of MTL-based methods over state-of-the-art approaches. Meanwhile, we conduct an in-depth analysis on the impact of task definitions and method selections in MTL on the prediction performance, and demonstrate that the impact of task definitions on prediction performance far exceeds that of method selections.

CCS Concepts: • **Information systems** → **Location based services**; *Data analytics*; • **Human-centered computing** → *Accessibility technologies;*

Authors' addresses: G. Gao, Jiangsu Police Institute, Department of Computer Information and Cyber Security, Nanjing, China; email: gaoguangliang@jspi.cn; Z. Bao, RMIT University, School of Computing Technologies, Melbourne, Australia; email: zhifeng.bao@rmit.edu.au; J. Cao (corresponding author), Nanjing University of Finance and Economics, College of Information Engineering, Nanjing, China; email: caojie690929@163.com; A. K. Qin and T. Sellis, Swinburne University of Technology, Department of Computer Science and Software Engineering, Melbourne, Australia; emails: {kqin, tsellis}@swin.edu.au.

ACM Transactions on Intelligent Systems and Technology, Vol. 13, No. 2, Article 32. Publication date: January 2022.

32

1 INTRODUCTION

With the improvement of people's living standards, the demand for houses increases. In the United States,[1] house sales have grown by 34% in the last decade and reached a record high of 5.64 million last year. In Australia,[2] house sales have increased by 36% since 2013. House price prediction [22, 44] has therefore attracted widespread attention because the prediction outcomes can help various real estate stakeholders to make more informed decisions. For example, buyers would use house price prediction to search for candidate houses that match their financial capabilities. Similarly, house owners would need it to keep monitoring the market and seek the best opportunity for house selling. Moreover, real estate sales agents also rely on house price prediction to help customers better find out market trends, and the accuracy of prediction has become an important criterion for measuring the credibility of house sales agents.

House price is considered to be related to various house features. In general, these features can be grouped into two categories: non-geographical features, such as the number of bedrooms and floor space area, and; geographical features, such as the distance to the city center and the quality of nearby schools. Therefore, research on house price prediction defines appropriate models to fit various features to predict house price. The hedonic price model [1, 27, 42] proposed from the perspective of economics is the most typical representative, and has been studied extensively in the literature of house price prediction [5, 23, 53, 58]. However, it is primarily used for analyzing the relationship between house price and house features, where it typically adopts regression methods. In recent years, with the extensive application of machine learning in various fields, house price prediction through more machine learning methods, such as **ANN (Artificial Neural Network)** [45], **SVM (Support Vector Machine)** [18, 48], and **AdaBoost (Adaptive Boosting)** [39], has also received more and more attention.

By carefully examining these studies, it reveals that modeling is usually carried out from a global view, i.e., implemented on the entire house data directly. However, the prediction performance of such approaches is not satisfactory, especially as the scale of the house data increases. The primary reason is that the weight of various house features is always assumed to be constant in the formulation of house price from the global view. In fact, however, the impact of house features on house price varies from house to house. For example, transportation features may have a greater impact on house price in the suburbs than that in urban centers, the difference in house price between school districts and non-school districts will be mainly concentrated on education features.

Numerous studies [4, 7, 16, 35] have also demonstrated that the different locations of houses and the surrounding communities of houses have a significant impact on the price of houses, though at a very coarse granularity. As a consequence, the researchers began splitting the entire house data into several partitions and designing prediction models for each partition individually. However,

---

[1]https://www.statista.com/statistics/226144/us-existing-home-sales/.
[2]https://tradingeconomics.com/australia/new-home-sales.

no matter which strategy is used for splitting, there are still two challenges that limit further improvements in the prediction performance along this line: (1) independent modeling ignores the relatedness among the partitions; (2) each model uses only the data of its corresponding partition. Data sparsity, especially in partitions where the number of original data is insufficiently small, is a serious problem.

To address the aforementioned issues, we exploit the framework of **Multi-Task Learning (MTL)** to model the house price prediction problem. MTL [2, 6, 14, 54] is a type of transfer learning, when there are multiple related tasks and each task has limited training samples, the model can enhance the performance by extracting and utilizing shared information among different tasks. MTL has been applied in many applications, including time series analysis [10], stock selection [17], event forecasting [21, 55], disease progression [13, 57], water quality prediction [34], and driving behaviour analysis [9, 19]. By using MTL, we inherit the principle of modeling house price prediction problem from local views because the entire house data is grouped into multiple tasks. Moreover, MTL considers the relatedness among tasks, and the shortcomings of insufficient samples and independent modeling can also be addressed.

There are two key points in MTL: (i) how to define tasks and (ii) how to characterize the relatedness among tasks. In general, the first point is data-driven, different strategies for defining tasks will result in different task sets, and the relatedness among tasks will also be determined. We split the entire house data and define each partition as a task after splitting, where the splitting strategy is equivalent to the strategy for defining tasks in MTL. The second point is method-driven – once a multi-task problem is formulated, the design of a specific method will indicate the level of learning. We select MTL-based methods with regularization terms to capture and utilize the relatedness among tasks, and different regularization terms represent different levels of learning. To this end, in this paper, we present an organized study on exploiting MTL for the problem of house price prediction. The specific contributions of our work are listed as follows:

- We formulate house price prediction as an MTL problem. To the best of our knowledge, this work is the first to consider using MTL to solve the, house price prediction problem, which provides a new perspective on the study of house price prediction. Additionally, our work also enriches the application fields of MTL.
- We define and capture a fine-grained location profile powered by a diverse range of location data sources. We observe that the location of house plays a critical role in house price prediction. Therefore, we focus on enriching the location-driven house features and grouping them into four profiles for further fine-grained, namely house, education, transportation, and facility.
- We demonstrate the superiority of MTL-based methods on real-world house data. Based on our house data, we evaluate the prediction performance of MTL-based methods and six state-of-the-art approaches. Experimental results show that MTL-based methods consistently outperform these competing approaches.
- We conduct an in-depth analysis on the impact of task definitions and method selections on prediction performance. We design two categories of strategies to define tasks and select three MTL-based methods with different regularization terms. By comparing their performance, we reveal that the impact of task definitions on prediction performance far exceeds that of method selections.

The remainder of this paper is organized as follows. We review the related work in Section 2, and Section 3 provides a comprehensive profiling of our location-centered house data. The MTL-based house price prediction is described in Section 4, and Section 5 shows our experimental results. Finally, we conclude the paper and give guidelines for our future work in Section 6.

2 RELATED WORK

A house is usually treated as a heterogeneous good, defined by a bundle of utility bearing features [15, 26]. Therefore, the house price can be considered as a quantitative representation of a set of these features. Over the past decades, a large number of studies have examined the relationship between house price and house features. For example, Król [23] investigated the relationship between the price of an apartment and its significant features based on the results of hedonic analysis in Poland. The work of [53] discussed which house features have negative or positive effects upon the value of the house in Turkey. Kryvobokov and Wilhelmsson [24] derived the weights of the relative importance of location features that influence the market values of apartments in Donetsk, Ukraine. Ottensmann et al. [36] compared measures of location using both distances and travel time, to the CBD, and to multiple employment centers to understand how residence location relative to employment location affects house prices in Indianapolis, Indiana, USA. Ozalp and Akinci [37] determined the housing and environmental features that were effective on residential real estate sale prices in Artvin, Turkey. Xue et al. [52] analyzed the house price in Xi' an, China. They found that urban public transportation (walking accessibility, bus accessibility, and metro accessibility) has an important impact on house price changes, and the importance of these three accessibility features to house price prediction is more than 20%.

Based on a broad study of the relationship between house price and various house features, house price prediction approaches output the estimated house price by inputting house features. According to the basic idea whether it relies on the global model or not, one can divide the existing house price prediction methods into two categories.

The global model predicts the house price on a range of its constituent features, and is usually modeled directly across the entire house data. Much work has been done along this line. Selim [45] examined the determinants of house prices in Turkey by using the hedonic model and demonstrated that artificial neural network can be a better alternative for prediction of the house prices in Turkey. Gu et al. [18] proposed a hybrid of **genetic algorithm and support vector machine approach (G-SVM)** to predict house prices, the cases from China showed the prediction ability of the method. Wang et al. [48] proposed a novel model based on SVM to predict the average house price in different years, meanwhile, the authors demonstrated that the PSO algorithm can effectively determine the parameters of SVM. Park and Bae [39] developed a general prediction model based on machine learning methods such as C4.5, RIPPER, Naive Bayesian, and AdaBoost and compared their classification accuracy performance. Peng et al. [41] collected second-hand house data in Chengdu, China, and thus demonstrated that the XGboost method has good generalization ability and robustness in second-hand house price prediction. Pai and Wang [38] employed a genetic algorithm to determine parameters and introduced feature selection, which improved the prediction performance of the four machine learning methods: least squares support vector regression, classification and regression tree, general regression neural networks, and backpropagation neural networks. The reason why global modeling can be widely recognized in house price prediction is obvious, because it is easy to apply and can reveal the comparative size of effects of various features on house price. However, global modeling ignores the impact of house location and surroundings on house price, so the prediction performance is often unsatisfactory as the scale of the house data increases.

Recent studies have focused on house price prediction from local views and have gradually become a serious alternative and extension of conventional house price modeling approaches. Among these studies, Bourassa et al. [4] compared alternative methods for taking spatial dependence into account in house price prediction, and concluded that a geostatistical model with disaggregated submarket variables performed the best. Case et al. [7] investigated the hedonic

model and three spatial models, and out-of-sample prediction accuracy was used for comparison purposes. Their prediction results indicated the importance to incorporate the nearest neighbor transactions for predicting housing values. Gerek [16] designed two different adaptive neuro-fuzzy approaches for prediction, namely **ANFIS with grid partition (ANFIS-GP)** and **ANFIS with sub clustering (ANFIS-SC)**, and the results indicated that the performance of ANFIS-GP was slightly better than that of ANFIS-SC. Montero et al. [35] considered parametric and semi-parametric spatial hedonic model variants to reflect the spatial effects in house price. The proposed model is represented as a mixed model that accounts for spatial autocorrelation, spatial heterogeneity, and (smooth and nonparametrically specified) nonlinearities using penalized splines methodology. The results obtained suggest that the nonlinear models are the best strategies for house price prediction. Law et al. [28] first proposed that images of the urban environment at both the street and aerial level can capture urban qualities that are difficult to directly quantify, such as the prestige of the neighborhood or the aesthetics of the street, and then improved the estimation of house price by using non-linear and linear methods to fuse these cues with conventional models of house pricing. Das et al. [12] developed **geo-spatial network embedding (GSNE)** to learn the geo-spatial context of the neighborhood of a house. The results showed that GSNE can consistently and significantly improve the prediction performance regardless of the downstream regression model.

Although the house price prediction problem has been widely studied, our work is significantly different from most of the existing work in the following aspects. First, as shown in Table 1, our house data (more details will be provided in the next section) is more adequate in terms of house transaction records and house features than the house data used in the literature, which enables us to better explore the impact of various house features on house price and study the house price prediction problem. Second, the existing studies, whether global or local, can be summarized into traditional **Single Task Learning (STL)**, while we use MTL to study house price prediction. MTL has been applied in various fields. For example, Emrani et al. [13] employed a multi-task learning regression framework for prediction of Parkinson's disease progression. Liu et al. [34] forecasted the water quality of a station over the next few hours, and used a multi-task multi-view learning method to fuse multiple datasets from different domains. Chen et al. [9] designed a multi-task learning framework consisting of graph representation learning and semi-supervised learning to identify the fine-grained driving styles in the temporal dimension accurately. However, to the best of our knowledge, our work is the first one adopting MTL for house price prediction. In STL, each task is considered to be independent and learn individually. In MTL, tasks learn simultaneously by using relatedness between each other. Combined with our previous statements, the principle of MTL can help us better model house price prediction problem from a finer-grained local view.

3 HOUSE DATA PROFILING

In this section, we elaborate a comprehensive real estate dataset used in this study. We first describe four location-centered profiles, each of which contain a variety of features. Then we analyze the house data to further demonstrate the motivations for introducing MTL to predict house prices.

3.1 Data Description

We utilize house data from Melbourne[3] [29–32], one of the largest cities in Australia, as an example to understand the domain situation. The data includes houses sold in Melbourne's metropolitan area since 2011. We extract 136,394 house transaction records from Jan. 2015 to Jan. 2018 to generate the dataset for this study. The house features are divided into four profiles

---

[3]http://www.realestate.com.au/.

Table 1. Summary of Our Data Profiles and Comparisons with Most of the Existing Work

	Data Examples	References																Our
		[5]	[23]	[53]	[45]	[39]	[4]	[7]	[35]	[15]	[24]	[36]	[37]	[25]	[1]	[11]	[38]	
Scale of data	100	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✔	✗	✗	✗	✗
	1,000	✗	✗	✔	✔	✗	✗	✗	✔	✗	✔	✗	✔	✗	✔	✗	✗	✗
	10,000	✗	✔	✗	✗	✔	✔	✔	✗	✔	✗	✗	✗	✗	✗	✔	✔	✗
	100,000	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔
House profile	floor area, number of bedrooms	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
	geo-information, address, suburb	✗	✗	✔	✔	✔	✔	✔	✗	✗	✔	✔	✗	✔	✔	✔	✔	✔
	air condition, water, heating, views	✔	✔	✔	✔	✔	✔	✔	✔	✔	✗	✗	✔	✗	✔	✗	✔	✔
Education profile	nearby schools	✗	✗	✔	✗	✔	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✗	✔
	school districts	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✗	✗	✔
	school rankings	✗	✗	✗	✗	✔	✗	✗	✗	✗	✗	✗	✔	✗	✗	✗	✗	✔
Transportation profile	nearby public transport	✗	✗	✔	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✔	✗	✗	✔
	travel time to work	✗	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✔	✗	✗	✗	✔
Facility profile	hospitals, shops	✗	✗	✔	✗	✗	✗	✗	✗	✔	✔	✔	✔	✔	✗	✗	✗	✔
	distance to nearest hospitals	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✔	✔	✔	✗	✗	✗	✔

Our location-centered house data is more comprehensive in terms of house transaction records and house features than those used in the literature.

to better observe the impact of different types of features on house price. In particular, this comprehensive dataset contains four location-centered profiles: house profile, education profile, transportation profile, and facility profile.

House profile. In this group, we choose seven relevant features about the house itself. The number of bedrooms, the number of bathrooms, and land size are the most common basic features. The number of parking spaces is gradually related to price with the booming house market, thus we introduce this feature. Considering the possible correlation between price and income, we also include family weekly income as an independent feature. Moreover, geographical information has a great impact on house price. In the vast majority of cases, consumers are more concerned with general locations than with detailed addresses, so we choose two regional features, SAID[4] (Statistical Area Level, the geographical areas for the processing and release of Australian census data) and postal code, to reflect the impact of geographical information. In particular, SAID has four levels, from SA4 to SA1, which can continuously refine the area where the houses are located. Table 2 reports the number of partitions at different statistical area levels and postal code.

Education profile. In recent years, educational resources have received more and more attention, thus we either find the exact primary and secondary school districts[5] (top 20% schools) that each house belongs to, or map the house with its nearest primary and secondary schools (not top 20% schools). The corresponding school rankings provided by BetterEducation[6] based on academic results (number of students enrolled in Victorian Certificate of Education, percentage of students securing 40% marks, etc) are used as four features to examine the impact on house price. There are two reasons for choosing 20% as the boundary: (1) The higher the school ranking, the clearer the concept of school district, which helps us to better analyze the impact of education profile on house

---

[4]http://www.abs.gov.au/.
[5]http://melbourneschoolzones.com/.
[6]http://bettereducation.com.au/.

Table 2. Number of Partitions at
Different Statistical Area Levels

Statistical area level	#Partitions
SA4	17
SA3	65
SA2	100
Postal code	547
SA1	10703

SA4 has the coarsest-grained split of an area
and the smallest number of partitions, while
SA1 has the finest-grained split of an area
and the largest number of partitions.

price. (2) With the decline in school rankings, people may pay more attention to other aspects of schools, such as the distance to the school. The ranking loses its meaning to a certain extent.

Transportation profile. Since transportation networks have always been of great concern, we set up six features: (1) the distance and walking time from each house to its nearest train station using Google Maps API,[7] (2) the distance and travel time between each pair of train stations based on the **GTFS**[8] (**General Transit Feed Specification**) data, and (3) the distance and self-driving time from the location of the house to the city center, i.e., to the main **Central Business District (CBD)**, using Google Maps API.

Facility profile. Proximity to facilities such as shops, hospitals, clinics, and supermarkets may affect the house price as well. Therefore, we introduce four different features based upon these four facilities to describe the distance between a given house and the nearest four facilities, where the distance calculation uses Google Maps API.

Table 3 summarizes all selected house features and their definitions. Meanwhile, some important statistics for our dataset can be also found in Table 3. The value we want to predict, and the house price at a given sales time, is the target. Figure 1(a) describes the trend of average house prices in each month during the three years. It is clear that average house prices fluctuate over time, which indicates that house price is time-sensitive. Considere that location is one of the important factors in shaping the price of a house. Without loss of generality, we choose the partitions based on SA4 here. As shown in Figure 1(b), the average house price in each partition is different. Therefore, spatial dependence also clearly exists in house price.

3.2 Data Insight

Because of the time sensitivity and spatial dependence of house price as described above, the researchers [4, 7, 25, 35] intuitively split house data and model each partition individually. However, such modeling usually does not deal well with the house price prediction problem. The two challenges that affect the prediction performance have been mentioned in the previous sections, and we now elaborate on them by analyzing the house data.

The first reason is that no matter which splitting strategy is adopted, it is difficult to ensure that the number of samples allocated to the generated partitions is optimal. We choose partitions based on SA4 and POSTCODE as two cases, respectively. Figure 2 shows the number of samples for each partition during the three years. We can find that the number of samples in the partitions

---

[7]http://developers.google.com/maps/.
[8]http://www.data.vic.gov.au/data/dataset/.

Table 3. List of Selected House Features and Statistics of Our House Data

| Category	Name of features	Descriptions	Min.	Max.	Median	Std. Dev.
House	#BEDROOM	The number of bedrooms	1	5	–	–
	#BATHROOM	The number of bathrooms	1	3	–	–
	#PARKING	The number of parking spaces	1	5	–	–
	LANDSIZE	The land size of the house (m^2)	340	2500	708.79	291.73
	INCOME	Family weekly income (K)	935	2836	1553.91	387.96
	SAID	Statistical area level	SA1	SA4	–	–
	POSTCODE	Postal code	3000	3996	–	–
Education	SCH_DISTRICT	School district where the house is located	1	100	–	–
	SCH_NEAREST	School closest to the house	1	500	–	–
	PSCH_RANK	The ranking of primary school	3	500	–	–
	SSCH_RANK	The ranking of secondary school	1	500	–	–
Transportation	STN_DIS	Distance to the nearest train station (m)	23	5040	2026.22	1186.78
	STN_TIME	Walking time to the nearest train station (min)	1	126	34.75	20.36
	CBD_PUB_DIS	The train distance to city center (m)	1300	82600	35239.52	14594.42
	CBD_PUB_TIME	The train time to city center (min)	6	101	48.33	16.64
	CBD_PRI_DIS	The self-driving distance to city center (m)	1245	83497	35109.47	14678.34
	CBD_PRI_TIME	The self-driving time to city center (min)	10	120	47.98	18.58
Facility	SHOP_DIS	The distance to nearest shopping center (m)	5	4999	1643.19	961.48
	HOSPITAL_DIS	The distance to nearest hospital (m)	15	5000	1832.83	1102.66
	GP_DIS	The distance to nearest clinic (m)	8	4999	957.21	745.01
	MARKET_DIS	The distance to nearest supermarket (m)	25	5000	1452.63	847.71
SALES_DATE		The sales time of the houses	Jan. 2015	Jan. 2018	–	–
HOUSE_PRICE		The sales price of the houses (K)	262	2090	680.54	353.97
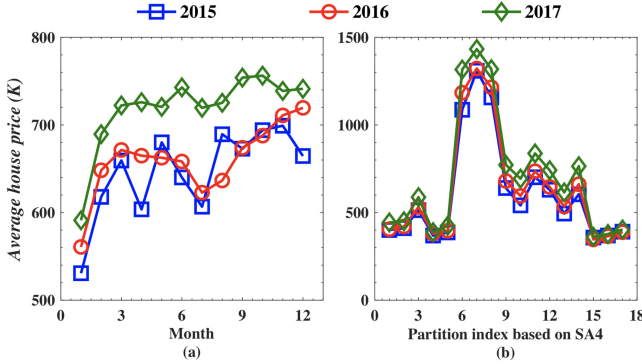
Fig. 1. Three-year average house price trend by Month and SA4, respectively. Average house price fluctuates over time and varies by partition.

varies greatly. Moreover, with the further refinement of the split, the number of samples in each partition is generally small, especially in areas where the number of original samples is insufficient, the impact of splitting is more pronounced, which reduces the prediction performance (please refer to the empirical results for the STL-based approaches in Tables 6 and 7).

Another important reason is that independent modeling ignores the relatedness between partitions. In fact, partitions are not completely independent, and there are many explicit or implicit connections between them. For example, two independent partitions in the geographical space may belong to the same school district. As a result, when analyzing the impact of the school on house price, the two partitions should be merged rather than separated. In order to better illustrate our intuition, we count the number of features that belong to multiple partitions based on SA4 and POSTCODE, respectively. The results are summarized in Table 4. The phenomenon that features belong to multiple partitions is very common and becomes more apparent with further splitting.
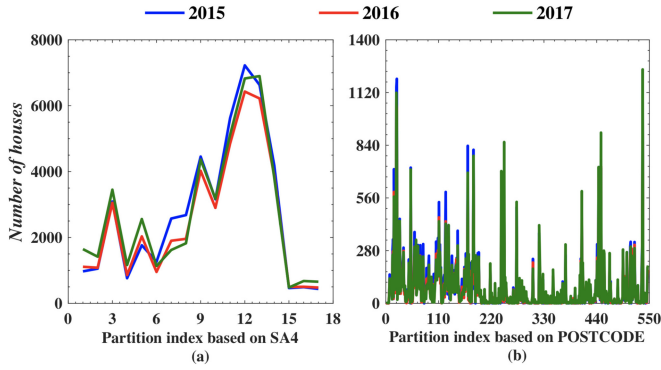
Fig. 2. Number of samples per partition based on SA4 and POSTCODE, respectively. The number of samples in the partitions varies greatly. As the split is further refined, the number of samples is generally small.

Table 4. Number of Features That Belong to Only One Partition (=1), Two Partitions (=2), and at Least Three Partitions (≥3) Based on SA4 and POSTCODE, Respectively

Category	SA4			POSTCODE		
	=1	=2	≥3	=1	=2	≥3
PRIMARY SCHOOL	46	50	225	34	30	257
SECONDARY SCHOOL	40	32	148	24	23	173
STATION	123	70	25	11	33	174
SHOP	289	64	17	84	84	202
HOSPITAL	328	83	13	94	87	243
GP	2133	112	7	1386	633	233
MARKET	361	62	12	65	101	269

It also indicates the importance of preserving the relatedness between partitions for modeling the problem of house price prediction.

Based on the above data analysis, we conclude the two challenges that affect the prediction performance of existing approaches from the data view. Meanwhile, considering that the inherent relatedness of the partitions due to the consistency of the house features, we cast the house price prediction problem into an MTL problem. By using the framework of MTL, we can not only reflect the relatedness between partitions well, but also solve the dilemma of insufficient samples in some partitions.

4 MTL-BASED HOUSE PRICE PREDICTION

In this section, we first describe our house price prediction problem and provide preliminaries about MTL. Then we formulate the problem of MTL for house price prediction. To facilitate our illustration, the notations used throughout this paper are presented in Table 5.

4.1 Problem Description and Preliminaries

Consider that the house data contains P partitions and in each time interval t (e.g., month, quarter), each partition p has $m_t^p \in \mathbb{Z}$ house transaction records. Therefore, given a timestamp τ, the objective

Table 5. Notations and Explanations

Notations	Explanations
p, P	a task (partition), all tasks (partitions)
t, τ	a time interval, a timestamp
h	prediction horizon
m_t^p	number of samples for task p in a time interval t
$m_{k,t}^p$	number of samples for task p in k time intervals
D	number of features
x_p, y_p	training input and output for task p
w_p, W	feature weight parameter for task p, weight matrix for all tasks
L, Ω	empirical loss and regularization error
$r_{p,q}$	Ratio of average house price for two tasks p and q

of our house price prediction is to predict the price of houses that appear in each partition from τ to $\tau + h$ based on the historical transaction data collected before τ, where h is a specific prediction horizon. It is not difficult to find that multiple timestamps refer to multiple predictions.

In this paper, we comprehend each house price prediction as a problem that is jointly learned by multiple tasks. There are two key points in the formulation of such an MTL: defining tasks and characterizing the relatedness among tasks. In terms of the first point, it is usually determined by the information in the data used and the specific application scenario. For example, in the widely used case of predicting student performance in schools, the 139 schools involved were defined as 139 tasks. In terms of the second point, methods with regularization terms are generally used to formulate the relatedness among tasks, and different regularization terms represent different ways of formulation. For example, the $l_{2,1}$-norm means that all tasks share a common set of representations.

Here, we define the partitions contained in the house data as tasks, and each partition is aligned with a task. Thus, the two key points in MTL-based house price prediction become how to construct the P tasks and what methods are used to model the relatedness among these P tasks. Next, we will introduce our strategies for each of these two issues.

4.2 Task Definition

In the literature [34, 55, 57], tasks are usually uniquely identified and given directly, thus the impact of task definitions on performance has also not received sufficient attention. However, as an essential element of using MTL, there are various ways to define tasks, even in the same application scenario. For example, in the case of predicting student performance in schools as described above, in addition to defining each school as a task, we can also group adjacent schools into one task. Obviously, this change in task definition will affect the subsequent steps of the MTL formulation. Therefore, for our house price prediction problem, we propose two categories of strategies for task definition, and explore the impact of different task definitions on prediction performance.

4.2.1 Defining Tasks Based on One Single Profile. Existing STL-based house price prediction approaches [4, 25, 35] can be grouped into this category, usually from the perspective of geographical factors. For example, we can define the area of a postal code as one task. One reason for this

definition is that geographical factors are the most intuitive expression of house price differences, so task definition along this line is the most common one. Another reason is that the house data used by the above studies includes limited features that affect house price, making it difficult to find more ways to define tasks. Considering that our data set contains a (much richer) variety of house features, we conduct a wide range of task definition attempts and select the following four cases as representatives of this category.

In the house profile, we use the statistical area levels to split the house data. The four area levels indicate four splitting strategies. One partition in each level is defined as a task in the corresponding level. For example, there are 17 partitions at the SA4 level, so we can get 17 tasks at this level. Similarly, we also consider task definition based on postal code, where one postal code partition corresponds to a task.

In the education profile, we employ the concept of school districts to split the house data, and each school district serves as one task. The primary and secondary school districts lead to two splitting strategies. In addition, we note that the attention to the school district is closely related to the ranking of the school. Therefore, we mainly focus on the school districts of top 20% schools.

In the transportation profile, there seems to be no obvious perspective to define the tasks compared to the previous two feature profiles. Considering that distance/time is an important criterion for measuring the situation of transportation, we define each train station as the centroid of a task and determine the scope of the task by specifying the distance/time threshold to the train station. Therefore, houses with the same train station and the distance/time to the train station (that do not exceed a pre-specified threshold) belong to the same task.

In the facility profile, we group houses according to the similarity of the facilities. Specifically, there are four types of facilities in our data set, so we give four criteria for measuring similarity, namely SHARED1 (share one type), SHARED2 (share two types), SHARED3 (share three types), and SHARED4 (share four types). Thus, given a criterion and the names of the facilities, such as SHARED1, market, we group houses that have the same market into one task, SHARED2, shop and hospital, we group houses that have the same shop and hospital into one task.

4.2.2 Defining Tasks Based on Multiple Profiles. The above task definitions extract one feature profile at a time as a guideline. Such definitions not only constrain the differences in the specific feature profiles of the houses in each task, but also ensure the relatedness among tasks in terms of these profiles. However, the relatedness among tasks that depend on one feature profile are relatively weak. By introducing more feature profiles as a guideline to defining tasks, the relatedness among the resulting tasks can be strengthened, but with the refinement of the definition, the number of houses in each task may be insufficient meanwhile.

In order to guarantee a sufficient number of houses necessary for each task and to enhance the relatedness between tasks, we consider six cases by combining any two of the above four task definitions, where each partition obtained corresponds to a task:

(1) statistical regions and school districts;
(2) statistical regions and transportation areas;
(3) statistical regions and neighbor facilities;
(4) school districts and transportation areas;
(5) school districts and neighbor facilities;
(6) transportation areas and neighbor facilities.

4.3 The MTL Model

In this paper, we regard the MTL-based house price prediction problem as a multi-task regression problem. Given a timestamp τ, we extract the transaction records in the previous k time intervals

to construct the training input $x_p \in \mathbb{R}^{m^P_{k,t} \times D}$ and output $y_p \in \mathbb{R}^{m^P_{k,t} \times 1}$ for each task p. Here, $m^P_{k,t}$ is the number of transaction records in k time intervals, D is the number of house features, and y_p includes the actual house price. Thus, for each task, we want to infer a linear function f_p where $f_p(x_p) = x_p w_p$ and $w_p \in \mathbb{R}^{D \times 1}$. Let us denote $W = \{w_1, w_2, \ldots, w_P\} \in \mathbb{R}^{D \times P}$ as the weight matrix over P tasks. One typical MTL model for estimating W is to minimize the following objective function:

$$\arg \min_W L(W) + \Omega(W), \tag{1}$$

where $L(W) = \sum_{p=1}^{P} ||x_p w_p - y_p||_F^2$, and $\Omega(W)$ is the regularization term that controls the information shared among tasks.

There can be various choices of regularization terms to fit the above objective function, and the specific choice is based on the identification of the relatedness among the defined tasks. In our house price prediction problem, it is too strict or even unrealistic to use only one type of regularization because there are various task definitions. Moreover, the purpose of this paper is to study the application of MTL for the house price prediction problem, rather than designing a sophisticated MTL-based method to fit all task definitions. Therefore, we choose three different regularization terms to model the relatedness among tasks, and thus investigate the impact of different MTL-based methods on prediction performance.

The first way is to constrain the models of all tasks to be close to each other. The l_1-norm regularization is widely used because it can reduce model complexity and feature learning by introducing sparsity into the model, a common simplification of l_1-norm in MTL is that the parameter controlling the sparsity is shared among all tasks. Then the objective function can be defined as:

$$\arg \min_W \sum_{p=1}^{P} \left\| x_p w_p - y_p \right\|_F^2 + \theta_1 \left\| W \right\|_1, \tag{2}$$

where θ_1 is the parameter that controls sparsity.

The second way is to assume all tasks share a common yet latent representation, such as a common set of features, a common subspace. This motivates the group sparsity, the $l_{2,1}$-norm regularization is usually used to implement this assumption. The objective function can be expressed as:

$$\arg \min_W \sum_{p=1}^{P} \left\| x_p w_p - y_p \right\|_F^2 + \theta_1 \left\| W \right\|_{2,1}, \tag{3}$$

where θ_1 is the parameter controlling the group sparsity.

Besides these two most common methods, we also consider ensuring the relatedness between tasks by adding graph regularization. Specifically, the structural relatedness among P tasks is represented by a graph, each task is defined as a node, and two nodes are connected by a weighted edge. The overall objective function can be described as:

$$\arg \min_W \sum_{p=1}^{P} \left\| x_p w_p - y_p \right\|_F^2 + \theta_1 \sum_{p,q=1}^{P} r_{p,q} \left\| w_p - w_q \right\|_F^2 + \theta_2 ||W||_{2,1}, \tag{4}$$

where $r_{p,q}$ is the connection strength between nodes (tasks) p and q, θ_1 and θ_2 are parameters for graph regularization and group sparsity, respectively.

We define $r_{p,q}$ as the ratio of the average house price for the two nodes to measure the structural relatedness between tasks p and q. Intuitively, the larger the $r_{p,q}$, the graph regularization term will force w_p to be closer to w_q. Meanwhile, the closer w_p and w_q are, the more similar the average

house price for these two nodes should be, i.e., $r_{p,q}$ tends to 1. Thus, we compute $r_{p,q}$ as follows:

$$r_{p,q} = \frac{\min\{averageprice_p, averageprice_q\}}{\max\{averageprice_p, averageprice_q\}}. \tag{5}$$

All the optimization problems above can be solved by using the accelerated gradient descent method [43]. In this paper, we apply the implementation of accelerated gradient descent method included in the MALSAR [56] package to efficiently solve the optimization. As for the trade-off parameters in Equations (2), (3), and (4), MALSAR uses 5-fold cross-validation to determine them.

5 EXPERIMENTS

In this section, we comprehensively evaluate the performance of our methodology in comparison with six STL-based approaches for house price prediction, and make an in-depth analysis of the impact of task definitions and method selections on our MTL-based house price prediction, and examine the prediction performance of each task individually to demonstrate:

- MTL-based methods can significantly outperform the STL-based approaches. (Section 5.3)
- The impact of task definitions on prediction performance far exceeds that of method selections. (Section 5.4)
- The advantages of using MTL to preserve the relatedness among tasks. (Section 5.5)

5.1 Training and Test Sets

The data set studied in this experiment has been detailed in Section 3. Given the three-year records of sold house data, we first obtain several tasks based on a strategy for defining tasks. Then we evaluate the performance of our methodology by predicting the price of the houses in each task per month, i.e., h is one month. In each prediction, we create a training set and a test set for each task. On the training set, the data samples are the information tabulation (features, sales price) of the houses that appeared in the previous months of the selected month, i.e., t is month. On the test set, we use the data samples in the selected month. Meanwhile, we use the semi-logarithmic form of the sales price to fit the data in the training and test sets.

In order to analyze the impact of the number of months used for training on prediction, we use the split of the whole data set according to SA3 as a case and define the December of each year as the prediction horizon. We first count the number of samples for each task in December and extract those tasks whose number of samples exceeds the first quartile (1/4) of the data distribution as the test set. Then we collect data samples from the previous one month to the previous eleven months as different training sets and fit them using linear regression to predict the price of the houses in December. The prediction performance is shown in Figure 3(a), from which we have two main observations. (1) The prediction error increases with the number of previous months. This trend is not obvious in 2017, as mentioned earlier, the fluctuation of the average monthly house price in 2017 is not significant. (2) The previous three months are the most appropriate one, especially in 2017. In 2015 and 2016, this does not seem to be the best option. However, as shown in Figure 2 at Section 3, with a refinement of the split, the number of samples per partition continues to decrease. Obviously, if we further consider the number of samples per month in each partition, the data sparsity issue will be more serious. Meanwhile, considering that the shorter months may lead to contingency in prediction performance, it is necessary to use the data from the previous three months as training data. In the following experiments, unless stated otherwise, we choose to use the previous three months in the training set, i.e., k is 3.

In summary, we have a total of 36 predictions, each of which is one month in three years. We supplemented the house data for the last three months of 2014 to ensure that the first three months
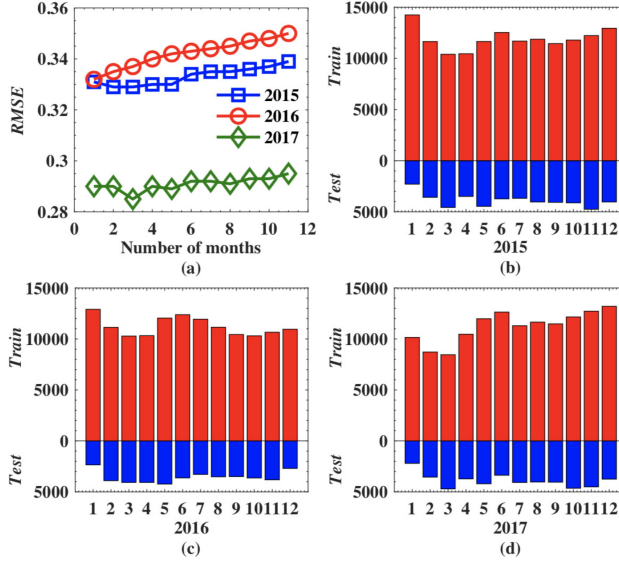
Fig. 3. Figure (a) shows the impact of the number of previous months used in the training set on the prediction performance for December each year. Figures (b)-(d) show the number of samples in the training and test sets for each prediction.

of 2015 are predictable. The number of samples in the training and test sets for each prediction is shown from Figure 3(b) to Figure 3(d). All experiments were performed on a PC equipped with an Intel quad-core i7 processor (2.6 GHz) and 16GB memory. The running time for training is less than 110 seconds.

5.2 Performance Metrics

To evaluate the prediction performance of different methods, we employ two categories of evaluation metrics.

The first category is two widely used prediction evaluation measures, **Root Mean Squared Error** [51] **(RMSE)** which considers the square root of the average of all prediction error, and **Mean Absolute Error** [8] **(MAE)** which calculates the average of absolute error for each predicted result. For each task, these two measures are defined as follows:

$$RMSE_p = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} (y_i - \hat{y}_i)^2},$$

$$MAE_p = \frac{1}{N_p} \sum_{i=1}^{N_p} |y_i - \hat{y}_i|, \tag{6}$$

where y_i $(\hat{y}_i)$ is the actual (estimated) house price, N_p is the number of observations for each task p. It is expected that a better prediction will result in a smaller value for both measures selected.

Given a task definition, we record the metric values for the methods in all tasks and use the mean as the performance of the methods in a prediction. Similarly, the mean of performance in all predictions is used as the overall performance of the methods under this task definition. To make performance comparison in a statistically sound way, we also use Wilcoxon's rank sum test at the significant level of 0.05 [50].

The second category is the Win-Loss-Draw records [49]. It is a comparative descriptive statistic. The three values, respectively, are the number of data sets for which method m_1 obtained better, worse, or equal performance outcomes than method m_2 on a given measure. These records compare the performance of the two selected methods on different data sets and indicate a systematic underlying advantage to one of the methods.

5.3 Evaluation on Overall Prediction Performance

We compare our methodology with the following six baseline approaches:

(1) Lasso Regression [47] (Lasso) is a linear model with l_1-norm regularization. It tends to prefer solutions that use a small number of features to improve the prediction accuracy and interpretability of the model. Increasing the penalty parameter in Lasso will produce more zeros in the feature coefficients.

(2) Ridge Regression [20] (Ridge) is also a linear model, which gives more penalties by adding l_2-norm regularization. Although such a formulation loses some information and reduces the accuracy of the fit, the regression coefficients obtained are more realistic and reliable. The penalty parameter in Ridge controls the extent of the loss.

(3) Support Vector Regression [3] (SVR) is the natural extension of large margin kernel methods of support vector machine used for classification to regression analysis. It seeks to minimize an upper bound of the generalization error instead of the empirical error. The choice of kernel function, kernel coefficient, and penalty parameter determines the prediction performance of SVR.

(4) AdaBoost [46] is an algorithm to improve performance by using an ensemble of weak learners to create a strong learner. The output of the weak learners is combined into a weighted sum that represents the final output of the boosted learner. There are many variants of the AdaBoost algorithm depending on the choice of weak learners. Here, we intuitively use AdaBoost.R2 (AdaR2).

(5) Random Forest [33] (RF) also belongs to the category of ensemble learning algorithms. Each base learner has the same weight of influence, reducing overfitting by combining them together. Decision tree is often used as the base learner of the ensemble. In this study, the base learner is actually a regression tree. The number of trees in the ensemble is considered to be the design parameter of RF.

(6) Hedonic Price Model [1] (Hedonic) is a revealed preference model of valuation. It regresses the house price on a range of its constituent features and is usually calibrated using the ordinary least squares method. Hedonic has become a popular approach to modeling house price as it is easy to apply and can reveal the comparative size of effects of various features on house price.

We use the Python scikit-learn library [40] to implement the above approaches. There are various parameters and options in each approach. Specifically, we set the kernel function of SVR to a Gaussian function, and the weak learner in AdaR2 is defined as a decision tree regressor. As for the remaining parameters of each approach, we use 5-fold cross-validation to determine them.

Tables 6 and 7 present the values of RMSE and MAE for all tested approaches under different task definitions, and the last row shows the mean of the RMSE and MAE for these approaches. We use the best performing approach under each task definition as a benchmark and compare the performance of the approaches by Wilcoxon's rank sum test, from which we have four main observations. (1) The RMSE and MAE values for all approaches fluctuate insignificantly. This indicates that the performance of these approaches is stable. (2) In the comparison of MTL-based methods, the impact of method changes on overall performance is limited. Although MTL-Graph

Table 6. Evaluation of RMSE Among All Tested Approaches Under Two Categories of Task Definitions

Category	Task definition strategies	MTL-L1	MTL-L21	MTL-Graph	Lasso	Ridge	SVR	AdaR2	RF	Hedonic
House	SA4	0.219	0.226	**0.192***	0.268	0.260	0.338	0.257	0.244	0.272
	SA3	**0.191**	**0.191**	**0.189***	0.306	0.280	0.278	0.272	0.225	0.294
	SA2	**0.203**	**0.207**	**0.190***	0.350	0.310	0.273	0.303	0.223	0.341
	POSTCODE	**0.203**	**0.206**	**0.191***	0.383	0.366	0.244	0.365	0.214	0.544
Education	PSCH_RANK [1, 10]	0.244	**0.228**	**0.209***	0.361	0.326	0.325	0.318	0.255	0.376
	PSCH_RANK [1, 20]	0.259	**0.244**	**0.219***	0.345	0.296	0.351	0.299	0.281	0.363
	PSCH_RANK [1, 30]	0.267	0.253	**0.207***	0.347	0.296	0.372	0.295	0.260	0.347
	PSCH_RANK [1, 40]	0.247	**0.239**	**0.202***	0.345	0.302	0.364	0.297	0.263	0.343
	PSCH_RANK [1, 50]	0.262	**0.252**	**0.210***	0.348	0.316	0.353	0.308	**0.257**	0.347
	SSCH_RANK [1, 10]	0.263	**0.248**	**0.229***	**0.252**	**0.250**	0.377	**0.249**	0.274	**0.265**
	SSCH_RANK [1, 20]	**0.242**	**0.229**	**0.213***	0.255	0.300	0.337	0.285	0.257	0.271
	SSCH_RANK [1, 30]	**0.235**	**0.227**	**0.211***	0.350	0.302	0.340	0.268	**0.247**	0.365
	SSCH_RANK [1, 40]	0.262	0.251	**0.218***	0.351	0.307	0.341	0.285	0.277	0.367
	SSCH_RANK [1, 50]	0.258	0.247	**0.216***	0.354	0.314	0.339	0.290	0.258	0.369
Transportation	STN_DIS [0, 1000]	**0.213**	**0.211***	0.221	0.351	0.294	**0.245**	0.270	0.255	0.378
	STN_DIS [0, 2000]	**0.202**	**0.201***	**0.204**	0.314	0.313	0.244	0.311	0.230	0.327
	STN_DIS [0, 3000]	**0.197***	**0.197**	0.198	0.323	0.322	0.243	0.322	0.221	0.331
	STN_DIS [0, 4000]	**0.194***	0.195	0.194	0.295	0.293	0.250	0.292	**0.217**	0.300
	STN_DIS [0, 5000]	**0.195***	0.196	0.197	0.306	0.303	0.248	0.303	**0.220**	0.310
Facility	SHARED1_S (SHOP)	**0.199**	**0.200**	**0.193***	0.328	0.323	0.262	**0.224**	**0.221**	0.328
	SHARED1_H (HOSPITAL)	**0.193**	**0.192***	0.192	0.329	0.324	0.263	**0.224**	**0.221**	0.330
	SHARED1_G (GP)	**0.201**	0.201	**0.194***	0.371	0.361	0.229	0.260	**0.226**	0.376
	SHARED1_M (MARKET)	**0.189**	**0.188***	0.193	0.316	0.314	0.249	0.216	0.222	0.312
	SHARED2_S, H	**0.193**	**0.192***	0.194	0.350	0.332	0.243	**0.221**	**0.222**	0.365
	SHARED2_S, G	**0.186**	**0.185***	0.192	0.338	0.336	0.223	0.232	0.228	0.347
	SHARED2_S, M	**0.192**	**0.191***	0.194	0.341	0.359	0.240	0.235	**0.222**	0.354
	SHARED2_H, G	**0.188**	**0.187***	0.194	0.328	0.338	0.224	0.234	0.228	0.332
	SHARED2_H, M	**0.190**	**0.189***	0.193	0.349	0.337	0.237	0.225	0.223	0.358
	SHARED2_G, M	**0.188**	**0.187***	0.194	0.335	0.337	0.225	0.234	0.231	0.343
	SHARED3_S, H, G	**0.186**	**0.185***	0.192	0.347	0.332	0.219	0.228	0.230	0.357
	SHARED3_S, H, M	**0.190**	**0.189***	0.194	0.346	0.364	0.232	0.260	0.227	0.357
	SHARED3_S, G, M	**0.186**	**0.185***	0.192	0.332	0.332	0.221	0.219	0.230	0.344
	SHARED3_H, G, M	**0.186**	**0.185***	0.194	0.346	0.334	0.221	0.230	0.230	0.358
	SHARED4_S, H, G, M	**0.185**	**0.184***	0.192	0.351	0.323	0.217	0.240	0.231	0.367
House Education	SA3, PSCH_RANK [1, 40]	**0.194***	0.201	0.194	0.476	0.476	0.266	0.476	**0.206**	0.499
	SA3, SSCH_RANK [1, 30]	**0.195**	0.205	**0.193***	0.443	0.378	0.273	0.377	**0.208**	0.462
House Transportation	SA3, [0, 4000]	**0.190***	0.192	0.195	0.373	0.353	0.240	0.350	**0.196**	0.398
House Facility	SA3, M	**0.190***	0.190	0.191	0.356	0.339	0.242	0.337	**0.194**	0.372
	SA3, S, M	**0.193**	**0.190***	0.195	0.452	0.382	0.237	0.378	**0.197**	0.483
	SA3, S, H, M	**0.191**	**0.188***	0.194	0.544	0.483	0.230	0.478	0.197	0.579
Education Transportation	PSCH_RANK [1, 40], [0, 4000]	**0.190***	0.191	0.195	0.587	0.559	0.234	0.454	**0.197**	0.604
	SSCH_RANK [1, 30], [0, 4000]	**0.191***	0.194	0.197	0.562	0.437	0.240	0.431	**0.199**	0.591
Education Facility	PSCH_RANK [1, 40], M	0.191	**0.189***	0.195	0.562	0.437	0.240	0.431	0.198	0.591
	PSCH_RANK [1, 40], S, M	**0.193**	**0.190***	0.196	0.582	0.584	0.234	0.433	0.200	0.608
	PSCH_RANK [1, 40], S, H, M	**0.191**	**0.187***	0.195	0.639	0.621	0.226	0.418	0.200	0.665
	SSCH_RANK [1, 30], M	**0.193***	0.193	0.193	0.617	0.614	0.246	0.470	**0.197**	0.633
	SSCH_RANK [1, 30], S, M	**0.193**	**0.191***	0.195	0.614	0.612	0.237	0.479	**0.198**	0.630
	SSCH_RANK [1, 30], S, H, M	**0.192**	**0.190***	0.195	0.629	0.617	0.231	0.445	0.200	0.654
Transportation Facility	[0, 4000], M	**0.192**	**0.188***	0.198	0.444	0.394	0.228	0.388	0.198	0.443
	[0, 4000], S, M	**0.193**	**0.188***	0.199	0.685	0.516	0.224	0.462	0.200	0.685
	[0, 4000], S, H, M	**0.193**	**0.188***	0.199	0.678	0.661	0.219	0.471	0.203	0.676
Mean of overall performance		0.205	0.203	0.200	0.403	0.378	0.263	0.320	0.220	0.422

The first two columns show the specific task definitions, such as House, SA4 means that one partition at the SA4 level is a task. Education, PSCH_RANK [1, 20] means that the top 20 primary schools, one school district is a task. Similarly, SSCH_RANK [1, 40] means that of the top 40 schools, one is school district is a task. Transportation, STN_DIS [0, 4000] means that one station is a task, and houses within 4, 000 meters belong to each task. Facility, SHARED2_S, M means that houses with the same shop and market belong to the same task. The last nine columns show the RMSE for all tested approaches. Those in bold and asterisk indicate the benchmark under each task definition, and those in bold-only indicate that the p-value for the rank sum test is greater than 0.05.

Table 7. Evaluation of MAE Among All Tested Approaches Under Two Categories of Task Definitions

Category	Task definition strategies	MTL-			Lasso	Ridge	SVR	AdaR2	RF	Hedonic
		L1	L21	Graph						
House	SA4	0.169	0.177	**0.148***	0.208	0.203	0.268	0.200	0.190	0.208
	SA3	**0.149**	**0.150**	**0.147***	0.231	0.215	0.220	0.208	0.175	0.233
	SA2	**0.157**	**0.161**	**0.147***	0.261	0.236	0.216	0.230	0.173	0.268
	POSTCODE	**0.156**	**0.161**	**0.148***	0.288	0.278	0.193	0.277	**0.166**	0.401
Education	PSCH_RANK [1, 10]	**0.178**	**0.171**	**0.150***	0.273	0.235	0.257	0.236	0.185	0.284
	PSCH_RANK [1, 20]	0.193	0.187	**0.164***	0.270	0.225	0.278	0.235	0.205	0.279
	PSCH_RANK [1, 30]	0.199	0.194	**0.157***	0.272	0.228	0.295	0.231	0.197	0.268
	PSCH_RANK [1, 40]	0.187	0.184	**0.154***	0.269	0.269	0.289	0.232	0.202	0.267
	PSCH_RANK [1, 50]	0.197	0.192	**0.160***	0.275	0.242	0.281	0.236	0.195	0.271
	SSCH_RANK [1, 10]	0.197	**0.192**	**0.178***	0.195	**0.191**	0.299	**0.190**	0.205	0.205
	SSCH_RANK [1, 20]	**0.184**	**0.178**	**0.166***	0.194	0.215	0.267	0.207	0.195	0.208
	SSCH_RANK [1, 30]	**0.181**	**0.178**	**0.166***	0.276	0.225	0.277	0.209	0.189	0.286
	SSCH_RANK [1, 40]	0.198	0.194	**0.167***	0.274	0.230	0.270	0.219	0.206	0.286
	SSCH_RANK [1, 50]	0.195	0.192	**0.165***	0.278	0.239	0.267	0.226	0.197	0.289
Transportation	STN_DIS [0, 1000]	**0.167***	0.167	0.174	0.260	0.229	0.197	0.222	0.199	0.290
	STN_DIS [0, 2000]	**0.158***	0.159	0.160	0.237	0.238	0.195	0.238	0.180	0.243
	STN_DIS [0, 3000]	**0.155***	0.155	0.155	0.239	0.243	0.194	0.244	**0.173**	0.246
	STN_DIS [0, 4000]	0.153	0.154	**0.152***	0.222	0.223	0.197	0.223	**0.170**	0.225
	STN_DIS [0, 5000]	**0.153***	0.155	0.154	0.229	0.230	0.196	0.230	**0.172**	0.226
Facility	SHARED1_S (SHOP)	**0.154**	**0.156**	**0.149***	0.252	0.243	0.208	**0.175**	**0.172**	0.253
	SHARED1_H (HOSPITAL)	**0.150**	**0.151**	**0.149***	0.252	0.243	0.209	**0.175**	**0.171**	0.255
	SHARED1_G (GP)	**0.155**	**0.156**	**0.149***	0.277	0.275	0.184	0.203	0.178	0.280
	SHARED1_M (MARKET)	**0.150***	**0.150**	0.152	0.230	0.233	0.197	**0.170**	**0.171**	0.226
	SHARED2_S, H	**0.151***	0.152	0.151	0.275	0.264	0.195	**0.172**	**0.174**	0.286
	SHARED2_S, G	**0.150***	0.150	0.153	0.266	0.265	0.181	0.184	0.181	0.273
	SHARED2_S, M	**0.151***	0.151	0.151	0.270	0.287	0.191	0.188	**0.173**	0.281
	SHARED2_H, G	0.151	**0.150***	0.155	0.259	0.267	0.182	0.187	0.181	0.264
	SHARED2_H, M	**0.149***	0.150	0.151	0.275	0.268	0.189	0.178	**0.174**	0.285
	SHARED2_G, M	**0.150***	0.150	0.155	0.265	0.268	0.182	0.187	0.184	0.272
	SHARED3_S, H, G	**0.151***	0.151	0.155	0.272	0.264	0.180	0.180	0.185	0.284
	SHARED3_S, H, M	**0.151**	0.151	**0.135***	0.271	0.277	0.187	0.207	0.178	0.283
	SHARED3_S, G, M	**0.150***	0.150	0.155	0.264	0.264	0.180	**0.171**	0.185	0.272
	SHARED3_H, G, M	0.151	**0.150***	0.156	0.272	0.265	0.181	0.181	0.185	0.284
	SHARED4_S, H, G, M	**0.151**	**0.150***	0.156	0.274	0.243	0.178	0.190	0.186	0.287
House Education	SA3, PSCH_RANK [1, 40]	**0.151***	0.158	0.151	0.397	0.396	0.212	0.396	0.164	0.412
	SA3, SSCH_RANK [1, 30]	**0.151**	0.160	**0.150***	0.378	0.312	0.217	0.311	0.165	0.381
House Transportation	SA3, [0, 4000]	**0.149***	0.152	0.153	0.281	0.269	0.192	0.266	**0.156**	0.304
House Facility	SA3, M	**0.148***	0.149	0.149	0.263	0.255	0.192	0.253	**0.152**	0.280
	SA3, S, M	**0.151***	0.151	0.152	0.357	0.298	0.189	0.295	**0.156**	0.401
	SA3, S, H, M	0.152	**0.151***	0.153	0.480	0.390	0.185	0.387	**0.157**	0.506
Education Transportation	PSCH_RANK [1, 40], [0, 4000]	**0.151***	0.152	0.154	0.511	0.492	0.187	0.358	**0.157**	0.535
	SSCH_RANK [1, 30], [0, 4000]	**0.151***	0.154	0.155	0.504	0.371	0.191	0.366	**0.158**	0.523
Education Facility	PSCH_RANK [1, 40], M	**0.150***	0.150	0.153	0.504	0.371	0.192	0.366	**0.156**	0.523
	PSCH_RANK [1, 40], S, M	0.153	**0.152***	0.155	0.509	0.508	0.189	0.368	0.159	0.537
	PSCH_RANK [1, 40], S, H, M	0.153	**0.151***	0.155	0.541	0.521	0.184	0.337	0.160	0.561
	SSCH_RANK [1, 30], M	**0.151***	0.152	0.151	**0.518***	0.515	0.196	0.380	**0.155**	0.548
	SSCH_RANK [1, 30], S, M	**0.152***	0.152	0.154	0.514	0.511	0.191	0.387	**0.158**	0.544
	SSCH_RANK [1, 30], S, H, M	0.153	**0.152***	0.155	0.525	0.518	0.187	0.347	0.159	0.550
Transportation Facility	[0, 4000], M	**0.152**	**0.150***	0.156	0.345	0.307	0.184	0.302	**0.156**	0.342
	[0, 4000], S, M	**0.159**	**0.152***	0.159	0.573	0.418	0.181	0.374	0.160	0.571
	[0, 4000], S, H, M	**0.156**	**0.154***	0.161	0.569	0.556	0.179	0.383	0.164	0.567
Mean of overall performance		0.160	0.158	0.155	0.323	0.300	0.210	0.254	0.175	0.340

The first two columns show the specific task definitions, such as House, SA4 means that one partition at the SA4 level is a task. Education, PSCH_RANK [1, 20] means that the top 20 primary schools, one school district is a task. Similarly, SSCH_RANK [1, 40] means that the top 40 secondary schools, one school district is a task. Transportation, STN_DIS [0, 4000] means that one station is a task, and houses within 4,000 meters belong to each task. Facility, SHARED2_S, M means that houses with the same shop and market belong to the same task. The last nine columns show the MAE for all tested approaches. Those in-bold and asterisk indicate the benchmark under each task definition, and those in bold-only indicate that the p-value for the rank sum test is greater than 0.05.

enhances the relatedness between tasks through graph regularization, which may help improve performance, we can see from the experimental results that this improvement is not significant. (3) In the comparison of STL-based methods, the performance of SVR and RF are similar and better than others, followed by AdaR2, Ridge, Lasso, and Hedonic. Intuitively, as all regressors it tries to fit a line to data by minimizing a cost function. In this scenario, SVR and RF deploy kernel functions and decision trees to make the fit more appropriate. (4) The first three MTL-based methods outperform all the six baseline approaches. It indicates that the tasks are not independent and capturing their relatedness can improve the learning performance.

5.4 Performance Evaluation on Different Task Definitions

We try two categories of task definitions, i.e., single-profile one and multiple-profile one. For each task definition, we evaluated three MTL-based methods to capture the relatedness between tasks in house price prediction. Given an MTL method, we choose the task definition that makes it perform the best as a benchmark, and compare its performance under various task definitions by Wilcoxon's rank sum test.

5.4.1 Task Definitions Based on One Single Profile. The results are summarized in Table 8. It can be clearly seen that no matter which method is involved, the rank sum test results between its performance under various task definitions and its best performance are similar. For MTL-L1, compared to the task definition with the best performance, there are significantly different task definitions that exist in all four single profiles, and are mainly concentrated in the education profile. For MTL-L21, the experimental phenomenon is very similar to MTL-L1. For MTL-Graph, although the differences between the education profile and the best scenario are still evident, the performance of other profiles has improved.

Comparing these experimental results, we can find: (1) Task definitions in the house price prediction can be diverse. Traditional prediction approaches are usually considered from geographical factors, but it can be seen from the experimental results that other factors can also ensure good prediction performance, even beyond geographical factors. (2) Given one single profile to define tasks, the lower the RMSE and MAE values indicate the more accurate the house price prediction based on the remaining profiles. In other words, the profile used for task definition has a relatively limited impact on house price. Conversely, the higher the RMSE and MAE values, the greater the impact of a given profile on house price. As a result, in four categories of single profiles, Education has the greatest impact, followed by House, Transportation, and Facility. (3) The optimal MTL-based methods that correspond to the various task definitions are different. Although MTL-Graph enhances the relatedness between tasks by adding graph regularization, such a setting is too strict in some task definitions, which reduces the prediction performance. For example, in the transportation profile, compared with MTL-Graph, the performance of MTL-L1 and MTL-L21 is available. (4) In terms of prediction performance, the choice of task definitions has a greater impact than the choice of MTL-based methods, such as, the performance of the facility profile is superior to the education profile regardless of which MTL-based method is used.

5.4.2 Task Definitions Based on Multiple Profiles. We choose the following task definitions as representatives of each profile and assemble them in pairs to redefine tasks. The first is SA3 in the statistical regions. The second is PSCH_RANK [1, 40], SSCH_RANK [1, 30] in the primary and secondary school districts, respectively. The third is STN_DIS [0, 4000] in the transportation areas. The fourth is M (MARKET) in SHARED1, S (SHOP), M in SHARED2, and S, H (HOSPITAL), M in SHARED3 in the neighbor facilities. The prediction performance of MTL-based methods under each of the above task definitions is also summarized in Table 8.

Table 8. Evaluation of RMSE and MAE Among Three MTL-based Methods Under Two Categories of Task Definitions

Category	Task definition strategies	RMSE			MAE		
		MTL-L1	MTL-L21	MTL-Graph	MTL-L1	MTL-L21	MTL-Graph
House	SA4	0.219	0.226	**0.192**	**0.169**	**0.177**	**0.148**
	SA3	**0.191**	**0.191**	**0.189***	0.149	0.150	0.147
	SA2	**0.203**	**0.207**	**0.190**	0.157	0.161	0.147
	POSTCODE	**0.203**	**0.206**	**0.191**	0.156	0.161	0.148
Education	PSCH_RANK [1, 10]	**0.244**	**0.228**	**0.209**	0.178	0.171	0.150
	PSCH_RANK [1, 20]	0.259	0.244	0.219	0.193	0.187	0.164
	PSCH_RANK [1, 30]	0.267	0.253	**0.207**	0.199	0.194	**0.157**
	PSCH_RANK [1, 40]	0.247	0.239	0.202	0.187	0.184	0.154
	PSCH_RANK [1, 50]	0.262	0.252	0.210	0.197	0.192	**0.160**
	SSCH_RANK [1, 10]	0.263	0.248	0.229	0.197	0.192	0.178
	SSCH_RANK [1, 20]	**0.242**	**0.229**	**0.213**	**0.184**	**0.178**	**0.166**
	SSCH_RANK [1, 30]	**0.235**	**0.227**	**0.211**	**0.181**	**0.178**	**0.166**
	SSCH_RANK [1, 40]	0.262	0.251	0.218	0.198	0.194	**0.167**
	SSCH_RANK [1, 50]	0.258	0.247	0.216	0.195	0.192	**0.165**
Transportation	STN_DIS [0, 1000]	0.213	0.211	0.221	**0.167**	**0.167**	0.174
	STN_DIS [0, 2000]	**0.202**	**0.201**	**0.204**	0.158	0.159	0.160
	STN_DIS [0, 3000]	**0.197**	**0.197**	**0.198**	0.155	0.155	0.155
	STN_DIS [0, 4000]	**0.194**	**0.195**	**0.194**	0.153	0.154	0.152
	STN_DIS [0, 5000]	**0.195**	**0.196**	**0.197**	0.153	0.155	0.154
Facility	SHARED1_S (SHOP)	0.199	0.200	**0.193**	0.154	0.156	0.149
	SHARED1_H (HOSPITAL)	**0.193**	**0.192**	**0.192**	0.150	0.151	0.149
	SHARED1_G (GP)	0.201	0.201	**0.194**	0.155	0.156	0.149
	SHARED1_M (MARKET)	**0.189**	**0.188**	**0.193**	0.150	0.150	0.152
	SHARED2_S, H	**0.193**	0.192	**0.194**	0.151	0.152	0.151
	SHARED2_S, G	**0.186**	**0.185**	**0.192**	0.150	0.150	0.153
	SHARED2_S, M	**0.192**	**0.191**	**0.194**	0.151	0.151	0.151
	SHARED2_H, G	**0.188**	**0.187**	0.194	0.151	0.150	0.155
	SHARED2_H, M	**0.190**	**0.189**	**0.193**	0.149	0.150	0.151
	SHARED2_G, M	**0.188**	**0.187**	**0.194**	0.150	0.150	0.155
	SHARED3_S, H, G	**0.186**	**0.185**	**0.192**	0.151	0.151	0.155
	SHARED3_S, H, M	**0.190**	**0.189**	**0.194**	0.151	0.151	**0.135***
	SHARED3_S, G, M	**0.186**	**0.185**	**0.192**	0.150	0.150	0.155
	SHARED3_H, G, M	**0.186**	**0.185**	**0.194**	0.151	0.150	0.156
	SHARED4_S, H, G, M	**0.185***	**0.184***	**0.192**	0.151	0.150	0.156
House Education	SA3, PSCH_RANK [1, 40]	**0.194**	**0.201**	**0.194**	**0.151**	**0.158**	**0.151**
	SA3, SSCH_RANK [1, 30]	**0.195**	0.205	0.193	**0.151**	0.160	**0.150**
House Transportation	SA3, [0, 4000]	**0.190**	**0.192**	**0.195**	**0.149**	**0.152**	0.153
House Facility	SA3, M	**0.190**	**0.190**	**0.191**	**0.148***	**0.149***	**0.149**
	SA3, S, M	**0.193**	**0.190**	**0.195**	0.151	0.151	0.152
	SA3, S, H, M	**0.191**	**0.188**	**0.194**	0.152	0.151	0.153
Education Transportation	PSCH_RANK [1, 40], [0, 4000]	**0.190**	**0.191**	**0.195**	0.151	0.152	0.154
	SSCH_RANK [1, 30], [0, 4000]	**0.191**	**0.194**	**0.197**	0.151	0.154	**0.155**
Education Facility	PSCH_RANK [1, 40], M	**0.191**	**0.189**	**0.195**	0.150	0.150	0.153
	PSCH_RANK [1, 40], S, M	**0.193**	**0.190**	**0.196**	0.153	0.152	0.155
	PSCH_RANK [1, 40], S, H, M	**0.191**	**0.187**	**0.195**	0.153	0.151	0.155
	SSCH_RANK [1, 30], M	**0.193**	**0.193**	**0.193**	0.151	0.152	0.151
	SSCH_RANK [1, 30], S, M	**0.193**	**0.191**	**0.195**	0.152	0.152	0.154
	SSCH_RANK [1, 30], S, H, M	**0.192**	**0.190**	**0.195**	0.153	0.152	0.155
Transportation Facility	[0, 4000], M	**0.192**	**0.188**	**0.198**	0.152	**0.150**	0.156
	[0, 4000], S, M	**0.193**	**0.188**	0.199	0.159	**0.152**	**0.157**
	[0, 4000], S, H, M	**0.193**	**0.188**	0.199	0.156	**0.154**	0.161

The first two columns show the specific task definitions, such as House, Transportation: SA3, [0, 4000] means that the tasks are defined based on these two profiles, and the same SA3 and houses within 4000 meters of the same station belong to the same task. The last six columns show the RMSE and MAE for three MTL-based methods. Those in bold and asterisk indicate the benchmark under each method, and those in bold-only indicate that the p-value for the rank sum test is greater than 0.05.

We have the following observations. (1) The performance difference between various task definitions is not significant. This is because the definition based on multiple profiles is fine-grained, hence the tasks obtained are similar, which makes the difference in performance small. (2) Combining any two profiles to define tasks ensures lower RMSE and MAE values. This indicates that once the houses show advantages in multiple aspects, such as school districts and convenient transportation, even if the remaining profiles are unsatisfactory, such as lack of facilities, the impact on house price is also weak. (3) The performance of MTL-L1 and MTL-L21 is closer to MTL-Graph, even better in some task definitions. This somehow indicates that optimizing the performance by enhancing the relatedness among tasks is limited when the task definitions are sufficient. (4) Compared with one single profile, the overall performance based on multiple profiles is generally better. This also indicates that the task definitions have a greater impact on prediction performance than that of the MTL-based methods.

Discussion. Based on the above analysis, we not only validate the impact of task definitions and method selections on prediction performance, but also demonstrate that the impact of task definitions on prediction performance far exceeds that of method selections. In addition, we observe that the impact of various profiles on house price is not constant. There are thresholds for the impact of Transportation and Facility. When the distance from the house to its nearest train station exceeds 3,000 meters, or there are more than two types of facilities nearby, the impact of these two profiles on house price will be weakened. The impact of House and Education is limited. Although the features of the house itself and the school district have always received widespread attention, the impact of these two profiles on house price will be significantly reduced if the combination of profiles is considered. Moreover, we note that the MTL-based method using graph regularization usually performs well when the task definitions are not sufficient (definition based on one single profile). However, when the task definitions are very refined (definition based on multiple profiles), the performance of the MTL-based methods using general regularization (l_1-norm, $l_{2,1}$-norm) is good enough. These deserve further investigations via more data sources.

5.5 Performance Evaluation on Each Task

Why MTL-based prediction methods can achieve good performance? We here take a further look into the MTL model by investigating the quality of prediction for each task in a prediction. In particular, we extract the task definitions based on SA3, SSCH_RANK [1, 30], STN_DIS [0, 4000], and M (MARKET) in SHARED1 as four cases, respectively. In each task definition, we first count the distribution of the number of data samples for all tasks, and then divide the tasks into four groups according to the first quartile (1/4), the second quartile (1/2), and the third quartile (3/4). Finally, we choose MTL-Graph as a benchmark and choose RMSE as the measure to count the Win-Loss-Draw records for all tested approaches in each group.

The results are summarized in Table 9. We can find: (1) In groups with fewer samples, i.e., (0, 1/4) and (1/4, 1/2), the performance of MTL-based methods is significantly better than that of STL-based approaches. This confirms the advantages of the MTL model. (2) In groups with enough samples, i.e., (1/2, 3/4) and (3/4, 1], the performance of MTL-based methods is still good, even in some task definitions, SVR and RF outperform MTL-based methods. However, STL-based approaches ignore the relatedness among tasks, which result in well-behaved approaches in these groups that do not fit into other groups. Therefore, the overall performance of MTL-based methods is superior to them.

In order to better understand the conclusions of the above quantitative evaluation, we illustrate the performance of Lasso, SVR, RF, and MTL-Graph for each task based on four task definitions. As

Table 9. Win/Loss/Draw Records Obtained by Comparing All Tested Approaches with MTL-Graph Based on Four Different Task Definitions

Task definition strategies	Group	MTL-			Lasso	Ridge	SVR	AdaR2	RF	Hedonic
		L1	L21	Graph						
SA3	(0, 1/4]	6/11/0	5/12/0	**9/8/0**	0/17/0	0/17/0	1/16/0	0/17/0	0/17/0	0/17/0
	(1/4, 1/2]	6/9/1	4/12/0	**9/6/1**	0/16/0	0/16/0	0/16/0	0/16/0	0/16/0	0/16/0
	(1/2, 3/4]	**12/4/0**	3/13/0	4/12/0	0/16/0	0/16/0	0/16/0	0/16/0	0/16/0	0/16/0
	(3/4, 1]	**10/4/2**	0/16/0	4/10/2	0/16/0	0/16/0	0/16/0	0/16/0	0/16/0	0/16/0
SSCH_RANK [1, 30]	(0, 1/4]	1/4/0	0/5/0	**3/2/0**	0/5/0	0/5/0	1/4/0	0/5/0	0/5/0	0/5/0
	(1/4, 1/2]	3/2/0	**4/1/0**	1/4/0	0/5/0	1/4/0	0/5/0	1/4/0	1/4/0	0/5/0
	(1/2, 3/4]	**2/1/0**	**2/1/0**	0/3/0	0/3/0	1/2/0	0/3/0	1/2/0	0/3/0	1/2/0
	(3/4, 1]	3/2/0	4/1/0	0/5/0	2/3/0			**4/1/0**	2/3/0	1/4/0
STN_DIS [0, 4000]	(0, 1/4]	**24/29/1**	**24/29/1**	18/35/1	1/53/0	2/52/0	14/39/1	1/53/0	17/36/1	2/52/0
	(1/4, 1/2]	20/34/0	20/34/0	**22/30/2**	0/54/0	0/54/0	6/48/0	0/54/0	21/31/2	0/54/0
	(1/2, 3/4]	25/26/3	25/26/3	15/38/1	0/54/0	0/54/0	2/52/0	0/54/0	**34/20/0**	0/54/0
	(3/4, 1]	18/36/0	18/36/0	**28/24/2**	0/54/0	0/54/0	1/52/1	0/54/0	9/41/4	0/54/0
MARKET	(0, 1/4]	52/53/2	**61/42/4**	17/85/5	1/106/0	2/105/0	26/80/1	3/104/0	47/57/3	2/105/0
	(1/4, 1/2]	39/68/1	**47/57/4**	33/73/2	0/108/0	1/107/0	6/101/1	1/107/0	46/60/2	0/108/0
	(1/2, 3/4]	33/71/2	39/65/2	**44/59/3**	0/106/0	0/106/0	2/104/0	2/104/0	33/69/4	0/106/0
	(3/4, 1]	29/72/5	28/76/2	**51/48/7**	0/106/0	0/106/0	1/105/0	0/106/0	22/78/6	0/106/0

The first two columns show the selected task definitions and the quantile-based groupings, such as SA3, (1/4, 1/2] means that the tasks are defined based on SA3, and the number of samples for the tasks in this group is between the first quartile and the second quartile of the distribution of the number of samples for all tasks. The last nine columns show the Win/Loss/Draw scores for all tested approaches. Those in bold correspond to the best scenarios.

(a) Task definition based on SA3

(b) Task definition based on SSCH_RANK [1, 30]

(c) Task definition based on STN_DIS [0, 4000]
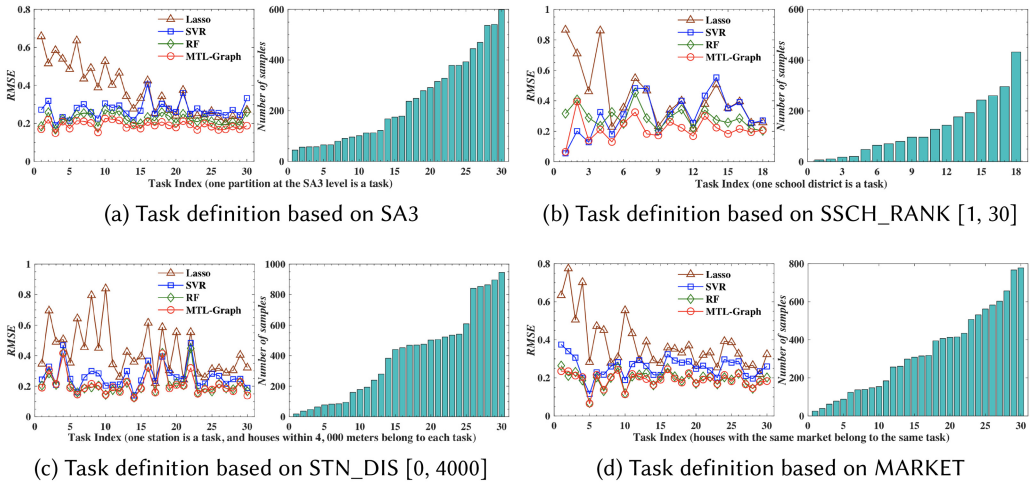

(d) Task definition based on MARKET

Fig. 4. Evaluation of the prediction performance for four methods in each task. The x-axis corresponds to the task indexes under the specified task definition. The y-axis corresponds to the RMSE and the number of samples per task, respectively.

shown in Figure 4, we extract 30 tasks under each task definition and give the number of samples for each task and the prediction performance for four approaches in each task. Taking Figure 4(a) as an example, it can be clearly seen: (1) In tasks with fewer samples, such as task index less than 10, the performance of MTL-based methods is consistently better than that of other methods. (2) The performance gap is narrowed as the number of samples in the tasks increases, such as task index greater than 22, but the performance of MTL-based methods is still acceptable. Note that the results of some SECONDARY SCHOOL cannot be described in Figure 4(b) due to the absence

of house transactions. However, the trends shown are basically consistent with the other three, and since the number of samples for each task is generally small, the advantages of MTL-based methods are more obvious. The only regret is that it is difficult to determine when the performance gap will narrow.

In summary, our MTL-based house price prediction is robust. It guarantees the prediction performance when data samples are sufficient, and when the samples are insufficient, it optimizes the prediction performance by exploiting the relatedness between tasks. As a result, the overall performance is improved.

6   CONCLUSIONS AND FUTURE WORK

In this paper, we carried out a deep research on the application of MTL for the house price prediction problem. In terms of data profiling, we define and capture a fine-grained location profile powered by a diverse range of location data sources. In term of prediction model, there are two key points in the implementation of MTL-based house price prediction: task definitions and method selections. Therefore, we designed two categories of strategies to define tasks based on various house features, and selected three general MTL-based methods with different regularization terms to capture and utilize the relatedness between tasks. By extensive experimental evaluations, we first demonstrated that modeling based on MTL can significantly improve the overall performance of house price prediction. Then we illustrated that the diversity of task definitions is conducive to the MTL formulation for the house price prediction problem. Finally, we revealed that the impact of task definitions on prediction performance far exceeds that of method selections.

In the future, we will not only continue to supply profiles that usually do not change over time, but also explore profiles that are strongly related to time and affect house price, such as year of construction, number of days for sale, and surrounding average price per square meter. The proposed method has preliminarily considered the time dimension, that is, using the previous three months to predict the next month. Based on the above two categories of profiles, we will further employ multi-task feature learning models to enhance the applicability of the proposed method. In addition, we will also try the house recommendations based on the outcomes of the price prediction.

REFERENCES

[1] Alastair S. Adair, Jim N. Berry, and W. Stanley McGreal. 1996. Hedonic modelling, housing submarkets and residential valuation. *Journal of Property Research* 13, 1 (1996), 67–83.

[2] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. 2006. Multi-task feature learning. In *Proceedings of the 20th Annual Conference on Neural Information Processing Systems*. 41–48.

[3] Debasish Basak, Srimanta Pal, and Dipak Chandra Patranabis. 2007. Support vector regression. *Neural Information Processing-Letters and Reviews* 11, 10 (2007), 203–224.

[4] Steven Bourassa, Eva Cantoni, and Martin Hoesli. 2010. Predicting house prices with spatial dependence: A comparison of alternative methods. *Journal of Real Estate Research* 32, 2 (2010), 139–160. DOI:https://doi.org/10.1080/10835547.2010.12091276

[5] Ayse Can. 1992. Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics* 22, 3 (1992), 453–474.

[6] Rich Caruana. 1997. Multitask learning. *Machine Learning* 28, 1 (1997), 41–75.

[7] Bradford Case, John Clapp, Robin Dubin, and Mauricio Rodriguez. 2004. Modeling spatial and temporal house price patterns: A comparison of four models. *Journal of Real Estate Finance and Economics* 29, 2 (2004), 167–191. DOI:https://doi.org/10.1023/B:REAL.0000035309.60607.53

[8] Tianfeng Chai and Roland R. Draxler. 2014. Root mean square error (RMSE) or mean absolute error (MAE)?– arguments against avoiding RMSE in the literature. *Geoscientific Model Development* 7, 3 (2014), 1247–1250. DOI:https://doi.org/10.5194/gmd-7-1247-2014

[9] Chao Chen, Qiang Liu, Xingchen Wang, Chengwu Liao, and Daqing Zhang. 2021. semi-Traj2Graph: Identifying fine-grained driving style with GPS trajectory data via multi-task learning. *IEEE Transactions on Big Data* (2021), 1–15. DOI:https://doi.org/10.1109/TBDATA.2021.3063048

[10] Boris Chidlovskii. 2017. Multi-task learning of time series and its application to the travel demand. *CoRR* abs/1712.08164 (2017).

[11] Jakob A. Dambon, Fabio Sigrist, and Reinhard Furrer. 2021. Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. *Spatial Statistics* 41 (2021), 100470. DOI : https://doi.org/10.1016/j.spasta.2020.100470

[12] Sarkar Snigdha Sarathi Das, Mohammed Eunus Ali, Yuan-Fang Li, Yong-Bin Kang, and Timos Sellis. 2020. Boosting house price predictions using geo-spatial network embedding. *CoRR* abs/2009.00254.

[13] Saba Emrani, Anya McGuirk, and Wei Xiao. 2017. Prognosis and diagnosis of Parkinson's disease using multi-task learning. In *Proceedings of the 23rd International Conference on Knowledge Discovery and Data Mining*. ACM, 1457–1466. DOI : https://doi.org/10.1145/3097983.3098065

[14] Theodoros Evgeniou and Massimiliano Pontil. 2004. Regularized multi-task learning. In *Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining*. ACM, 109–117. DOI : https://doi.org/10.1145/1014052.1014067

[15] Gang-Zhi Fan, Seow Eng Ong, and Hian Chye Koh. 2006. Determinants of house price: A decision tree approach. *Urban Studies* 43, 12 (2006), 2301–2315. DOI : https://doi.org/10.1080/00420980600990928

[16] Ibrahim Halil Gerek. 2014. House selling price assessment using two different adaptive neuro-fuzzy techniques. *Automation in Construction* 41 (2014), 33–39. DOI : https://doi.org/10.1016/j.autcon.2014.02.002

[17] Joumana Ghosn and Yoshua Bengio. 1996. Multi-task learning for stock selection. In *Proceedings of the 10th Annual Conference on Neural Information Processing Systems*. 946–952.

[18] Jirong Gu, Mingcang Zhu, and Liuguangyan Jiang. 2011. Housing price forecasting based on genetic algorithm and support vector machine. *Expert Systems with Applications* 38, 4 (2011), 3383–3386. DOI : https://doi.org/10.1016/j.eswa.2010.08.123

[19] Suiming Guo, Chao Chen, Jingyuan Wang, Yaxiao Liu, Ke Xu, Zhiwen Yu, Daqing Zhang, and Dah Ming Chiu. 2020. ROD-revenue: Seeking strategies analysis and revenue prediction in ride-on-demand service using multi-source urban data. *IEEE Transactions on Mobile Computing* 19, 9 (2020), 2202–2220. DOI : https://doi.org/10.1109/TMC.2019.2921959

[20] Arthur E. Hoerl and Robert W. Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12, 1 (1970), 55–67.

[21] Natasha Jaques, Sara Taylor, Ehimwenma Nosakhare, Akane Sano, and Rosalind Picard. 2016. Multi-task learning for predicting health, stress, and happiness. In *NIPS Workshop on Machine Learning for Healthcare*.

[22] John F. Kain and John M. Quigley. 1970. Measuring the value of housing quality. *J. Amer. Statist. Assoc.* 65, 330 (1970), 532–548.

[23] Anna Król. 2015. Application of hedonic methods in modelling real estate prices in Poland. In *Data Science, Learning by Latent Structures, and Knowledge Discovery*. 501–511. DOI : https://doi.org/10.1007/978-3-662-44983-7_44

[24] Marko Kryvobokov and Mats Wilhelmsson. 2007. Analysing location attributes with a hedonic model for apartment prices in Donetsk, Ukraine. *International Journal of Strategic Property Management* 11, 3 (2007), 157–178. DOI : https://doi.org/10.1080/1648715X.2007.9637567

[25] Michael Kuntz and Marco Helbich. 2014. Geostatistical mapping of real estate prices: An empirical comparison of kriging and cokriging. *International Journal of Geographical Information Science* 28, 9 (2014), 1904–1921. DOI : https://doi.org/10.1080/13658816.2014.906041

[26] Hakan Kusan, Osman Aytekin, and Ilker Özdemir. 2010. The use of fuzzy logic in predicting house selling price. *Expert Systems with Applications* 37, 3 (2010), 1808–1813. DOI : https://doi.org/10.1016/j.eswa.2009.07.031

[27] Kelvin J. Lancaster. 1966. A new approach to consumer theory. *Journal of Political Economy* 74, 2 (1966), 132–157.

[28] Stephen Law, Brooks Paige, and Chris Russell. 2019. Take a look around: Using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology* 10, 5 (2019), 54:1–54:19. DOI : https://doi.org/10.1145/3342240

[29] Mingzhao Li, Zhifeng Bao, Farhana Murtaza Choudhury, and Timos Sellis. 2018. Supporting large-scale geographical visualization in a multi-granularity way. In *Proceedings of the 11th International Conference on Web Search and Data Mining*. ACM, 767–770. DOI : https://doi.org/10.1145/3159652.3160587

[30] Mingzhao Li, Zhifeng Bao, Timos Sellis, and Shi Yan. 2016. Visualization-aided exploration of the real estate data. In *Databases Theory and Applications*. Springer, 435–439. DOI : https://doi.org/10.1007/978-3-319-46922-5_34

[31] Mingzhao Li, Zhifeng Bao, Timos Sellis, Shi Yan, and Rui Zhang. 2018. HomeSeeker: A visual analytics system of real estate data. *Journal of Visual Languages and Computing* 45 (2018), 1–16. DOI : https://doi.org/10.1016/j.jvlc.2018.02.001

[32] Mingzhao Li, Farhana Murtaza Choudhury, Zhifeng Bao, Hanan Samet, and Timos Sellis. 2018. ConcaveCubes: Supporting cluster-based geographical visualization in large data scale. *Computer Graphics Forum* 37, 3 (2018), 217–228. DOI : https://doi.org/10.1111/cgf.13414

[33] Andy Liaw, Matthew Wiener, et al. 2002. Classification and regression by RandomForest. *R News* 2, 3 (2002), 18–22.

[34] Ye Liu, Yu Zheng, Yuxuan Liang, Shuming Liu, and David S. Rosenblum. 2016. Urban water quality prediction based on multi-task multi-view learning. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*. 2576–2581.

[35] José-María Montero, Román Mínguez, and Gema Fernández-Avilés. 2018. Housing price prediction: Parametric versus semi-parametric spatial hedonic models. *Journal of Geographical Systems* 20, 1 (2018), 27–55. DOI : https://doi.org/10.1007/s10109-017-0257-y

[36] John R. Ottensmann, Seth Payton, and Joyce Man. 2008. Urban location and housing prices within a hedonic model. *Journal of Regional Analysis and Policy* 38, 1 (2008), 19–35. DOI : https://doi.org/10.22004/ag.econ.132338

[37] Ayse Yavuz Ozalp and Halil Akinci. 2017. The use of hedonic pricing method to determine the parameters affecting residential real estate prices. *Arabian Journal of Geosciences* 10, 24 (2017), 535–548. DOI : https://doi.org/10.1007/s12517-017-3331-3

[38] Ping-Feng Pai and Wen-Chang Wang. 2020. Using machine learning models and actual transaction data for predicting real estate prices. *Applied Sciences* 10, 17 (2020), 5832–5843. DOI : https://doi.org/10.3390/app10175832

[39] Byeonghwa Park and Jae Kwon Bae. 2015. Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications* 42, 6 (2015), 2928–2934. DOI : https://doi.org/10.1016/j.eswa.2014.11.040

[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, Oct (2011), 2825–2830.

[41] Zhen Peng, Qiang Huang, and Yincheng Han. 2019. Model research on forecast of second-hand house price in Chengdu based on XGboost algorithm. In *Proceedings of the 11th International Conference on Advanced Infocomm Technology*. IEEE, 168–172. DOI : https://doi.org/10.1109/ICAIT.2019.8935894

[42] Sherwin Rosen. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* 82, 1 (1974), 34–55.

[43] Sebastian Ruder. 2016. An overview of gradient descent optimization algorithms. *CoRR* abs/1609.04747.

[44] Rainer Schulz and Axel Werwatz. 2004. A state space model for Berlin house prices: Estimation and economic interpretation. *Journal of Real Estate Finance and Economics* 28, 1 (2004), 37–57. DOI : https://doi.org/10.1023/A:1026373523075

[45] Hasan Selim. 2009. Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications* 36, 2 (2009), 2843–2852. DOI : https://doi.org/10.1016/j.eswa.2008.01.044

[46] Dimitri P. Solomatine and Durga L. Shrestha. 2004. AdaBoost.RT: A boosting algorithm for regression problems. In *Proceedings of the 17th International Joint Conference on Neural Networks*, Vol. 2. 1163–1168. DOI : https://doi.org/10.1109/IJCNN.2004.1380102

[47] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), 267–288.

[48] Xibin Wang, Junhao Wen, Yihao Zhang, and Yubiao Wang. 2014. Real estate price forecasting based on SVM optimized by PSO. *Optik-International Journal for Light and Electron Optics* 125, 3 (2014), 1439–1443. DOI : https://doi.org/10.1016/j.ijleo.2013.09.017

[49] Geoffrey I. Webb. 2000. MultiBoosting: A technique for combining boosting and wagging. *Machine Learning* 40, 2 (2000), 159–196. DOI : https://doi.org/10.1023/A:1007659514849

[50] Frank Wilcoxon. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1, 6 (1945), 80–83.

[51] Cort J. Willmott and Kenji Matsuura. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* 30, 1 (2005), 79–82. DOI : https://doi.org/10.3354/cr030079

[52] Chao Xue, Yongfeng Ju, Shuguang Li, Qilong Zhou, and Qingqing Liu. 2020. Research on accurate house price analysis by using GIS technology and transport accessibility: A case study of Xi'an, China. *Symmetry* 12, 8 (2020), 1329–1350. DOI : https://doi.org/10.3390/sym12081329

[53] Rüştü Yayar and Derya Demir. 2014. Hedonic estimation of housing market prices in Turkey. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* 43 (2014), 67–82. DOI : https://doi.org/10.18070/euiibfd.42448

[54] Yu Zhang and Qiang Yang. 2021. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering* (2021), 1–20. DOI : https://doi.org/10.1109/TKDE.2021.3070203

[55] Liang Zhao, Qian Sun, Jieping Ye, Feng Chen, Chang-Tien Lu, and Naren Ramakrishnan. 2015. Multi-task learning for spatio-temporal event forecasting. In *Proceedings of the 21st International Conference on Knowledge Discovery and Data Mining*. ACM, 1503–1512. DOI : https://doi.org/10.1145/2783258.2783377

[56] Jiayu Zhou, Jianhui Chen, and Jieping Ye. 2012. Malsar: Multi-task learning via structural regularization. *Arizona State University* 21 (2012), 1–36.

[57] Jiayu Zhou, Lei Yuan, Jun Liu, and Jieping Ye. 2011. A multi-task learning formulation for predicting disease progression. In *Proceedings of the 17th International Conference on Knowledge Discovery and Data Mining*. ACM, 814–822. DOI : https://doi.org/10.1145/2020408.2020549

[58] Nor Hamizah Zulkifley, Shuzlina Abdul Rahman, Nor Hasbiah Ubaidullah, and Ismail Ibrahim. 2020. House price prediction using a machine learning model: A survey of literature. *International Journal of Modern Education and Computer Science* 12, 6 (2020), 46–54. DOI : https://doi.org/10.5815/ijmecs.2020.06.04