Semi-Supervised Regression for Evaluating Convenience Store Location

Xinxin Bai[1], **Gang Chen**[1,2], **Qiming Tian**[1], **Wenjun Yin**[1], **Jin Dong**[1]
[1]IBM China Research Laboratory, Beijing, China
[2]Automation of Department, Tsinghua University, Beijing, China
[1]{baixx, tianqim, yinwenj, dongjin}@cn.ibm.com
[2]g-c05@mails.thu.edu.cn

Abstract

Location plays a very important role in the retail business due to its huge and long-term investment. In this paper, we propose a novel semi-supervised regression model for evaluating convenience store location based on spatial data analysis. First, the input features for each convenience store can be extracted by analyzing the elements around it based on a geographic information system, and the turnover is used to evaluate its performance. Second, considering the practical application scenario, a manifold regularization model with one semi-supervised performance information constraint is provided. The promising experimental results in the real-world dataset demonstrate the effectiveness of the proposed approach in performance prediction of certain candidate locations for new convenience store opening.

1 Introduction

Location selection for a convenience store is an interesting and challenging task that will evidently affect its subsequent success in business. The reason lies in mainly two folds. From the profitability point of view, a location with good accessibility can attract a large number of customers and achieve high turnovers or profits. From the loss point of view, it is very difficult to make up once an inappropriate location has been established since location is the top priority before any other decisions. [Jain and Mahajan, 1979; Craig *et al.*, 1984; Kuo *et al.*, 2002].

The conventional methods of location selection for a convenience store usually demand the expertise of analysts and hence their capability and experience may influence significantly the final evaluation results. This also means high cost on human resources [Achabal *et al.*, 1982; Arnold *et al.*, 1983; Goodchild, 1984]. Due to the ever-growing uses of spatial systems such as Geographical Information Systems (GIS), there are already huge amounts of spatial data, providing ample opportunities to handle convenience store location selection problems via spatial data analysis and knowledge discovery techniques. The goal of this paper is to provide a completely automatic solution framework for evaluating convenience store location based on spatial data mining.

The problem of evaluating a convenience store location can be simply put as follows: Given a candidate location, how to evaluate or predict new convenience store's performance (in this paper, we use 'turnovers' as the key performance indicator that mean income that a store receives from the sale of goods and services to customers.) if opened at this location according to some designated information extracted from the surrounding geographic elements such as the residential areas, roads, schools, etc. The evaluation results could help retail executives' decision making on whether a location is appropriate for a new convenience store.

From the above viewpoint, the convenience store location selection problem can be seen as a regression problem. The dependent variables (also called response variable or measurement) is the turnovers of the convenience store, and the independent variables (also known as explanatory variables or predictors) are the features extracted from the geographic elements around candidate store location. However, in the practical applications, store performance data is very limited because of commercial confidentiality. For example, in general a convenience store retailer only knows the turnovers of its own stores without knowing the competitors'. In other words, the number of training data is quite small. To improve the prediction accuracy, the unlabeled (no performance information) data can be introduced to train a model in some way that is called semi-supervised learning. Compared to supervised learning which uses only labeled data to train, semi-supervised learning refers to the use of both labeled and unlabeled data for training so that it may perform better in many cases. In the past years, semi-supervised learning has drawn more and more studies [Zhu *et al.*, 2003; Zhou *et al.*, 2003; Belkin *et al.*, 2005]. A detailed survey on the semi-supervised learning can be found in [Zhu, 2006].

In this paper, we formulate a novel semi-supervised regression model for the convenience store location selection based on spatial data analysis. First, the input features for each convenience store can be extracted by analyzing its neighbor geographic elements based on a GIS platform and the corresponding output value is its 'turnovers'. Second, our model embodies the use of semi-supervised information in two aspects

- A basic assumption in the semi-supervised learning problems is that nearby points are likely to have the similar output. The geometric consistency assumption can

be expressed by the manifold regularizer [Belkin *et al.*, 2005].

- In practice, although the performance information of each convenience store from the competitors is unknown, their average performance can be given a reasonable estimation. This is based on another assumption that most or a certain percent of existing ones are rational or profitable. For example, their average performance should be above some predefined value that as a constraint is incorporated into our semi-supervised regression model.

The rest of the paper is organized as follows. Section 2 gives some preliminary knowledge about spatial feature extraction. In Section 3, we elaborate our semi-supervised regression model for the convenience store location selection. A case study on evaluating the convenience store location is presented in Section 4, followed by our conclusions in Section 5.

2 Preliminaries

In this section, we describe how the data features are generated.

2.1 Basic Features

A geographic information system captures, stores, analyzes, manages and presents data that refers to or is linked to location. GIS applications allow us to create interactive queries, analyze spatial information, edit data, maps, and present the results of all these operation. Our data extraction is conducted via our GIS-based spatial decision support platform called iFAO [Bai *et al.*, 2007; Yin *et al.*, 2008].

For a convenience store, we can easily get the information about its surrounding geographic elements such as road, office building, residential area, etc. Each element has two kinds of depiction data: pertinent data (depicting location of geographic element objects) and attribute data (describing physical characteristics of each element object). Intuitively, the neighbor element information reflects whether a convenience store location is appropriate or not. For instance, if there are large amounts of residential areas or office buildings around some location, we may conclude that the location is appropriate for new convenience store openning.

In order to uniformly present the vectorizable features of each convenience store, we should aggregate its neighborhood elements information. All elements have one common denominator,location of geographic objects, i.e. xy coordinates. Based on this common denominator, all data points can be related to others via location as well as other physical characteristics. In other words, the relationship between each convenience store and each element can be quantitated. For example, we could count the number of the objects in one element which appears close to a given convenience store. Similarly, the contribution of one element's physical characteristics to each store could also be calculated. More detail will be listed as follows.

Given p elements and one fix convenience store as input, we aim to find the relationship between them. Assume that the jth element has $O_j(1 \leq j \leq p)$ objects and each object should be represented with the same physical characteristics. For instance, we obtain an element *school*, and there are O_s schools, i.e. O_s objects. Meanwhile, each school object is described with xy coordinates, number of students (#*student*), school area, etc. Now given a convenience store C_i, the relationship between *school* element and C_i could be calculated by the following equation

$$\mathrm{NO}_{ij} = \|O_j|\delta_{D1} \leq D(C_i, O_j) \leq \delta_{D2}\| \tag{1}$$

Eq. (1) gives the formula for counting the number of objects in the jth element which are in the neighbor $[\delta_{D1}, \delta_{D2}]$-distance area around the ith convenience store. δ_{D1} and δ_{D2} are predefined by some criteria which which are different in different countries or even different cities of a same country [Yao, 2002] (say, 0m, 500m, 1000m, 2000m, etc.).

Similarly, the physical characteristics contribution of the elements to each convenience store can be calculated as follows (here, we take *school* as an example)

$$S_{i,\#\text{student}} = \begin{cases} \frac{1}{\mathrm{NS}_i} \sum_{k=1}^{\mathrm{NS}_i} S_{\mathrm{Map}(k),\#\text{student}} & \mathrm{NS}_i > 0 \\ 0 & \mathrm{NS}_i = 0 \end{cases} \tag{2}$$

where $\mathrm{NS}_i = \|S|\delta_{D1} \leq D(C_i, S) \leq \delta_{D2}\|$ is the number of schools around the ith convenience store. $S_{i,\#\text{student}}$ indicates the #*student* attribute's contribution of school element to the ith convenience store. Map(*k*) is the table for *school* object index. $S_{\mathrm{Map}(k),\#\text{student}}$ is the attribute #*student* value of the Map(*k*)th school.

2.2 The Weighted Walk Distance Feature

In practice, the most valuable physical characteristic for one element may be the *human flow*. For example, the *human flow* for *school* element means the number of students, and for the residential area it is the amount of residents. Empirically, the larger the *human flow* around one convenience store, the more appropriate its location. However, we argue that with this attribute it is still insufficient for evaluating the convenience store location. What's more,it is often quite difficult to obtain the *human flow* attribute for some elements such as busline, highway, etc. Intuitively, the walk distances along roads from its neighbor elements to itself should have a significant influence to its location selection. Generally, there should be some elements with large human flows and close walk distances around one good convenience store location. In the past, no doubt it is rather hard to obtain the walk distances between one convenience store and its neighbor elements. Fortunately, the GIS applications make us relatively easily do this.

In order to compute the walk distances, we first need to manually mark enough amounts of accessible points on the roads shown on the GIS maps. Then the computation problem of the walk distance are transformed into the shortest path search problem based on the marked points. Once the walk distance from each convenience store to its any neighbor element object is achieved, the corresponding weighted walk distance features can be calculated as the following formula

(we still take *school* as an example)

$$S_{i,\text{walkdis}} = \begin{cases} \dfrac{\sum_{k=1}^{\text{NS}_i} S_{\text{Map}(k),\#\text{student}} \times S_{\text{Map}(k),\text{walkdis}}}{\sum_{k=1}^{\text{NS}_i} S_{\text{Map}(k),\#\text{student}}} & \text{NS}_i > 0 \\ 0 & \text{NS}_i = 0 \end{cases} \tag{3}$$

where $S_{\text{Map}(k),\text{walkdis}}$ represents the walk distance from the Map(k)th school to C_i.

Based on the above aggregating process, in this paper, each convenience store data will be represented as a vector with $3p$ components. Concretely, the extracted input features of the convenience store O_i can be expressed as

$$F_i = \{z_{i1}^{num}, z_{i1}^{flow}, z_{i1}^{dis}, \cdots, z_{ip}^{num}, z_{ip}^{flow}, z_{ip}^{dis}\} \tag{4}$$

where z_{ip}^{num}, z_{ip}^{flow} and z_{ip}^{dis} represents the number of objects, the human flow and the weighted walk distance belonging to the ith element around O_i respectively.

3 Semi-supervised Regression for Evaluating Convenience Store Location

In this section, we will describe a semi-supervised regression for evaluating convenience store location.

3.1 Basic Framework

As mentioned in Section 1, the convenience store location selection problem can be viewed as a regression problem. In the recent ten years, the regularization algorithms such as Support Vector Machines (SVM), Regularized Least Squares (RLS) and Support Vector Regression (SVR) [Vapnik, 1995; Joachims, 1998] [Chang and Lin, 2001; Rifkin *et al.*, 2003] have achieved great successes in both theoretical analysis and practical applications. Recently, [Belkin *et al.*, 2005] provided a unified manifold regularization framework that exploits the geometry of the marginal distribution, and many standard methods including SVM and RLS can be obtained as special cases. Motivated by the manifold regularization algorithm, we provide a novel semi-supervised regression model for evaluating convenience store location.

Given l labeled convenience store data $(\boldsymbol{x}_i, y_i), i = 1, \cdots, l$ from one retailer where $\boldsymbol{x}_i$ is the extracted feature vector of the ith convenience store and y_i is its turnover, and u unlabeled convenience store data $\boldsymbol{x}_{l+1}, \cdots, \boldsymbol{x}_{l+u}$. Consider a linear regression function $f(\boldsymbol{x}) = \langle \boldsymbol{w}, \boldsymbol{x} \rangle + b$, where $\langle , \rangle$ is inner product. One often deal with this term by appending each instance with an additional dimension

$$\boldsymbol{x}_i^T \leftarrow [\boldsymbol{x}_i, 1] \quad \boldsymbol{w}^T \leftarrow [\boldsymbol{w}^T, b] \tag{5}$$

Thus $f(\boldsymbol{x}) = \langle \boldsymbol{w}, \boldsymbol{x} \rangle$.

Based on the manifold regularization, a direct semi-supervised regression model can be formulated as the following optimization problem

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2 + \frac{1}{2}\lambda \boldsymbol{f}^T L \boldsymbol{f} + \frac{1}{2} C \sum_{i \in \mathcal{L}} (f_i - y_i)^2 \tag{6}$$

where $\boldsymbol{f}$ represents the predicted income vector $(f_1, \cdots, f_{l+u})^T$, $\mathcal{L}$ is the index set of labeled data and L is the graph Laplacian given by $L = D - W$ where W_{ij} are the edge weights in the data adjacency graph. Here, the diagonal matrix D is given by $D_{ii} = \sum_{j=1}^{l+u} W_{ij}$. The first margin term $\frac{1}{2}\|\boldsymbol{w}\|^2$ imposes smoothes conditions on possible solutions. The second term $\frac{1}{2}\lambda \boldsymbol{f}^T L \boldsymbol{f}$ is a manifold regularizer that assumes the convenience store data with the similar geographic features tend to have the similar turnovers. The final squared term $\frac{1}{2} C \sum_{i \in \mathcal{L}} (f_i - y_i)^2$ reflects the empirical loss. The predefined nonnegative parameters λ and C tradeoff various regularization terms. Clearly, if $\lambda = 0$, the semi-supervised regression model called LapRLS in [Belkin *et al.*, 2005] reduces to RLS.

Just as referred in Section 1, in practice, for a retailer, besides the data from itself, we can also get many unlabeled data from its competitors. On the one hand, these unlabeled data can be used in the manifold regularizer to enforce the geometric smoothness. On the other hand, they also give some prior knowledge. Based on the expertise, a reasonable constraint is that the average turnover of these competitors should be above some predefined value, i.e.

$$\frac{1}{|\mathcal{O}|} \sum_{i \in \mathcal{O}} f_i \geq h \tag{7}$$

where $\mathcal{O}$ is the index set of unlabeled data from the competitors and h is some predefined positive constant. In the business applications, based on the open business information from the competitors and domain knowledge, we can give a rough estimation of h. Thus, incorporate the semi-supervised constraint (7) into Eq. (6), we can obtain a novel semi-supervised regression model with one constraint (In order to contrast against LapRLS, we call the proposed model LapRLSC) for evaluating convenience store location

$$\begin{aligned} \min \quad & \frac{1}{2}\|\boldsymbol{w}\|^2 + \frac{1}{2}\lambda \boldsymbol{f}^T L \boldsymbol{f} + \frac{1}{2} C \sum_{i \in \mathcal{L}} (f_i - y_i)^2 \\ \text{s. t.} \quad & \frac{1}{|\mathcal{O}|} \sum_{i \in \mathcal{O}} f_i \geq h \end{aligned} \tag{8}$$

The above optimization problem is a linearly constrained quadratic programming problem, whose dual problem is (Please see Appendix for details)

$$\begin{aligned} \max \quad & g(\mu) = -\frac{1}{2} \boldsymbol{x}_0^T A^{-1} \boldsymbol{x}_0 \mu^2 + \\ & \left(C \sum_{i \in \mathcal{L}} y_i \boldsymbol{x}_i^T A^{-1} \boldsymbol{x}_0 - h \right) \mu + \text{constant} \\ \text{s. t.} \quad & \mu \leq 0 \end{aligned} \tag{9}$$

where $A = I + \lambda X^T L X + C \sum_{i \in \mathcal{L}} \boldsymbol{x}_i \boldsymbol{x}_i^T$ and $\frac{1}{|\mathcal{O}|} \sum_{i \in \mathcal{O}} \boldsymbol{x}_i = \boldsymbol{x}_0$.

It can be easily seen that the solution of the dual problem (9) is

$$\mu = \begin{cases} \frac{C \sum_{i \in \mathcal{L}} y_i \boldsymbol{x}_i^T A^{-1} \boldsymbol{x}_0 - h}{\boldsymbol{x}_0^T A^{-1} \boldsymbol{x}_0} & C \sum_{i \in \mathcal{L}} y_i \boldsymbol{x}_i^T A^{-1} \boldsymbol{x}_0 - h < 0 \\ 0 & C \sum_{i \in \mathcal{L}} y_i \boldsymbol{x}_i^T A^{-1} \boldsymbol{x}_0 - h \geq 0 \end{cases}$$

Once the dual problem is solved, the optimal primal variables $\boldsymbol{w}$ can be calculated by Eq. (15).

3.2 Discussions

So far, we have proposed a semi-supervised regression model for evaluating convenience store location. In a typical application scenario, given N candidate locations, their turnovers can be predicted based on our model. Then we could select top k locations with the highest prediction turnovers as the future new convenience store sites.

Different from the conventional methods of location selection that need much executives' judgment based on their knowledge and experience, our approach is almost automatic. That means it costs less to evaluate large number of candidate locations in the GIS maps via spatial data analysis at once. This could help the retailer find some easily neglected good locations. Actually, in practice, since the factors influencing the convenience store location selection are quite complicated, it is more appropriate to combine our approach and the conventional ones to make the final decisions.

In addition, considering a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$ with kernel $k(,)$ and the regression function $f \in \mathcal{H}$, the margin term $\|\boldsymbol{w}\|^2$ in the objective function of the problem (8) will become RKHS norms $\|f\|^2_{\mathcal{H}}$ of $\mathcal{H}$. Since the optimization problem (8) only depends on point evaluations and RKHS norms of $\mathcal{H}$, by Representer Theorem [Schlkopf and Smola, 2001], its solution has the following form

$$f(\boldsymbol{x}) = \sum_{r=1}^{n} \alpha_i k(\boldsymbol{x}, \boldsymbol{x}_i) \;\; (i = 1, \cdots, n) \tag{10}$$

where α_i are the coefficients. We thus easily obtain the following nonlinear kernel version of the proposed algorithm

$$\begin{aligned} \min \quad & \frac{1}{2}\alpha^T K\alpha + \frac{1}{2}\lambda\alpha^T KLK\alpha + \frac{1}{2}C\sum_{i\in\mathcal{L}}(f_i - y_i)^2 \\ \text{s. t.} \quad & \frac{1}{|\mathcal{O}|}\sum_{i\in\mathcal{O}} f_i \geq h \end{aligned} \tag{11}$$

Similar to the linear case (8), the above optimization problem (11) is still a linearly constrained quadratic programming problem that also has close solutions. However, in this paper, we only focus on the linear regression model.

4 A Case Study on Evaluating Convenience Store Location

In this section, we use a real-world application on the convenience store location selection to examine the proposed semi-supervised regression model.

4.1 Data Preparation and Experimental Setup

In this case, total 272 convenience stores data are used for study. Among them, 46 convenience stores belonging to one major retailer in China have performance knowledge, i.e. turnovers. For the other 226 stores belonging to competitors, we cannot get the performance information, but we can easily obtain their input features with the method in Section 2. We randomly select 36 from the major retailer and the other 226 as the training set, the left 10 from the major retailer as the test set.

Table 1 shows the detailed elements around the convenience stores obtained from one China city. Among them, column #Objs indicates the number of objects in the corresponding element. Column #Atts indicates the number of attributes which describe the geographical location and physical characteristics of each objects. From Table 1, we found that most of elements only contain two attributes which describe the geographical location, i.e. xy coordinates. For the last ten elements {airport, entertainment, hospital, shoppingmall, supermarket, trainstation, residentarea, walkstreet, college and phmschool} with three attributes, they contain not only xy coordinates but also the *human flow* attribute. Besides these basic attributes, we also obtain the walk distance information as the method provided in Section 2.2. Then by aggregating the data features as described in Section 2, we have the input feature vectors of convenience store data. Finally, in order to avoid attributes in greater numeric ranges dominate those in smaller numeric ranges, each attribute feature is scaled to [0,1].

In our model, it is necessary to identify the best value of model parameters such as C, λ on the training data. Here, the grid search method with 5-fold cross validation is used to determine the best parameter values. In order to find good parameters C and λ , select different values $C = 2^{-6}, 2^{-5}, 2^{-4}, \cdots, 2^{0}, 2^{1}, 2^{2}, \cdots, 2^{13}$ and $\lambda = 2^{-6}, 2^{-5}, 2^{-4}, \cdots, 2^{0}, 2^{1}, 2^{2}, \cdots, 2^{13}$. For each pair of (C, λ), do 5-fold cross validation on the training data and compute the corresponding performance measure. Finally, select the ones with the best performance as the values of C and λ.

To evaluate the performance of our semi-supervised method, for each labeled set size l (5, 10, 20, 30), we perform 10 trials. In each trial we randomly sample l labeled data from the training data, and use the unlabeled data from the competitors as the unlabeled data. We adopt the average Mean Absolute Error (MAE) as the performance measure in 10 trials, and the MAE is given by

$$\text{MAE} = \frac{1}{t}\sum_{i=1}^{t}|f_i - y_i| \tag{12}$$

where t is the number of the remaining labeled data on the training set.

In addition, for all experiments, we constructed adjacency graphs with 7 nearest neighbors to compute the adjacency matrix W and

$$W_{ij} = \begin{cases} \exp[-\beta\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2] & \text{if } \boldsymbol{x}_i \text{ and } \boldsymbol{x}_j \text{are the neighbors} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

where β is a nonnegative constant. In our experiments, we set $\beta = 8$.

4.2 The influence of Parameter h

As discussed in Section 3, the parameter h reflects the average turnover information from the competitors and influences the prediction performance of the proposed method.

Table 1: Elements description around the convenience stores in one China city

Elements	Detailed elements	#Objs	Atts
Road	busline	168	2
	highway	3	2
	quickroad	1	2
	roadline	1126	2
GeoEnvironment	riverarea	193	2
	riverline	523	2
	greenarea2	36	2
	greenarea3	29	2
Building	airport	1	3
	busstation	8	2
	company	1055	2
	entertainment	34	3
	government	123	2
	harbor	1	2
	hospital	52	3
	offbuliding	88	2
	shoppingmall	48	3
	stockinsurance	41	2
	supermarket	11	3
	trainstation	3	3
Area	olypCourt	1	2
	residentarea	252	3
	structuremap	4	2
	walkstreet	4	3
	welcomeroad	1	2
Pop	college	60	3
	village	516	2
	pmhschool	131	3

Fig. 1 shows how the value of parameter h affects the performance of our model under different number of labeled data. To facilitate the parameter tuning, here we set $h = \eta \times \frac{1}{l}\sum_{i=1}^{l} y_i$. From Fig. 1, when η, the ratio of the performance of unlabeled competitors' store to that of the labeled studied retailers' store, varies from 0 to 1.5, the performance first increasingly grows better, and achieves the best around 0.6, then increasingly becomes worse. This indicates that competitors' store is a little worse performed than the studied retailer, which is accord with the fact that in this city, the studied retailer is the no. 1 retailer in terms of turnover and market share. Specially, when the value of η is small, the semi-supervised constraint information is relatively weak that cannot lead to large performance improvement. Conversely, if the value of η is too large, the prediction performance will be hurt and even be worse than that in the case without the constraint ($\eta = 0$). For all the later experiments, we set $\eta = 0.6$.

4.3 Performance Comparison

We constrasted four different methods under different number of labeled data as follows

- LapRLS without the weighted walk distance features, denoted as LapRLS_n.

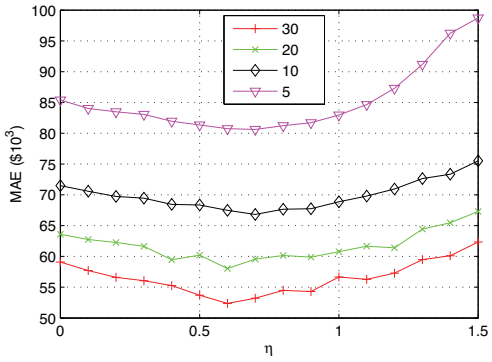

Figure 1: The performances of our proposed approach when the parameter h varies under different number of labeled data

- LapRLS with the weighted walk distance features, denoted as LapRLS.
- LapRLSC without the weighted walk distance features, denoted as LapRLSC_n.
- LapRLSC with the weighted walk distance features, denoted as LapRLSC.

Fig. 2 indicates the performances of all the four methods on the test data under different number of labeled data. From Fig. 2, we can find

- All four approaches show a same trend of decreasing MAE when the number of labeled data increases. It is consistent with the general expectation.
- Under the same number of labeled data, LapRLS always performs better than LapRLS_n, and LapRLSC always performs better than LapRLSC_n. The facts verifies the weighted walk distance features are indeed effective to improve the prediction accuracy.
- Likely, under the same number of labeled data, the performance of LapRLSC is always better than that of LapRLS, and the performance of LapRLSC_n is always better than that of LapRLS_n. This manifests that the semi-supervised constraint information from the competitors does improve the prediction performance.

Fig 2 shows that LapRLSC achieves the best performance among the above four methods under the same number of labeled data. To sum up, the experimental results sufficiently illustrate that incorporating the weighted distance features and the semi-supervised constraint information into LapRLS does significantly improve the prediction performance.

5 Conclusions

In this paper, we have presented a novel semi-supervised regression model for evaluating the convenience store location via spatial data analysis. The extracted spatial features for one convenience store data include not only the number of objects belonging to each element around the store and the *human flow* attribute of each object, but also the weighted walk distances. Our model embodies the use of semi-supervised information in two aspects: one is the manifold regularizer that

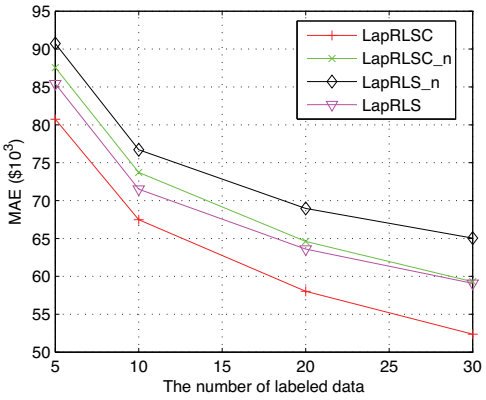

Figure 2: The performances of four methods on the test data under different number of labeled data

assumes that the convenience stores with similar spatial features tend to achieve the similar performance, the other is one constraint that reflects the average performance of the competitors is above certain level of labeled stores'. The experimental results on a real-world dataset verify the effectiveness of our proposed approach.

A Computation of the Dual Problem

To solve Eq. (8), we introduce the dual variable $\mu \leq 0$ related to the constraint $\frac{1}{|\mathcal{O}|}\sum_{i\in\mathcal{O}} f_i \geq h$. The Lagrangian can then be computed

$$\text{Lag}(\boldsymbol{w},\mu) = \frac{1}{2}\|\boldsymbol{w}\|^2 + \frac{1}{2}\lambda \boldsymbol{f}^T L \boldsymbol{f} + \frac{1}{2}C\sum_{i\in L}(f_i - y_i)^2$$
$$+\mu\left(\frac{1}{|\mathcal{O}|}\sum_{i\in\mathcal{O}} f_i - h\right) \quad (14)$$

Setting $\partial\text{Lag}/\partial\boldsymbol{w} = 0$ yields

$$\boldsymbol{w} = A^{-1}\left(C\sum_{i\in\mathcal{L}} y_i\boldsymbol{x}_i - \mu\boldsymbol{x}_0\right) \quad (15)$$

where $A = I + \lambda X^T L X + C\sum_{i\in\mathcal{L}} \boldsymbol{x}_i\boldsymbol{x}_i^T$.

Using Eq. (15) , the dual of Eq. (8) can be expressed by

$$\begin{aligned} \max \quad & g(\mu) = -\frac{1}{2}\boldsymbol{x}_0^T A^{-1}\boldsymbol{x}_0\mu^2 + \\ & \left(C\sum_{i\in\mathcal{L}} y_i\boldsymbol{x}_i^T A^{-1}\boldsymbol{x}_0 - h\right)\mu + \text{constant} \\ \text{s. t.} \quad & \mu \leq 0 \end{aligned} \quad (16)$$

References

[Achabal *et al.*, 1982] D.D. Achabal, W.L. Gorr, and V. Mahajan. Multiloc: a mulitiple store location decision model. *Journal of Retailing*, 58(2):5–25, 1982.

[Arnold *et al.*, 1983] D.R. Arnold, L.M¿ Capella, and G.D. Smith. *Strategic Retail Management*. Addison-Wesley, 1983.

[Bai *et al.*, 2007] Xinxin Bai, Wei Shang, Wenjun Yin, and Jin Dong. A service-oriented solution for retail store network planning. In *Proc. of IEEE International Conference on Services Computing*, 2007.

[Belkin *et al.*, 2005] M. Belkin, P. Niyogi, and V. Sindhwani. On manifold regularization. In *Proc. of AISTATS*, 2005.

[Chang and Lin, 2001] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines, 2001.

[Craig *et al.*, 1984] C.S. Craig, A. Ghosh, and S. McLafferty. Model of the retail location process: a review. *Journal of Retailing*, 60(1):5–36, 1984.

[Goodchild, 1984] M.F. Goodchild. Ilacs: a location-allocation model for retail site selection. *Journal of Retailing*, 60(1):84–100, 1984.

[Jain and Mahajan, 1979] A.K. Jain and V. Mahajan. Evaluating the competitive environment in retailing using multiplicative competitive interactive model. *Research in Marketing*, 2, 1979.

[Joachims, 1998] Thorsten Joachims. Text categorization with support vector machines: Learning with many relevant features. In *Proc. of ECML*, pages 137–142. Springer Verlag, 1998.

[Kuo *et al.*, 2002] R.J. Kuo, S.C. Chi, and S.S. Kao. A decision support system for selecting convenience store lcoation through integration of fuzzy ahp and artificial neural network. *Computers in Industry*, 47:199–214, 2002.

[Rifkin *et al.*, 2003] R. Rifkin, G. Yeo, and T. Poggio. Regularized least squares classification. In *Advances in Learning Theory: Methods, Models and Applications*. IOS Press, 2003.

[Schlkopf and Smola, 2001] Bernhard Schlkopf and Alexander Smola. *Learning with Kernel: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 2001.

[Vapnik, 1995] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.

[Yao, 2002] Y. Yao. Beijing downtown-mapping customer research in an urban core. *Column of GIS Retail*, 12(4), 2002.

[Yin *et al.*, 2008] Wenjun Yin, Xinxin Bai, Minghua Zhu, Ming Xie, and Jin Dong. ifao: Spatial decision support services for facility network transformation. In *Proc. of IEEE International Conference on Services Computing*, 2008.

[Zhou *et al.*, 2003] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Scholkopf. Learning with local and global consistency. In *Proc. of NIPS*, 2003.

[Zhu *et al.*, 2003] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian random fields and harmonic functions. In *Proc. of ICML*, 2003.

[Zhu, 2006] X. Zhu. Semi-supervised learning literature survey. Techinial Report TR 1530, University of Wisconsin-Madision, 2006.