Branch Reconfiguration Practice Through Operations Research in Industrial and Commercial Bank of China

Xiquan Wang, Xingdong Zhang, Xiaohu Liu, Lijie Guo, Thomas Li, Jin Dong, Wenjun Yin, Ming Xie, Bin Zhang,

To cite this article:

Full terms and conditions of use: http://pubsonline.informs.org/page/terms-and-conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article’s accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2012, INFORMS

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org
Industrial and Commercial Bank of China (ICBC), the world’s largest publicly traded bank as measured by market capitalization, deposit volume, and profitability, has a network of over 16,000 branches and serves over 200 million individual customers and 3.6 million corporate clients. In China, ICBC has 38 provincial and large-city divisions and over 300 city units. (Provincial and large-city divisions report directly to headquarters; city units report to provincial divisions.) Internationally, ICBC has a global network of over 200 overseas institutions and more than 1,400 correspondent banks.

Branches are a bank’s most important service and marketing channel. Typically, they are more effective than other channels (e.g., Internet banking) for acquiring new customers. Customers in China, as in other developing markets, prefer using branches even for routine transactions and financial services. A large and efficient branch network is necessary for acquiring customers, expanding market share, increasing customer satisfaction, reducing overall operating costs, and improving operational efficiency. Such a network is fundamental to ICBC’s business development, an important core competency, and is a valuable asset for sustainable future growth.

With China’s increased economic level, modernization, urbanization, and globalization, its financial markets are changing rapidly. Many new urban districts and satellite cities are emerging, and personal wealth is increasing. In this fast-changing and
competitive market, ICBC needs to reconfigure its branch locations and service capabilities to match the regional economy and customer distribution; therefore, it has to quickly identify new high-potential market areas in which to open branches.

With such a large-scale branch network, complex competitive environment, and dynamic demographics, ICBC needs a rigorous methodology and system to objectively and efficiently optimize the branch network, increase the utilization of branch resources, and strengthen channel competency. A branch reconfiguration system must do the following:

- determine the number and sizes of branches that are needed to match a city’s economic situation and financial potential;

- model the financial potential of a market based on geographic and demographic data to identify new high-potential market areas and good branch locations—the system must calculate the geographic distribution and value of customers in a city, determine the right locations and personnel arrangements of new, relocated, and renovated branches, and help improve branch management;

- improve the decision making involved in branch reconfiguration to minimize capital costs, allow quicker decision making, and reduce the costly mistakes caused by poor branch investment choices.

In a typical city (see Figure 1), ICBC has a network of more than 500 branches and needs to consider four questions.

Figure 1: The figure shows ICBC branches and geographic entities (e.g., bank branches designated by circled I, office buildings designated by gray dots, and residential buildings designated by black dots) in a typical city in China.
(1) How many ICBC branches does the city need based on its economic conditions?

(2) Do the location and service capabilities of each existing branch match the customer volume and value in the neighborhood? If they do not match, how should the branch be reconfigured?

(3) Where are the high-potential market areas in which ICBC should consider opening new branches?

(4) How should ICBC use quantitative data to quickly answer these questions in an informed way?

To answer these questions, its branch reconfiguration system has to analyze the city’s branch performance and geographic data, which may include hundreds of thousands of various geographic entities (see Figure 1). Appendix A includes detail descriptions of the geographic data.

**Branch Reconfiguration (BR) System**

In 2006, ICBC partnered with IBM to develop a customized branch network optimization system, Branch Reconfiguration (BR). Optimizing a bank’s branch network, especially for a large number of branches, is complex. Some banks use a scoring method to determine whether a candidate site is suitable for opening a branch. With this approach, users (i.e., those who use the scoring method) identify several key evaluation metrics and set weights based on the experience of banking experts. A candidate branch site is scored by summing the product of the metrics times the weights. Some banks rely only on experts to determine the locations. Although the experience of these experts is important, they usually do not have a comprehensive view of the entire city’s market potential distribution. This scoring approach is subjective and ineffective in many cases, especially when scaled to a large number of branches.

**Challenges**

Reconfiguring ICBC’s branch network presents the following challenges.

- ICBC branches cover almost all of China. Customer behavior and financial conditions differ significantly in these diverse areas, necessitating customized parameters for these areas. However, the information for customizing model parameters for each area is insufficient.
- ICBC has many domain experts with years of branch network management experience; the BR system has to appropriately incorporate their knowledge.
- BR must search for an optimized network, which is a large-scale mixed-integer optimization problem (MIP). In most cases, many users will simultaneously run the solver for the MIP problem. Solving the MIP requires an efficient algorithm and a system capable of handling simultaneous users.

In this project, we developed a customized branch network optimization system based on operations research (OR) technologies, professional research in branch networks, and domain expertise. It includes three key models: a market potential prediction model, a branch network optimization model, and a branch site evaluation model. The system is built based on a geographic information system (GIS) platform to manage, analyze, and optimize ICBC’s branch network.

**Market Potential Prediction Model**

Estimating customer volume and value for each small geographic area in a city is a basic requirement of branch network optimization. Prior to using the market potential prediction model, BR performs a separate partitioning step in which it partitions a city into a grid of tens of thousands of cells, each measuring 100 meters by 100 meters. The market potential prediction model processes the geographic and demographic data in each cell to calculate a market potential value to indicate the cell’s customer volume and value. Because the cell’s actual market potential value is not known before using this model, this information cannot be available as training data to build the prediction model. Therefore, BR uses a quadratic optimization model to integrate two types of information: cell preference and metric preference, which domain experts provide, to create the model. To determine cell preference, an expert is given information on two cells and must indicate which cell has more and higher-value customers. A metric preference is a pair of metrics for which the expert indicates which
metric has a bigger impact on market potential. The model accuracy improves as more sample and metric-preference data are provided. Finally, BR computes the market potential value of each grid cell to obtain the distribution of customer volume and value in a city (see Figure 2). Appendix A provides details of this model.

**Branch Network Optimization Model**

The branch network optimization model searches the entire city for branch sites that will provide an optimal branch network for a specified number of branches. For example, in a city with 120,000 cells, ICBC might want to identify the 100 best branch locations from 12,000 candidate sites and determine the type of branch at each location. Based on the market potential provided by the market potential prediction model, the branch network optimization model maximizes the market potential covered by the branch network. The key challenge is to efficiently solve this large-scale MIP. To do this, we developed a hybrid nested-partitions (HNP) location network heuristic optimization algorithm (Xia et al. 2010) to reduce the computational complexity. Appendix B provides details.

**Branch Site Evaluation Model**

The branch site evaluation model assesses the value of opening a bank branch at a specific location by considering market potential, existing competitor locations, and other ICBC branches in the neighborhood. If ICBC has existing branches in a city, their performance is a measure of the value of opening a branch at the location; we can use this performance to train the model. Because we must build a separate branch site evaluation model for each city, the number of existing branches is sometimes insufficient to train the model; therefore, the branch site evaluation model needs additional information. In such a situation, we use expert knowledge; we give an expert a pair of site locations and ask him (her) to indicate the better of these two sites (i.e., specify the location preference), which is easy for the expert. BR integrates both existing branches information and location preferences information to create the branch site evaluation model. Appendix C provides details.

**System Design**

BR is a quantitative analysis and optimization system with customized decision support modules built on a GIS platform. The user interface has three interactive areas (see Figure 3):

1. GIS map interface: users can see branch locations, geographic data, and analytic results on this window;
2. analytic task interface: users enter data and perform all analytic functions;
3. function navigator: users can navigate all system functions by clicking corresponding links in this window.

The system has three layers: data layer, middle layer, and application layer (see Figure 4).

From a software point of view, the system must address three requirements: (1) integrating and synchronizing the heterogeneous information in the data layer; (2) simultaneously supporting a large number of users performing GIS operations; and (3) managing long-running CPU-intensive threads, while ensuring good system response time. In BR, many OR analytic tasks are computationally intensive and time intensive. Some run for over 10 hours. We introduced an analytic task management module into BR to schedule all analytic tasks submitted by users.
Based on the above OR models and system design, BR can manage all ICBC branch information, including the location and performance data. It computes a city’s market potential distribution by using the market potential prediction model to identify high- and low-potential market areas. Next, the branch network optimization model searches the entire city for the optimal branch site locations. Then users leverage the branch-site evaluation model to assess the generated branch sites. By comparing the optimal site locations and existing branch locations, BR generates suggestions of the final branch network optimization actions (i.e., recommendations).

**Implementation**

Because we needed an elaborate project implementation schedule to ensure BR’s successful deployment at ICBC, we devised a two-stage implementation plan (see Figure 5). The project team consisted of experts from the ICBC human resources department, selected functional departments, selected city units, and IBM Research.

**The Pilot Stage**

In 2006, ICBC selected the city of Suzhou as the first pilot city to test BR’s implementation. It successfully
completed the pilot in early 2007. At the beginning of the pilot, the project team collected branch performance data and geographic data to train the models with the preference data provided by experts. The project team then interviewed over 30 executives and experts within ICBC Suzhou to verify its strategy and to understand their experiences with BR. We used this information, especially their experiences, to formulate and tune the model parameters. Based on the market potential distribution result computed by BR, ICBC Suzhou’s project team saw a large discrepancy between the high market potential areas and branch locations. In the next step, BR generated the optimized branch locations using an HNP algorithm, and then used the branch-site evaluation model to assess all 150 existing branch locations and the new locations, which both BR and local experts provided. Finally, BR produced the branch reconfiguration recommendations, including opening more than 40 new branches, upgrading more than 10 branches, relocating approximately 30 branches, and closing approximately 20 branches. These results helped ICBC Suzhou to thoroughly understand the branch network, the market environment, and its competition. The ICBC Suzhou project team made a detailed branch development and rebuilding plan to implement the results.

To further test BR’s applicability and calibrate the models in other regions, we piloted six additional cities between 2007 and 2010: Beijing, Guangzhou, Xiamen, Wenzhou, Shaoxing, and Changzhou. During these pilots, ICBC and IBM worked closely to customize and improve the algorithms, templates, and parameters to improve their accuracy and capabilities.

The results from the six pilots received high recognition from the ICBC city unit executives (e.g., the president of the ICBC Suzhou city unit) and headquarters executives.

The Rollout Stage
Based on the successful pilot and proven customized branch network optimization methodology, ICBC started deploying the system to all cities at the end of 2009. During this stage, each city formed a local project team. Each team followed the branch network optimization methodology and used the system to perform optimization analysis, with guidance from ICBC headquarters and IBM Research. As of this writing, BR has been implemented in 100 major cities including Shanghai, Tianjin, Shenzhen, Chongqing, Nanjing, and Hangzhou (see Figure 6). The deployments will continue until all 16,000 branches have been analyzed.

Benefits and Impact

Direct Benefits
The optimized branch network has greatly enhanced ICBC’s ability to reach new individual and small-to-medium commercial customers. Additionally, because the services provided by each branch better match local customer needs, branch performance has significantly improved. The increase in deposits attributable to BR in a typical major city like Suzhou was US $1.04 billion. In future years, such deposits are expected to increase by tens of billions of US dollars; these deposits do not include those attributable to the improved wealth of customers.

Over the past three years, ICBC has invested significant amounts of money to upgrade its branches. This investment will continue based on business development requirements and expected return on investment. BR provides a quantitative base for making branch reconfiguration decisions. It helps minimize costs and risks by ensuring that ICBC does not place branches in low-profit areas. ICBC has achieved a competitive advantage by quickly identifying the most profitable market locations.

Organizational Impact
BR helps ICBC improve its branch management in the following areas.
(1) Strategic planning. ICBC decision makers can gain insights into the market to improve its branch resource investments and support its strategic planning. Its headquarters staff has a complete view of the branches in all provincial divisions and a better understanding of the market. Provincial divisions have a similar view within their cities. Bank management within the city units can understand the market in each branch’s neighborhood. Thus, strategic planning is supported on multiple levels.

(2) Branch network optimization. Because the market in China is changing rapidly, ICBC must frequently reexamine its branch network. BR provides an explicit and standardized process of branch network analysis, which ICBC analysts can easily follow to optimize the branches.

(3) Branch operations. BR integrates economic data, demographic data, and branch business data, and displays this data on a map in a visual and user-friendly way. A branch analyst can monitor branch configurations, track branch performance, identify low-efficiency branches, and take necessary corrective actions.

(4) Employee growth. As part of its BR deployment, ICBC has trained more than 500 employees to use the system for branch reconfiguration. These employees are now experts with a solid foundation in ICBC branch management.

(5) Customer satisfaction. BR provides analysts with a view of a city’s customer distribution. This information is useful for identifying high-density areas in which ICBC should target customers. ICBC now allocates more resources in these areas to better fulfill customer demand and improve customer satisfaction.

**External Impact**

ICBC is one of the banking industry’s most important banks; therefore, its initiatives and practices attract the attention of both peer institutions and the public. Because BR has improved ICBC’s branch management in a rigorous and systematic way, as the Suzhou
city pilot demonstrates, the ICBC Suzhou BR implementation has been cited as a best practice by the China banking community and external media.

Transportability

BR: Applicability to Other Banks
In the competitive markets and especially in emerging markets, banks that are facing the challenge of optimizing their branch networks can also use BR. Banks that are in a growth stage or large banks that want to open new branches in emerging markets can use it to forecast their market potential and quickly identify and evaluate the locations with the highest market potential. Thus, they can gain a competitive advantage. Banks with a large existing branch network can determine the efficiency of their existing network and identify possible improvements. They can identify poorly performing branches that are negatively affecting competitiveness. Banks that are involved in mergers and acquisitions need to rationalize the new larger merged branch network. BR can help these banks to evaluate the performance of existing branches and provide improvement recommendations.

For BR to be applicable to a bank, some model parameters and templates must be adapted based on that bank’s specific business requirements. For example, when evaluating a candidate branch, a bank might focus on specific geographic or demographic factors: large foreign banks targeting high-income consumers might focus on large-scale business districts, shopping malls, and A-level office buildings, whereas local smaller banks targeting a mass market might focus on residential areas. Thus, the models and parameters will need to be adjusted to handle these differences.

BR: Applicability in Other Consumer-Oriented Industries
Many businesses (e.g., supermarkets, convenience stores, hotels, and gasoline stations) have physical facilities in which they provide services or products to customers and which usually serve as the most important channel for delivering services. The owners must maintain their facility networks and make adjustments to them over time. A facility network optimization methodology and system are necessary to diagnose and optimize an existing facility network, and (or) build a network for a new market entry.

In the retail industry, brick-and-mortar stores are the most important, and most costly, channels for many retailers. Some large retailers use several store formats, including hypermarket, club membership store, and community store. The optimization of a store network, which could involve opening new stores or renovating existing stores at a cost of hundreds of millions of dollars, is a critical strategic decision. Zhang et al. (2009) discuss how these retailers can adapt BR to their requirements.

Conclusion
The BR implementation was possible because of the close cooperation between ICBC and IBM. The ICBC BR project is an example of using OR and management sciences to successfully transform the service channels of a large bank. We believe that it will continue to improve ICBC’s business development, decision making, and strategic initiatives.

Appendix A. Market Potential Prediction Model
In the market potential prediction model, we divide a city into a grid with tens of thousands of cells (see Figure A.1). The model estimates the market potential value of each cell based on customer quantity and value in the cell, which it estimates from the geographic entities inside the cell; these are obtained from the geographic information system (GIS) data. It contains the geographic locations of various categories of entities and their demographic attributes (see Table A.1).

The model computes a list of metrics for each cell from the GIS data. For example, the metric “Total households of residential buildings” is the summation of the numbers of households of all residential buildings inside the cell. Other examples of metrics include the number of supermarkets and the number of companies. The model uses the metrics as independent variables to predict the market potential value.

A medium-size city of approximately 400 square kilometers, of which China has 300, has 40,000 cells. Therefore, the system has over 12 million cells. Such
a large grid requires a highly efficient model. Consequently, we selected the linear model described below.

\[ f = w^T x, \]

where \( x \) is a vector of the metrics of the cell, \( w \) is a vector of weights, and \( f \) is the predicted market potential value. Because \( f \) correlates positively with the metrics, each element of \( w \) should be larger than zero.

The actual market potential value of a cell cannot be obtained; therefore, this information is not available as training data to build the prediction model. BR uses a quadratic optimization framework to integrate two new types of information: cell preference and metric preference, which experts provide, to create the market potential prediction model.

- **Cell preference.** Given a pair of cells (data samples), experts can know which cell has more customer volume and value (i.e., the expert prefers one cell over the other). Given a pair of cells \( x_i, x_j \) in the compared set \( E_s \), then \( x_i \) is preferred over \( x_j \) if

\[ w^T x_i - w^T x_j \geq 0. \]

- **Metric preference.** Similar to cell preference, given a pair of metrics, experts can know which metric is more important (i.e., the expert prefers one metric over the other). Given a metric pair \( (\text{Metric}_m, \text{Metric}_n) \) with weights \( (w_m, w_n) \) in the metric compared set \( E_m \), when the expert prefers \( \text{Metric}_m \) over \( \text{Metric}_n \), then

\[ w_m \geq w_n, \quad w_m \geq 0, \quad w_n \geq 0, \]

where \( w_m \) and \( w_n \) are the \( m \)th and \( n \)th elements of vector \( w \), respectively.

The framework to estimate \( w \) is given by

\[
\begin{align*}
\min_w & \left\{ \lambda_1 \|w\|^2 + \lambda_2 f^T L f + \lambda_3 \sum_{ij} \xi_{ij} + \lambda_4 \sum_{ij} \xi_{ij}^* \right\} \\
\text{s.t.} & \forall (x_i, x_j) \in E_s, \quad w^T x_i - w^T x_j \geq 1 - \xi_{ij} \\
& \forall (w_i, w_j) \in E_m, \quad w_j - w_i \leq \xi_{ij}^* \\
& w \geq 0, \quad \xi_{ij} \geq 0, \quad \xi_{ij}^* \geq 0,
\end{align*}
\]

where \( \lambda_1, \lambda_2, \lambda_3, \) and \( \lambda_4 \) are the weights of the four terms in the objective function and are predetermined parameters.

The variables \( \xi \) and \( \xi^* \) are soft margin variables. When \( \xi_{ij} \) tends to zero, the difference between \( x_i \) and \( x_j \), \( w^T x_i - w^T x_j \), tends to be larger. When \( \xi_{ij}^* \) tends to zero, the weight \( w_i \) tends to be closer to \( w_j \).

In the second term, the Laplacian matrix \( L \) is defined as the difference between the degree matrix \( D \) and the similarity (weight) matrix \( S \), which are defined as

\[ L = D - S, \]
with
\[
D = \{d_{ij}\}, \quad d_{ij} = \begin{cases} 
0 & \forall i \neq j \\
\sum_k s_{ik} & \forall i = j,
\end{cases}
\]
\[
S = \{s_{ij}\}, \quad s_{ij} = \frac{x_i^T x_j}{\|x_i\| \|x_j\|}.
\]

\(S_{ij}\) is the similarity between \(i\)th and \(j\)th data. Thus, the second term \(f^T Lf\) can be
\[
f^T Lf = f^T (D - S) f
\]
\[
= f^T D f - f^T S f
\]
\[
= \sum_i d_{ii} f_i^2 - \sum_{ij} s_{ij} f_i f_j
\]
\[
= \sum_{ij} s_{ij} f_i^2 - \sum_{ij} s_{ij} f_i f_j
\]
\[
= \frac{1}{2} \sum_{ij} s_{ij} (f_i^2 + f_j^2 - 2f_i f_j)
\]
\[
= \frac{1}{2} \sum_{ij} s_{ij} (f_i - f_j)^2.
\]

It measures the smoothness of \(f\). Minimizing the smoothness ensures that the predicted values of two similar samples are close.

The above optimization can be converted to a quadratic programming problem, which can be solved using existing methods (Boyd and Vandenberghe 2004). Finally, we can compute the market potential value of each cell (i.e., the market potential distribution in a city). As in Figure 2, we label the color of each cell according to its market potential value. White cells have large market potential, and gray cells have small market potential.

**Appendix B. A Hybrid Nested Partitions (HNP) Algorithm for Branch Network Location Optimization**

The branch network is optimized by locating branches in candidate sites that cover the largest market potential.

In this problem, \(N\) branches are to be located in \(J\) candidate locations, where a candidate location is a cell; \(I\) is the set of all cells.

Each branch can be exactly one of \(K\) formats. \(N^k\) is the maximum number of \(k\)-type branches allowed; therefore, the maximum number of candidate branches is \(\sum_{k=1}^{K} N^k\). \(S^k\) is the maximum coverage distance of a \(k\)th-type branch (i.e., the radius of its neighborhood area); \(w_i\) is the market potential value in an \(i\)th cell, which is obtained from the market potential prediction model; \(d_{ij}\) is the distance between cell \(i\) and cell \(j\); \(f^k(d_{ij}) = \max\{1 - (d_{ij}/S^k), 0\}\) is the coverage function, which specifies the maximum covered percentage of cell \(j\) by a \(k\)th-type branch at candidate site \(i\).

Then, the total covered percentage of the \(i\)th cell is
\[
y_i = \max \left\{ 1, \sum_{j \in I, k \in K} f^k(d_{ij}) x_j^k \right\}, \quad \forall i \in I,
\]
where \(x_j^k\) is the decision variable,
\[
x_j^k \in [0, 1], \quad \forall j \in J, \quad k \in K.
\]
\(x_j^k = 1\) if a \(k\)-type branch is located at candidate location \(j\); otherwise, \(x_j^k = 0\). We can see that \(y_i\) is a non-linear function of \(x_j^k\).

This can be modeled as an extended maximal coverage location problem (MCLP), as Church and ReVelle (1974) and Berman and Krass (2002) describe, where the objective function is to maximize the total covered market potential:
\[
\max \left\{ \sum_{i \in I} w_i y_i \right\}
\]
\[
s.t. \sum_{k \in K} x_j^k \leq 1, \quad \forall j \in J
\]
\[
\sum_{k \in K} x_j^k \leq N^k, \quad \forall k \in K.
\]

If the nonlinear function \(y_i\) is removed by considering \(y_i\) to be a decision variable, then
\[
\max \left\{ \sum_{i \in I} w_i y_i \right\}
\]
\[
s.t. \sum_{k \in K} x_j^k \leq 1, \quad \forall j \in J
\]
\[
\sum_{k \in K} x_j^k \leq N^k, \quad \forall k \in K
\]
\[
y_i \leq 1, \quad \forall i \in I
\]
\[
y_i \leq \sum_{j \in J, k \in K} f^k(d_{ij}) x_j^k, \quad \forall i \in I.
\]

This is a large-scale optimization problem with \(J \times K\) binary decision variables and \(I\) continuous
decision variables. A typical midsize city has three branch types \((K = 3)\), 40,000 grids \((I = 40,000)\), and 4,000 candidate locations \((J = 4,000)\). The number of candidate locations is approximately 10 percent of the number of cells. In the largest cities, such as Beijing and Shanghai, BR must handle four branch types, 12,000 candidate locations, and 120,000 grids. Because of the problem’s complexity and size, we developed an HNP method with two steps to efficiently solve it (Xia et al. 2010).

In the first step, we designed a problem-specific heuristic to reduce the infeasible and redundant candidate locations. We found that cells with a small market potential were not locations in the optimal solution; therefore, we could remove them in a preprocessing step. This reduced the search space but preserved the optimal solution.

In the second step, we extended the nested partitions method to solve the optimization problem. Because the objective function is the sum of the market potential covered by each cell and a typical city usually has thousands of cells, changing one decision variable usually has a minor impact on the solution’s overall performance. That is, the neighbors of good solutions are also good, and good solutions tend to cluster together. Therefore, we used the nested partitions framework to solve this type of problem (Shi and Olafsson 2000).

The nested partitions method is a sampling-based method. Its key concepts are to systematically partition the feasible decision space or region into subregions, to identify the most promising subregion—the subregion that is most likely to contain the optimal solution—through sampling, and to concentrate the computational efforts on this subregion. Sampling the solution space is a key step in this method. We constructed a relaxed problem by relaxing the binary variable \(x^i_j\) to a continuous variable. This problem is easier to solve and can generate acceptable solutions. We use these solutions to guide the sampling process and thus to enhance the efficiency of nested partitions.

### Appendix C. Branch Site Evaluation Model

Branch site evaluation determines the quality of a site (location) for opening a banking branch. It considers the geographic information around the site (e.g., a circular area with a radius of 500 meters around the site) to assess the desirability of opening a branch at the site.

Similar to market potential forecasting, the first step is to compute the metrics of the candidate site from the GIS data in its neighborhood area. The metrics include those used in the market potential prediction model and some additional ones, such as the number of competitors, the average distance from competitors, and the number of ICBC branches.

The next step is to use a framework, such as semisupervised learning (Chapelle et al. 2010), to integrate partial training samples and location preference information. \(f_1, \ldots, f_n\) are the predicted values of \(n\) locations, including \(l\) existing branches and identified locations. Existing branches can be used as partial training samples, denoted as \(E_r\) with size of \(l\). Each existing branch is indicated by \(x_i \in E_r\), and the deposit of each existing branch can be the objective value \(y_i\). Given a pair of sites \((x_u, x_v)\) in the compared set \(E_r\), we can denote their predicted values as \((f_u, f_v)\), and experts can know which site is better than the other; then the location preference can be \(f_u \geq f_v\).

Following the semisupervised regression method, we can solve \(f_1, \ldots, f_n\) from the model:

\[
\begin{align*}
\min_f & \quad \left\{ \sum_{x \in E_r} (f_i - y_i)^2 + \lambda_1 \|Lf\|_2^2 + \lambda_2 \|f\|_2^2 \right\} \\
\text{s.t.} & \quad f_u - f_v \geq 0, \quad \forall (x_u, x_v) \in E_r.
\end{align*}
\]

Here, \(\lambda_1\) and \(\lambda_2\) are the predetermined weights of the last two terms in the objective function. The first term models the regression of partial training samples, and the second and third terms measure the smoothness and complexity of \(f\), respectively. The constraint models location preferences. We do not use a soft margin for the location preference constraint, because we use only the reliable location preference pairs here as complementary information to partial samples.

In the second term, the Laplacian matrix \(L\) is the difference between the degree matrix \(D\) and the similarity matrix \(S\), as shown below.

\[L = D - S,\]
with

\[ D = \{ d_{ij} \}, \quad d_{ij} = \begin{cases} 0 & \forall i \neq j \\ \sum_k s_{ik} & \forall i = j \end{cases} \]

\[ S = \{ s_{ij} \}, \quad s_{ij} = \exp \left( -\frac{\| x_i - x_j \|^2}{2\sigma^2} \right). \]

Without the loss of generality, we assume that \( f = Ka \) from the representer theorem (Kimeldorf and Wahba 1971, Schölkopf and Smola 2002), where \( K \in \mathbb{R}^{n \times n} \) is a semidefinite kernel matrix constructed on the data set, and \( \alpha \) is the expansion coefficient. We use a radius basis function as the kernel function, which is common in similar problems. Therefore, the semidefinite kernel matrix \( K \) is the same as \( S \). We can reformulate the above problem as

\[
\begin{align*}
\min_f & \quad \| JK\alpha - y \|^2 + \lambda_1 \alpha^T KL\alpha + \lambda_2 \alpha^T K\alpha \\
\text{s.t.} & \quad e_u^T K\alpha - e_v^T K\alpha \geq 0, \quad \forall (u, v) \in E,
\end{align*}
\]

where \( e_u \) is a vector with its \( u \)-th element as 1; the others are 0. \( y \) is an \( n \times 1 \) vector, \( y = [y_1, \ldots, y_l, 0, \ldots, 0]^T \). \( J \) is an \( n \times n \) diagonal matrix, \( J = \text{diag}(1, \ldots, 1, 0, \ldots, 0) \).

We can convert this to a quadratic programming problem and solve it using existing methods (Boyd and Vandenberghe 2004).

**Acknowledgments**

We thank Xiao Fang Wang, Ya Bing Zong, Zhi Liang Hou, Li Ying Jiang, Qiang Fan, Lei Zhang, Tao Song, Ran Jie Li, Zhen Yu Song, and Ke Ye from ICBC, and Jin Sun, Jinyan Shao, Xin Xin Bai, J. P. Fasano, and Brenda Dietrich from IBM Research for their support and constructive comments, and our coach, Grace Lin, for her guidance and advice.

**References**


