Visual analysis of retailing store location selection

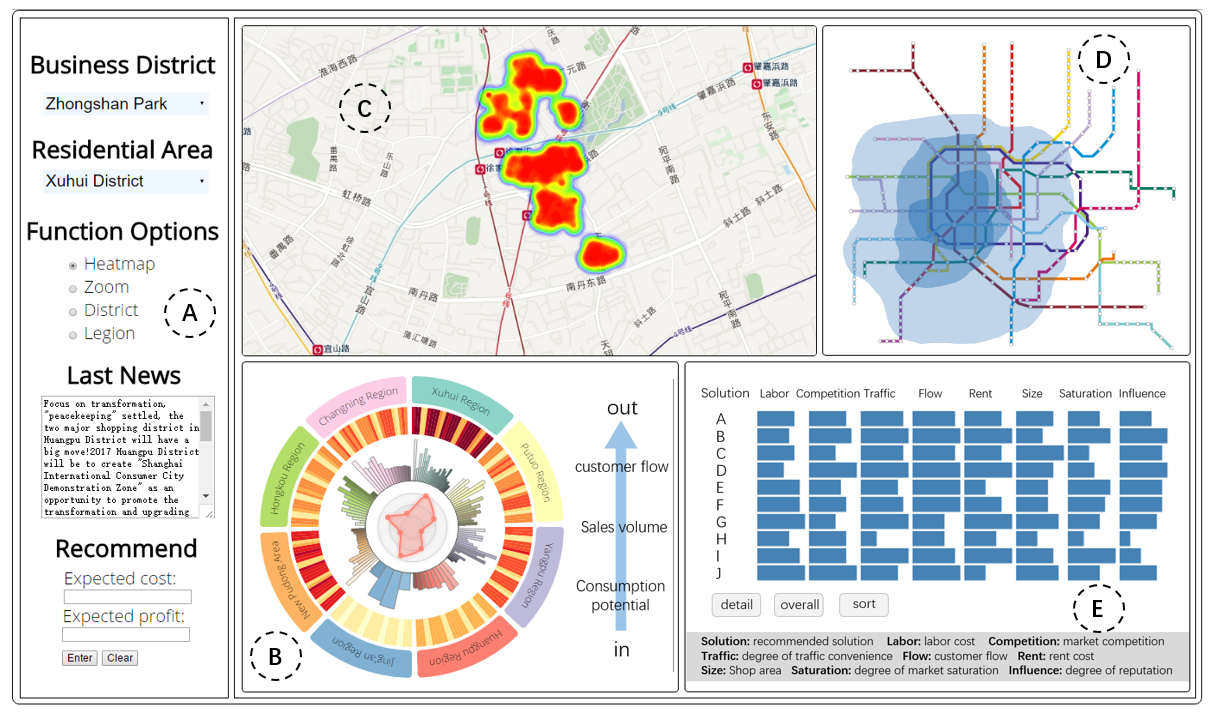


Fig. 1. Location recommendation system. (A) A navigation window provides functions of selection and input. (B) Statistical analysis view shows flow information, sales information and the market expectation from expert experience. (C) Location recommendation view provides the solutions and market prosperity. (D) Business influence view shows the influence area of business districts. (E) Visual comparison view shows the details and ranking of the solutions.

**Abstract**—Locations of retailing stores is vital for business success. Choosing a right location requires a huge amount of complex information, such as the attributes of business districts, customer flow and current business performances. Unlike location recommendation approaches based on statistical sampling, we establish the model of business district attractiveness based on customer flow and provide a method of data-driven visual comparisons. We build an interactive visual analytics system and user-friendly interface for an interactive visual query about business district attribute, customer flow, location-based information and transportation durations. Our system can recommend store locations, support interactive visual queries, and display rich information to facilitate decision making.

**Index Terms**—location recommendation, visual analytics, data mining, model optimization, visualization system

Introduction

Choosing a right location for a retail store has major influences on business successes. Managers have to integrate complex information including population, traffic and accessibility, zoning and neighbors, location cost, existing sales performance, public transit durations and distances, and geographical attributes [36]. The traditional approaches of retailing store location selection are basically based on statistical random sampling and probabilistic inference, and provide results with large deviations due to the volatility of customer flow. The results are challenging to understand, and therefore elicit difficulties in achieving consensus in collective decision-making. Furthermore, the heterogeneity of managers’ needs among diversified industries requires interactive queries and user-friendly interfaces for exploration.

To support the retailing store location selection, we provide a convenient and promising method for managers to choose locations for retailing stores. We synthesize pertaining economic knowledge and experts’ experience and propose an optimized market attractiveness model based on the analysis of customer flows. Among influential factors, customer flow brings sales and profits [37]. We analyze public transit passes data to present customer flow, and establish business district attractive model by modifying huff’s model [43] using machine learning and linear regression equations.

Then we recommend candidate locations based on data of sales performance and business districts, propose a visual-driven visual comparison method, and provide a user-friendly interface for interactive queries and persuasive communication.

The major challenges in analysis and visualization include (1) optimizing the model of business districts attractiveness based on customer flow, (2) designing a multi-factor analysis model to calculate the advantage of a location, (3) providing multidimensional business information including labor costs, shop rents, and competitors, and (4) creating intuitive visualization views to facilitate comparative analysis.

Our research contributes to the literature and practice in visual analysis by providing

* an optimized attractiveness model measures customer flow through a business district
* a profit driven location recommendation model based on multivariate analysis, and
* a business intelligence visualization system supporting interactive query and exploration.

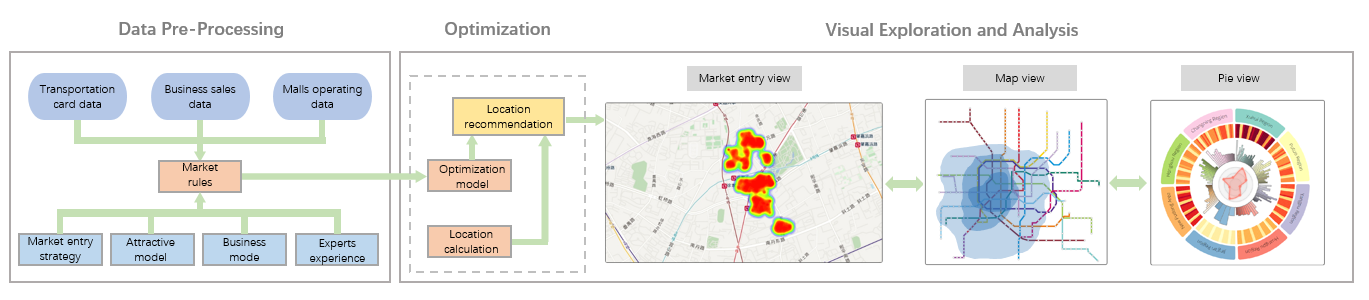


Fig. 2. Overview of our system workflow. In the data pre-processing stage, we clean up and cluster the data, and then, we combine the location recommendation strategy, business model and the experience of experts to optimize the attractiveness model. Finally, we use the mature location selection method to complete the location recommendation. Moreover, we use multiple visualization modules to present this information.

# Related Work

This section discusses the prior studies in the following four categories:

**Attractiveness model** of the business district has been well documented in the literature of economics. Scholars have built models of attractiveness respectively for urban and suburban [1, 2], retail stores [4] and shopping malls [1, 3]. Those models are generally based on Reilly’s rule [5] and Huff's model [6]. Comparing with Reilly’s rule, Huff's model can provide information on the number of customers. Therefore, one of our studies is to optimize Huff's model to calculate the attractiveness of the business district. In our work, we need to derive accurate customer flow, in order to a further assess to the maximum possible profit. However, traditional economic methods cannot meet our requirements for accuracy, so we use the machine learning and linear regression with visual analysis methods to optimize this model. To our knowledge, the problem has not been systematically using data mining and visual analysis methods. In this study, we present an optimized model of business district attractiveness.

**Choosing Location** has been assisted by information systems based on social network analysis, such as Foursquare [27] [28] Heckman-style model [47, 48]. However, existing method failed to provide overview for exploration, or insufficient to support comparison between candidate solutions. In our study, we use k-location method [23, 28] that is a calculation way use greedy algorithm. Unlike others, we first quantify the location-influencing factors and calculate the weight for each, and then get the advantages of each possible location, finally, recommending the best location. Besides, our research has good applicability and expansibility.

**Visual comparison** is a fundamental visualization task [20, 21]. Researchers have explored [20, 22] efficient method to provide clear view for quick and intuitive comparison [23]. Most visual comparisons take juxtaposition approach or superposition approach [17, 18]. We use a juxtaposition method in our work like previous work [23, 24], and put all recommended location by side, so that the users can get the pros and cons of each method under different conditions. In our system, we provide three interactive views (Fig.8) to the users in order to a better experience.

**Business intelligence** has been extensively studied and applied [25, 26] in retail stores [10], customer behavior analysis [11], and market research [12, 13]. However, most of these studies through a series of field investigation completed, with a certain degree of subjectivity. Our research focuses on solving the problem in market entry strategies and particularly recommending retailing stores location by utilizing visual analytics techniques. Based on heatmap [14, 15] and stream diagram [16], we develop time-series-based customer flow visualization techniques to intuitively reveal the pattern of consumer behavior.

# Overview

We conduct a focus group composed of experts in business and marketing to analyze the steps and concerns in retailing store location selection. Then we propose a workflow to achieve the solution of retailing store location, and specify the tasks in our analysis scenario.

## Background Knowledge

Choosing a suitable location has to be supported by the understanding of specific business scenario and multidisciplinary knowledge, including economics, statistics, marketing theory and scientific computing. Our invited experts in marketing practice, marketing research, and information visualization propose their needs and concerns in choosing a location for retailing stores. Experts reach a consensus on the procedures in the store location selection, including

* **Zooming in a business district.** Based on the zoning knowledge of a city, managers have to learn the attributes of a business district such as the attraction of customer flow, radiation scope of business district, et al.
* **Location selection in a specified area.** Managers have to take operating costs (store costs, labor costs, etc.) and competing neighbors into account, and choose between retailing spaces in shopping malls and department stores. It is a demanding task to collect above-mentioned information. Managers seek for support from information systems. Among the influential factors, experts believe that the customer flow is vital for sale and profit, and should be highlighted in the decision-making.
* **Assessment of selections.** The decision of a retailing location is usually made collectively by multiple stakeholders. It is very hard to reach consensus on location selection without scientific assessment and intuitively represented information. A solution for convenient comparison is highly needed.
* **Providing users with a variety of solutions.** Diversifiedstores may need customized solutions to meet the heterogeneity of business operations. The visualization system has to provide a personalized approach for explorations.

## Input Data Sets

We mainly use three types of data as follows:

**Transit card data** include the time, location and amount of fare payment of each passages for buses, subways and taxis. Totally 470 million records cover 30 days in April 2016.

**Sales data for clothing retail store** contains information on products original prices, discount prices, category and quantity of sale of more than a hundred stores in the last six months. We collected 118 department stores and retailing spaces in shopping malls in the center business district of Shanghai.

**Store category and location data** contains the exact locations of stores (including department stores and retailing space in shopping mall) in Shanghai, the surface area of each store, the cost of rent et al. In the study, we also use a number of market vitality indexes develop by Urban Development Institute of Fudan University to measure the attributes of business districts [46].

## Task Analysis

According to the four steps of retailing store location selection, our research focuses on providing support for the following tasks:

* T1: **Optimization of business attractiveness model:** *Does the model can predict the customer flow? If it cannot support our analysis tasks, how to optimize?* This helps us obtain the predicted value of customer flow.
* T2: **Location model design:** *How to reconcile the relationship between the factors? How to choose locations to meet the expected?* Too many factors are not very useful, so getting the relevant factors is one of the tasks to choose locations.

The basic tasks above focus on data mining and model optimization. Moreover, in order to recommend the appropriate commercial location to users, we need a visualization system:

* T3: **Business information display:** *what location is right for the market? Where is the best location?* Users want to obtain useful information, and for us, it is important to statistics and clustering.
* T4: **Location recommendation:** *Where are the recommendation locations? Whether there is a greater development prospect?* A variety of options will display in the view to satisfy the user requirement.
* T5: **Location comparison, assessment and ranking:** *What are the benefits of different store locations? What is the basis for these rankings?* It is necessary to provide a detailed explanation to convince the users.

# Model

## Market Attraction Model

We measure the customer flow using Transit card data, include the time, location and amount of fare payment of each passage for buses, subways, and taxis. Totally 470 million records cover 30 days in April 2016.

Basically, mall charm and consumption resistance is used to calculate the business district attractiveness in our model, and market cost in the most important influencing factor in consumption resistance. Then, we will get a predict value of customer flow by number of shoppers and the attractiveness.

* *Customer flow* refers to the number of customers visiting a shopping mall or department store during a period. We predict customer flow using a different method for working days and holidays.
* *Business district attractiveness* usually expressed as the possibility that residents go shopping in the district. It is an important measure of the prosperity of a certain place. Business district attractiveness commonly determined by the business area, traffic and geographical location and so on.
* *Mall charm* refers to the competitive advantages like the store area, fame or goods abundance.
* *Consumption resistance* refers to the factors diminishing the residents shopping, such as long commuting time and distance.
* *Market cost* includes rent, payment to sales, advertising, et al.
* *Shopper number* is the number of people start a trip from a station that have a shopping tendency.

In our study, a major study focuses on the analysis of the attractiveness for business districts. In this section, we introduce the analysis and optimization of the models to complete Task 1.

### Model Evaluation

The analytic model of the attraction of the business district mainly includes Reilly’s [41, 44] rule and Huff’s model. Since Reilly’s rule is exclusively applicable to a single market competition environment that means it is suitable the region only has one business district, we use Huff’s model [42] to describe the attraction of business district. In the original’s model

(1)

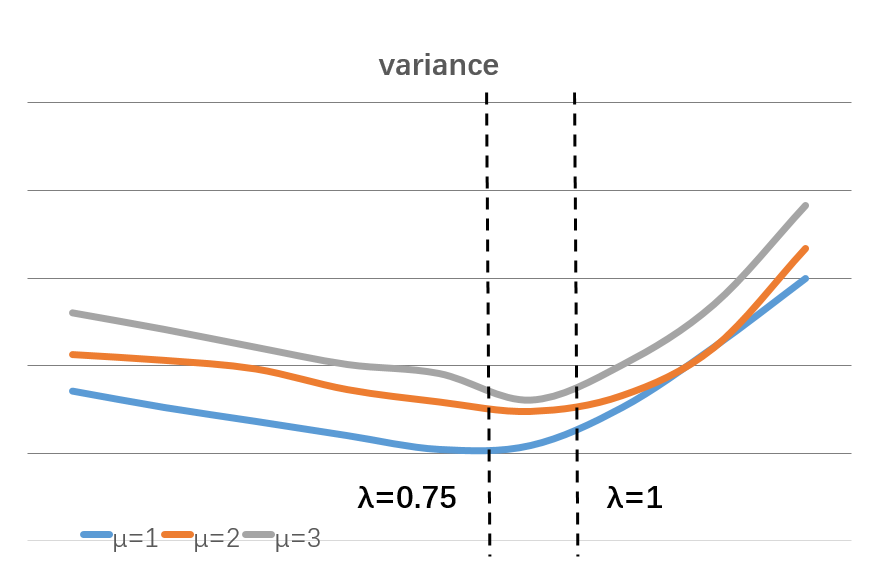
Where refers to the probability of consumer shopping for i to j, and pertains to the attraction of the business district j, is the resistance where i to business district j. and are variables that are estimated on the basis of experience, and in our research, these variables are obtained by big data analysis and machine learning. Moreover, n is the number of business district in competition.

Then we can get the number of people who from *i* to the business district *j* as

(2)

is the number of customers in region *i*.

### Data pre-processing



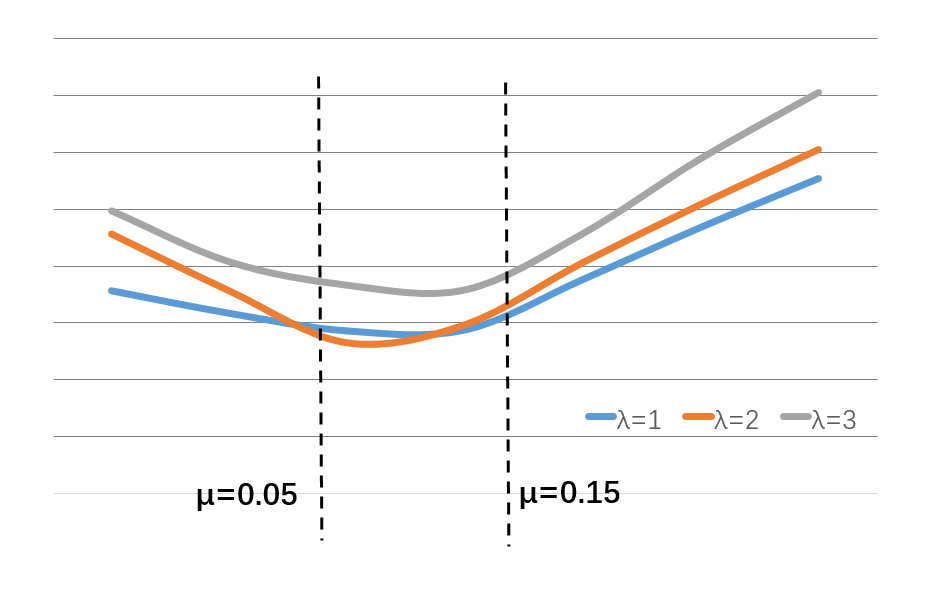


Fig. 3. The error calculated by Huff’s model, where has the least error from 0.05 to 0.15 (μ) and 0.75 to 0.1 (λ).

According to the report of the Shanghai Municipal Government, we have selected a total of 125-rail transit stations in Shanghai downtown and nine city-level shopping district for analysis and research. Nine city-level business district are North Sichuan Road, Zhenbei Road, Zhongshan Park, Wujiaochang, Nanjing East Road, Xujiahui, Nanjing West Road, small Lujiazui and the new Hongqiao.

As shopping schedule is mainly concentrated in the holidays, we first carried out statistical analysis of the card records data and removed the stations that have fewer customers flow. As the business district has a great function of shopping and after discussion with professionals in related fields, we assume fixed proportion of potential customers that is the larger customers flow number has larger potential customers. We carried out the classification and statistical analysis of the data, calculated the number of people who from station to the business district, and further got the actual attractive value.

We use the k-means clustering algorithm to cluster the 112 data sets, and we have nine clustering points because there are nine business districts. Then, we found that some sites had a very significant preference, that is more than 50% of people choose to go to the same business district for shopping. Besides, we found that most of these sites have the same characteristics: the resistance value to one business district is very small (the residential area around business district). We removed these sites, which may have a large effect on the results. Finally, we have a further study for the remaining 82 sites.

We use Huff’s model which have different and to calculate the value of the attraction, as shown in Fig.3.

Through the above figure, we can get the error of preference value by comparing the difference about and. Moreover, we infer that the most suitable is 0.1 and is 0.85. Subsequent experiments show that our results are more reliable in calculating the attractiveness of Shanghai business districts.

Through the above study, we get a targeted parameter, which can calculate the value of attraction. Then, we have a further optimization because we find that there are still errors.

### Model Optimization

However, the above results still have some errors, which have great influence on our following research. Therefore, we take the attraction and resistance as the point of penetration for further model optimization. Fig.4 is the model optimization process.

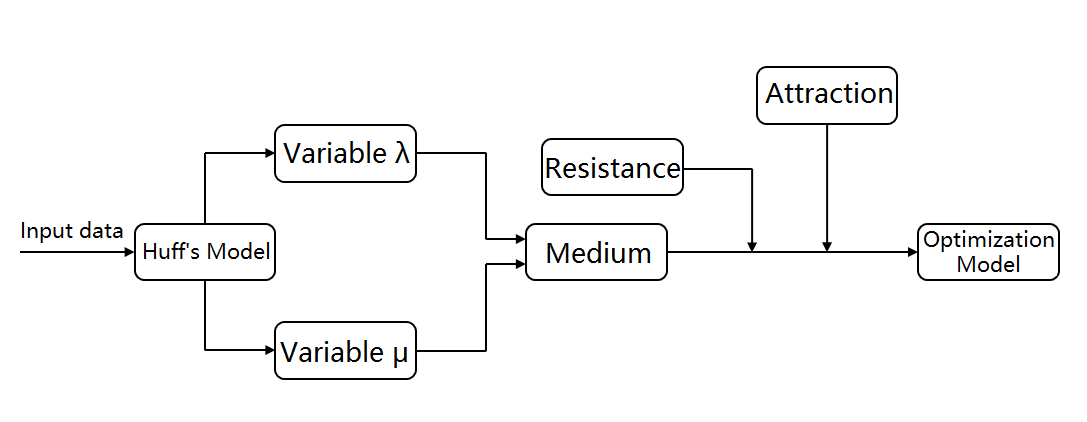


Fig.4. Flowchart of model optimization.

After many discussions with Expert A and Expert B, we believe that the resistance and attractiveness of Huff’s model can optimize further, Expert B put forward that, the more shopping centre near the place, the greater resistance of shopping to big business district and Expert A suggest that we add the degree of abundance in the calculation of attraction.

Hence, we initially proposed the formula, as follows:

(3)

Where *A* and *R* are the correction value of the attraction and resistance, therefore, the experts give the value of A and R. Simultaneously is the number of stores, is the number of malls around the residence and *m* is the correction value. Some validation and explanations will be given in case study.

## Proposed the Commercial Location

The study shows that [38], the main factors affecting the location choice include shop diversification, shop competitive, shop interaction and traffic flow. We explored the location model to solve the problem raised by Task 2.

Regardless the difference between business districts and neighbour competing or supporting stores, our research focuses on the accessibility of a store, which is predicted using traffic flows.

We compared the attractiveness of different malls to users, and finally, we compare the advantages and disadvantages between different malls in the same business district, to assist users in the development of location recommendation strategies.

In the calculation process, through the statistical analysis of the location of malls and the sales data in Shanghai, we found that there is a great relation between sales volume and the distance from station to mall (sales volume is proportional to the customer flow), which is the closer the distance, the higher the sales volume. In addition, the scale of the mall also has an impact, but its impact is small. After discussion, we believe that we can ignore it in our research.

We summarize the eight variables that affect the location (5.4) as follows: labour, competition, traffic, flow, rent, size, saturation and influence. Firstly, we calculate the influence factor about these eight variables and quantify these as a value for one to ten.

(4)

In the calculation of recommendation value, is the weight of each variable and is the value of the influence factor.

For the calculation of the weight, it is difficult to use the data-mining model due to the insufficient data. Therefore, we get the weight through the expert scoring and fix it by iterative correction.

Because of the stability of the mall, the scope of the location recommendation is limited. We select the most advantageous ten positions sorted by recommended to the users based on more than one hundred malls.

# Visual Design

This module describes a set of visualization techniques that assist users to develop market entry strategies. Section 5.1 and 5.2 display the business information for Task 3, including administrative divisions, the influence scope of the business district, sales changes, et al. For Task 4 and Task 5, we use some visualization techniques in Section 5.3 and 5.4.

## Business Influence View

The map view is common in the study of multidimensional geographic information visualization, and in our study, we use the subway map view to show the influence scope of business districts and administrative division.

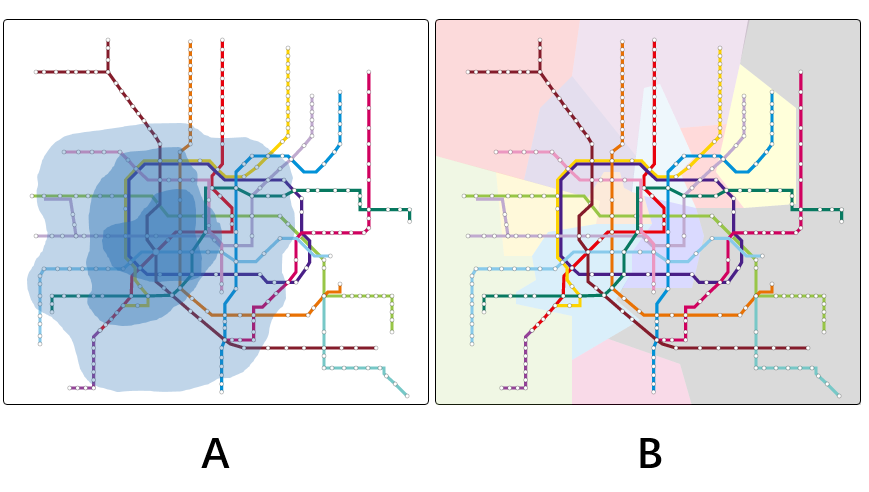


Fig. 5. (A) An influence scope view of Xujiahui business districts. (B) An administrative division view.

In this view, we use different degrees of colour to represent the range of influence (Fig.5A), where the deepest colour indicates that the inhabitants here have more than 30% going to this business district to shop, the shallower is 20%, and the shallowest is 15%. We get this data by clustering all stations (cluster center is the business district in center city). We found that there is a clear regular pattern about the consumption preferences of residents: the residents prefer to go to the mall that has less time cost, meanwhile, the area and visibility of shopping mall also have a great impact. For example, Xujiahui is one of the largest business districts in Shanghai, after analysing and comparison, we find that residents still like to come here even there is higher time cost.

The influence calculation is in Fig.6a. Where A, B, C are the stations and A is target station, a, b, c are the midpoint of two stations and the black line is the normal line through target station, and the green area is the influence area of the target station. First, we calculated the midpoint between the target station and adjacent station, and then we get the normal line through the target station. Finally, we draw semicircle according to the distance between the two stations. If there are multiple stations, we can also use the same method. After calculating the approximate range of influence, we connect these graphs to form an influence area like Fig.6b.

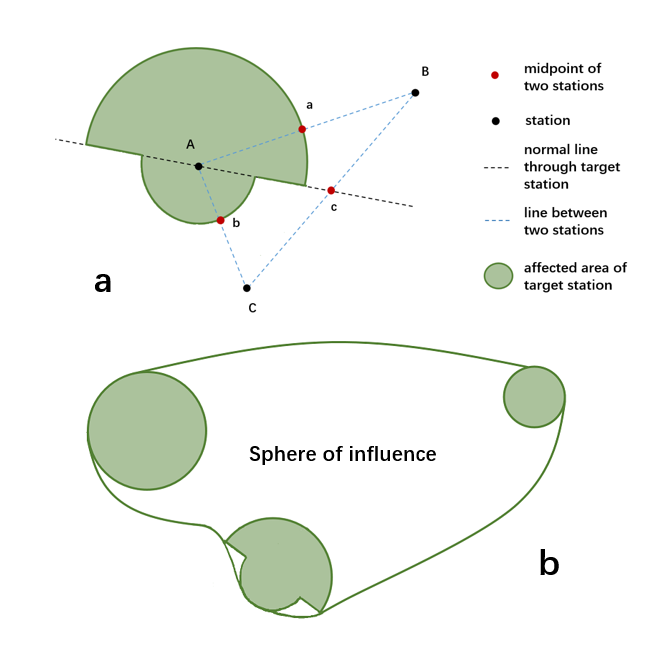


Fig. 6. The diagram about the influence area of the subway station.

Meanwhile, we divided all stations according to the administrative region (Fig.5B), where each colour represents one region, and users can select the information through interaction. In addition, we provide a zoom function for better satisfaction.

## Statistical Analysis View

This part is a multi-layer circular diagram that is used to show the inherent pattern through detailed statistical information.

In this section, we provide the main presentation of the statistics data. Fig.7A is the diagram shown in our system, and we provide two views to the users, one is the district view (Fig.7D) and another is the legion view (Fig.7C).

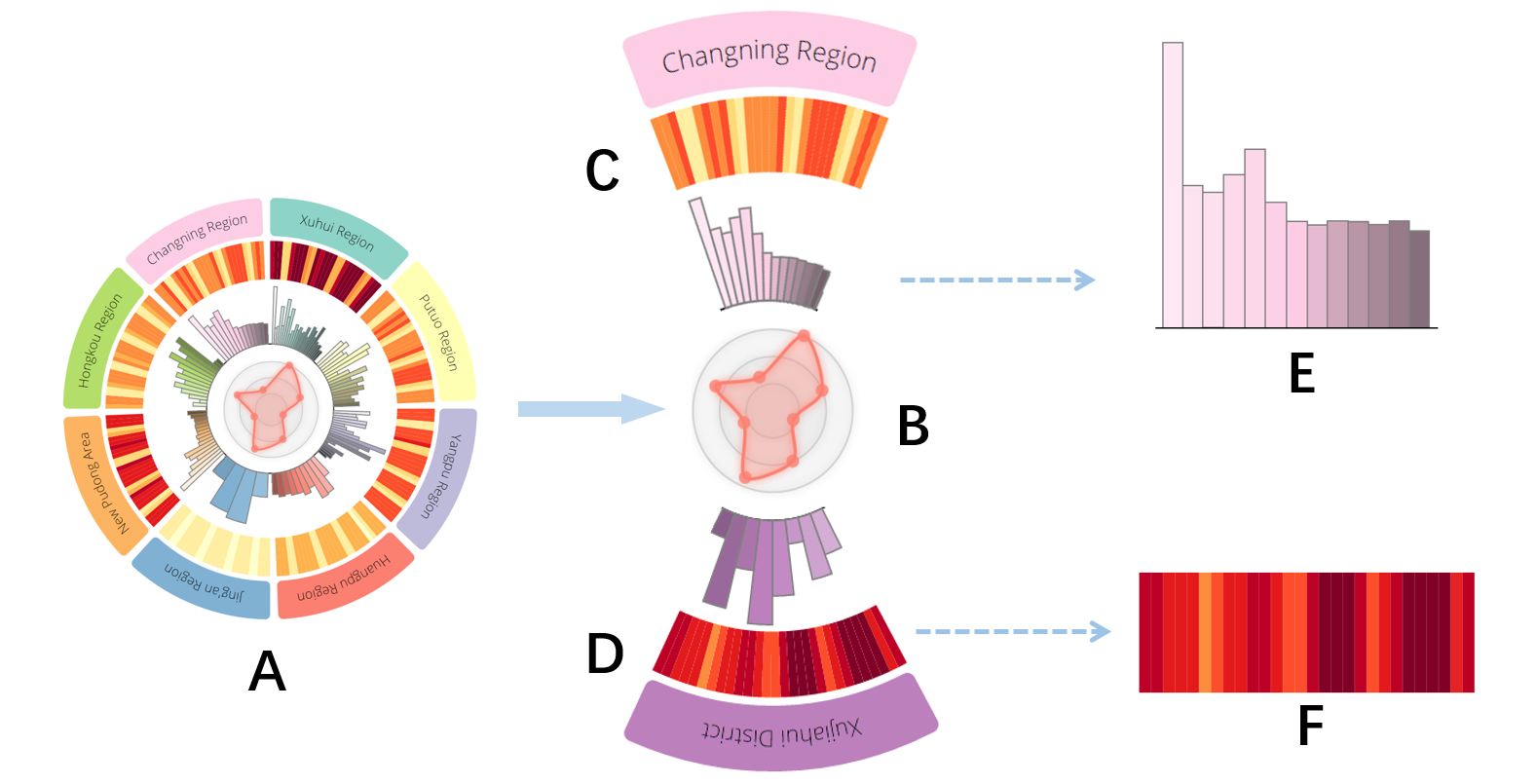


Fig.7. Statistical analysis view. (A) The overall in our system. (C, D) A pie shows the information for one area. (B) A grade about business prospects. (E, F) Bar chart and heat map for the details in one area.

In this view, each sector of the pizza chart represents a business district or an administrative region. One pie provides a variety of information, like the customer flow in one month (Fig.7F) and sales volume of each mall in this business district in district view, and the number of the traveler in one month and the shoppers of each station (Fig.7E) in region view. Using transit passes data we assume that the transit passes holders who regularly pass between 6:00-9:00 and 5:00 - 7:00 pm each weekday are employees working in the business area, and we predict the customer flows by excluding the employees. We obtain the sales volume through the sales data from the company and the customers flow by Huff’s model. Moreover, the traveler expressed the approximate number of residents in this area.

In addition, we also predict the business prospects of the current region through expert experience (T5), which in the range 1 to 10 (Fig.7B). However, the accuracy has not yet to verify, because it is based on experience.

As result, the number of passengers monthly remains constant, and the number of shoppers is relatively stable. This result is different from our expectation. After discussing with the experts, we believe that the reason for this result is mainly due to the prosperous economy and the large population in the big city, which results in a smooth number of shoppers. In addition, the main reason for the regular changes in the number of travellers is that most people do not expect to leave home on holidays.

## Location Recommendation View

Our system can visualize as many as 10 recommended locations on a heatmap displaying market prosperity (Fig.10).

We also provide the input box for the expected cost and profit, and through the user interacts, the system will display ten locations that best meet the expectations.

Except for the solutions (Fig.8), we use one-year data to reduce the error of the predicted propensity.

In addition, we also provide a zoom function to better display our research.

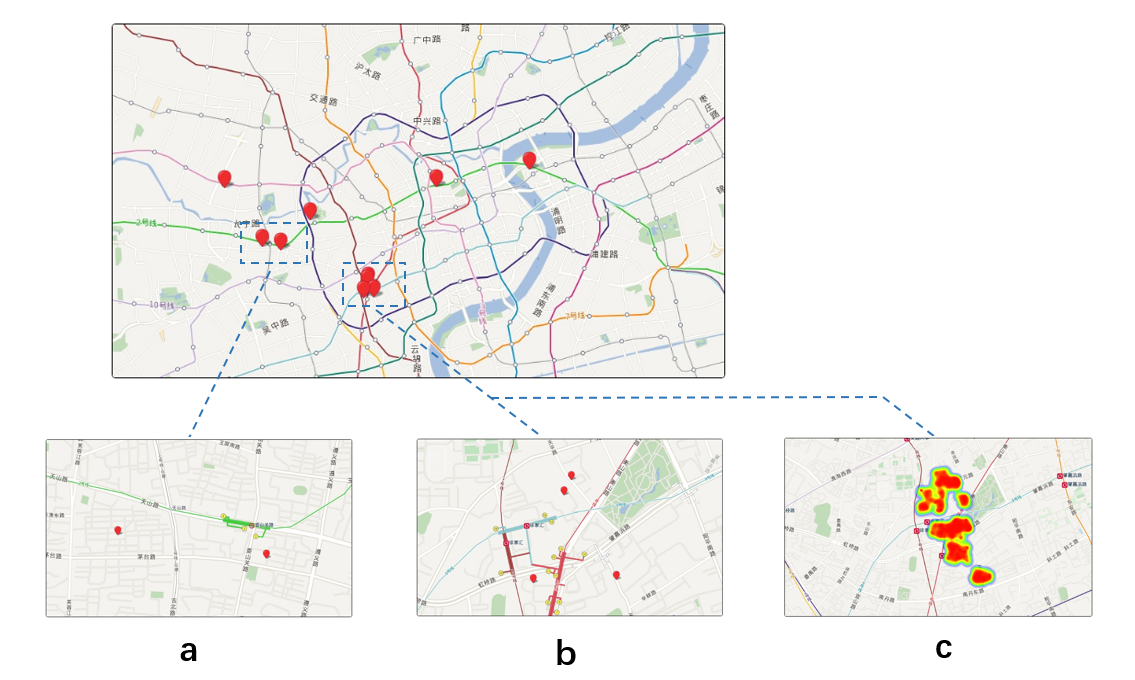


Fig.10. Recommended location of the top ten. This view contains a zoom function and the heatmap of business districts.

## Visual Comparison View

In order to solve the T5 and T6, we design the visual comparison view (Fig.10), it makes a good comparison of the advantages of different locations. In Fig.10A, we provide the users with a detailed comparison of each solution, and an overall comparison in Fig.10B, finally, users can rank the solutions (Fig.10C).

We chose eight details from more than twenty factors. First, we get more than twenty factors that can be classified into four categories: consumer factors (business district preference, time cost on transit), mall factors (labour cost, rent, advertising cost or accessibility), market factors (industry competition or market saturation) and social factors (economic basis, economical policy or fashion trend).

In these, the factors of market saturation, industry policy have the same value, and the main commodity, sales style only have little impact on our research. Therefore, after several studies and discussions with experts, we chose ten factors for further study.

Finally, we found the two factors situation cost and behaviour preference are not available, so although they have an impact on the results, we have to get rid of them.

We use the range one to ten to represent the degree of these factors, one is the worst, and ten is the best. Such as one represents the strongest competitive pressure or the smallest customer flow, and ten represents no competitive pressure or countless customers.

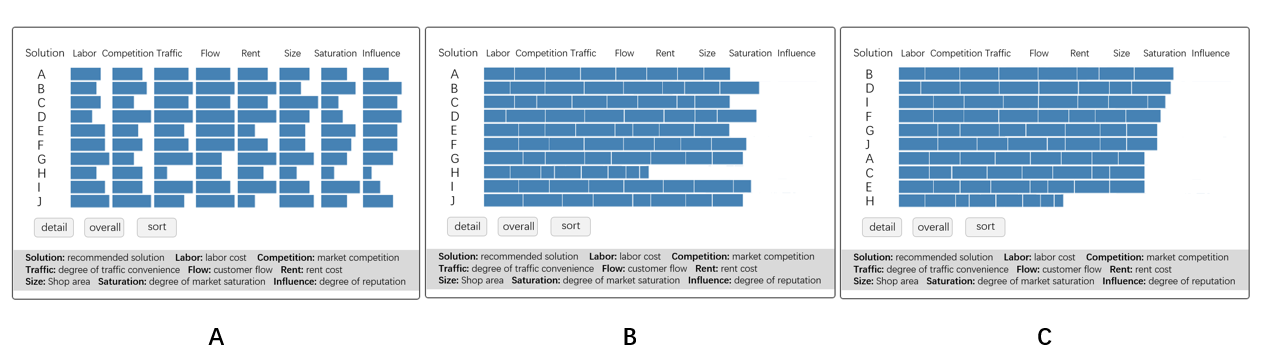


Fig. 8. Advantage comparison about recommended location. This view included the detailed comparison of eight factors (A), overall comparison (B), and the ranking on recommended position (C).

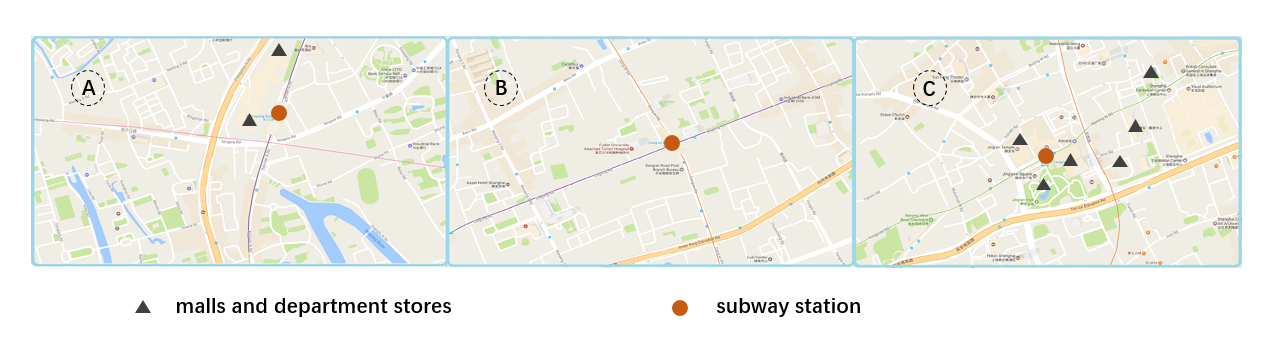


Fig. 9. Three subjects with significant differences have a different market structure.

# Case Studies

We conducted case studies with experts and business managers, and each of them was familiar with some part of our designs.

## Exploring the Advantages of Our Model

Furthermore, we optimized the model of customer flow by conducting quantitative studies.

We use twenty-two days (16 weekdays and 6 holidays) public transportation data to optimize the model, and then we use the additional nine days (working day six, holiday three) data to test the validity of the optimization. The results of the comparison are shown in Fig.11 (A-I are consistent in 4.1.2) and part of the details of the business districts is shown as Tab.1.

From the results, we can see that our optimization model is able to achieve our customer flow predict task basically, although some deviation. We can see the business district D, E and H have a greater error. On the business district D, experts believe that the result is due to a unique geographical location, this district is located near the university town. As the proportion of students here is far greater than the rest districts, and college students are more favour of using personal transport (bicycles, walking), resulting in the deviation. Moreover, another possible reason is that the geographical location is remote, the customers in where has no public transport will choose the personal traffic (car). This speculation is also reflected in two other remote business district A and B.

Tab.1. Attribute description about business districts. The number of malls only include a shopping center and the distance to city center is represented by the travel time (min).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| District | A | B | C | D | E | F | G | H | I |
| Number of Malls | 4 | 4 | 9 | 7 | 12 | 11 | 9 | 3 | 3 |
| Distance to Centre | 24 | 33 | 14 | 37 | 5 | 11 | 6 | 21 | 16 |

Business district E has too many tourists with the complex situation in using the public transport, and so far, we do not have a good solution yet. For the business district H, the main result is due to the uncertainty time in using public transport, because there are too many companies.

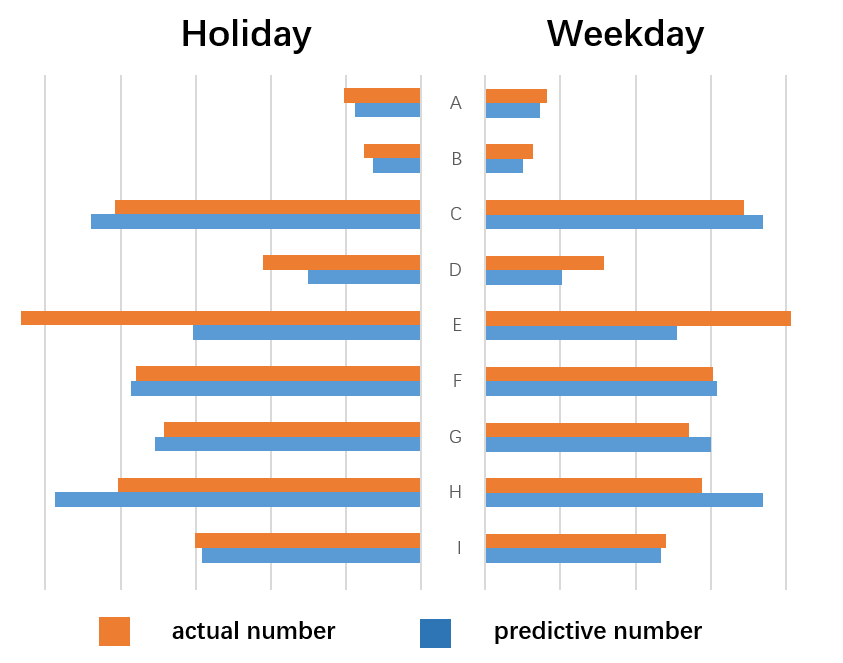


Fig.11. Comparison of actual and predicted numbers in weekdays and holidays perspective.

At the same time, the experts give the possible situations: (1) the closer to the city center, the more accurate, (2) the higher accuracy of the holiday than the working day（3）the more malls within the business district, the more accurate. Therefore, we analyse the results from these three perspectives. Our results are as Fig.12.

The result is sorted by the number of malls in the business district and the distance from the city center. It is clear that there is no obvious correlation between whether it is a holiday and the accuracy, and the number of malls and the distance to the city center show a clear pattern. Of course, this does not include the business district E, because it is very special in our study.

For these results, experts believe that the more close to the city center, the wider radiation range of the business district, because of urban traffic congestion, more people tend to choose public transport as a way to shopping. The more away from the city center, the greater number of customers choose travel themselves, which result in the deviation. The number of shopping malls has an impact on the predict results, which showing a clear ladder in the figure. Greater number of malls means greater attraction [45], so the result will stable, and others to attract customers most rely on advertising. However, whether in holiday have no obvious impact on the result, we expect the next work will find some new rules.

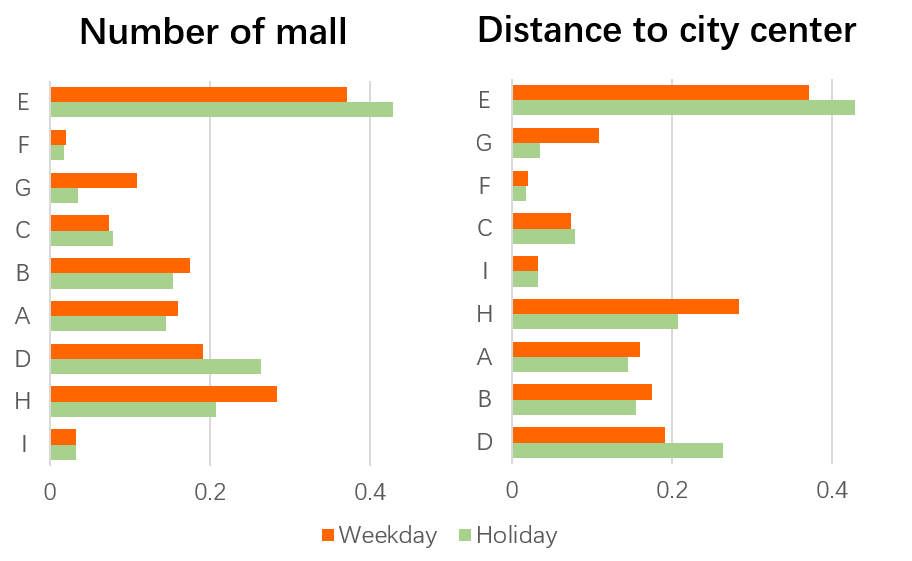


Fig.12 Two different sorties according to the number of malls and the distance to the city center. Each business district provides a comparison of the average error for weekdays and holidays.

## Comparing the Solution

Because the accuracy of the results is difficult to determine, we study it in the following way: first, enter the expected value according to the discussion with the experts; and then provide the results to the business managers; finally, verify our work through feedback.

We entered the cost of 500,000 ¥, the profit of 700,000 ¥, which cost include labour costs, rents, the goods cost and so on, and profit cost is pre-tax profit. Both are in one month. We get ten solutions (Fig.8A), which has four in Xujiahui, two in New Hongqiao and one on others. We can see the specific location (Fig.10b) and the ranking of the solutions (Fig.8C).

We take the results to a business manager, and he gives us some helpful advice. He thinks that except the four solutions in Xujiahui, our solutions are quite feasible. He mentioned that because of the large scale of Xujiahui district, there are several stores already. Therefore, except the similar goods, the competition of same goods is also very intense. Hence, a new store will have a higher risk and lower profit than expected.

He suggested that we consider the density about the same brand, for example, two adjacent Nike stores will greatly affect the profit. He also suggested us focus on the scope of the middle ring (a road divided the areas of the city). Because he believes that, there is a more development potential in a certain distance from the center.

He also suggested that we join the analysis of pure profits because it can make the system more clear.

# Discussion

Our evaluation demonstrates the effectiveness of retailing store location selection system. Nevertheless, there is still space for improvement.

This system focuses on location recommendation in the centre business district, and our research does not spread to the medium and the small due to the limitation of data. Nevertheless, we still believe our optimization model is suitable for predicting the customer flow, but there is still space for optimization. A possible solution to the problem is hierarchical research and optimization, because there are a number of medium-size and small-size business districts. Furthermore, we try to apply our optimization model in non-central business districts. After comparison and verification, we find the results not particularly obvious advantages compare with other models. Therefore, we plan to collect more business information to improve the applicability of our model.

Regarding the location recommendation model itself, due to the complexity of the multidimensional influencing factors, we just focus on the main factors, it is necessary to adopt a more complex method to improve the accuracy of the result. One feasible way is to use data mining (neural network) with the computer computable ability for detail analysis about correlation. This is a very common problem in the field of high-dimensional data. We plan to cooperate with business managers to discuss deeper problems for improving the analysis mechanism in order to get a more accurate location recommendation model.



Fig.13. Standard deviation and the error on the most preferred business district.

This research is based on the existing business district for choosing locations, which has very high saturation in the market and it will cost enterprise resources too much. Therefore, company managers prefer to open a new store in a brand new market, we predict the customer flow by market attraction model. However, different business regions have different attributes and we have not conducted the research for choosing locations in the new market. In this aspect, we believe one efficient solution is to do a deep study about market operation rules and speculate several effective market attributes according to analysis results.

It is necessary to get sale preferences of different commodities for companies. We apply the market attraction model to the research of attractions of different commodities. Nevertheless, since there is not a highly scalable, we do not get a result that can achieve our purpose. Therefore, it is still a considerable question and we are interested to conduct a deep research about the sale preferences later.

Our research focuses on the large-scale market and the system can be efficiently applied in different types of companies like catering or clothing. However, it requires manual works for pre-processing and analysing data. Therefore, we need a more effective visual driven model to put our research forward.

# Conclusion and Future Work

We systematically study the traffic data and business data, and then solve the problem that gets the value about customers flow in the business district. We work closely with experts and company manager to complete our tasks and challenges facing managers in big companies, namely, provide helpful location recommendation strategies and a variety of solutions. Hence, we design a visualization system that integrates statistical analysis, location recommendation and solutions comparison. It has several well-designed visualization and interaction techniques. Our case study validates the usefulness and scalability of our technique.

There are many elements regarding this work that can be further investigated and addressed, either in future versions of our system or in work done by others. For example, we have only implement a set of parameter focus the urban business districts, but other district options exist. There are also other models can be used to calculate customer flow, except from business attractiveness.

Our study results, analysis, and discussion should be noted with the following considerations: The foci of our studies are optimize the attractiveness model where need more analysis about the suburbs, and design visual views that need establish a comprehensive multidimensional factor analysis mechanism to provide better location recommendation.

For deployment in a different domain, such as shopping behaviour analysis about white-collars or elderly, there are many new factors to consider. Different ages of people will have certainly varied tendencies. In different area, the same type may also have different tendencies. In this context, a tool for location recommendation must ensure a detailed and abundant business analysis.

References

1. Dennis C, Marsland D, Cockett T. Central place practice: shopping center attractiveness measures, hinterland boundaries and the UK retail hierarchy ☆ [J]. Journal of Retailing & Consumer Services, 2002, 9(4):185-199.
2. Wahlberg O. Small town center attractiveness: evidence from Sweden [J]. International Journal of Retail & Distribution Management, 2016, 44(4):465-488.
3. Yao Lizhen, Yue Yang. Factor analysis of shopping center attractiveness based on principal component logistic model [J]. Journal of Geo-Information Science, 2016.
4. Singla V, Rai H. Investigating the effects of retail agglomeration choice behavior on store attractiveness [J]. Journal of Marketing Analytics, 2016:1-17.
5. Reilly T. Reilly's rules for value-added selling [J]. Official Board Markets, 2006.
6. Honda M, Ushizawa K. Quantitative geographic study of buying behavior: Huff's model and its parameter estimation [J]. Sanno College Bulletin Department of Management & Informatics, 1982, 2:81-104.
7. Anne Marie Doherty. Factors influencing international retailers' market entry mode strategy: qualitative evidence from the UK fashion sector [J]. Journal of Marketing Management, 2000, 16(1):223-245.
8. Doherty A M. The internationalization of retailing: factors influencing the choice of franchising as a market entry strategy [J]. International Journal of Service Industry Management, 2007, 18(2):184-205.
9. Lee H S, Griffith D A. Transferring corporate brand image to local markets: governance decisions for market entry and global branding strategy [J]. Advances in International Marketing, 2012, 23:39-65.
10. Khakzar K, Blum R, Kohlhammer J, et al. Interactive product visualization for an In-Store sales support system for the clothing retail [C]. Human Interface and the Management of Information. Methods, Techniques and TOOLS in Information Design, Symposium on Human Interface 2007, Held As. 2007:307-316.
11. Yaeli A, Bak P, Feigenblat G. Understanding customer behavior using indoor location analysis and visualization[J]. Ibm Journal of Research & Development, 2014, 58(5/6):3:1-3:12.
12. Nesbitt K V, Barrass S, Stephen Barrass@csiro. Evaluation of a multimodal sinification and visualization of depth of market stock Data [C]// R. Nakatsu and H. Kawahara. 2002:2--5.
13. Macer T. Data visualization in market research [J]. Research World, 2014, 2014(47):10–19.
14. Pryke A, Mostaghim S, Nazemi A. Heatmap visualization of population based multi objective algorithms[C]// International Conference on Evolutionary Multi-Criterion Optimization. Springer-Verlag, 2007:361-375.
15. Hashimoto Y, Matsushita R. Heat map scope technique for stacked Time-series data visualization[C]// International Conference on Information Visualization. IEEE, 2012:270-273.
16. Shibahara N. Realtime Panoramic Mobile Streaming Application for Assisting Visualization to Remote Person [J]. 2015.
17. C. Shi, Y. Wu, S. Liu, H. Zhou, and H. Qu. LoyalTracker: visualizing loyalty dynamics in search engines. IEEE TVCG, 20(12):1733–1742, 2014.
18. Y. Wu, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou, and H. Qu. OpinionSeer: interactive visualization of hotel customer feedback. IEEE TVCG, 16(6):1109–1118, 2010.
19. Gleicher M, Albers D, Walker R, et al. Visual comparison for information visualization [J]. Information Visualization, 2011, 10(4):289-309.
20. Borowski K, Soh J, Sensen C W. Visual comparison of multiple gene expression datasets in a genomic context [J]. Journal of Integrative Bioinformatics, 2016, 5(2):94-103.
21. Kim M C, Zhu Y, Chen C. How are they different? A quantitative domain comparison of information visualization and data visualization (2000–2014) [J]. Scientometrics, 2016, 107(1):123-165.
22. Liu D, Di W, Li Y, et al. SmartAdP: Visual analytics of large-scale taxi trajectories for selecting billboard locations [J]. IEEE Transactions on Visualization & Computer Graphics, 2016:1-1.
23. Collins C, Carpendale S. VisLink: revealing relationships amongst Visualizations [J]. IEEE Transactions on Visualization & Computer Graphics, 2007, 13(6):1192.
24. Koleva, Vyara. Interactive visualization of the building of university of economics – varna via 3D modeling [J]. Digital Presentation & Preservation of Cultural & Scientific Heritage, 2013, 263-266(III):8.
25. Kuang Y, Liang Z, Lei Y. Scientometrics research on spatial economics based on information visualization [J]. Science & Technology Management Research, 2012.
26. Ye M, Yin P, Lee W C. Location recommendation for location-based social networks[C]// ACM Sigspatial International Symposium on Advances in Geographic Information Systems, Acm-Gis 2010, November 3-5, 2010, San Jose, Ca, Usa, Proceedings. DBLP, 2010:458-461.
27. Bao J, Zheng Y, Mokbel M F. Location-based and preference-aware recommendation using sparse geo-social networking data[C]// International Conference on Advances in Geographic Information Systems. 2012:199-208.
28. U. Feige. A threshold of inn for approximating set cover. Journal of the ACM (JACM), 45(4):634–652, 1998.
29. Wills G. Visualizing network data [J]. IEEE Transactions on Visualization & Computer Graphics, 2016, 1(1):16-28.
30. Huang X, Zhao Y, Yang J, et al. TrajGraph: a graph-based visual analytics approach to studying urban network centralities using taxi trajectory data [J]. IEEE Transactions on Visualization & Computer Graphics, 2016, 22(1):160.
31. Quinan P S, Meyer M. Visually comparing weather features in forecasts [J]. IEEE Transactions on Visualization & Computer Graphics, 2016, 22(1):389-398.
32. Goodwin S, Dykes J, Slingsby A, et al. Visualizing multiple variables across scale and geography [J]. IEEE Transactions on Visualization & Computer Graphics, 2016, 22(11):599-608.
33. Guo H, Gomez S, Ziemkiewicz C, et al. A case study using visualization interaction logs and insight [J]. IEEE Transactions on Visualization & Computer Graphics, 2016, 22(1):51-60.
34. Aldohuki S, Wu Y, Kamw F, et al. SemanticTraj: a new approach to interacting with massive taxi trajectories [J]. IEEE Transactions on Visualization & Computer Graphics, 2017, 23(1):1-1.
35. Reuveni E. System architecture and method for customer flow management. US, US20050259653 [P]. 2005.
36. Beatty S E, Mayer M, Coleman J E, et al. Customer-sales associate retail relationships[J]. Journal of Retailing, 1996, 72(3):223-247.
37. Hongyang Q, Zhiwen Y, Miao T, et al. Research and implementation of commercial site recommendation system based on LBSN. School of Computer Science. Xi’an, China, 2015, 1002-137X.
38. Takam M J, Park B J. Entry strategy of korean construction SMEs in the african market [J]. Nature, 2015, 428(6983).
39. Kamau A. Market entry strategy [J]. Journal of Global Marketing, 2013:89-110.
40. Kang J S, Downing S. Keystone effect on entry into two-sided markets: An analysis of the market entry of WiMAX [J]. Technological Forecasting & Social Change, 2015, 94:170-186.
41. Reilly T. Reilly's Rules for Value-Added Selling [J]. Official Board Markets, 2006.
42. Shu C, Wenwei J, Yuan W. Huff model used for green space layout of Guanhaiwei Town. Journal of Zhejiang A & F University, 2013, 30(6): 887-895.
43. Huff D L. Defining and Estimating a Trading Area [J]. Journal of Marketing, 1964, 28(28):34.
44. Reilly W J. The Law of Retail Gravitation [J]. American Journal of Sociology, 1932.
45. Zhao F, Sun H, Wu J, et al. Analysis of Road Network Pattern Considering Population Distribution and Central Business District: [J]. Plos One, 2016, 11(3):e0151676.
46. Yina Z. Urban Development Institute of Fudan University. <http://www.linkshop.com.cn/web/archives/2016/341823.shtml>.
47. Branikas I, Hong H G, Xu J. Location Choice, Portfolio Choice [J]. Social Science Electronic Publishing, 2017.
48. Song Y, Jin E, Kim I. Evaluation of elderly mobility based on transit card data in seoul [j]. Promet-Traffic & Transportation, 2014, 26(4).