NATIONAL UNIVERSITY OF SINGAPORE

INSTITUTE OF SYSTEMS SCIENCE

Intelligent shipping recommendation system

Author:
Shu WanYang
Yan JiaHuan
Xiao ChangWei
Supervisior:
Mr. GU ZHAN

Submitted in partial fulfillment of the requirements for the Mtech degree in Intelligent System of University of Singapore

October 2022

Contents

Chapter 1

Executive Summary

In 2021, the number of newly registered cross-border e-commerce enterprises in China reached an all-time high of 1,486. This means that there are more and more online orders from overseas in China. On the other hand, through the observation of classmates around us, we found that many people still use Taobao for shopping even after they have left China. Even though there are shopping sites like Shopee or Lazada in Singapore, Chinese students are still used to using Taobao. There are many reasons for this. For example, Taobao has a Chinese interface, the habit of using Taobao for many years in China, and it is easier to find the goods you want on Taobao.

But as more overseas buyers appear, there is an obvious problem. Unlike shopping in China, overseas buyers need to receive their goods by air or sea. However, no matter which mode of transportation is chosen, the uncertainty of transportation time is a big pain point for overseas customers.The existence of this problem is not conducive to the development of the international trade market.

A healthy and sustainable international trade market requires a robust logistics system to solve the problems aforementioned. A Logistics system refers to the organic assembly composed of two or more logistics functional units, with the purpose of completing logistics services. The ”input” of the logistics system is the labor, equipment, materials and resources of procurement, transportation, storage, and distribution processing. Thus, one of our main idea of doing this project is to optimize the existing logistic system of China to Singapore.

The optimization of logistic systems can be divided into several separate parts which include policy optimization, warehouse distribution optimization, or logistics and transportation modeling. Policy optimization refers to reducing entry and exit procedures at Customs. Generally, the cargo will be stuck in customs for a few days even a few months during the epidemic. Warehouse distribution optimization strategy refers to scientifically choosing the location of warehouses, and ensuring the cargo storage capacity and types of goods every warehouse storage align with local market needs.

Under the realistic conditions that policy is determined by goverments and warehouse distribution is determined by logistic coperation, we introdeuc a more prac-

tical optimization methods for the customers: to select the commodities whose dispatch locationes are close to each other and close to location of consolidation. Hence, we propose a state-of-the-art method to design an intelligent shipping recommendation system to automatically collect information and make decisions.

Using the techniques imparted to us in lectures, our project team first set out to build a knowledge base as a local file of transport centers in china and their transport time and fee from them to Singapore via searching reports on Internet and time cost of transport from every city in china to these transport centers via Amap API. While building the system, aim to reach the goal of easy to use and use to deploy, we decided to utilize the framework like react to build the user interface and make sure the business logic is simple enough, flask which is based on python to build backend API and no database to make the backend easy to deploy, web crawler based on python to get the significant information from amazon and Taobao extract the feature of data by semantic analysis. The highlight of the system is the recommendation function, which we use Genetic Algorithm implemented by python to get the most optimized selection of goods and transport center to reach the goal of the lowest price and least transport time.

Our project team hopes that with our solution, individual customers could have a better overseas shopping experience and cross-border e-commerce merchants in Singapore could minimize the cost of time and money spent on transportation.

Chapter 2

Business Justification

Our group and other Chinese students will still have the demand for shopping on Taobao after arriving in Singapore. Even though there are shopping sites like Shopee or Lazada in Singapore, Chinese students are still used to using Taobao. There are many reasons for this. For example, Taobao has a Chinese interface, the habit of using Taobao for many years in China, and it is easier to find the goods you want on Taobao.

Taobao is the China's largest digital retail platform. In our survey, we found that not only Chinese but also foreigners use Taobao for shopping. During the Chinese New Year in 2021, more than 20 million overseas users shopped on Taobao.

However, there is one obvious difference between using Taobao in Singapore and China. Goods need to be shipped or flown to reach shoppers. It is difficult for users to judge the time from the place of origin to the collection point when selecting commodities, which leads to the inability to select the fastest merchant. This problem is more obvious when purchasing multiple commodities at the same time.

Not only that, the comparison of commodity prices is also of great concern to buyers when shopping.

We realized these two problems and decided to develop this Intelligent shipping system. Users can paste the Amazon link to the product they want into the web page we provide. The system can obtain commodity information through web links and search for similar commodities in Taobao through crawlers in the back end. Users can input multiple items, and the system will automatically generate the optimal time and price.

The use of our system can effectively help users to screen commodities and select the right merchants. It greatly reduces the time from ordering to receiving the goods.

Chapter 3

Project Deliverables

3.1 Project Goal

A robust, self-abductive China-to-Singapore consolidation system will be seen as the standard ideal deliverable in this project. In our vision, the system can automatically find solutions to minimize the cost of customers on multiple dimensions at least including time spent on transportation and money spent on purchasing. As international students from China, we found the inconvenience when we try to purchase goods from Taobao. It normally costs us much more money on the process of centralized shipping, and the shipping time can be unpredictable when one of the commodities is sent from medium/high risk of epidemic areas. Hence, we want to develop a smart consolidation optimization application, to help customers have a better overseas shopping experience. The management and implementation of this intelligent shipping recommendation system will have significant leverage over order leadtime and fill-rate reliability. If utilized properly, it will boost the foreign trade economics of both Singapore and China, and more importantly, it will stimulate the shopping consumption of people and promote economic globalization.

3.2 Project Objectives

This section will discuss what we want to achieve in the process of system designing and target function modeling. And we will divide the total objectives into three parts: data acquisition, feature extraction, and the representation of recommendations.

3.2.1 objectives in data acquisition:

- The data acquisition of consolidation locations in china(e.g. Shanghai, Yixing).
- The data acquisition of commodities (e.g. name, feature, price, location...).
- The data acquisition of the transportation cost between cities (e.g. Shanghai to Yixing).

3.2.2 Objectives in feature extraction:

- Having the reliable features of the target commodity.
- Conducting feature reduction (e.g. combination, drop, dividing).
- Constructing the structured data storage method (e.g. MongoDB, Mysql, local file...).

3.2.3 Objectives in recommendation representation:

- Try to implement a self-build fitness function.
- Try to use the Genetic Algorithm, Informed Search or other methodologies to minimize the target value.
- Try to construct a representation of the performance of the recommendation system.

3.3 Project Milestones

This project is under the organization and supervision of NUS-ISS, thus the working schedule will follow the course requirements:

- At least 30 man-day workload.
- The workload should be evenly distributed every week or day.
- The workload should be fairly allocated to every person.

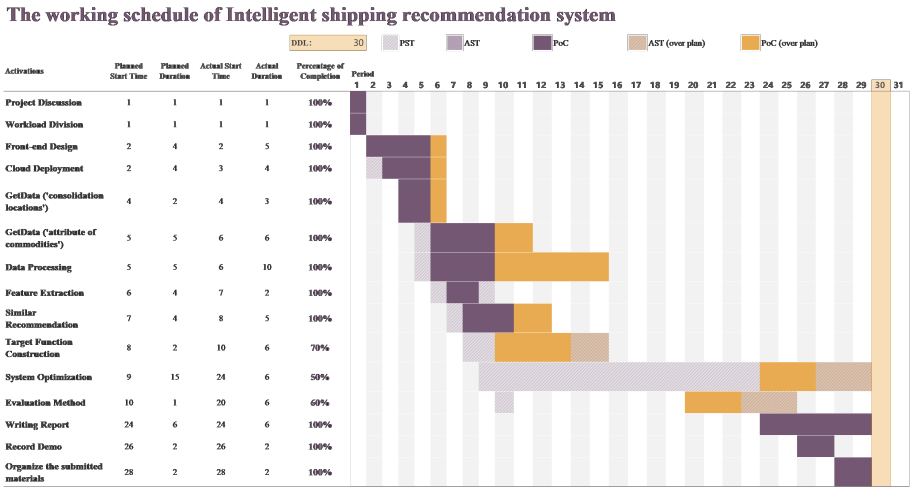

Figure 3.1: The Gantt Chart about workload

Chapter 4

Projects Solution

4.1 Application Features

The intelligent shipping recommendation system aims to make good use of the application programming interface (API) from Taobao to find similar commodities. The information of similar commodities will be collected and processed, the prices and the send-off location of goods will be further crawled from Taobao webpage. The system will calculate the cost function and return recommended solutions to users, thus the customer will know the least delivery time and the relatively least total expenses. Below shows the diagram where the system consists of main features.

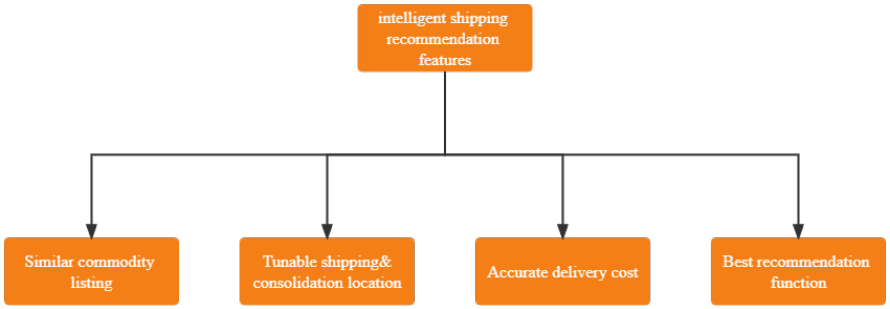

Figure 4.1: The application features of system

Similar commodity listing: The system will automatically find the most relative commodities from Taobao according to the attributions of the original good. Users can manually pick the desired goods from the list otherwise the system will pick the most suitable one for users.

Tunable shipping/consolidation location: The system is outer-connected to our database, where all the location information is storage. The system will search for the nearest consolidation location and calculate the necessary delivery time. The users can filter certain consolidation locations if necessary.

Accurate delivery cost: The total delivery expenditures are calculated according to the instant single delivery cost. The users can see the prices of goods and the

delivery cost of each.

Best recommendation function: Our system will calculate the cost of both time and money according to the cost function we developed, and use certain optimization algorithms to find the best solution in every case. The user can easily pick the best solution which is recommended by our system.

4.2 Knowledge Representation

4.2.1 Data Acquisition

This system needs all the replacable goods' information and the main consolidation locations to build a reliable recommendation model. In this section, we will discuss type of the main information we need and introduce the methodologies we used, including using API functiona and Web crawlers methods. The table below shows the list of web pages where data are to be extracted.
Below shows the flow chart on how the extraction of data is done in our recommend system. The system is designed to have an embedded web scraping function to

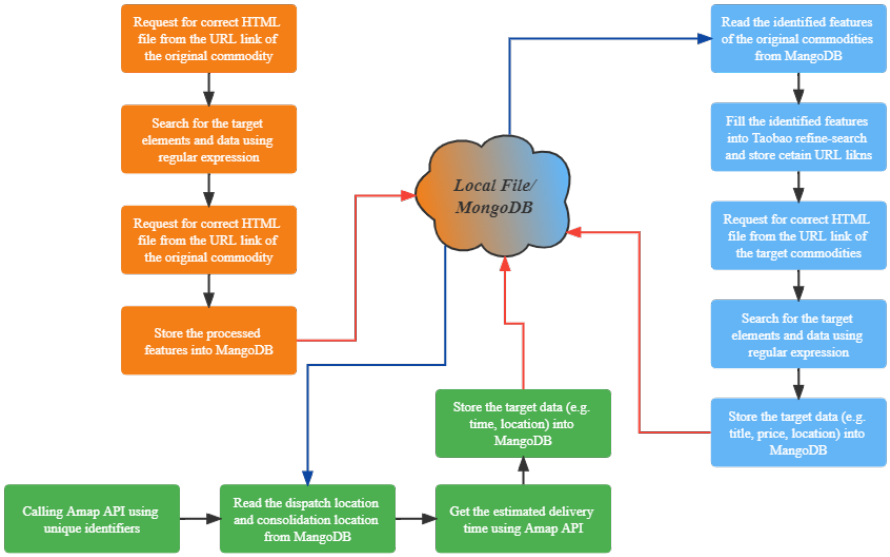

Figure 4.2: The flow chart of data extraction and processing

extract resources and data from online web pages. Our original program is designed by using the Selenium library to fully collect the information from the target web, but Taobao is equipped with login verification to conduct anti-crawler. Thus One of the best ways to search for information we're interested in is to use regular expressions. According to the target web pages' HTML, corresponding patterns can be defined to extract needed texts. We use the "findall" method in the RE library to do the most extracting job, which makes use of elements and tags in the HTML file for the searching.

4.2.2 Data processing

It is inevitable that there will be circumstances that the data crawled from webpages are inaccurate and incomplete, then we need to implement data processing or cleaning strategy to improve the quality of data. The regular expressions method is used to detect inaccurate data, then the data will be manually corrected and replaced. There will be cases that replaceable commodities are out of stock, and the system will automatically remove the commodities in that situation.

During the data extraction procedure, it is possible that the information our model crawled from TaoBao is a long list of words. And we will use certain NLP techniques like NLTK to remove unimportant words, stop words and punctuations, the entity and descriptive words will be reserved and further transferred to next module to be constructed in the form of the storage type.

Chapter 5

Project implementation

5.1 Overall System Architecture

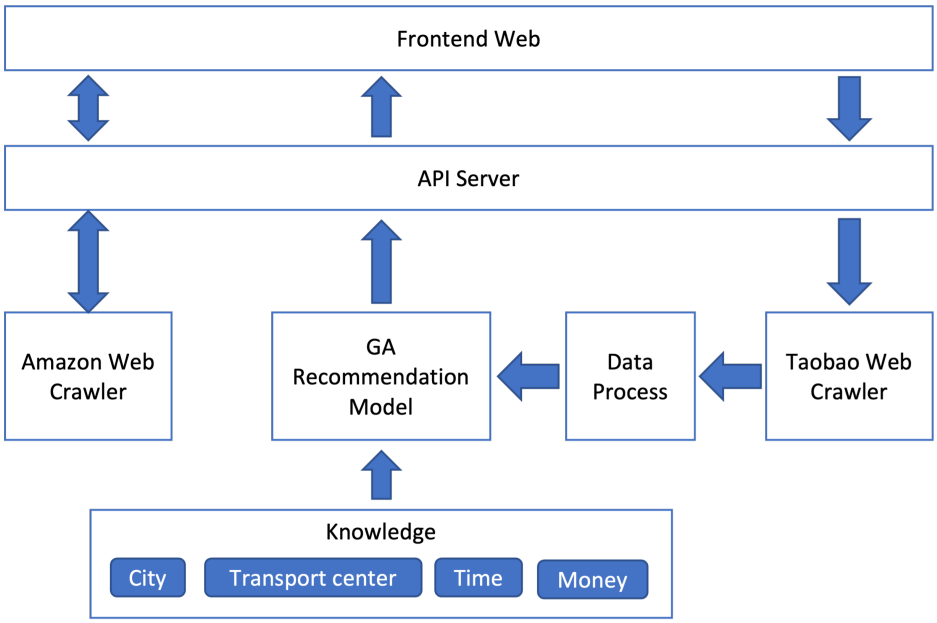

Figure 5.1: Architecture

The whole project is divided into 5 modules illustrated as figure. Module 1 is the establishment stage of knowledge base of time-cost of each city in China to each collect center and time and money cost of each center point to Singapore. This module has been implemented by python and free API provided by Gaode. Module 2 is Amazon web crawler module to do the good information and keywords extraction. In this module, we use python to do web crawler from amazon.cn to get the name, url and price of this good and use jieba to do Keywords extraction. Module 3 is the Taobao web crawler Module to do keywords to good List. This module is aimed to get goods list from Taobao by the keywords of good we want. In this module, we use python to do web crawler from Taobao to get the list of name, url, location and price of good by the keywords list. Then do data cleaning by selecting the most

suitable and cheapest the good of each location. Module 4 is GA recommendation Model Module. In this module we use genetic algorithm to optimize the goods and transport center selection. And the last module, module 5 is the UI and Backend API module, the user interface is implemented by next.js which is based on react and the Backend API is implemented by Flask based on python.

5.2 System modules

5.2.1 Amazon web crawler

In the module, we will implement a web crawler from amazon.cn to get the wanted good information and keywords of user.

The process flow of how to extract the information and keywords from the URL is like below:

1. Request the HTML file from the URL link provided by User Interface;.
2. Search the target elements such as title, imgUrl, price by using regular expression;
3. Using jieba which is a Chinese Semantic Analysis Library to get the keywords of the good;
4. Organize the information as a JSON object and return to web application;
5. Add the Object to goodList in web application

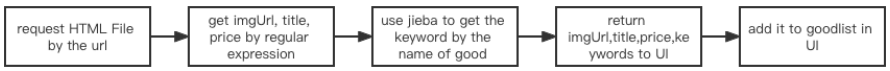

Figure 5.2: getweb

5.2.2 Taobao web crawler

In the module, we will implement a web crawler from taobao.com by searching the keywords to get 100 goods and do data clean to filter the unsuitable goods from the list and select the cheapest good from each location.

Web crawler

The process flow of how to get 100 good by searching the keywords in Taobao.com is like below:

1. Get keywords from keywords list from web application;.

2. Request Taobao API and set keywords as search parameter;
3. Append to goodList from the json response;
4. Repeat 1-3 until complete the iteration of keywords list
5. Return goodList

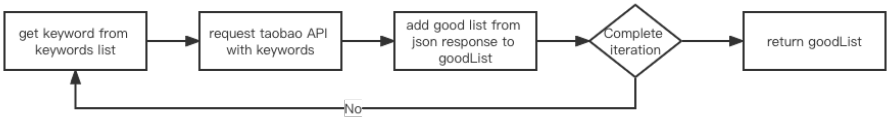

Figure 5.3: web crawler

Data cleaning

The process flow of how to do data cleaning from goodList from web crawler like below:

1. Get goods from goodList from web crawler;
2. Calculate the average value of the goods' prices;
3. Delete good whose price lower 20% than medine from goods;
4. Repeat 1-3 until complete the iteration of keywords list
5. Return goodList

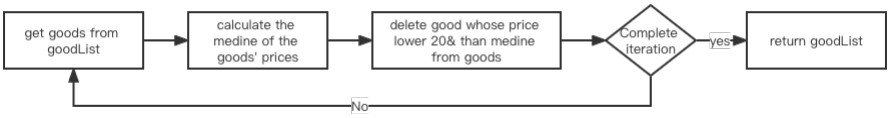

Figure 5.4: Data cleaning

Data transformation

The process flow of how to do data transformation from goodList from Data cleaning like below:

1. Get goods from goodList from Data cleaning;
2. Get good from goods;
3. Check if dict has key named good's location and dict[good's location] ¿ price;

4. If so, set dict[good's location] as good's price, if not go to next good;
5. Repeat 2-4 until complete the iteration of goods
6. Transfer dict to list and append to goodLocList
7. Repeat 1-7 until complete the iteration of goodList
8. Return goodLocList

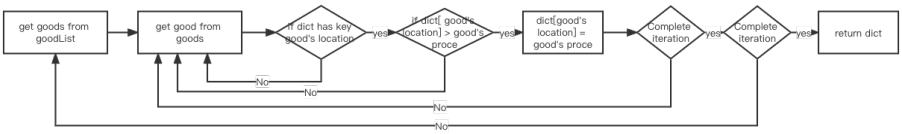

Figure 5.5: Data transformation

5.2.3 GA recommendation Model

In the module, we will adopt gene algorithm to optimize the selection of good of each goodList and transport center.

Set parameter

In gene algorithm, we should set cross possibility, mutation possibility, number of generation, popularity size and low and high bind of each gene which is the list of goodLocList get from Keywords to good list module.

Parameter	Value
cross possibility	0.8
mutation possibility	0.2
number of generations	1000
popularity size	1000
Low and high bond	Low is 0, High is the length of list of goodLocList

Figure 5.6: Parameter

Set evaluation function

The process flow of how to do data transformation from goodList from Data cleaning like below:

1. Get transport center and transparent mode from the last gene from geneinfo;
2. set cost = 0 and time = 0
3. get rest gene from geneinfo

4. cost = cost + gene.price
5. time = max(time, dur[gene.loc][transport center]
6. Repeat 3-5 until complete the iteration of the rest geneinfo
7. get transport fee and time by mode from knowledge-based rule
8. calculate evaluation value by time and cost and return

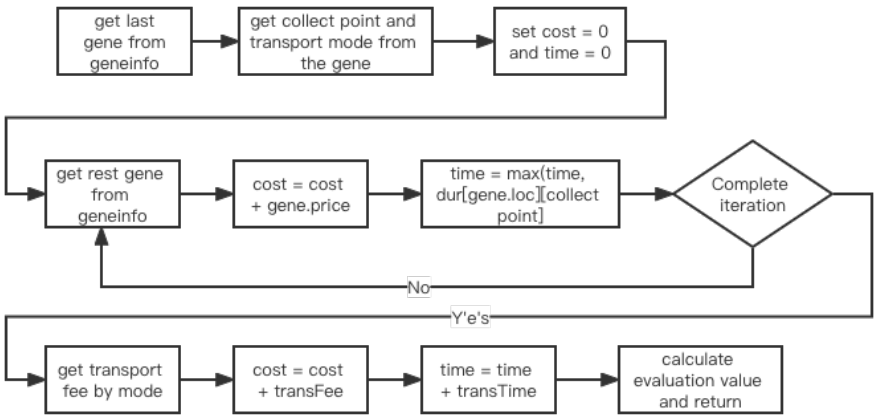

Figure 5.7: GA

5.2.4 UI and Backend API

API	Description
/getWeb	Get goods' name, price, imgUrl and keywords using url provided by user
/recommend	Get recommended goodList and collect point and time cost by the keywords list from the good list added by user

Figure 5.8: Backend API

The web user interface is a Single-page Application implemented by Next.js which is an optimized framework based on react and antd components library.

The backend API server is based on flask framework using python in case the other modules are all implemented by python.

Chapter 6

Project Performance

6.1 System Validation

6.1.1 Get wanted good information

By filling the URL of product page of amazon.cn and click ADD button, then the web page will show the picture, name and keywords of the good.

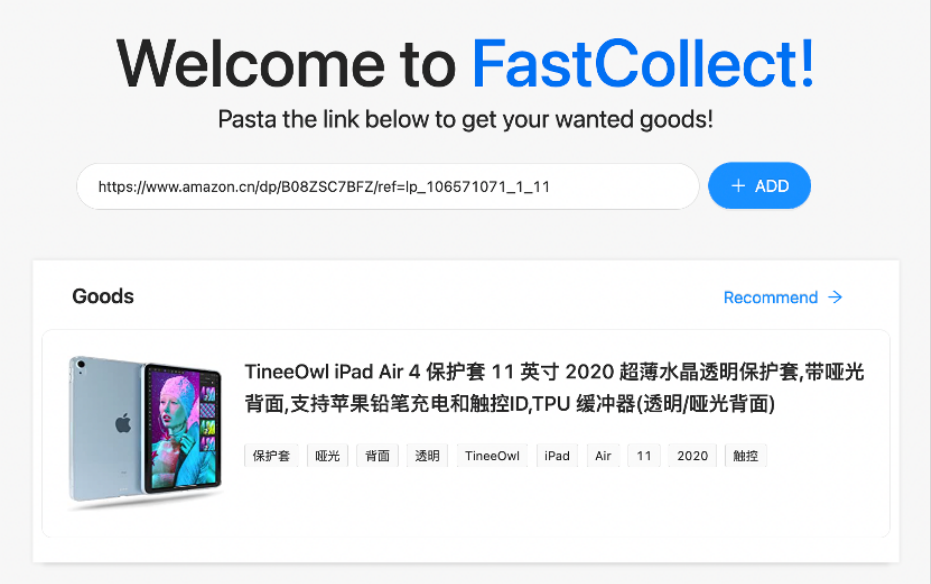

Figure 6.1: Get Good Page

6.1.2 Get recommended good List and transport center

After we add all the good we want in the goods, we click Recommend Button. Then, after a few seconds, the result will show the transport time, transport center, goods from Taobao recommended by the system.

Figure 6.2: Get Recommend Page

Chapter 7

conclusion

Our group successfully completed the coding of the intelligent system and the construction of the web page. However, the whole project still has some problems and can be improved in the future.

7.1 Limitation

7.1.1 Commodity search results are not good

In our system, the keywords used when calling Taobao search engine are extracted from the web pages provided by users. This can sometimes lead to problems where the system finds completely irrelevant content if the user offers something that is too unpopular. And because of Taobao search engine itself, when read too many search results will also appear irrelevant content.

7.1.2 The response time is slightly longer

Due to the limitation of the algorithm itself, it takes more than 10 seconds for the user to input the product to produce the result. And the problem is hard to fix. Although it still has the speed advantage compared to the user searching on Taobao in person, it will also affect the user experience. However, on the other hand, it is also an advantage that the time required by the system does not increase when the user increases the number of goods, also due to the algorithm.

7.2 Future Improvements

7.2.1 Better search results

The results of commodity keywords obtained by web crawler in Taobao search are not perfect. Due to changes in the web page itself and the characteristics of the search engine lead to poor search results. In the future, we will try to improve the accuracy of search by asking users to provide links to products on Taobao instead of Amazon.

7.2.2 More countries and modes of transport

Right now our system only supports goods from China to Singapore, but in fact Taobao can send goods to many countries in the world. And besides sea transport and shipping, there is also road transport and rail transport, which we have not yet supported. In the future, we will gradually expand the supported countries and modes of transportation, so that more users can use our system.

7.2.3 Support more shopping sites

While Taobao is the biggest shopping site in China, many people also use shopping platforms such as JD.com and Pinduoduo, which also support shipping goods overseas. In the future, we will add more shopping platforms. The products users need can be searched on multiple platforms to get the best price and shipping time.

7.2.4 Support for ios and android platforms

Nowadays people use mobile phones instead of computers. In 2021 Taobao 11.11 Online Shopping Festival, about 70 percent of users use smartphones to make purchases. The web can also be opened on the phone, but the experience is not good. In the future, the development of mobile applications is imminent. No matter on iphone or android phones, users can have a good experience using our system.

Chapter 8

Appendix

8.1 Project Proposal

GRADUATE CERTIFICATE: Intelligent Reasoning Systems (IRS)
PRACTICE MODULE: Project Proposal

Date of proposal:

26 October 2022

Project Title:

ISS Project – Intelligent Shipping Recommendation System

Sponsor/Client: *(Name, Address, Telephone No. and Contact Name)*

Institute of Systems Science (ISS) at 25 Heng Mui Keng Terrace, Singapore
NATIONAL UNIVERSITY OF SINGAPORE (NUS)
Contact: Mr. GU ZHAN / Lecturer & Consultant
Telephone No.: 65-6516 8021
Email: zhan.gu@nus.edu.sg

Background/Aims/Objectives:

The proposed intelligent Shipping Recommendation System will make use of various advanced web crawler techniques and genetic algorithm to generate a recommended goods and transport center for individual customers and cross-border e-commerce merchants in Singapore to minimize the cost of time and money spent on transportation from China to Singapore.

Requirements Overview:

- Research, data extraction & elicitation ability - from website, JSON
- Programming ability - Python, JavaScript
- Flask and Next.js framework understanding
- Clean data crawled from website

Resource Requirements (please list Hardware, Software and any other resources)

Hardware proposed for consideration:
- PC with browser

Software proposed for consideration:
- Reasoning systems: jieba
- Pertained machine learning models: genetic algorithm
- Machine learning use cases: Python 3.8
- Application framework: Next.js, Flask
- Cloud computing/server: Tencent cloud.

Number of Learner Interns required: (Please specify their tasks if possible)

a team of three project members

Shu Wanyang:

- Data Acquisition – crawl by Amap API to get the time cost of every city in China to every transport center in China
- System architecture design – system/data flow design, system modules design
- User Interface development – get good info pages and recommend page
- Backend API: POST and GET API of getWeb and recommend
- Data cleaning: filter the unsuitable goods from the result of web crawler of taobao
- Debug & troubleshooting
- Project management
- Project report writing

Yan JiaHuan:

- Data Acquisition – information of alternative commodities
- Data Processing – process location, identification number, price data
- Recommendation system modeling
- Genetic Algorithm programming
- Debug & troubleshooting
- Project management
- Project report writing

Xiao Changwei:

- Project idea generation
- Data Processing – Web crawler design and API interface, Capture user requirements
- Network server setup – Cloud storage database
- Data acquisition - China's prefecture-level city information and wharf information
- Project management
- Debug & troubleshooting
- Project report writing

Methods and Standards:

Procedures	Objective	Key Activities
Requirement Gathering and Analysis	The team should meet with ISS to scope the details of project and ensure the achievement of business objectives.	1. Gather & Analyze Requirements 2. Define internal and External Design 3. Prioritize & Consolidate Requirements 4. Establish Functional Baseline
Technical Construction	· To develop the source code in accordance to the design. · To perform unit testing to ensure the quality before the components are integrated as a whole project	1. Setup Development Environment 2. Understand the System Context, Design 3. Perform Coding 4. Conduct Unit Testing
Integration Testing and acceptance testing	To ensure interface compatibility and confirm that the integrated system hardware and system software meets requirements and is ready for acceptance testing.	1. Prepare System Test Specifications 2. Prepare for Test Execution

		3. Conduct System Integration Testing 4. Evaluate Testing 5. Establish Product Baseline
Acceptance Testing	To obtain ISS user acceptance that the system meets the requirements.	1. Plan for Acceptance Testing 2. Conduct Training for Acceptance Testing 3. Prepare for Acceptance Test Execution 4. ISS Evaluate Testing 5. Obtain Customer Acceptance Sign-off
Delivery	To deploy the system into production (ISS standalone server) environment.	1. Software must be packed by following ISS's standard 2. Deployment guideline must be provided in ISS production (ISS standalone server) format 3. Production (ISS standalone server) support and troubleshooting process must be defined.

Team Formation & Registration

Team Name: Group 13
Project Title (repeated): Intelligent Shipping Recommendation System
System Name (if decided):
Team Member 1 Name: Shu Wanyang
Team Member 1 Matriculation Number: A0261754B
Team Member 1 Contact (Mobile/Email): e0983148@u.nus.edu
Team Member 2 Name: Yan Jiahuan
Team Member 2 Matriculation Number: A0261968M
Team Member 2 Contact (Mobile/Email): E0983362@u.nus.edu
Team Member 3 Name: Xiao Changwei
Team Member 3 Matriculation Number: A0226757U
Team Member 3 Contact (Mobile/Email): E0641610@u.nus.edu

For ISS Use Only

Programme Name:	**Project No:**	**Learner Batch:**
Accepted/Rejected/KIV:		
Learners Assigned:		
Advisor Assigned: Contact: Mr. GU ZHAN / Lecturer & Consultant Telephone No.: 65-6516 8021 Email: zhan.gu@nus.edu.sg		

8.2 Individual Report

Master of Technology in Intelligent Systems

Individual Project Report

Your Name:	Shu Wanyang
Certificate:	Graduate Certificate in Intelligent Reasoning System

1. Your personal contribution to the project.

Throughout whole work of this project, my contribution could be introduced in point:
The first is to bring out the project idea. At the very beginning of this project discussion, I proposed the idea of the recommendation system which is an existing problem in our daily life as an oversea Chinese student in Singapore and briefly provide my solution to prove the technical feasibility of the project.
The second is the architecture of this project. During the system design phase, I was responsible for the functional design of the modules and the technical selection of each module. This included the application front-end, the back-end API, the knowledge representation and the recommendation system.
The third is the front-end work of this project. I use Next.js to design and implement the get good info page and recommendation page.
The fourth is the back-end work of this project. I use Flask which is based on python to design and implement the API of get good info and recommendations.
The fifth is the knowledge representation acquisition of this project. I use python to write a tool which based on Amap API to get the time cost between each city in China and the transport center.
The sixth is the data cleaning work of this project. After we find the result of the web crawler from Taobao has a lot of unreasonable goods, I design a data cleaning function to filter these goods by their price.

2. What you have learnt from the project.

The most useful knowledge I learnt from this project is when to meet the not ideal result of recommend system, how to improve it by data preprocessing. At the beginning of implementing this recommend system, we always get the result not we want just because the good's name has the keyword, and its price is lower than other goods. After a few analyzing we work, we notice that this kind of good's price is lower than 20% of the average price of all goods. So, we add a data cleaning function after the web crawler to filter this not-ideal result. Which brings a better performance of this to recommend system.
In addition, I also learn how to design and implement an easy-to-use web page by writing the front end using the Next.js framework. Through this project, I learnt how to build a whole frontend project by Next.js including route setting, server-side service, and webpack.
Last but not least, during the whole project process. I learn how to collaborate effectively with my teammates by the tools of GitHub, CI/CD, Kanban and google meet in accordance with the principles of agile development. We have meetings every morning and night to share our process and problem in daily work.

3. How you can apply this in future work-related projects.

As I mentioned above, my work in this project includes system design, project management, and recommend system optimization. The knowledge and experience I gained through this project will definitely be useful to my future professional life.

First, the experience of frontend work in Next.js and backend work in Flask will help me become a good software engineer because they are widely used in almost every software and Internet company. I could help my teammates build an initialized frontend project with all significate set-up to help them do the frontend programming work.

In addition, the knowledge of adopting recommend algorithm in a software system will also be a widely used solution to a lot of real-word problems. Such as movie, logistic and daily shopping.

Finally, goal setting and schedule management using agile development principles during the tea work will definitely be applied in my future work. Modern software engineering is based on advanced project management skills, and having this knowledge and experience will help the projects to process more smoothly

Master of Technology in Intelligent Systems

Individual Project Report

Your Name:	YAN JIA HUAN
Certificate:	Graduate Certificate in Intelligent Reasoning System

1. Your personal contribution to the project.

The Contribution to this whole project of mine can be divided into several parts: Core function determination, methodology determination, data acquisition, data processing and storage, feature extraction, and recommendation module programming. I can be seen as a backend developer to build the bridge between users' feature input and predict outcomes. And if focus on this specific task, my job is to extract the feature from the original commodity, then provide a list of similar alternatives, and finally construct a model to provide the best purchase strategy.

At first, I have been collaborating closely with my teammates to extract the features from the original commodity. I further created web scrapping scripts to extract important data from TaoBao to find all the alternatives of original commodities. I used several web scrawler methods including regular expression and Selenium library.

Then I write a data processing script to further standardize the information received from web scrawler. I removed the unrecognizable characters from the text gathered from HTML files. The numbers (prices, stock) are transferred to floats and stored in MongoDB. The URL of sample pictures and identification numbers are also built an inner connection to dispatch locations and joint storage in MongoDB.

I also construct the recommendation module based on Genetic Algorithm. I wrote the functions including crossover, mutation, and population iteration according to the courseware written by Dr. Zhu FangMing.

2. What you have learnt from the project.

There are so many things to learn during this whole project. The other two teammates are all equipped with rich working experience, and they taught me many transferrable developing metrics (e.g. fast development, development accountability, and progress arrangement) which are used by internet companies. I also learned how to apply Cloud deployments to a team project.

As to what I have learned from developing an intelligent shipping recommendation system, I become more familiar with using web crawlers to get the data I need. In my past learning journey, I did not have experience in using the API function from public platforms. This project involves using the API function from Amap and TaoBao developing platforms. After this project, I can actually build some easy scripts to call the API functions of platforms to make programming easier.

Besides, all the past projects I conducted are actually research-oriented. These projects will apply some state-of-the-art algorithms or architectures, which will provide the model with good theoretical performance. But the data I used are all from the public dataset. And the idea is also not original. This project is not about simply duplicating and improving the old models, but to open a new field of modeling and optimization. We collect the data by ourselves and construct the models by ourselves, which is significant to every one of us in our team.

Last but not least, I also learned how to fully tap into the developing talent of Github. It is an

efficient way to collaborate with teammates in terms of software development. I learned how to create the branches in Github for testing new functions of our system, and I also learned how to distribute the project workflow in a more systematic and efficient way during the project development.

3. How you can apply this in future work-related projects.

We can not deny that our daily life is closely connected to all kinds of recommendation systems. It helps us to make decisions easier especially when we meet a trade-off situation (a classic example will be finding the best choice under the constraints of time and money cost). Normally, recommendation systems are among the most powerful machine learning systems that online retailers implement in order to drive sales. In other words, they are designed to help corporations make more money. But our recommendation system is determined to save the cost of customers, it is a program developed from the customer's point of view.

This application can be further embedded in smartphones and out-connected to online shopping applications like TaoBao and Amazon, which will make the customers feel more convenient to use. There are two ways to improve the performance of our model. One is to use deep learning models to give all the alternatives a score, and filter the list of replaceable commodities according to a certain level. This method will help improve the accuracy of our model. Another way to optimize our system is to find a better way of constructing the fitness function during the recommendation module. Our current method is to calculate the square root of the normalized time and money. But I am sure there will be more scientific ways to do so. For example to build an expert system to make decisions according to the exact circumstances. This is a meaningful project, i enjoyed the development process.

Master of Technology in Intelligent Systems

Individual Project Report

Your Name:	XIAO CHANGWEI
Certificate:	Graduate Certificate in Intelligent Reasoning System

1. Your personal contribution to the project.

I played my part in every phase of the project.
At the beginning of the project, we need to set a goal. In discussing what we need to do and how to do it, I put forward some suggestions and ideas.
My main responsibility is to crawl the required content from user-provided web pages. Learn python crawler code and help other team members understand it after learning. Use python to write code to achieve the extraction of commodity names and keywords. Use python to write code to achieve the user needs the product picture extraction.
In terms of data acquisition, I obtained the names of prefecture-level cities and the names of major ports in China because it was an analysis of the place of origin in China.
As for the storage of project data, I built a virtual host from Tencent Cloud. By changing the virtual host firewall and setting the port, I assisted the team members to build the mongoDB database. The required data can be easily uploaded and downloaded through the mongoDB database.
In the final stage of the project, I worked with other team members to complete project testing and debugging. I worked with other team members to complete the final report and video recording.

2. What you have learnt from the project.

In terms of technology, I learned how to use crawler in order to obtain the information of web pages provided by users. I had never tried crawler before, although I had learned to use many programming languages.
I learned how to use cloud hosting to help the project run. Most of my past projects were run on my own computer, where the database was stored. This is not convenient in group work and can cause a lot of problems. This project I learned how to use cloud host to build database.
In terms of group cooperation, because my team members have rich working experience, I have learned a lot of skills in teamwork in Internet enterprises. Like how to use lists to identify each person's tasks and how well they were accomplished. Like agile development processes. Like how the overall idea and framework of the project is generated.

3. How you can apply this in future work-related projects.

I learned a lot in this whole project and a lot that I can apply to my future work.
The most obvious is the experience of teamwork. Nowadays, all walks of life emphasize the importance of teamwork. The teamwork experience I have accumulated in this project can provide a lot of reference for my future work. I learned a lot from this project about how to communicate with others and how to improve the working efficiency of the whole team.

Of course, technology is also important. In the future work, it will also be applied to crawler or genetic algorithm. This project enables me to have more knowledge and practice of these technologies.

8.3 Knowleged Mapping

In this section, we will list all the mapped system functionalities against knowledge, techniques and skills of modular courses.

8.3.1 Machine Reasoning (MR)

Knowledge Elicitation and extraction:

- Web crawling from websites & repositories.
- Manual extraction from websites/internet.

Knowledge Representation:

- Table of origin city, destination city and the time cost between them.
- Table of transport center, transport mode, and its cost of money and time.

8.3.2 Reasoning System (RS)

Search & Optimization:

- Genetic Algorithm search technique applied on result from web crawler from Taobao to optimize the results.

8.3.3 Cognitive Systems (CGS)

Slot Detection:

- Use jieba to get the entity which is the keywords from the name of good.

8.4 Use Guide

The code can be executed from python and node.js environment. Below is the software requirement:

- python 3.8
- node.js 16.17.0
- flask 2.2.2

Steps to run the system:

1. Download the files from GitHub.
2. run command in the terminal:

- cd SystemCode/front
- npm install
- npm run dev

3. Create a new conda environment with python 3.8
4. Activate your conda environment and install the below packages: flask, jieba
5. run app.py in /back file in spider
6. open http://127.0.0.1:3000 and start use