DAT410/DIT728 Design of AI systems
Module 1

Assignment 1: Group 54

January 23, 2024

Group members:

Member 1:
Name: Hongguang Chen
Personal number: 19961019-5795
Program: MPHPC
Uni email: chenhon@chalmers.se
Time spent: 12 hrs

Member 2:
Name: Sanjaya Thilak Bandara Asurasinghe Rajamantrilage
Personal number: 19890711-4733
Program: MPDSC
email: arstbandara@gmail.com
Uni email: sanraj@chalmers.se
Time spent: 12 hrs

The declaration:
"We hereby declare that we have both actively participated in solving every exercise. All solutions are entirely our own work, without having taken part of other solutions."

1 Predict the temperature

1.1 For the next weekend:

Member 1 Solution:

The expected output is the temperature of the next weekend. This implies that the inputs consist of recent historical weather data (form past seven days or more), such as temperature, humidity, wind speed, cloud condition and so on. And the constraints are evident; the result temperature should fall within a reasonable range(for example, between 0 and 40.0 degrees).
Therefore this problem can seems as a typical regression problem. or to be more precise, a time series prediction problem.
And the prediction might can be updated continually when we get the new data for the coming day.(For instance, if we make a prediction on Monday, and then we can update it on Tuesday with the inclusion of new data).

One possible approach to solve this problem is by using auto-regressive integrated moving average (ARIMA) algorithm. Alternatively, LSTM or even more advanced techniques like Neural Network.

Member 2 Solution:

We can use linear or polynomial regression techniques for this predictions.

Steps need to perform:

1. Collect the Data: This steps involve collect temperature values over a period of time. We have to consider the location, date and time as well as other factors which can affect the temperature of a location. All these factors can be consider as the features for the dataset.
2. Preprocess the collected data: this steps is involved about handling missing data, scale or normalize the data, make a training and test set and other things which can improve the data value for the model.
3. Visualize the dataset: This steps is the best way to start with where we plot the data using tools to understand about the dataset better.
4. Train a Model: Use linear or polynomial regression model as needed. For this case we can start using linear regression model and extend or fine tune with polynomial regression model later. To train the model we need to feed the historical data and find the best fit values.
5. Predict: Use the trained model to predict the temperature.

1.2 For the same date as today but next year:

Member 1 Solution:

Predicting the temperature for specific day next year is more challenging, (and perhaps an impossible task) compared to predicting for the next weekend, This is because weather is influenced by numerous variables, and even a slight change can lead to different results. However, climate shows a relatively stable pattern. Thus, as the very least, predicting the temperate for a month, a week, still appears feasible, and base on this prediction, then to predict a specific day. But the result might not only be a single value, but rather a distribution indicating the possibility of a temperature range. Similar to the previous sub-question, this task requires more historical data (as well as additional dimensions (heat map, data of ocean current)). It is a more complex dynamic system problem that cannot be solved with a simple regression algorithm. Perhaps, a Large Neural Network module or Simulation would be a suitable choice for addressing this this task.

Member 2 Solution:

For this to make an attempt of predicting we have to capture seasonal trends or some pattern. It will not be more accurate and less meaningful to predict next year same day temperature as the temperature can be even increase with global warming as well and other human activities. It will be quite difficult to consider all the factors and predict a temperature for the next year same day, but we can try some seasonal trend analysis and predict some reasonable average value for a period of dates in next year. It is better in that case we choose a model which can track seasonal patterns of the dataset. We can plot the measured historical temperature values in a time series data and understand the dataset with some statistical values. Later We can use ARIMA(Auto regressive integrated moving average) method to predict temperature which usually capture the seasonal trend patterns.

2 Bingo lottery problem

Member 1 Solution:

This task is going to draw a set of numbers that results in exactly 120 winners.
The constraints is clear:
1. 1000 tickers.
2. each one has 25 randomly placed number range from 1 to 25.
3. 120 winners.
4. a winning ticket has a complete horizontal,vertical or diagonal row.
5. no ticket have more than one winning row.

We can use a matrix with binary values (0,1) to represent a ticket, where 1 means a drawn number and 0 means undrawn number. The goal is to find a combination of numbers that lead to 120 winner.

There are several method to solve this problem.
1. Brute force is a way but not a feasible method because of a huge searching space, A_{25}^{25} possible combinations.
2. Heuristic algorithms: can find a good solution that may not be perfect but is close to the desired outcome.
3. constraint programming.

Member 2 Solution:

Assumption: Number can be repeated in a single ticket.
In the problem statement, we have 1000 tickets and we need to maximize winning combinations up to 120 tickets. We can formulate solution as below,
1. Randomly select a ticket from the 1000 tickets.
2. Create a function to evaluate wining combination criteria. (Same set of numbers cannot be in horizontal,vertical or diagonals) and evaluate current selected ticket met the criteria. If not select other.
3. Select each distinct row, column or diagonal and search in the other tickets. Make a count on the maximum winning counts gives in the selected sample.
4, Do multiple feasible iterations and find the best winning combination, giving maximum number of winning tickets under 120.
We might have to select the tickets randomly and will do feasible number of iterations, so, this will not give the optimal 120 tickets. But we can maximize the chance to satisfy the condition.

3 Public transport departure forecast

Member 1 Solution:

For this question, the expected result is the estimated departure time predicted based on some inputs like (last one departure time, road condition, the location of last bus and the next bus, the weather condition, the historical departure time). It is a simple question, not relying on much data, not constrains, not in a dynamic system. Thus, regression algorithm may can get a pretty good result. Of course, if we have data with more dimensions, like other vehicles' speed on road, the traffic light, pedestrian, using ML models might yield a more precise result.

Member 2 Solution:

As this is mainly depends on the road situations, weather and other factors such as traffic, road closures which affect usually for this departure time, it is best to use statistical method to predict the departure time.

Let say we are collecting large amount of historical data for the specific route at a selected stop, we can collect the data such as data of week, date time(morning, evening, ..), is a holiday, previous delays, if possible weather condition as a categorical value,(sunny, rainy, snowing, ...).
After that we can clean and fix any missing values and formulate statistical summary using mean, median, standard deviations. Then we can visualize the data using tool like python matplotlib, seaborn with heatmaps, box plots, line or box charts.
This will give good indication of any patterns and trends. Based on statistical values, we can categorize, peak hours, off peak hours, previous delays as such. We can split the dataset to train and test set and make use of it to test the model.
This approach we can integrate real time data and continuously enhance the systems with feedback loops.

4 Film festival problem

Member 1 Solution:

This problem is a Combinatorial Optimization Problem. In this task, we have to select an optimal combination from a group of feasible combinations allowing as many as audiences can watch their selected movies.
Now, let assuming:
+.There are $M, M \in Z$ movies.
+.There are $N, N < M$ movies can be played at the same time.
+.Ignore the seat number.
+.A person can watch more than 1 movies.
+.There are K shows (total $K * N$ films will be played)
+.There are m person
+.The priority of a person to a movie P_{ijk}, the i th person, j th show, k th movie
+ A_{ijk} indicate i th person if will attend to j th show k th movie.

$$\begin{cases} A_{ijk} = 1, true \\ A_{ijk} = 0 \end{cases} \tag{1}$$

Then, we will have to maximize the following equation:

$$S = \sum_{i=1}^{m} \sum_{k=1}^{K} \sum_{n=1}^{N} P_{ikm} * A_{ikn} \tag{2}$$

And the constraints:
1. Each person can only attend one same time: $\sum_{n=1}^{N} A_{ikn} = 1$.

2. Each person can only attend one time for each movie: $M[map(A_{ki})] <= 1$.

Then we can use Integer Programming or Heuristic Algorithm to solve the problem.

Member 2 Solution:

This problems mainly in the form of Optimizations problem. We have to make a optimal schedule for film enthusiasts who want to attend their desired films based in the ranking. We can formulate integer constrained linear programming problem as below. Let say we have m people and n films showing,

Decision variables:
$x_{ij} = 1$, if attendee i th person attend j th the film
$x_{ij} = 0$, if not
Variables:
p_{ij} := priority of the i th person for j th film
Objective function:

$$\text{Maximize } Z = \sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} \times x_{ij}$$

Constraints:
1. Each person should attend only one showing of their prioritization.

$$\sum_{i=1}^{n} x_{ij} = 1, \forall i \in \{1, 2, \ldots, m\}$$

2.It should not conflict for each person.

$$\sum_{i=1}^{m} x_{ij} \leq 1, \forall j \in \{1, 2, \ldots, n\}$$

As above formulation and using some standard integer programming solvers we can achieve optimal value for the problem.

5 Product rating in consumer test

Member 1 Solution:

For this task, it is trying to get a score based on the response of consumers. And there will be many factors that should be considering, like Cleaning performance (CP), drying performance (DP), energy efficiency(EE), water efficiency(WE), noise level(NL), ease of use (EU), price-to-performance ration (PP) and so on. Then we can define the scoring equation like below:

$$SCORE = CP + DP + EE + WE + NL + EU + PP \tag{3}$$

What's more, the consumers' background also important to the rating, like different age group might have different preference which will lead to a different result. Thus, the evaluating function become:

$$SCORE = w_1 * CP + w_2 * DP + w_3 * EE + w_4 * WE + w_5 * NL + w_6 * EU + w_7 * PP \tag{4}$$

Where w are the preference weight. Different target consumer group might have different w setting.

Member 2 Solution:

We can define a statistical perspective model to determine score based on the input features. Let's assume we are calculating dish washer rating index(DWRI) based on cleaning efficiency, energy consumption, noise and outer look. Collect the user feedback data with all the feature values as mentioned above for the dishwashers of interest. Then preprocess the dataset and try to visualize statistical data for the dataset. Next step is to make a model with some weighted feature selection based on the importance. As an example we can have more weight for cleaning efficiency and less value for the noise of the machine. Let's define simple scoring system as below. cleaning efficiency (CE), energy consumption (EC), and noise level (NL) We need to consider some normalize scale for the values calculated above. Otherwise the index can be bias in some perspective.

$$DWRI = w_{CE} \cdot \text{CE} + w_{EC} \cdot \text{EC} + w_{NL} \cdot \text{NL}$$

We can give weights for these system based on the previous statistical data where it will show clear importance for the end user. We can continuously improve the model by user feedback and ranking the dishwashers based on DWRI index we defined.

6 Constraint satisfaction and constraint programming (read and explain)

Member 1 Solution:

1.What is "Constraint satisfaction problem" and how does it work?
It is a kind of problem that its' solution must satisfy some constraints or limitations. For example, problems like Eight queens puzzle, map coloring problem. And with finite domain, there are three classical method (algorithm) to solve it: Backtracking, Constraint propagation, and Local search.

2. What is "constraint programming"
It is a programming paradigm and a set of techniques used for solving problems that can be formulated in terms of constraints on variables. In constraint programming, the relationships between variables are expressed using constraints, and the goal is to find a solution that satisfies all these constraints Constraint programming is particularly useful for solving combinatorial optimization problems, scheduling problems, and problems with complex logical constraints. It provides a high-level abstraction for problem-solving, allowing programmers to focus on the structure of the problem rather than the specific algorithmic details of how to solve it.

Member 2 Solution:

1. Constraint satisfaction problems(CSPs):
CSPs are the form problems where we define set of objects with states and the states need to be satisfy certain set of constraints or limitations. Examples of such problems are map coloring problem, Sudoku, Chess moves, Crossword puzzels, etc. We can use backtracking, constraint propagation, and local search algorithums to find a better solution for these type of problems. We will use to define these problems with variables, domains and constraints.

Let's take a practical example to understand it easily.

Scheduling courses in a university:
Variables : list of courses
Domain: list of time slots
Constraints: availability of professors, lecture halls, hall capacity, etc

2. Constraint programming(CP):
This is a technique use to solve combinatorial problems by modeling them as constraint satisfaction problems. So, in this case users declare the constraints on the feasible solution set of the decision variables. Other than that users need to specify a method to solve the constraints. Usually use backtracking, constraint propagation, minimum conflict approach to reach the final solution. This is high level declaration of a complex problems which will make easy to describe and solve the problem.

Constraint optimization problem is also can be consider as a constraint satisfaction problem associated to an objective function. We either minimize or maximize objective function based on certain set of constraints.

7 References

Regression analysis
Support vector machine
Integer Programming
Linear Programming
ARIMA
LSTM
RNN
The Forecast Process: Observing and Analysis
Data analytics approach for train timetable performance measures using automatic train supervision data
Combinatorial optimization
Mathematical Modeling and Analysis of Product Rating with Partial Information
Bingo
Heuristic
Constraint satisfaction problems