CS6220 BDS Workshop Presentation

Team members: Jichuan Zhang, Anbang Ye, Guanchen Meng, Qiying Mu, Shuran Wen

Team 15

Motivation and Objective

- Location-based recommendation systems.
- Make use of the user and restaurant coordinates to provide a more delicate recommendation.
- Build an application based on the restaurant recommendation model system which matches the customers’ profile including user coordinates, historic order details, preview feedback as well as the restaurants’ information.
- So that, customers are allowed to get some recommended restaurants based on their location.

Related Work

- Personalized recommendation (Davidson et al., 2010) (Shepitsen et al., 2008) (Guy et al., 2009) (Qian et al., 2013).
- Traditional recommendation methods include collaborative, content-based filtering (Basilico et al., 2004) (Pazzani, 1999), etc.
- With the development of social media, users might share their feedback wherever they are, and there is some research about recommendation based on geo-tagged information (Memon et al., 2015) (Majid et al., 2013).
- As for the geo-tagged information, some research used geospatial coordinates (Silva et al., 2011) information.
- Explainable recommendation (Zhang et al., 2018) (Wang et al., 2018).

Project Plan

Dates	Subjects
Sep 27 - Oct 3	Project Proposal Due
Oct 4 - Oct 10	Prepare restaurant dataset and investigate system architecture
Oct 11 - Oct 24	Design and implement the recommendation system
Oct 25 - Oct 31	Model training and system optimization
Nov 1 - Nov 7	Testing, evaluating and debugging
Nov 8 - Nov 21	Workshop preparation
Nov 22 - Dec 3	Finalize project and prepare demonstration

Related Resources:

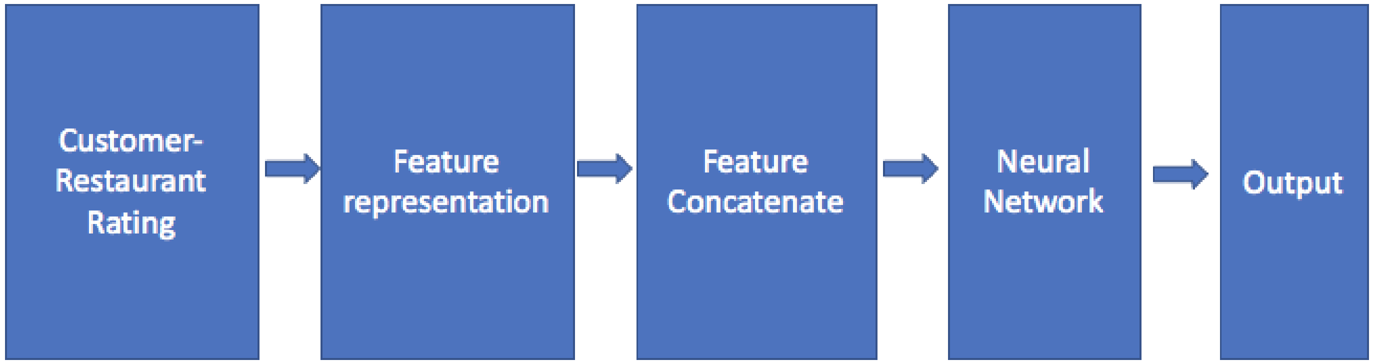
- Software requirement:
    - Flume: for data gathering
    - Apache Kafka: for buffering data
    - Spark Streaming: for streaming process data
    - Hadoop HDFS: database
    - URORA: push recommendation notification

Recommendation Methods-Neural Network

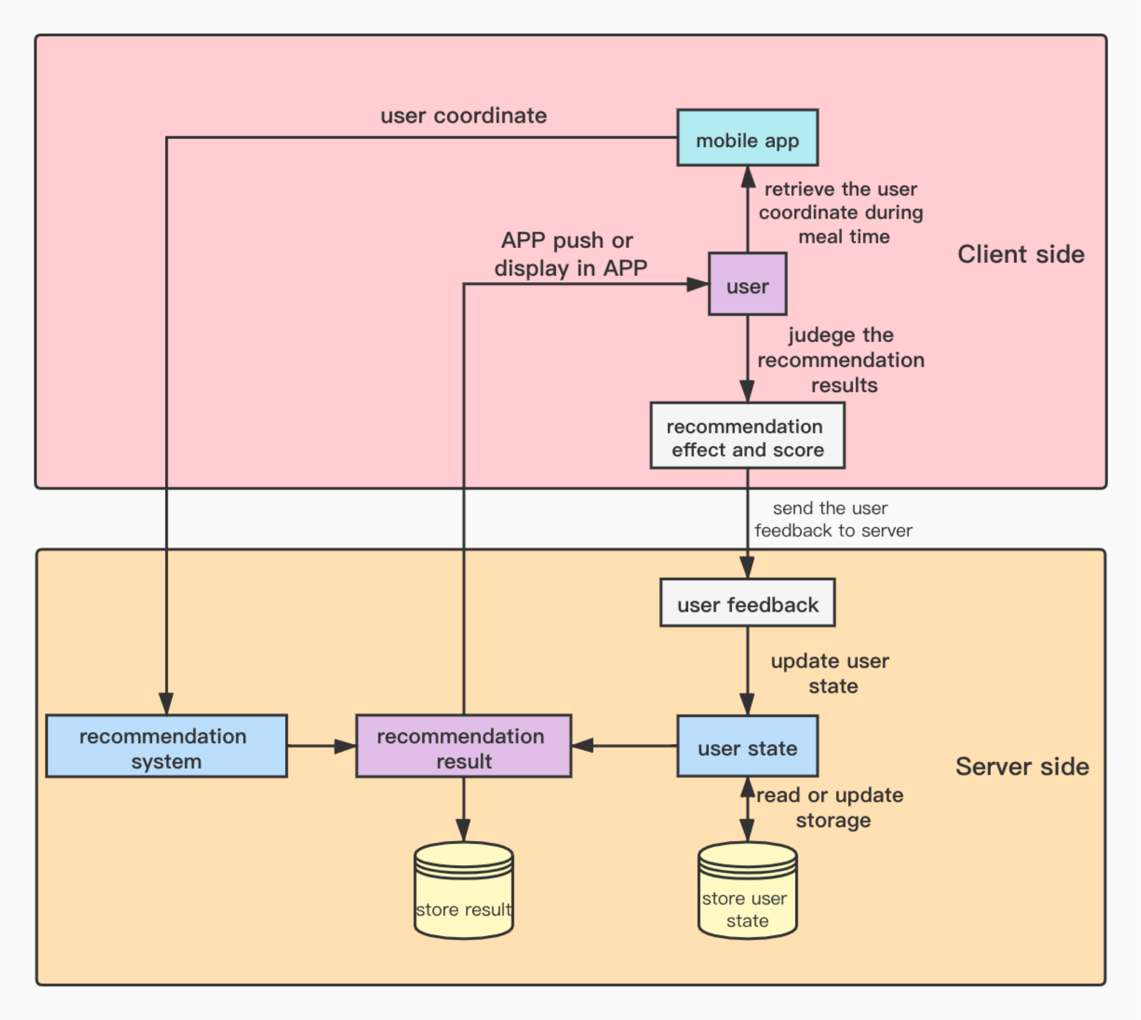
- Shallow ML models & potential Deep Neural Network

Customer feature vector: the user vector is a 99 dimensional float vector that represent the feature of each user.

Restaurant feature vector: the restaurant feature vector is a 100 dimensional vector that represent the feature of each restaurant.



Data-pipeline



Our project use a traditional pipeline for information-based system.

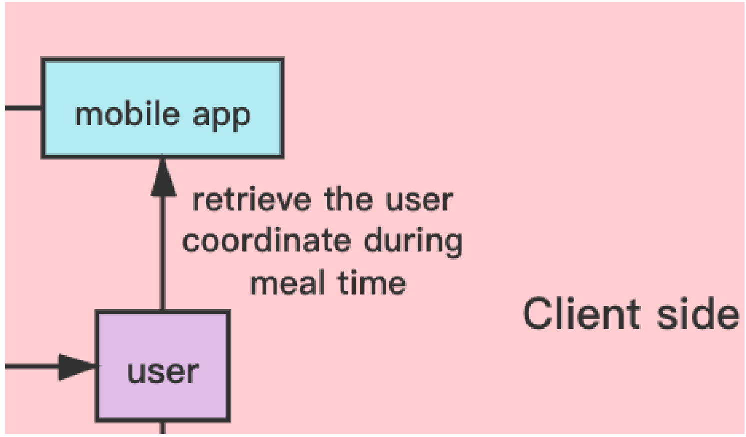
The server side maintain a database and generate recommendation result and push the recommendation to users devices

The client side collect user data and display the recommendation to the user

The database stores the user records which feed the recommendation system with useful user informatio

Data-pipeline – Client – Location Generator

- Since we cannot access the user's true location in our application, we use a location generator named gt-mobisim to generate the user coordinate. It can generate mobility traces and query traces for large numbers of mobile agents moving in a road network.



Data-pipeline – Client Server Interaction

- To send the user requests for restaurant recommendations, we use Spring Boot to implement the http request handler. When user request for recommendations, a http post request containing the user id and coordinate will be sent to the server side, and then get the recommendation results as response.

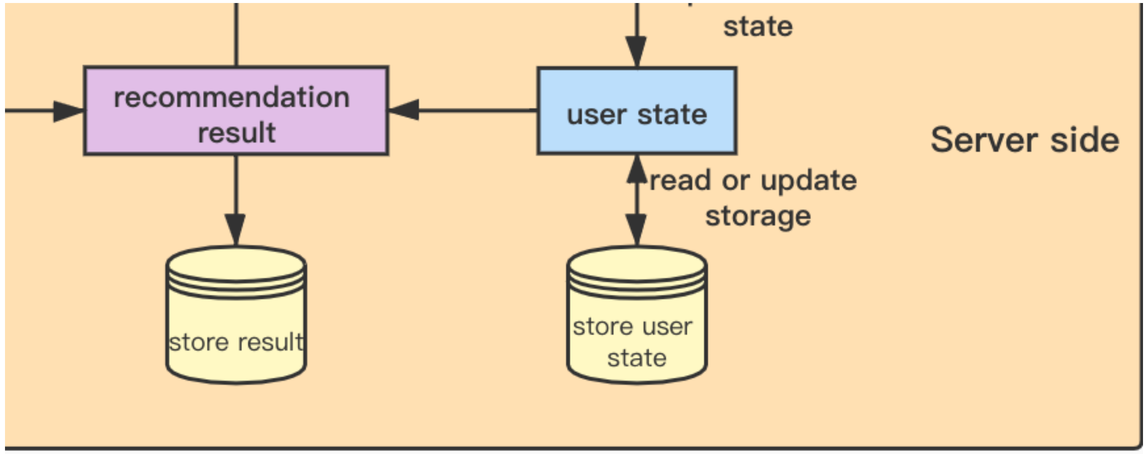
Data-pipeline – Server – Message Forwarding

- Users' rating of the recommendation results can also be stored for future use, so an interface is also designed to send the rating results to server. Here we use Apache Flume, a distributed, reliable, and available service for efficiently collecting, aggregating, and moving large amounts of log data.
- Since the request amount may be large during mealtime, we also use a message queue, Apache Kafka, to buffer the user requests.

Data-pipeline - Server – Database

The database maintain a feature vector for each user and each restaurant.

We use hbase to store the user states and recommendation results, since it's column-based database, it can provide more flexible storing ability and use the space more effective.



Num of Requests	Average Delay
1000	1.15ms
5000	1.32ms
10000	1.44ms

We get a average throughput of about 20MB/s

Evaluation

- Two sets of data:
  - The first is the training and test data that we use to build the model.
  - The second has the same data structure as our training dataset but was collected from a different city like Austin.
- In the first evaluation, we evaluate the accuracy of the model on the first data set to determine how well our model reflected the preferences of customers in San Francisco.
- In the second evaluation, we evaluated the model's transferability in another city. We will use the data gathered from Austin to evaluate our model that's trained on San Francisco's data and get the F1 score. If the F1 score is close to the F1 score we get in the first evaluation, we then are confident that our model has good transferability and can be used to give restaurant recommendations in another American city, Atlanta, for instance.

$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{\mathrm{tp}}{\mathrm{tp} + \frac{1}{2}(\mathrm{fp} + \mathrm{fn})}$$

Data-pipeline –Server –Data Filtering

customer_id	gender	location_number	location_type grid_3x3	latitude_x	longitude_x	country_id	city_id	**91 d user preference vector**
TCHWPBT	0	0	1	-96.44	-67.2	1	1	0,0.14,0,0.02,0,0 ...

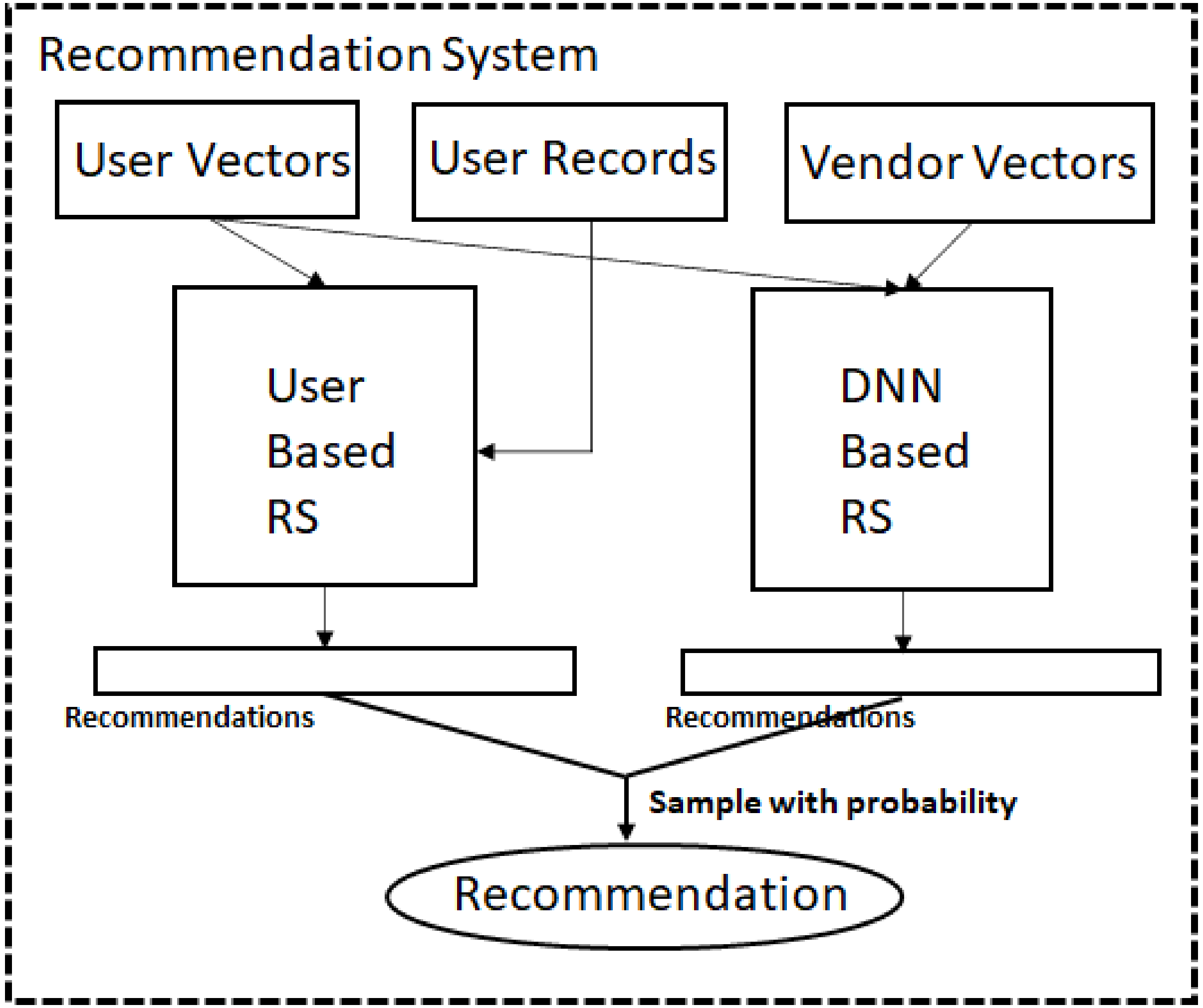
id	latitude	longitude	rank	Vendor_rate	Average_cost	country_id	city_id	**91 d vector tagger**
89	-96.44	-67.2	11	4.4	-13.5	1	1	0,1,0,1,0,0...

The user vector is a 99 dimensional float vector that represent the feature of each user.

The restaurant feature vector is a 100 dimensional vector that represent the feature of each restaurant.

Arabic,Breakfast,Burgers,Desserts,Free.Delivery,Grills,Lebanese,Salads,Sandwiches,Shawarma…

Data-pipeline – Server – Recommendation System



The Recommendation system consists of two parts

A traditional user based RS

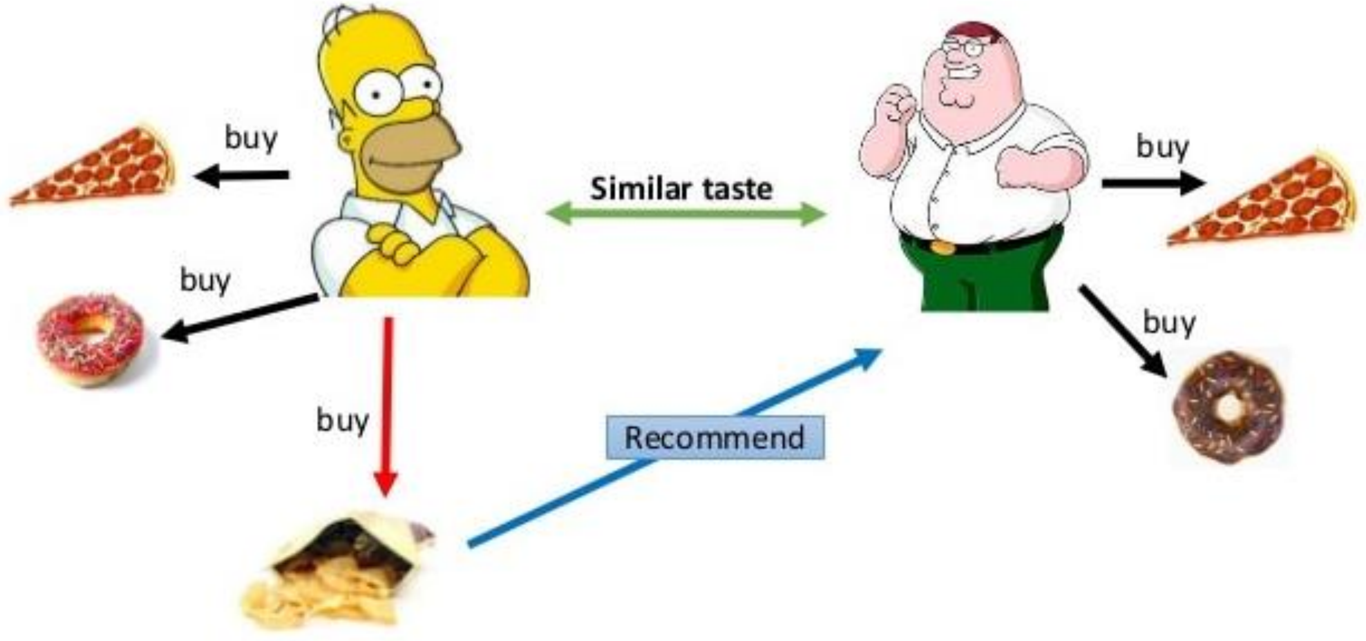
A neural network based RS

Each RS run independently and generate a recommendation

The final recommendation is sampled from the user based RS with probability p=0.5 and from from the DNN based RS with p=0.5.

Data-pipeline – Server – Recommendation System

Profile-based Recommendation: Motivation



Since our user feature space is less sparse than the restaurant feature space. For the non-neural network recommendation method, we choose to use user based recommendation.

References

- Shepitsen, A., Gemmell, J., Mobasher, B., & Burke, R. (2008, October). Personalized recommendation in social tagging systems using hierarchical clustering. In *Proceedings of the 2008 ACM conference on Recommender systems* (pp. 259-266).
- Guy, I., Zwerdling, N., Carmel, D., Ronen, I., Uziel, E., Yogev, S., & Ofek-Koifman, S. (2009, October). Personalized recommendation of social software items based on social relations. In *Proceedings of the third ACM conference on Recommender systems* (pp. 53-60).
- Qian, X., Feng, H., Zhao, G., & Mei, T. (2013). Personalized recommendation combining user interest and social circle. *IEEE transactions on knowledge and data engineering*, *26*(7), 1763-1777.
- Zhang, Y., & Chen, X. (2018). Explainable recommendation: A survey and new perspectives. *arXiv preprint arXiv:1804.11192*.
- Wang, X., Chen, Y., Yang, J., Wu, L., Wu, Z., & Xie, X. (2018, November). A reinforcement learning framework for explainable recommendation. In *2018 IEEE international conference on data mining (ICDM)* (pp. 587-596). IEEE.
- Basilico, J., & Hofmann, T. (2004, July). Unifying collaborative and content-based filtering. In *Proceedings of the twenty-first international conference on Machine learning* (p. 9).

References

- Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. *Artificial intelligence review*, *13*(5), 393-408.
- Memon, I., Chen, L., Majid, A., Lv, M., Hussain, I., & Chen, G. (2015). Travel recommendation using geo-tagged photos in social media for tourist. *Wireless Personal Communications*, *80*(4), 1347-1362.
- Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I., & Woodward, J. (2013). A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science*, *27*(4), 662-684.
- Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., ... & Sampath, D. (2010, September). The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 293-296).
- Silva, A., & Martins, B. (2011, November). Tag recommendation for georeferenced photos. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks* (pp. 57-64).
- Yelp dataset: https://www.yelp.com/dataset/documentation/main
- Restaurant Recommendation Challenge: https://www.kaggle.com/mrmorj/restaurant-recommendation-challenge/tasks