

Uncertain Data Reduction Based on Demographic Analysis for Tourist Place Recommendations

Abdulwadood Sabeeh Abdulwadood^a, Mohammed Abdal Baqee Mahmood^b, Duraïd T. Salim^c, ^aImam A'adham University College, ^bCollege of Education for Human Science , Department of Quranic Sciences, ^cAl – Mustansiriyah University , college of Basic Education , Computer Science ,Iraq, Email: abdulwadood_83@yahoo.com, moh_ameernet@yahoo.com, draid.edbs@uomustansiriyah.edu.iq

In recent years, millions of tourist photos are tagged into the online social network like Flickr, Panoramio. The social user shares their photos, travel experience, location, weather conditions, comments and queries into the Flickr. In that time, the social network contains some uncertain data that means some unrelated information is stored. So, they have many problems that occur while loading the data. To avoid the uncertain data we have to implement the pruning and clustering. In the proposed system, to implement the LSTM and dbSCAN algorithm. The LSTM algorithm is used to remove the uncertain data and dbSCAN algorithm is used for clustering and recommendation. The LSTM is one of the concepts in the recurrent neural network. We have to collect the data from the Flickr dataset. The dataset contains 5000 images. In clustering, we have to categorise the persons according to age. By this way, we have to recommend the tourist place for the people. For example, we analyse the four categories: children, family, bachelor and senior citizen. The tourist places are grouped into this category.

Key words: *Data Preprocessing, Demographic clustering, DBSCAN Algorithm, LSTM Algorithm.*



Introduction

In recent years, the camera phone and digital camera are most helpful to share the media on web service and also social media such as Flickr, Facebook and YouTube. The people share their travel photos and videos on social media websites. To increase the content in social media, one should define the different information such as geo-tagged information, videos, photos and types of textual information. The all geo-tagged facilities are most helpful for developing social media sites. The more than 50 million of geotagged photos are uploaded into the Flickr social media site, and more than 2 million of geotagged information stored into the Wikipedia. The multimedia is helpful for to share the textual metadata and temporal updates. Moreover, the multimedia is used for, to find the graphical information and collect some knowledge about social media and implementation of new opportunities. The people share the image in social media, which include where the image is taken and also contains the image contents.

The people have to share the information with others anywhere and anytime in social media. The web service and mobile apps have improved social media and also improved human knowledge capabilities. The social network has so many applications such as Email, LinkedIn, etc; they are most helpful for to communicate to others for the purpose of business, work and so on. There are so many photo sharing websites are there in social media such as Flickr, Panoramio and so on. These are most helpful for travellers in travel time. It provides the latitude, longitude, travel route, multiple geotagged photo and location service. The travel photo and travel information and text tags are most helpful for travellers in travel time. The researchers attention to the interest of the geo-tagged photo, demographic analysis and travel recommendation system. The geotagged photo is used to detect the landmark and geographical information and also it detects the errors up to 40 meters. It is also helpful to identify the video, mobile device, electronic games and landmarks. The location-based search and image quality provide the original digital image for to improve the milestone. The metadata and tags is to produce the image visual appearance and its helpful to travellers. Nowadays, the websites contain multimedia such as video, photo, textual information and also tagged the temporary context. It also contains the spatial context information, its nothing but where the photo was taken and which time the photo was taken. Nowadays, tourists are most interested in taking a photo in their cameras and sharing the photo and the place they are visiting into their social media. The GPS is the important thing for travelling time, it provides automated navigation and place recommendation, and also it provides the accurate route image that contains longitude, latitude and travel path. The data mining technique to provide a correct accurate result. The people can view travel information and share their own travel experience to others by using GPS. This system collects peoples travel history and location details to recommend new places. This system provides the demographic-based recommendation. The Demographic analysis means, the peoples are categorised into some

group based on age. In our system, this consists of four categories such as children, family, bachelor and senior citizen. The researcher analyse the peoples' history, and to categorise people and location. For example, children are mostly linked to the zoo and park.

In LSTM(Long Short Term Memory) is one of the types in Recurrent Neural Network. The LSTM purpose is to remove uncertainty. The users upload their tourist photos in the website, at that time many captions are created for the same image. The LSTM predict that caption and save the memory. For example, the user visits the Tajmahal, the user creates a different caption for the same image, but the place is the same. At that time, we have to implement the LSTM. The LSTM collect the all caption, after that classify them all, and predict the correct caption for that one. The DBSCAN algorithm purpose is clustering and recommendation. The User Data are collected from the website after that categorised based on demographics such as children, family, bachelor and senior citizen. Then recommend the tourist place for each category.

Related Work

To develop the milestone in social media based on collaborative users, they have so many ways is there. The people to find the geotagged photo and tag information are collected from Flickr (Crandall et al., 2009). The Place landmark is to provide the location of an image and to provide the unknown geo-tagged information into known geo-tagged information (Zheng et al., 2011). The clustering is used to detect the people interest location, and then grouping the people interest then to share the geographical knowledge (Lee et al., 2014). This system provides the location path details to the people. It calculates the distance of the location and location landmark. It estimates the direct view of where the photo was taken (Huang et al., 2012). The landmark contains the three types of information, they are angle, feature vector and landmark position. Some landmark image is similar to another landmark image based on the angle and features (Han, 2013). The people interest are improved based on geographical information, images and location shared on social media (Kennedy and Naaman, 2008). They provide the travel route recommendation and route information to the people by using GPS (Joshi and Luo, 2008). The HITS is used, get the user experience, landmark information, user search query information, user history from the social media (Zheng and Xie, 2011).

The matrix factorisation method is used, this method suggests the location according to people activity and trace the people current location based on GPS (Zheng et al., 2010; Yin et al., 2011). The various method to be implemented to predict the photo geo-location, and each location cell is represented by a grid and to estimate the efficiency of, which place the image was taken (Serdyukov et al., 2009). In k-means the algorithm is not suitable for more population place for to find the clustering geo-tagged photo. The k value calculation is very difficult. So, its not suitable for clustering (Chen, 2009). The geotagged photos and their features predict the image location and nearest neighbour place (Hays and Efros, 2008). The

people not considered the temporal information of location such as weather information, trip duration and so on. To resolve this problem, to implement the automatic route planning problem, its to compute the visit time and trip plan (Adrian and Gregory, 2009; Falah et al., 2019).

The recommendation system is used for, provide the best travel place to the user as well as route. The geo-tagged photo is to provide the best route in the city. The route distance and road conditions also provide the route recommendation. The route recommendation provides the best route and to maximise the tourist popularity and minimise the distance. The user have good travel plans, they recommend the best route (Sun, 2013; Falah et al., 2019). The Bayesian method used to compute the individual user favourites and to calculate the similarity of geo-tagged photos. The geotagged photos are collected from the Flickr website (Clements et al., 2010). We discover the two kinds of people one is who have most interested in to visit the location and routes and another one who interest on only location (Yin et al., 2011; Falah et al., 2019).

Implementation

Data Preprocessing

Data Preprocessing is a data mining technique. The real dataset contains some incomplete data, error data and null values. To overcome this issue, to implement the preprocessing. The Data preprocessing contains the following steps: Data Cleaning, Data Transformation and Data Reduction. The large dataset contains irrelevant and missing data. The data transformation work is to normalise, aggregate and generalise the data. The data Reduction aim is, to reduce the representation of the data in the data warehouse. The data preprocessing is mostly used for customer relationship management and neural network. For example worst, worsts, worstful all have the same meaning thus can be grouped into one word worst. The data preprocessing is completed, after that move to data classification.

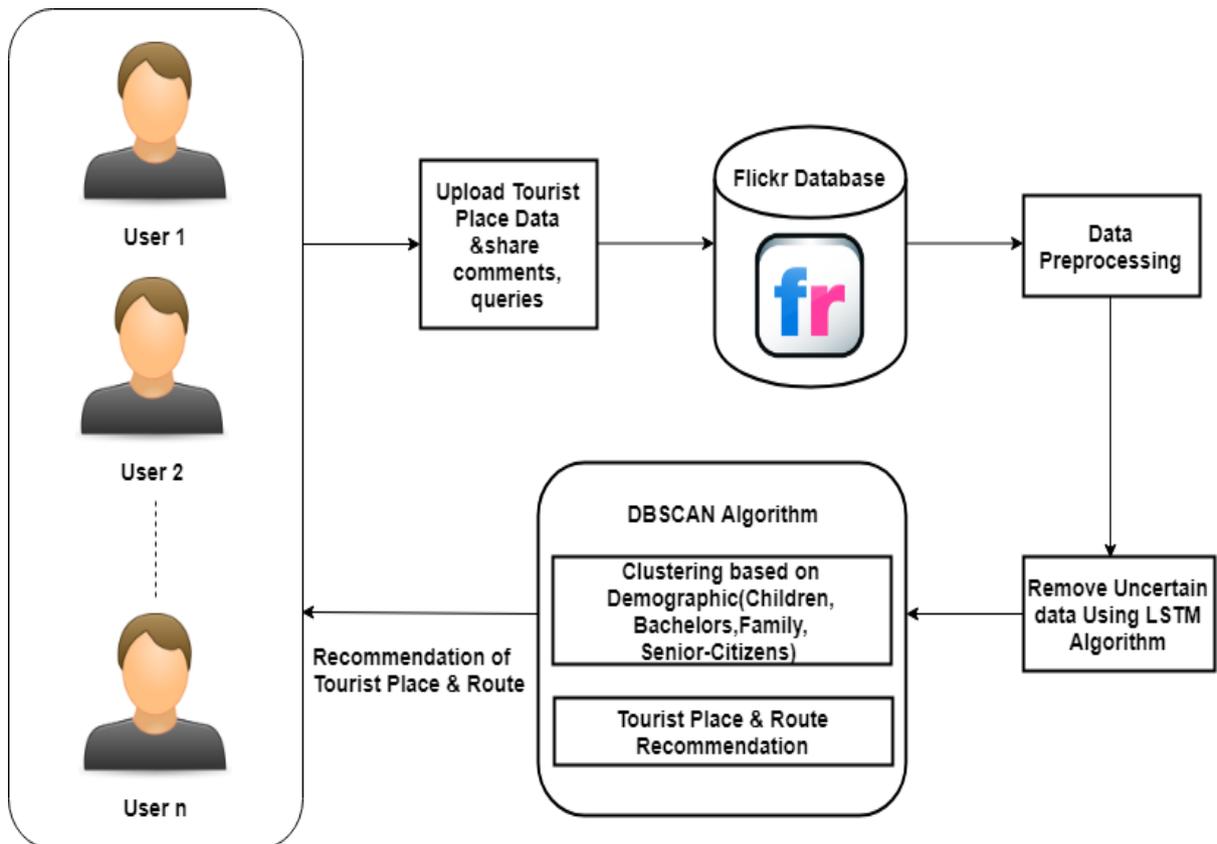
Demographic Clustering and Recommendation

In clustering, the group of object, grouped into a similar class. First, partition the set of data into groups based on the similarity and assign the label for each group. The Demographic analysis means an age, gender and so on. In this paper, we have to categorise the four category of pepole such as children, family, bachelor and senior citizen. After categorization we have to implement the recommendation. The user will have to give their current location and based on the location, he will be given with all the related details that are all the nearby places. Furthermore, the details will be displayed with a reference photo and the route required in reaching the place. Moreover, the suggests season, cost and the type of the place such as the museum, mountain or a library.

Demographic Category	Tourist Place
Children	Zoo,park
Family	Museum,garden
Bachelor	Club,movie
Senior citizen	Temple,Palace

Uncertain Data

The uncertain data means, the irrelevant data. It take more memory and also some time give error also. For example, the people went for the one zoo, and take photo and upload to the site. Each and every photo contain the same zoo but differnt caption is occur. So we have to face the memory error, repeated data error. To overcome this problem we have to use the LSTM technique. The LSTM get all the information, and analyse the data after that predict the uncertain data. LSTM is a long short term memory. LSTM extends that idea by creating a short-term and long-term memory component. Therefore, LSTM is a great tool for anything that has order. Since the meaning of a word depends on what is in front of it.



Algorithm Implementation

DBSCAN Algorithm

Input: DS - dataset, ϵ - neighborhood radius, Adf - adaptive density flag, Addh - adaptive density drop threshold

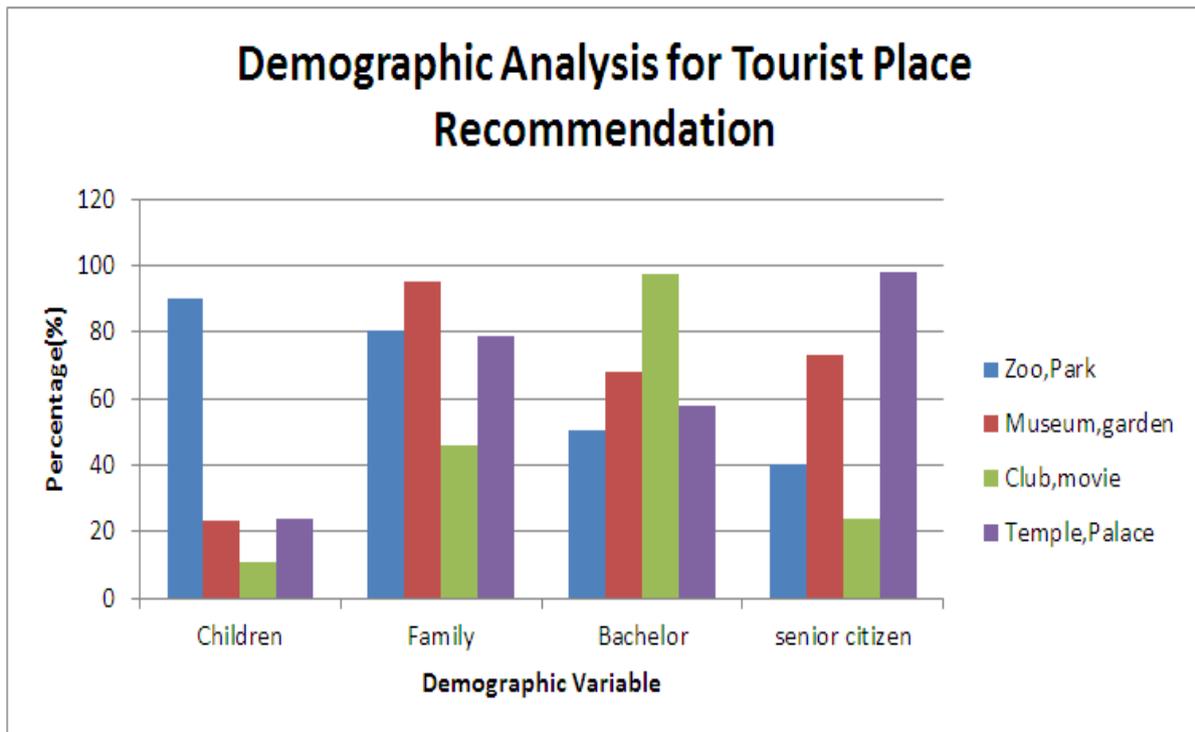
Output: Set of clusters

```
1 cid = 0
2 while ((A = getUPhoto(DS))  $\in \emptyset$  / ) do
3 CD = Addh
4 if ( $|\text{Nh}(A)| < M$ ) then
5 MPAN(p)
6 else
7 cid = cid+ 1
8 AssignCluster(A,cid)
9 UQ(R,GetNhPhotos(A))
10 while (R is not empty) do
11 A = DeQueue(R)
12 AssignCluster(A,cid)
13 if ( $|\text{Nh}(A)| \geq M$ ) then
14 if (Adf == true) then
15 AD(...)
16 else
17 UQ(R,GetNhPhotos(A))
18 end
19 end
20 end
21 end
22 end
```

Result Analysis

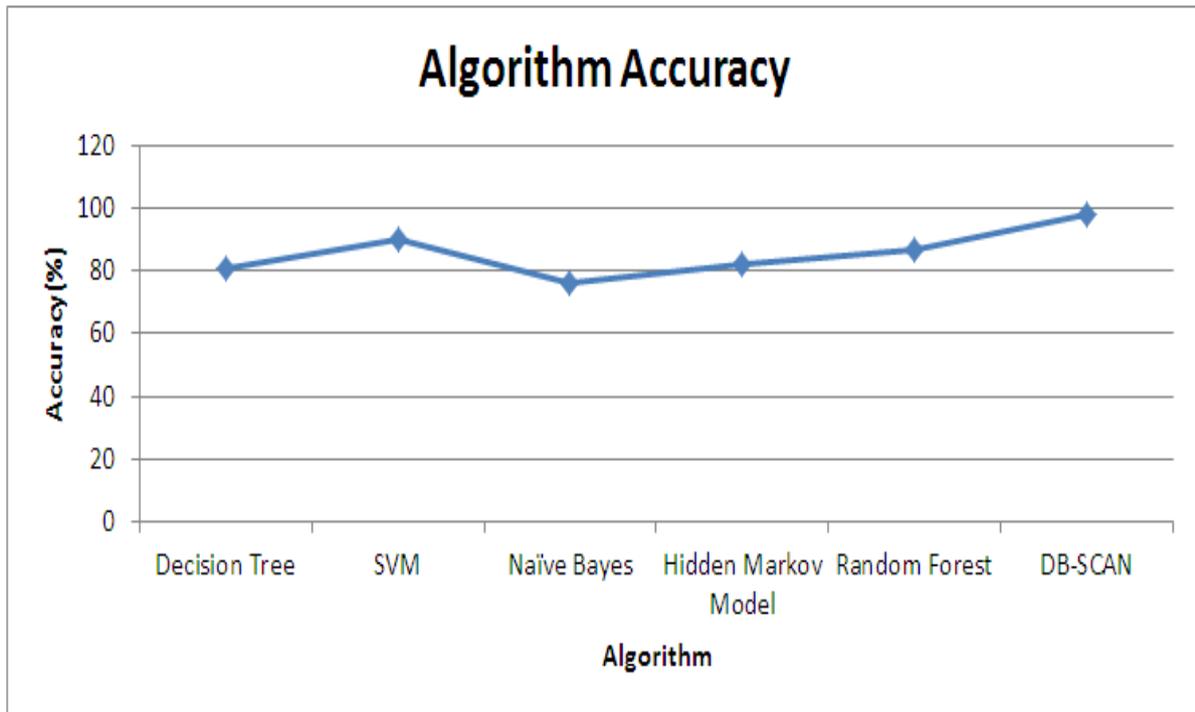
Demographic Analysis for Tourist Place Recommendation

The Flickr dataset contains many data. The data can be clustering based on demographic. And the recommend the tourist place and travel route. In categorisation and recommendation, we have to use for DBSCAN algorithm.



Algorithm Accuracy

Many algorithms used for the clustering process. We have to analyse, in which the algorithm is given more accurate result for clustering as well as recommendation. In that DBSCN Algorithm only give the correct accuracy for clustering and recommendation.



Conclusion and Future Work

In this paper, we have implemented the correct recommendation for tourism based on people who share photo, tags and geographical information in the social media. We have provided a methodology that applies Clustering and remove uncertain data, user's publicly provided photos and obtaining tourism preferences, and taking into account the current context of the consumer for the tourist recommendation system. We have presented the evaluation of our method on a publicly available model. The Flickr data set and results, which include photos taken in many cities, preferences in the new city can be more accurately evaluated and better tailored to other sophisticated landmark recommendation methods. We have to cluster the people based on the demographic. We conclude the best result by using DBSCAN and LSTM. In future, we have research the people most visited place are stored into the cache memory. The cache memory purpose is to store the repeated information.

REFERENCES

- Adrian, P. & Gregory, G. (2009). Deducing trip related information from flickr.
- Chen, W.-C., et al. (2009). Visual summaries of popular landmarks from community photo collections.
- Clements, M., Serdyukov, P., Vries, A. & Reinders, M. (2010). Using
- Crandall, D., Backstrom, L., Huttenlocher, D. & Kleinberg, J. (2009). Mapping the world's photos. (pp. 761–770
- Falah Y.H Ahmed, Khairul Farisb, Muhammad Irsyad Abdullah (2019).The University, Social Networks Media Sites, and Integration of Students' Most Popular. 20.
- Falah Y.H Ahmed, Rivishaa A/P Kanakarathanm, Marwan D. Saleh, Kevin Loo Tow Aik (2019). Enhance smart sales forecast and budgeting management. IEEE 9th International Conference on System Engineering and Technology (ICSET), 7 October 2019.
- Falah, Y. H. Ahmed., Amal Abulgsim, Masli and Ali, M. Mansoor. (2019). Review of Downlink Scheduling Over LTE Cellular Network. International Journal of Advanced Science and Technology Vol. 28, No.6, (2019), pp. 36-44.
- Han, Y. (2013). Recognize objects with three kinds of information in landmarks. Pattern Recognition.
- Hays, J. & Efros, A. A. (2008). IM2GPS: Estimating geographic information from a single image.
- Huang, Y.-T., Chen, K.-T., Hsieh, L.-C., Hsu, W. H. & Su, Y.-F. (2012). Detecting the directions of viewing landmarks for recommendation by large-scale user-contributed photos.
- Joshi, D. & Luo, J. (2008). Inferring generic activities and events from image content and bags of geo-tags.
- Kennedy, L. S. & Naaman, M. (2008). Generating diverse and representative image search results for landmarks.
- Lee, I., Cai, G. & Lee, K. (2014). Exploration of geo-tagged photos through data mining approaches. International Journal Expert Systems with Applications.



- Serdyukov, P., Murdock, V. & van Zwol, R. (2009). Placing flickr photos on a map.
- Sun, Y. et al. (2013). Road-based travel recommendation using geo-tagged images. Computers, Environment and Urban Systems.
- Yin, Z., Cao, L., Han, J., Luo, J. & Huang, T. S. (2011). Diversified trajectory flickrgeotags to predict user travel behaviour pattern ranking in geotagged social media.
- Zheng, Y. & Xie, X. (2011). Learning travel recommendations from user-generated gps traces.
- Zheng, Y.-T., Zha, Z.-J. & Engua, T.-S. (2011). Research and applications on geo-referenced multimedia: A survey. Multimedia Tools and Applications
- Zheng, V.W., Zheng, Y., Xie, X. & Yang, Q. (2010). Collaborative location and activity recommendations with gps history data.