Automatic User Profiling for Intelligent Tourist Trip Personalisation

Liam Attard [0299300L]
Department of Artificial Intelligence
University of Malta
liam.attard.18@um.edu.mt

***Abstract*—Planning for a holiday is a task which nowadays is associated with online planning and research. This paper will construct together a system which tries to automate the whole process. Given the increase in the amount of time each person spends posting on social media, the solution is aimed towards gathering the user's preferences from his profile's images automatically using current machine learning technologies. The paper then aims in using such information in generating an itinerary filled with activities to which the user's needs are met without the whole process of asking numerous questions.**

***Index Terms*—Automatic User Profiling, AI planning, Tourist Planner, Trip itinerary**

I. Introduction

A. Problem Definition

Producing an itinerary before a trip can be a demanding task which requires a substantial amount of research. Many times people rely on travel books, individual travel blogs and online websites to form a holiday plan, but these are not always tailored according to the traveller's preferences and opinions [1].

This paper focuses on creating a system which helps tourists automate the process of travel planning. An adequate automated trip planner application would consist of two parts,

1) the retrieval of user preferences
2) the generation of a custom itinerary

Numerous systems, which will be discussed in the Literature Review, are available and therefore building a working prototype is both possible and feasible Although these systems automate the process of producing the itinerary, they require a lot of end-user data and preferences to form a personalised itinerary. Can the user preference gathering be automated?

Given the amount of information a single user holds online, it is possible to automate and help the process of gathering personal preferences [2]. A deep learning model could be trained to classify a person's social media profile to determine what the user wants from a trip. This information alongside other parameters such as the user's budget and trip length could give out a very accurate personalised holiday plan.

B. Motivation

The immense amount of data generated by each user online [3] was the main motivation behind using this advantage in creating a unique system that benefits tourists by implementing something easy to use and does not bombard them with a lot of extra questions. Although planning itineraries can be a complex problem [4], if the users allows the system to gather preferences based on their social media profile, preferences can be collected automatically based on his posts.

C. Why the Problem is non-trivial

User Profiling based on social media has been an essential part of Personalized advertising. The advertisers can target their customers more accurately and earn more sales per viewer [5]. However, this paper aims in using such a technology to implement a different approach in automating the preference gathering.

D. Aims and Objectives

The aim of this project is to quickly generate a personalised itinerary by making use of preferences and parameters.

This system will aim to achieve the following Objectives:

1) Collect social media images to form a training and testing set which will be categoriesd by the activity. These can include images associated with events such as, nature, beach, sports, food, bars and clubs.
2) Design a model that classifies the images correctly.
3) Define a user profile based on the social media collection results and additional parameters.
4) Gather a list of places available and form scores for each activity based the user's parameters.
5) Generate quickly multiple itineraries each with different score levels.

II. Background Research and Literature Review

Several studies both on user profiling and on real-time automatic trip itinerary generation have been carried out throughout the years. There are many types of systems which help the travellers in their trips. Gavalas et al. [6] categorised these into **POI recommenders, Tourist Service Recommenders, Collaborative content from users and social media services, path recommenders** and **Personlised multiple-day tour planners**. The planning of a trip to a traveller introduces the Tourist Trip Design Problem (TTDP) which has recieved a lot of observation and heuristic contribution [7], [8]. Sylejmani et al. [9] have defined the TTDP as part of the Orienteering Problem(OP). OP problems contain a number of nodes each containing a score and try to solve the path containing the maximal score constrained with parameters such as time and budget [7]. Gunawan et al. [7] state that OP is a combination of the Knapsack problem and the Travelling Salesman Problem (TSP). There are many solutions to this problem which will be discussed in the next section.

A. Tourist Recommender Systems

In 2004, a paper by Dunstall et al. [4] was published using a prototype called the The Electronic Travel Planner (ETP). This system selects destinations by determining activities based on the user's preferences. Each activity is stored in a relational database with information such as duration, availability, date and time categorised as either tours, lodging or transportation. The requirements for forming such an itinerary include the number of children and adults, the location, the date range, budget and user preferences in the form of *mandatory, at least once, desired, forbidden and permitted* activities. Since examples given in the paper took 15-45 seconds to process the resulting running time was listed as an issue.

The Recommender System (RS) was provided by Sebatsia et al. [10] and Garcia et al. [11] to suggest tourist locations. User preferences are collected in the form of age, gender, nationality and ontology. The recommender is based on 5 techniques, *Demographic recommendation, collaborative recommendation, content based recommendation and knowledge based recommendation*.

A different approach using social media was presented by Choudhury et al. [1] in 2010 and Brilhante et al. [12]. Georeferenced Flickr [1] content alongside Wikipedia [2] information was used to gather information such as the date, location and popularity of the photos being uploaded. An OP algorithm was then used to generate the ideal number of Point of Interests (POI).

A tabu Search approach was proposed by Sylejmani et al. [13] as a Multi Constrained Team Orienteering Problem with Time Windows (**MCTOPTW**), an advanced form of the OP.In this algorithm, three steps were used in order to generate the activity plan. A new activity is added as a node to the trip using *Insert*, A node is exchanged with a new activity using *Replace* and two nodes are swapped using *Swap*. A pair of tabu lists structured frequently are used to avoid repeating solutions.

Recently, a solution towards presenting an itinerary solving conflicts between multiple tourists with different preferences by creating a group recommender system (GRS) [8], [14] was proposed by [9]. All tourists are split into groups by preference, during certain activities the itinerary splits up the groups to visit their specific POI. Before the trip one of the options is selected:

1) **Solo**: A trip for a single person.
2) **Subgroups**: The tourists are separated into smaller groups by preference and travel together.
3) **All Together**: One itinerary for all Tourists.
4) **Tourists Combined**: At certain times, tourists are separated to meet their personal preferences

An unique approach towards collecting the group's preferences for a GRS is offered by Nguyen et al. [15] An android group chat application called STSGroup was created to target conflicts between tourists. The idea is to collect the users' preferences when they are communicating with each other rather than individually. An example of a group of students travelling to South Tylrol (Italy) was given in the paper in which each person described its profile using certain tags such as the mood or parts of groups they form of. Upon the text conversation users send a selected POI in which other users can give it a thumbs up or down. Ranking lists and logistics are calculated in the background based on the group chat's data collection to determine the ideal preferences for the group.

Iterated Local Search could be used to generate the travel itineraries as seen vansteenwegen et al 2009 [16]. This approach is considered to be very suitable for real-time TTDP applications [17]. This approach finds a solution using the the best generated outputs from local search and repeats the procedure until a desired score is reached. In 2011, CityPlanner3 [18] integrated ILS with Greedy Randomised adaptive search Procedure (GRASP) [19]. This system allowed to alter POI durations and choose a starting and ending points

B. User Profiling for Travel Preferences

Recent years have shown how the average internet user has gone from being a passive content absorber to a content producer through the rise in social media [20]. This section describes several methods for user profiling and information gathering.

1) User Profiling based on textual methods: Textual data from comments and posts can be used to gather such user

[1]https://www.flickr.com/

[2]https://www.wikipedia.org/

preferences. In 2013, A system was shown by Ikeda et al. [20] which could perform sentiment analysis based on 100,000 Japanese user profiles and perform demographic estimation. Tags from social media posts can also be useful information to gather information about the user. Hung et al. [21] demonstrated a user profiling technique based on tags. Given an object and a user, the similarity between the tags is calculated. Both of these maximum similarities are summed up by the correlation between the set of object tags and the set of user tags.

2) User Profiling based on images: Terttunen [22] has shown how Instagram [3] has been of a major influence towards tourists. The ability to share photos of the amazing sights and landscapes has provided with more excitement than looking for inspiration in a tourism brochure.

Chen et al. [23] produced a system for automatically retrieving tags from images and incomplete tags called *FastTag*. The algorithm can be trained in $O(n)$ time and uses two simple linear mappings. Figure 1 shows an example of an input image used alongside the incomplete input tags *snow, lake, feet*. Given these two inputs the algorithm was able to produce the following tags, mountain, snow, sky, lake, water,feet, legs, boat, trees.

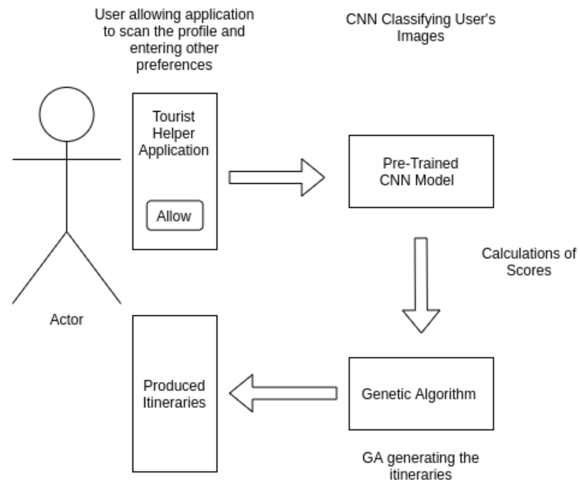
Fig. 1. The image shows an example given in the FastTag Paper. [23]

A technique which could be applied for user profiling could be classifying a person's social media images. Deep neural networks play an important role in this field of Image Classification [24]–[26]. In sections III,V, this paper will describe how this technology could be used to gather information for a tourist's user preference.

III. Proposed Solution

The Proposed solution will consist of four parts:

1) A dataset consisting of public Instagram images which will be discussed in section IV
2) A trained Convolutional Neural Network (CNN) model to predict the image category. A prototype of this model will be discussed in section V.
3) A method for calculating the scores of activities
4) A genetic algorithm which is able to produce itineraries in reasonable time and with a high score.



IV. The Dataset

A dataset of 2747 images was scraped from public Instagram posts from hashtags by using Puppeteer [4] (A javascript web-scraper). The dataset consists of 5 classes, beach, clubbing, nature, drinks, football with the potential to add more classes in the future. If a person's image post falls under a certain category, more activities leaning towards that activity will be suggested in the itinerary and therefore a person's preferences are collected automatically from Instagram posts and stories. The Beach dataset contains images of seaside, swimming and summer related activities, the Nature dataset contains images related to greenery and landscapes, the clubbing dataset contains crowds and people dancing in a club, the drinks dataset contains images of cocktails and people in a bar and the football dataset contains images of stadiums for sport fans.

[3]https://www.instagram.com/

[4]https://pptr.dev/

Fig. 2. The image shows a sample image form each of the categories.

V. Evaluation and Results

A. Convolutional Neural Network

1) Model Desciption: The following section will describe the TensorFlow Keras [27] model used in the prototype of the Image Classification tool alongside some results. After loading the images, the dataset was split into a training and testing set, 80% and 20% respectively. The model receives 180x180x3 standardized RGB images with a batch size of 32.

The dataset's small size will make it prone to overfitting. Data augmentation, a technique for generating additional training data by randomly transforming the existing data, is used to overcome this issue. Figure 3 shows an example of an image being rotated to generate additional images to the existing dataset. Overfitting is also avoided by using another technique called dropout. This technique randomly sets the activation function to 0. This forces the model to learn redundancy for everything.

Fig. 3. Data Augmentation

The model contains three convolutional layers with a rectified linear unit (ReLU) activation function, each followed by a pooling layer to lower down the spatial dimension of the input volume for next layers. The layer then ends with a flattening layer and two dense layers to reduce the outputs to the amount of classes, in this case 5. The following figure shows a summary of the whole model.

Fig. 4. model Summary

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
sequential_1 (Sequential)    (None, 180, 180, 3)       0
_________________________________________________________________
rescaling_2 (Rescaling)      (None, 180, 180, 3)       0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 180, 180, 16)      448
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 90, 90, 16)        0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 90, 90, 32)        4640
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 45, 45, 32)        0
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 45, 45, 64)        18496
_________________________________________________________________
max_pooling2d_5 (MaxPooling2 (None, 22, 22, 64)        0
_________________________________________________________________
dropout (Dropout)            (None, 22, 22, 64)        0
_________________________________________________________________
flatten_1 (Flatten)          (None, 30976)             0
_________________________________________________________________
dense_2 (Dense)              (None, 128)               3965056
_________________________________________________________________
dense_3 (Dense)              (None, 5)                 645
=================================================================
Total params: 3,989,285
Trainable params: 3,989,285
Non-trainable params: 0
_________________________________________________________________
```

2) Results: The results in the next figure show how the model has reached nearly 80% accuracy with very little data. More images could be added to make such a model more accurate.

The following shows an output of the prototype when given an image of a beach found on the internet: https://selfgrowth.info/photos/free-beach-photos-without-copyright/big-beach-illustrations-free-royalty6361.jpg

```
Downloading data from https://selfgrowth.info/photos/free-beach-photos-wi
1040384/1033413 [==============================] - 0s 0us/step
This image most likely belongs to beach with a 97.24 percent confidence.
```

B. Preference Gathering

This section will describe an example of the score calculation based on user criteria. Initially, the application asks the user certain preferences such as:

1) The trip budget
2) Moderation of activities (How busy the trip needs to be)
3) Users' characteristics (based on Instagram Results)
   A user's score for category **A** would be calculated by gathering the total number of **A** labelled photos over the total number of photos.
4) Number of people
5) Where the user is going
6) Date and time the user will be going

Fig. 5. The following figure includes some sample user preferences

€	4/5	30% Beach 20% Clubbing 50% Nature	4	Iceland Arrives at Airport	1 Nov - 7 Nov

€€€	2/5	50% Nature 50% Food	2	Paris Arrives at Airport	4 June - 14 June

€€€	4/5	No Instagram	5	New York Arrives by Car	13 August-20 August

Each Activity will contain its own details upon which a score is calculated. Some example of such parameters include:

1) Cost
2) How close is the place to the user's characteristics (from Instagram)
3) At what time is the place open
4) Approximately the amount of time people spend there (even based on user's moderation)
5) Place Importance
6) What type of Weather is the place accessible
7) Time to travel from the previous location
8) Place reviews

Fig. 6. The following figure includes some sample activity details

Example of a park

0	40%	0	1	8	Sunny	x	4/4

Example of a Restaurant:

20	70%	10am-11pm	1	5	0	x	3/5

VI. Conclusion

References

[1] M. De Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel, and C. Yu, "Automatic construction of travel itineraries using social breadcrumbs," in *HT'10 - Proceedings of the 21st ACM Conference on Hypertext and Hypermedia*, 2010, pp. 35–44. [Online]. Available: http://www.munmund.net/pubs/ht{_}10{_}long.pdf

[2] K. Buraya, A. Farseev, and A. Filchenkov, "Towards User Personality Profiling from Multiple Social Networks Prediction of Personality View project Multimedia Marketing AI View project," Tech. Rep., 2017. [Online]. Available: www.aaai.org

[3] J. Clement, "User-generated internet content per minute 2020 — Statista," aug 2020. [Online]. Available: https://www.statista.com/statistics/195140/new-user-generated-content-uploaded-by-users-per-minute/

[4] S. DUNSTALL, M. E. T. HORN, P. KILBY, M. KRISHNAMOORTHY, B. OWENS, D. SIER, and S. THIEBAUX, "an Automated Itinerary Planning System for Holiday Travel," *Information Technology & Tourism*, vol. 6, no. 3, pp. 195–210, may 2004.

[5] A. Hoppe, A. Roxin, and C. Nicolle, "Semantic User Profiling for Digital Advertising," *Economic computation and economic cybernetics studies and research / Academy of Economic Studies*, vol. 49, 2015.

[6] D. Gavalas, C. Konstantopoulos, K. Mastakas, and G. Pantziou, "Mobile recommender systems in tourism," pp. 319–333, mar 2014.

[7] A. Gunawan, H. C. Lau, and P. Vansteenwegen, "Orienteering Problem: A survey of recent variants, solution approaches and applications," pp. 315–332, dec 2016.

[8] A. Delic, J. Neidhardt, T. N. Nguyen, and F. Ricci, "An observational user study for group recommender systems in the tourism domain," *Information Technology and Tourism*, vol. 19, no. 1-4, pp. 87–116, jun 2018. [Online]. Available: https://doi.org/10.1007/s40558-018-0106-y

[9] K. Sylejmani, J. Dorn, and N. Musliu, "Planning the trip itinerary for tourist groups," *Information Technology and Tourism*, vol. 17, no. 3, pp. 275–314, sep 2017. [Online]. Available: https://link.springer.com/article/10.1007/s40558-017-0080-9

[10] L. Sebastia, I. Garcia, E. Onaindia, and C. Guzman, "E-Tourism: A tourist recommendation and planning application," *International Journal on Artificial Intelligence Tools*, vol. 18, no. 5, pp. 717–738, oct 2009.

[11] I. Garcia, L. Sebastia, S. Pajares, and E. Onaindia, "The Generalist Recommender System GRSK and its extension to groups," in *Lecture Notes in Business Information Processing*, vol. 75 LNBIP. Springer Verlag, 2011, pp. 215–229.

[12] I. Ramalho Brilhante, J. A. Macedo, F. M. Nardini, R. Perego, and C. Renso, "On planning sightseeing tours with TRIPBUILDER," 2014. [Online]. Available: http://dx.doi.org/10.1016/j.ipm.2014.10.003

[13] K. Sylejmani, J. Dorn, and N. Musliu, "A Tabu Search approach for Multi Constrained Team Orienteering Problem and its application in touristic trip planning," in *Proceedings of the 2012 12th International Conference on Hybrid Intelligent Systems, HIS 2012*, 2012, pp. 300–305.

[14] J. Castro, F. J. Quesada, I. Palomares, and L. Martínez, "A Consensus-Driven Group Recommender System," *International Journal of Intelligent Systems*, vol. 30, no. 8, pp. 887–906, aug 2015. [Online]. Available: http://doi.wiley.com/10.1002/int.21730

[15] T. N. Nguyen and F. Ricci, "A chat-based group recommender system for tourism," *Information Technology and Tourism*, vol. 18, no. 1-4, pp. 5–28, apr 2018. [Online]. Available: https://doi.org/10.1007/s40558-017-0099-y

[16] P. Vansteenwegen, W. Souffriau, G. Vanden Berghe, and D. Van Oudheusden, "Iterated local search for the team orienteering problem with time windows," *Computers and Operations Research*, vol. 36, no. 12, pp. 3281–3290, dec 2009.

[17] D. Gavalas, V. Kasapakis, C. Konstantopoulos, G. Pantziou, N. Vathis, and C. Zaroliagis, "The eCOMPASS multimodal tourist tour planner Keywords: Tourist Trip Design Problem Multimodal tour planning Urban transportation Orienteering Problem Time window Time dependent travel time Context awareness Web service Mobile application," 2015. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2015.05.046

[18] P. Vansteenwegen, W. Souffriau, G. V. Berghe, and D. V. Oudheusden, "The city trip planner: An expert system for tourists," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6540–6546, jun 2011.

[19] T. A. Feo and M. G. Resende, "Greedy Randomized Adaptive Search Procedures," *Journal of Global Optimization*, vol. 6, no. 2, pp. 109–133, mar 1995. [Online]. Available: https://link.springer.com/article/10.1007/BF01096763

[20] K. Ikeda, G. Hattori, C. Ono, H. Asoh, and T. Higashino, "Twitter user profiling based on text and community mining for market analysis," *Knowledge-Based Systems*, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.knosys.2013.06.020

[21] C.-C. Hung, Y.-C. Huang, J. Yung-jen Hsu, and D. Kuan-Chun Wu, "Tag-Based User Profiling for Social Media Recommendation," Tech. Rep., 2008. [Online]. Available: http://www.flickr.com/

[22] A. Terttunen, "The influence of Instagram on consumers' travel planning and destination choice," Tech. Rep., 2017. [Online]. Available: http://www.theseus.fi/handle/10024/129932

[23] M. Chen, A. Zheng, and K. Q. Weinberger, "Fast Image Tagging," Tech. Rep., 2013. [Online]. Available: http://tinyurl.com/9jfs7ut

[24] K. Balaji and K. Lavanya, "Medical Image Analysis With Deep Neural Networks," in *Deep Learning and Parallel Computing Environment for Bioengineering Systems*. Elsevier, jan 2019, pp. 75–97.

[25] A. Cufoglu, "User Profiling-A Short Review," Tech. Rep. 3.

[26] Francois Chollet, "Building powerful image classification models using very little data," pp. 1–12, aug 2017. [Online]. Available: https://deeplearning.lipingyang.org/wp-content/uploads/2016/12/Building-powerful-image-classification-models-using-very-little-data.pdf

[27] Mart\'\in~Abadi, Ashish~Agarwal, Paul~Barham, Eugene~Brevdo, Zhifeng~Chen, Craig~Citro, Greg~S.~Corrado, Andy~Davis, Jeffrey~Dean, Matthieu~Devin, Sanjay~Ghemawat, Ian~Goodfellow, Andrew~Harp, Geoffrey~Irving, Michael~Isard, Y. Jia, Rafal~Jozefowicz, Lukasz~Kaiser, Manjunath~Kudlur, Josh~Levenberg, Dandelion~Mané, Rajat~Monga, Sherry~Moore, Derek~Murray, Chris~Olah, Mike~Schuster, Jonathon~Shlens, Benoit~Steiner, Ilya~Sutskever, Kunal~Talwar, Paul~Tucker, Vincent~Vanhoucke, Vijay~Vasudevan, Fernanda~Viégas, Oriol~Vinyals, Pete~Warden, Martin~Wattenberg, Martin~Wicke, Yuan~Yu, and Xiaoqiang~Zheng, "{TensorFlow}: Large-Scale Machine Learning on Heterogeneous Systems," 2015. [Online]. Available: https://www.tensorflow.org/