International Journal of Engineering & Technology, 7 (4.19) (2018) 207-211

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Current Challenges and Approaches in Recommending Venues by Using Contextual Suggestion Track from TREC

KianLam Tan[1]*, Haseeb Ur Rehman Khan[2], ChenKim Lim[3]

Faculty of Art, Computing and Creative Industry,
Sultan Idris Education University, TanjongMalim, Perak DarulRidzuan 35900, Malaysia
**Corresponding Author Email: [1]tankianlam@fskik.upsi.edu.my*

Abstract

Contextual suggestion systems have been emerging as an entrance region of research, attributable to the innovative advances in smart connecting things and rapid growth of Big Data. In this regard, the primary purpose of contextual suggestion systems is to propose things that assist users to settle on choices from countless activities, for example, according to their specific context, the system may predict what place users would find interesting to visit or in what restaurant they would prefer to eat. In a smart environment using big data, users' current activity and past behavior could be incorporated into the suggestion process with an end goal that provides right suggestion at the right time with appropriate location on users personal preferences. The objective of this paper is to provide an overview of contextual suggestion system and a review of TREC's contextual suggestion track to investigate the approaches have been used in order to develop a model for contextual suggestion.

Keywords: *Internet of things (IoT); Contextual Suggestion Track; Recommendation System; Information Retrieval.*

1. Introduction

The Smart city has been very fashionable in the arena of recent years. This is one of the concepts that promoted from the merger of the Information and communication technology (ICT) and the Internet of things (IoT), this connects routine electronic devices and objects with live networking technologies. Since the rapid growth in urbanization, a number of applications have been introduced recently to support a smart environment for instance, smart transportation [6], smart grids [11], smart healthcare [13], smart homes [9] and smart cities [9][12]. Smart cities heavily depend on the availability and responsiveness of digitalization i-e information and communication network, social infrastructure and knowledge databases [9].

The dynamics and the increasingly strategic significance of smart cities lead to the evolution of databases. IoT provides the essential service to implicit this voluminous amount of data or big data. The magnitude of the Big Data perhaps differs from Petabytes to Exabytes. These databases are categorized by capacity, speed, and variation in data types that are coursing of online real-time data in voluminous amount have been growing on enormous rates [14]; [16]. Many challenges have been reported in the field of big data some are, security, searching, sharing, querying, and storing. The traditional applications cannot be used to manage, create, capture and process this enormous amount of datasets within a bearable elapsed of time. Approximately 2.5 quintillion bytes of data which is created daily that includes photos and written contents. [18].

To obtain a valuable insight of a city, big data offers the potential to pinpoint the considerable amount of data from various sources. [10]. Many techniques can be neutralized for managing the big data however we will like to focus only on Information Retrieval (IR) techniques that must have an ability to respond with appropriate information as per the needs of the user. It can be used for anticipating contextualization based on user interests and preferences. Apparently, users depend heavily on their cell phones when they are searching for occasions to take part, finding fascinating close by venues or things to do. To explicate contextualization further, consider a traveler in a new city given a set of the traveler's preferences for places and activities in his home city, the system suggests something based on his point of interest. For instance, If he drinks beer the system may suggest him a brand of beer that he drinks and suggests the nearest locations for the stores where the specific brand of drinks are available by mining into the user profile and big data[2]. In comparison to other conventional recommender frameworks, this idea might be an ideal fit to provide a suggestion within the context and it can produce information from various sources. The primary test in implementing even such essential contextual services lies in the capacity to contextualize web-scale IoT information based on available datasets and do this proficiently in concurrent time.

The purpose of this article is to review yearly development for TREC specifically in regards to the contextual suggestion. TREC offers a diverse set of data called tracks, the data has been collected by several teams working in different countries and use several approaches under the banner of the contextual suggestion track to explore approaches that are much subjected to context and user preferences by using complex data. In Section 3, we describe, the techniques and approaches used in contextual suggestion tracks. In Section 1, we presented a survey of the related literature, analyze, and evaluate key issues, and challenges. Section 2 highlights the overview of specifically TREC's contextual suggestion track. And we presented our final thought in Section 4.

2. Overview of Contextual Suggestion Track

The subject of contextual suggestion and personalize information depends on the data mining by using information retrieval techniques, by adapting this approach TREC, have been organizing yearly workshops in which several groups around the globe used to participate. The TREC series is sponsored by NIST Information Technology and the Contextual Suggestion Track is a part of TREC that manages complex data needs which are profoundly subject to context and information personalization.

Fig. 1: Comparison of Groups and Runs from 2012 to 2016

In the early development of contextual suggestion track, from 2012 to 2014 several groups participated in collecting the data and implementing the approaches called runs. Moreover, in 2015 to 2016 there were two types of experiment, one called a live run and other was batch run, in live run groups participated in live experiments for fourteen days. Individuals in the live task requested to enroll with contextual suggestion TREC. They sent a request to made suggestion, these suggestions included a city, a request related to point of interest, and details of the person, therefore information can be personalized. For the batch run, the request was made during the live run were tested, the request rated by most of the participant were included instead of all POIs [1] [2] [17].

2.1 Task Description

Sample suggestions consists of profiles, and contexts these sample's description were the results of detail suggestions that are returned by contributors tracked by a brief description. A single run consists of a single suggestion result. Two experiments were allowed for each candidate, a file was generated automatically for every experiment from suggestions, profile and context as a result.

2.2 Suggestions

Depictions of attractions were included in every suggestion on the bases of user's preferences or the users would find it interesting were intended to be proposed. Every suggestion comprises of a website's URL, short description and a title. To be utilized to make profiles sample suggestions were accumulated manually [2].

2.3 Profiles

Specific preferences were exhibit for a specific attraction. In every profile, sample suggestion list was comprised of preferences for the attractions, these profiles dispatched to the participants. To construct profiles for the task, according to preferences and POI of every participant, TREC team surveyed students from the University of Waterloo, and crowdsourced users from Mechanical Turk [1] [2].
Profiles were fragmented into two profiles which contain a list of suggestions and ratings. The list of each suggestion contains an id, a title, a description, and an URL [4].

2.4 Context

The context delineates a user's current location and the time usually a user is searching. Every context contains "a period of a day" (morning or evening), "a day of the week", "a seasonal search activity" and location I.e. City. In every context "a day of the week", "a period of a day" and "the season" were picked indiscriminately, with every alternative in the field had an equivalent possibility to be chosen. Cities with a population more than 100,000 were selected from geonames.org list of US cities. The probability for a city being picked was based on the population. If the procedure produced similar context twice then duplicate context was disposed of, and another context was created [2].

3. Description of Approaches

An overview of approaches used in TREC contextual suggestion track from the year of 2012 to 2016 are described in this section, these methods were actually utilized by the groups' participants, they mostly used the open web as their principle data source as shown in Table 1.

3.1 Team CSIRO

There were two runs used for juncture of venues called "csiroht" and "csiroth", Foursquare and Google Places provided the data. Ranking of venues was listed in two scores of a linear combination. One caught the venues' suitability specified by user's preferences, and the other one caught the venue's suitability at the particular time.
The description of every vector decreased to a weighted vector BM25. Fixed multipliers of 0.75 for positive cases and - 0.25 for negative cases used to sum the vectors. For scoring each suggestion's description the subsequent term weights were utilized. The period of a day scores mixed with the text-based scores in "csiroth" run by the ratio of 7:3 and the mixing ratio for "csiroht" run were 3:7 [2].

3.2 Team Fasilkom UI from Universitas Indonesia

For the TREC 2012, two runs were submitted by university Indonesia. In order for generating the place suggestions from the contextual data, geo locations and user models were combined by their contextual suggestion system. For producing ranked results they utilize the review rating for the places as for scoring. For improving the results they tried to make suggestion more interesting and wide-ranging. Techniques implemented in the both runs FASILKOMUI01 and FASILKOMUI02 were almost same however FASILKOMUI01 used diversity to its results while FASILKOMUI01 used the default settings. [2]

3.3 PITT at TREC

The Data from yelp used by the PITT's system for making participants suggestion and enlarging user profiles. To figure out the similarity between users and examples vector space model was used. To consolidate numerous attributes of user profiles a linear regression model was utilized. 5 fold cross validation technique was used for training and testing the system [4].

3.4 University of Lugano

Google places API was used by the system to get an initial list. Yandex Rich content and Google custom search APIs were utilized for producing pieces for description. The system extended the descriptions based on the examples and a positive and a negative model was created for each user. Suggestions' lists for every profile in regards to user's context were based on the models mentioned above [4].

3.5 Team BJUT

BJUT submitted two runs to TREC contextual suggestion track. A combination of venue ranking and category were defined by this group. To determine the probability of venues could be picked by the user was based on five categories (Activities, shopping, attraction, nightlife and restaurant). The probability determined the numbers of venues from top-ranked suggestions from each category. Those suggestions then ordered by either both rank and preferences of the user (BJUTB) or based on ranking specifically (BJUTa) [3].

3.6 Waterloo Clarke

The venues were categories in a hierarchical design, specific venues were on the lower and the most general venues were on the top of the hierarchy. K-L divergence model was implemented to normalize the terms from the venue descriptions and web pages and used it as a highlighted tags for sorting. Moreover, categories were also utilized as highlights. The attractions were picked from each tag based on the preferred attraction's example exist in them. Only positive ratings in the tags rank were used for the first run, WaterlooaA, and both positive and negative evaluations were utilized for the second run, WaterlooB [1] [3].

3.7 HP Labs China

Hp labs China submitted two runs, named, "hplcrating" and "hplcranking". Context-aware recommendation method was utilized to generate a list of suggestion. All the spots in every related city were crawled. For finding a latent factor in each profile Matrix factorization was implemented. For anticipating the scores for all suggestions, SVD++ was used in "hplcrating". While a pairwise positioning model was used to rank a list for all suggestion in "hplcrating". For each suggestion the depiction was produced in a human defined template, category location and numerous different highlights were included in it as well. [2]

3.8 Team USI (Phase I)

Virtually 600k venues were gathered from Foursquare. Based on user liked or disliked as their related normalized frequencies, they made a positive category and negative for profiles using the crawl data from Foursquare. For indicating the similarity between a specific user and a new venue the initial category from profiles were used. They picked the top 10 most ranked venues to create the initial rankings for each user to accumulate additional data. Additionally, a positive and negative frequency based on the profile's venue search keyword was made. They measured the similitude between the specific user and venues by extracted venue search keyword as a new set of venues. Then used the linear combination for the venue category and search keyword score to remake the top 10 venues for each user [17].

3.9 Team USI (Phase II)

In the phase II run they combined a set of multimodular scores that were computed based on Multiple Location-Based Social Networks (LBSNs) with a score that anticipated the level of suitability for venues in context of users. Positive and negative reviews were utilized to make a user profile to prepare the classifier which then predicts whether a specific user will or will not prefer another venue. Additionally, the venue categories and taste keywords were further calculated on bases of the frequency based scores. Then two data sets were made as for the prediction of context by using trained classifier and crowdsourcing with the highlights they extricated from the data sets. The final ranking of the user's suggestions was generated from a linear combination of all the scores. [17]

3.10 Bupt_pris_2016

Yelp APO and Foursquare API were used for data collection. For every tag, average ratings for users was calculated, from the list of preferences. Collaborative Filtering was used for the tags which were missing the users' rating. Mean function or a max function was performed to calculate the attraction ratings of the user. After ranking the candidates' ratings, the team tracked changes in the user's ranked list. The teams' best run, in which they applied higher weight for Foursquare (0.4) and lower weight for yelp (0.2) on ratings, and for calculating the attraction's rating the team applied max function. [17]

3.11 Laval Lake Head

In the beginning, the group retrieved 100 attractions according to a user profile by defining a customized query. For covering individual context preferences and worldwide pattern of interests, two independent ranking models were applied respectively to those 100 candidates. The proficient regressor was applied to the data from TREC's 2015, therefore, the categories and well-known spots could be prioritized based on the popularity, in the first model. To get the user preferences the second model acquainted the word embedding. A similar Euclidean space the word vectors were signified for both candidate places and the user profiles. To calculate comparison between user and attraction's scores vector distances was measured. In the last step, scores from these two models were summed up to get final rankings. [17]

3.12 U Amsterdam

The team used neural user profiling and neural category preference modelling to study the contextual suggestion by the assistance of suggestions released by contextual suggestion track in TREC 2016. In their best execution, they contemplated category preference models to foresee relevant context and suggestion for the users. Moreover, they tossed contextual recommendation issue to a binary classification issue. They used 123 suggestion category features as an input to a neural network with 4 concealed layers having 478 units, with the intention to study a user preference model. The model defined in two categories a train set and a test set. Train set consists of preferences in the profile of each user and the test set consists of candidates' suggestion presented in the phase 2 requests. [17]

4. Recommendations

For recommendations, it would be fascinating to direct research towards developing the contextual suggestion system to extract relevant information based on a deep user-based study and researchers might use this information for analyzing the impact of the contextual system on the user' preferences by using comparability measures for instance, Normalized Mutual Information [5], or Purity measure [8], Rand Index [20] and multi linear regression [15]. In the last few years variety of domains for recommendation system have been opening up for development in the mechanical research and applications, countless companies are adding contextual information to their recommendation systems to integrate reused components in different areas [7] [19]. Moreover, Researchers could study the privacy problem as well by developing tools that enable users to set consents of accessing to their data and control the mechanism for information sharing, however, it might affect the system quality. Hence it would be interesting to examine the effect of this control of the system [7].

We have identified in literature that we need to explore more models to achieve the contextual suggestions system based on the user's point of interests (POI). The problem in the models used in TREC (contextual suggestion tracks) are good in some perspective while powerless in numerous angles, moreover, a significant amount of solutions was noticed as overlapped in these experiments

Table 1: Contextual Suggestion Systems' Approaches and Results

Team	Year	Description of the Approaches	Results
Team CSIRO [2]	2012	Used Google Places and Foursquare Data. Linear Combination for two scores for ranking. Fixed multipliers of 0.75 for positive and 0.25 for negative used in BM25 Weight Vector.	P@5WGT: 0.0772 P@5 GT: 0.4516 P@5 G: 0.7579 P@5 T: 0.4734 P@5 W: 0.1531 P@5 D: 0.1438
Team Fasilkom UI from Universitas Indonesia [2]	2012	Yelp, Google Place, and Trip Advisor Data. Combines User Model and Geo location. Review Ratings for places. Diverse suggestion.	P@5WGT: 0.0667 P@5 GT: 0.4935 P@5 G: 0.7770 P@5 T: 0.5243 P@5 W: 0.1136 P@5 D: 0.1648
HP Labs China [2]	2012	User preferences and related places crawled from Yelp. Matrix Factorization for collaborative filtering. Pairwise ranking model. Contextual post filtering adjust the results.	P@5WGT: 0.2117 P@5 GT: 0.5725 P@5 G: 0.8815 P@5 T: 0.5833 P@5 W: 0.4124 P@5 D: 0.3802
University of Lugano [4]	2013	Obtain Initial suggestion from Google Places. Yandex Rich Content API and Google Custom Search APIs used for generating descriptions. Example description was expanded into positive and negative models. Ranked models into suggestions list for each profile.	P@5 Score: 0.4332 TBG Score: 1.8374 MRR Score: 0.5871
PITT [4]	2013	Used Yelp data for profiles and suggestion. Vector scale model for calculating similarity. Linear regression used for combining various attributes of user. 5-fold cross validation.	P@5 Score: 0.2601 TBG Score: 1.0495 MRR Score: 0.3816
Team BJUT [3]	2014	Used venue categories and venue ranking for suggestions. Calculate probability the venues user would like by using profile data to determine top 50 suggestions. The 50 suggestion were ordered in both ranked only or ranked and user's personal preferences.	P@5 Score: 0.5010 TBG Score: 2.2209 MRR Score: 0.6677
Waterloo Clarke [3] [1]	2014 - 2015	Categorized venues in hierarchy, Most general clusters of venues in the top and most specific in the bottom. Terms from webpages and description normalized with K-L divergence. The clusters sorted by most liked venues attraction. Both positive and negative rankings were used in the clustering ranking algorithm.	2014 P@5 Score: 0.4308 TBG Score: 1.8379 MRR Score: 0.6244 2015 Precision@5: 0.4716 MRR: 0.5929
USI (Phase I) [17]	2016	Crawled Foursquare for venues. Create positive and negative profiles based on categories user liked or disliked. Measured the similarity between user and venue by using Initial category. Created rankings, taste keyword with positive or negative frequency and picked the top 10 venues for each user. Re-ranked the top 10 venues by using linear combination of venue ranking and taste keyword score.	NDCG@5: 0.2826 P@5: 0.4295 MRR: 0.6150 NDCG: 0.2083 MAP: 0.0868 Bpref: 0.1772 P@10: 0.3148 Rprec: 0.1619
USI (Phase II) [17]	2016	Computed and combined multimodal scores from LBSNs with an appropriateness of a venue score. Used positive and negative reviews to train classifier. From the venue and taste keywords frequency based scores calculated. By using crowdsourcing and classifier two data sets were created for appropriateness prediction. Final ranking was produced by using linear combination of all scores.	NDCG@5: 0.3265 P@5: 0.5069 MRR: 0.6796 NDCG: 0.6804 MAP: 0.4590 Bpref: 0.4507 P@10: 0.4603 Rprec: 0.4177
bupt_pris_201[17]	2016	Data crawled from Foursquare and Yelp API. Calculated average user ratings for each marked tags. Tags without ratings filled by Collaborative Filtering. Mean and max function used to get user's rating. Tracked changed in the ranked list. Applied higher weight for foursquare (0.4) and lower weight for yelp (0.2) ratings. Used Max function for calculating user's attraction ratings.	NDCG@5: 0.2936 P@5: 0.4483 MRR: 0.6255 NDCG: 0.6625 MAP: 0.4318 Bpref: 0.4476 P@10: 0.3983 Rprec: 0.3956
LavalLakehead [17]	2016	Retrieve 100 initial attractions' query based on user's profile. The attractions divided into global trend of interests and contextual individual preference. Pre-trained regressor used from 2015 TREC data. Individual preferences captured by using word embedding. Calculated similarity score between user and attraction. At the end scores from both models was summed up.	NDCG@5: 0.3281 P@5: 0.5069 MRR: 0.6501 NDCG: 0.6770 MAP: 0.4536 Bpref: 0.4666 P@10: 0.4500 Rprec: 0.4168
UAmsterdam [17]	2016	Used neural user profiling and neural category preference modeling to study suggestions. Contemplated category preference models to foresee relevant context and suggestions. Tossed contextual recommendation issue to a binary classification issue. Used suggestion category features for an input to a neural network consists of 4 concealed layers having 478 units,	NDCG@5: 0.2824 P@5: 0. 4448 MRR: 0. 5924 NDCG: 0.66544 MAP: 0. 4168

		to study a user preference.	Bpref: 0. 4452 P@10: 0. 4310 Rprec: 0. 3881

5. Conclusion

This paper inspected approaches and models used in TREC contextual suggestion tracks' from 2012 to 2016 and discussed the importance of contextual suggestions in IoT and Big Data. One of the key necessities for contextual suggestion is analysis of real-time Big Data. The models used for the data collection and processing the approaches ought to be addressed properly. The potential advantages that can be gained by implementing the contextual suggestion system are recognizing user interests and providing personalized contextual support rapidly and effectively when required by the user.

Acknowledgement

"This research is fully supported by University Research Grant from Sultan Idris Education University under the grant number of 2018-0134-109-01."

References

[1] A. Dean-Hall, C. L. A. Clarke, J. Kamps, J. Kiseleva, and E. M. Voorhees. Overview of the TREC 2015 contextual suggestion track. In E. M. Voorhees and A. Ellis, editors, Proceedings of the Twenty-Fourth Text Retrieval Conference (TREC 2015). National Institute for Standards and Technology, NIST Special Publication 500-319, 2016.

[2] A. Dean-Hall, C. L. A. Clarke, J. Kamps, P. Thomas, and E. M. Voorhees. Overview of the TREC 2012 contextual suggestion track. In E. M. Voorhees and L. P. Buckland, editors, The Twenty-First Text Retrieval Conference Proceedings (TREC 2012). National Institute for Standards and Technology: NIST Special Publication 500-298, 2013.

[3] A. Dean-Hall, C. L. A. Clarke, J. Kamps, P. Thomas, and E. M. Voorhees. Overview of the TREC 2014 contextual suggestion track. In E.M. Voorhees and A. Ellis, editors, Proceedings of the Twenty-Third Text REtrieval Conference (TREC 2014). National Institute for Standards and Technology, NIST Special Publication 500-308, 2015.

[4] A. Dean-Hall, C. L. A. Clarke, J. Kamps, P. Thomas, N. Simon, and E. M. Voorhees. Overview of the TREC 2013 contextual suggestion track. In E. M. Voorhees, editor, The Twenty-Second Text Retrieval Conference Proceedings (TREC 2013). National Institute for Standards and Technology. NIST Special Publication 500-302, 2014.

[5] A. Strehl, J. Ghosh, Cluster ensembles —a knowledge reuse framework for combining multiple partitions, J. Mach. Learn. Res. 3 (2003) 583–617.

[6] Adeli, H., & Jiang, X. (2009). Intelligent infrastructure: neural networks wavelets, and chaos theory for intelligent transportation systems and smart structures. CRC Press.

[7] Ben Sassi, I., Mellouli, S. and Ben Yahia, S. (2017). Context-aware recommender systems in mobile environment: On the road of future research. Information Systems, 72, pp.27-61.

[8] C.D. Manning , P. Raghavan , H. Schütze , Introduction to Information Retrieval, Cambridge University Press, New York, NY, USA, 2008.

[9] Caragliu, A., Del Bo, C., & Nijkamp, P. (2011). Smart cities in Europe. Journal of Urban Technology, 18(2), 65–82.

[10] Chen, M., Mao, S., & Liu, Y. (2014). Big data: a survey. Mobile Networks and Applications, 19(2), 171–209.

[11] Chen, Song, Li, & Shen J. (2009). Survey on smart grid technology. Power System Technology, 33(8), 1–7.

[12] Chourabi, H., Nam, T., Walker, S., Gil Garcia, J. R., Mellouli, S., Nahon, K., & Scholl, H.J. (2012). Understanding smart cities: an integrative framework. Paper presented at the 2012 45th Hawaii International Conference on System Science (HICSS).

[13] Demirkan, H. (2013). A smart healthcare systems framework. It Professional, 15(5), 38–45.

[14] Gani, A., Siddiqa, A., Shamshirband, S., &Hanum, F. (2016). A survey on indexing techniques for big data: taxonomy and performance evaluation. Knowledge and Information Systems, 46(2), 241–284.

[15] I.B. Sassi , S.B. Yahia , S. Mellouli , User-based context modeling for music recommender systems, in: Proceedings of the 23rd International Symposium on Methodologies for Intelligent Systems - Foundations of Intelligent Systems, in: Lecture Notes in Computer Science, 10352, Springer, Warsaw, Poland, 2017, pp. 157–167.

[16] Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., &Gani, A. (2014). Big data: survey, technologies, opportunities, and challenges. The Scientific World Journal, 2014, 8. http://dx.doi.org/10.1155/2014/712826.

[17] S. H. Hashemi, C. L. A. Clarke, J. Kamps, J. Kiseleva, and E. Voorhees. Overview of the TREC 2016 contextual suggestion track. In Proceedings of TREC 2016, 2016.

[18] Shadab, B V Babu, (2016), Information Retrieval in Big Data Using Evolutionary Computation: A Survey, International Conference on Computing, Communication and Automation (ICCCA2016) 208-213.

[19] T. Hussein, T. Linder, W. Gaulke, J. Ziegler. Hybreed: a software framework for developing context-aware hybrid recommender systems, User Model. User-Adapted Interact. 24 (1-2) (2014) 121–174.

[20] W. Rand, Objective criteria for the evaluation of clustering methods, J. Am. Stat. Assoc. 66 (336) (1971) 846–850.