

Hybrid recommendation system for beauty & spas salons in Yelp

Qianwei Lin^{1,3} and Yuxin Zhang²

¹School of Mathematics, University of Edinburgh, Edinburgh, EH9 3FD, United Kingdom

²Department of Materials, Imperial College London, London, SW7 2AZ, United Kingdom

³qianweilin96@163.com

Abstract. The rising demand for beauty and healthy lifestyles makes the global beauty and spas salons market promising. How to choose appropriate salons for a specific customer becomes an issue and is underexplored. In this paper, we built two types of personalized recommendation system models, the pure model-based collaborative filtering model and the LightFM model (a kind of hybrid recommendation system model) to make recommendations for beauty and spas salons based on Yelp Dataset. The results showed the LightFM model had a better performance. Besides, by applying aspect-based sentimental analysis to extract new features from customer reviews, we further improved the performance of the LightFM model.

Keyword: hybrid recommendation system, aspect-based sentiment analysis, LightFM model.

1. Introduction

The word 'information overload' was created by Bertram Gross [1]. It is a situation when the amount of input information exceeds the system's processing capacity. To avoid overloading, a recommender system (RS), as a kind of information filtering system, was proposed and has been successfully applied in various fields, from social media platforms to online stores. Some websites have made such extensive functions as YouTube's video suggestions, Amazon's purchase recommendations, Netflix's movie recommendations, and LinkedIn's connection recommendations [2].

There are two common types of recommendation systems: content-based (CB) and collaborative filtering (CF). The CB method uses item metadata (item profiles) and users' historical ratings to build user profiles. It then compares the specific user profile with item profiles of all items to select the top k best-match items to make recommendations for each user. In this model, every user is regarded as a separate object. Differently, In the CF method, the behaviors of users are connected. This method uses the item-user interaction rating matrix and is generally classified into two types: memory-based and model-based. Both types predict the unrated items in the interaction matrix. The former uses the k-nearest neighbor (KNN) technique to find similar users or items by analyzing interaction rating data and then uses ratings of similar users or items to predict the unrated items. The latter uses a matrix factorization method to decompose the interaction rating matrix. After factorization, every user and item can be described as a latent vector (embedding). The predicted rating of item i was given by user u is

then the dot product of embeddings of the corresponding user and item. After rating prediction, the CF method selects the top k items with the highest values of ratings to a specific user and recommends them [3, 4].

There are two common issues with RS, data sparsity and the cold-start problem. Data sparsity refers to many unrated items of users in RS models affecting the accuracy and validity of predictions. CF model usually suffers serious data sparsity problems since it utilizes not only the rating information of a specific user but also other users. While CB model handles this problem to some extent as it only needs the specific user's rating information. Another issue for RS is the cold-start problem. It indicates the inability of RS to make accurate recommendations for new users or items since new users or items provide no rating information. CF model still suffers a severe cold-start problem since the values of a new row (representing a new item) or a new column (representing a new user) in the interaction matrix are totally empty. A common method is to fill empty values with specific values such as 0, but this leads to serious bias. CB model mitigates this problem to some extent, especially for new item conditions since metadata of the new item can be used to compare with user profile and then make accurate recommendations. Although the CB model performs better in the above two issues, this method has its own limitations. The most concerning one is that it only focuses on users' existing interests and cannot discover new types of items for a specific user. Since users are isolated, rating information from other users cannot be used to find specific users' potential interests [5].

To handle the above issues at the same time, a series of hybrid recommendation systems with a combination of various techniques and algorithms were proposed to get more accurate results [6-9]. A typical one is to combine CB and CF methods with or without other techniques among the hybrid filtering method. Many of these kinds of methods apply the weighting strategy, i.e., separately calculate the CB and the CF result and then use various weighting methods to put weight to combine them [10, 11].

Differently, the LightFM model, first proposed by Maciej Kula [12], integrates the idea of the CB and CF model into one part by using both item metadata and the interaction matrix. Besides, user metadata is allowed to be incorporated to improve the model efficiency. According to Maciej Kula [12], under cold start situations, this model performs at least as well as the content-based models. When there exists an ample amount of rating data (hot-start condition), it performs at least as well as the traditional matrix factorization CF model. LightFM also decomposes the user-item interaction rating matrix. Its core idea is to represent each feature as a d length embedding/ Latent vector. The input variables of this model are all Indicator variables (0 or 1), so each user and item is represented as the sum of their attributes.

Yelp is a typical example of user-generated media (UGM), like Facebook, YouTube, and Twitter. It provides a platform for customers to post their reviews and rate different products or services according to their gratification. At the same time, customers could find potentially liked products or services on Yelp [13]. So RS is necessary for Yelp to filter out appropriate products or services to specific customers. Among all categories of businesses in it, the focus of this paper is the Beauty & Spas category, indicating both beauty salons and spas salons. With the rising demand for people to pursue beauty and find ways to stay healthy or relax from busy lives, this market has vast potential currently and in the near future. According to the global spas and beauty salons market report [14], the global spas & beauty salons market is expected to increase from 143,829.43 million USD dollars in 2020 to 182,334.91 million USD dollar by the end of 2025. Besides, existing recommendation systems have been applied to a series of products, such as movies, books, restaurants, hotels, and videos, but almost no attention has been paid to the recommendations for beauty & spas salons. Based on this situation, we decided to build RS for the Beauty & Spas category in Yelp.

In addition, as there are tremendous reviews in Yelp that contain rich information about customers' attitudes, aspect-based sentiment analysis (ABSA) is introduced. It is a technique that leverages the natural language processing (NLP) technique to analyze unstructured text-based data and identifies the sentimental orientation of customers in some specific aspects about items such as location, price, and service attitude [15, 16].

In this paper, we firstly cleaned up data, chose relevant features, and made transformations if necessary. Then new features were extracted from customer reviews by ABSA technique as new item metadata. Next, we applied three models: pure model-based CF model, LightFM model without new extracted features, and LightFM model with new extracted features and then compared their performance. For the best performance model, we used it to make rating predictions for beauty & spas salons and selected those with high predicted ratings to recommend. The proposed process is shown in Figure 1.

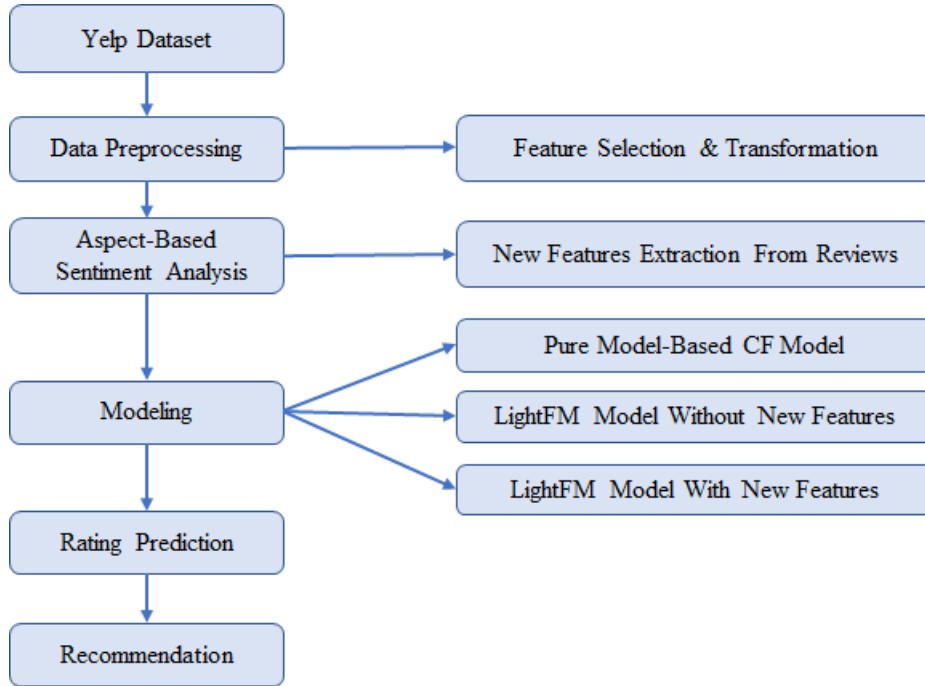


Figure 1. The proposed process of this paper.

Another contribution of this paper is that most recommendation systems do recommendations for a movie, hotel, music, restaurant, app, and almost nothing for Beauty & spas domain. In this paper, based on the characteristics of the Beauty & spas domain, we use the Aspect-based Sentimental analysis method to introduce two more abstract new features (professionalism and service attitude) as new information to improve the model effect. The construction of a recommendation system for the Beauty & spas domain is explored.

2. Data

The dataset is provided by the Yelp Dataset Challenge between the periods of 2010 to 2018. It includes information about different categories of businesses, customers, and their reviews across 11 metropolitan areas spanning four countries. Out of the 209393 businesses, the top 5 categories are Restaurants, Shopping, Food, Home Services, and Beauty & Spas (Table 1). Among all categories of businesses, Beauty & Spas, as a category of popular business, is the focus of this research. In addition, Las Vegas has the highest concentration of beauty & spas salons, so it is the focused location of this research.

Then this paper aims to build RS for beauty & spas salons based in Las Vegas. We firstly filtered out all closed beauty & spas salons as they were not meant to be recommended to customers. Besides, in terms of features, all features are structured except reviews (reviews of all customers). For structured data, only relevant features about RS were selected and classified into three types: business/item information, customer/user information, and item-user interaction information (Table 2). Here feature latitude and longitude were just used for plotting maps since all values of these two features were similar because all salons are in Las Vegas. After feature selection, some of these features were transformed as

necessary. Specifically, original feature date (the time when customers wrote reviews) were converted into four features (i.e., frequency_early_morning, frequency_late_morning, frequency_late_morning and frequency_late_morning) indicating frequency of customers wrote reviews in respective four periods: 00:00 - 05:59, 06:00 - 11:59, 12:00 - 17:59 and 18:00 - 23:59 to show customers' preference about time.

Table 1. Top 5 business categories in Yelp dataset.

Category	Count
Restaurants	63944
Shopping	34644
Food	32991
Home Service	22487
Beauty & Spas	20520

Table 2: Summary of features in Yelp dataset.

Type	Feature Name	Definition
Business/Item Information	business_id	The ID of the salon
	business_name	The name of the salon
	location_latitude	The latitude where the salon is located
	location_longitude	The longitude where the salon is located
	business_review_count	The number of reviews of the salon
	business_star	The average rating of the salon (from 1 to 5)
	professionality	The average VADER score of professionalism of the salon
Customer/User Information	service_attitude	The average VADER score of service attitude of the salon
	customer_id	The ID of the customer
	customer_star	The average rating of the customer (from 1 to 5)
	frequency_early_morning	The frequency of the customer writing reviews in the period of 00:00 - 05:59
	frequency_late_morning	The frequency of the customer writing reviews in the period of 06:00 - 11:59
	frequency_afternoon	The frequency of the customer writing reviews in the period of 12:00 - 17:59
	frequency_night	The frequency of the customer writing reviews in the period of 18:00 - 23:59
Item-User Interaction Information	review_id	The ID of specific review given by specific customer to specific salon
	review_star	The review rating of specific customer to specific salon (from 1 to 5)

Note: This table shows three types of used features in the Yelp dataset, their names, and definitions.

For unstructured review data, two new features professionalism and service_attitude that belong to the business information type were extracted from them by ABSA technique (Table 2). By achieving these, we further filtered out salons with less than 20 reviews which may not provide enough information. Besides, if one customer wrote reviews for the same salon more than once, the values of feature review_star were averaged. Finally, information about 1364 beauty & spas salons corresponding to 137823 reviews commented on by 100122 unique customers was used in this research.

3. Methodology

In this section, we gave a detailed description of the models and techniques applied in this paper.

3.1. *Aspect-based sentiment analysis (ABSA)*

In this part, an unsupervised ABSA technique was introduced to analyze all customer reviews to extract new features. Considering the beauty and spas salons are generally service-based rather than product-based, many customers concern about the service quality. In this sense, the professionalism of service provided by professionals (e.g., beauticians and stylists) and the service attitudes of staff are two important aspects to influence service quality. Based on these, two new features, professionalism, and service_attitude were created to reflect the sentimental orientations of customers in these two aspects. Especially, the features extraction process is divided into three steps: 1. Build representative word set of each aspect by applying the Word2Vec word embedding method. 2. Find all adverb - adjective-noun pairs in specific aspects. 3. VADER (Valence Aware Dictionary and Sentiment Reasoner) Dictionary is used to calculate the emotional orientation score of phrases from step 2 respectively and then aggregate the final scores as new features.

3.1.1. Word2Vec word embedding method. Before applying the Word2Vec model, raw review data were firstly preprocessed to be ready for this model. This process includes common NLP steps: 1. text tokenization: segment texts into individual tokens (words/punctuations). 2. Remove non-alphabetical tokens and stopwords. 3. word lemmatization: transform words to their base forms. 4. Apply the bigram model to connect specific individual words to two-word phrases.

Word2Vec method is a two-layer neural networks model with two categories: Skip Gram and Common Bag Of Words (CBOW). In this paper, the CBOW method was adopted. Its main idea is to take surrounding words of the centered word as input to predict the centered word. During the process, the original representations of words (one-hot encoded vectors) are mapped to lower-dimension vectors in the hidden layer that are essentially word embeddings reflecting words' meaning. In this way, two words with similar syntactic and semantic relationships have similar embeddings. Then given some initial words, we can find their synonyms by calculating similarities between words. Finally, the representative word sets were built with initial words and their synonyms [17].

3.1.2. Detection of necessary phrases. After getting representative word sets, we tried to find all adverb-adjective-noun pairs (adverbs can be existed or not) including words in the representative word sets. Here Spacy library [18] in Python was the primary tool. Firstly, the Part-of-speech (POS) Tagging technique was applied to label word types such as noun, adj, adv, and verb to every word for every sentence in each review. Next, dependency parsing (DP) was used to parse every sentence by assigning syntactic dependency labels such as nsubj (indicating nominal subject), acomp (indicating adjectival complement), advmod (indicating adverbial modifier) and neg (indicating negation modifier) to related pairs of words to reflect their relationship. Then looped through all words to find targets that are in representative word sets and their corresponding adverb-adjective-noun pairs. Besides, negation words such as not in raw sentences would also be extracted.

3.1.3. Context-aware VADER dictionary. VADER dictionary is a rule-based sentiment analysis tool to calculate the score of the individual word and their combinations (phrases, sentences, paragraphs). For an individual word, its score (ranging from -4 to 4) includes polarity (positive, neutral, or negative) and intensity (ranging from 0 to 4) components. For their combinations, what we need is the compound score (ranging from -1 to 1) calculated by the summing of all words, adjusting by rules, and then doing normalization [19, 20]. It is appropriate to use this dictionary to calculate the orientation score of phrases in a specific aspect because of its two advantages: 1. It identifies the degree of adverbs to increase or reduce the overall intensity of phrases. 2. It considers negation context.

Besides, to make the VADER dictionary more sensitive to our context – the professionalism and service attitude in beauty & spas salons, the top 100 high-frequency adjective words in each aspect were selected and then checked whether their scores should be modified or not. For example, 'skilled' in the original VADER dictionary is a neutral word with no sentimental orientation, but in the professionalism context, it should be a positive word, so its score was adjusted to +3. After calculating the sentimental

orientation scores of phrases, all of them relating to the same salon in specific aspects were aggregated to calculate average scores that were regarded as values of new features. In this way, two new features professionalism and service_attitude were created.

3.2. *LightFM model*

The LightFM model, FM referring to matrix factorization, is a variant of the matrix factorization model. In the traditional matrix factorization method of CF models, every user and item is represented by a latent factor (embedding) directly. In contrast, the LightFM method estimates a latent vector for every useful feature and item feature, then every user and item is described as linear combinations of corresponding feature embeddings. For example, a nail salon is described by features Beauty & Spa, location, and open time, then the embedding of this nail salon is the sum of embeddings of features Beauty & Spas, location, and open time. Besides, like the CB model in which users and items are described by functions of item metadata/features (i.e., item profile and user profile), items and users in the LightFM model are represented by functions of embeddings of item metadata and user metadata. In this sense, the LightFM model is also a variant of the CB model. So the LightFM is essentially a hybrid model of the CB and CF method [12].

The formal representation of the LightFM model is shown in the following 5 formulas. In its setting, the value in the item-user interaction matrix is 1(positive interaction) if the user buys the corresponding item and 0 (negative interaction) otherwise. In our case, star ratings ranging from 1 to 5 are given as weight to reflect users' preferences for items. Besides, in the model setting, e_f indicates the latent vector (embedding) of feature f with k -dimension. It belongs to either user feature set F_u or item feature set F_i . The specific user u is described by embedding m_u , which is a sum of corresponding user feature embeddings. In the same way, item i is represented by n_i , which is a sum of item feature embeddings. Besides, the overall bias terms β_u and β_i for user u and item i respectively are introduced as the summation of corresponding individual feature bias b_f .

$$m_u = \sum_{f \in F_u} e_f \quad (1)$$

$$n_i = \sum_{f \in F_i} e_f \quad (2)$$

$$\beta_u = \sum_{f \in F_u} b_f \quad (3)$$

$$\beta_i = \sum_{f \in F_i} b_f \quad (4)$$

Then the rating prediction $\hat{r}(u, i)$ for item i given by user u can be generated (Formula 5) by the dot product of the specific user and item embedding plus corresponding bias terms. In our research, the user indicates the customer and the item indicates the beauty & spas salons.

$$\hat{r}(u, i) = m_u \cdot n_i + \beta_u + \beta_i \quad (5)$$

Next, after the setting of the LightFM model, the asynchronous stochastic gradient method was then used to estimate all parameters of this model on 80% training data. After model training, the performance of the LightFM model was evaluated on 20 % test data by two metrics: precision@10 and Area under Curve (AUC). These two metrics show different aspects of model performance. For precision@10, it only evaluates the performance of the top 10 items with the highest predicted ratings. Specially, it calculates the proportion of positive interaction items in all top 10 items ranging from 0 to 1. Its maximum achievable value is based on the actual data. For example, for a user with only one actual purchase, the maximum precision@10 value is 0.1. Differently, AUC evaluates the performance of all items by calculating the probability of the predicted ratings of a randomly chosen positive interaction item higher than a negative interaction item. Its value range is also between 0 and 1, but its maximum achievable value is 1.

In our research, precision@10 and AUC can be estimated for each user, then the overall performance of this model is calculated by averaging precision@10 and AUC across all users. Besides, precision@10 is the primary metric since in RS the most important thing is to discover what users like rather than they

do not like, so the performance of items with low predicted ratings in RS setting do not matter. In this sense, the overall performance of all items is less important than the top 10 items. So AUC is just a complementary metric to reflect the overall performance of all items.

Finally, after LightFM RS is built, it was used to recommend appropriate beauty & spas salons to specific customers. Specially, for a specific customer whether existing or new, the top ten (or more) most appropriate salons will be selected and recommended. Suppose this customer is new entering into RS, we would use average values of user features as user metadata of this customer together with original item information and interaction information to make recommendations. If this customer provides any information about user features, this new information can be integrated into the model to improve model accuracy directly. Another condition is when a new item enters into RS, its metadata could be collected together with original user information and interaction information to make recommendations.

3.3. Optimization

In the process of optimization, we used the ASGD (asynchronous stochastic gradient descent), by Dean et al. in 2012. It consists of two main ingredients. First, the parameters of the model can be distributed on multiple servers, depending on the architecture. This set of servers are called the parameter servers. Second, there can be multiple workers processing data in parallel and communicating with the parameter servers. Each worker processes a mini-batch of data independently of the other ones.

4. Result

Firstly, all customer reviews cleaned by the bigram model were fed into the Word2Vec model to train our own dictionary. After experiments, important parameters of this model included: the context window (the number of surrounding words), the dimension of embedding, and the training epochs were 5, 200, 30 respectively. Then we set some initial words for the professionalism aspect and service_attitude aspect respectively (Table 3) and used a built dictionary to find their synonyms. For example, for an initial word 'trained', its synonyms (most similar word groups) included license, licensed, competent, well_trained, qualified, skilled, knowledgeable, capable, certify, and certified. In this way, representative word sets were formed.

Next, the Spacy library was applied to do Part-Of-Speech (POS) tagging for every token and then dependency parsing (DP) for every sentence. For example, the sentence 'This salon offers many awesome services.' was parsed (Figure 2). Further, we looped through all words. If any of them was in representative word sets and had matched pairs (i.e., adverb-adjective-noun pairs with or without adverb), then these matched pairs would be extracted. In the above example, the word 'services' was in the representative word set of service_attitude, so the specific pair 'many awesome services' would be extracted.

Table 3. Initial words of professionalism and service_attitude.

Feature	Initial words
professionalism	tech, technique, technician, skill, masterful, professional, unprofessional, skilled, unskilled, experienced, unexperienced, trained, untrained
service_attitude	attitude, manner, customer_service, friendly, unfriendly, polite, impolite, respectful, disrespectful, welcoming, unwelcoming, patient, impatient, enthusiastic, unenthusiastic, kind, rude

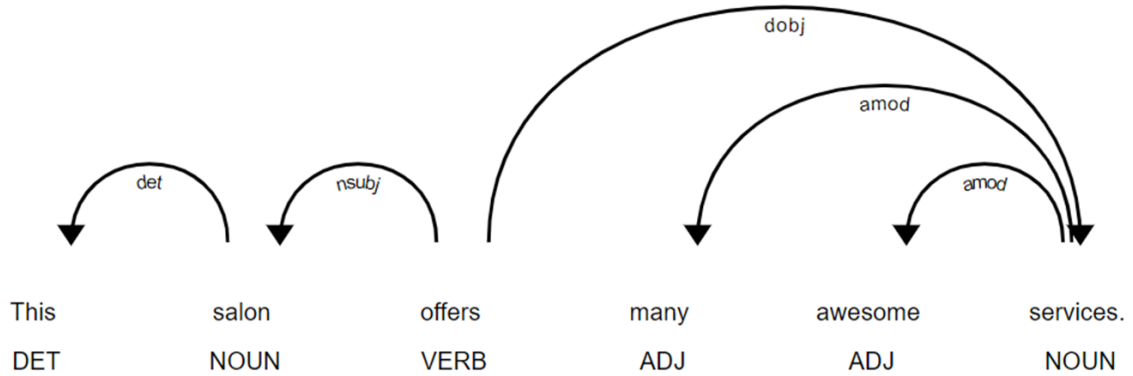


Figure 2. Visualization of POS and DP.

After extracting all relevant phrases, we then used the VADER dictionary to calculate the sentimental score. In the process, we firstly checked the top 100 frequent adjective words and their VADER score. Then based on the specific context, we adjusted the polarity and intensity of some adjective words. After modification, the VADER dictionary was applied to every phrase, and scores of the same salons of specific aspects were averaged to obtain the final score. Word clouds of the top 100 frequency adjective words and distributions of final scores in two aspects are shown in Figure 3 and Figure 4 respectively.

From Figure 4, it is easy to know major customers tended to give relatively positive feedbacks to both professionalism and service attitude aspects, but there were still existed some negative feedbacks.

After two features were created, then we fitted three RS models (i.e., pure model-based CF model, LightFM model without new extracted features and LightFM model with new extracted features) on 80% train dataset and evaluated their performances on 20% test dataset. During this process, important parameters were set based on experiments. Specifically, for loss function, the performance of Weighted Approximate-Rank Pairwise (WARP) loss was better than other loss functions (Logistic loss and Bayesian Personalized Ranking loss) in all three models, so it was selected. For the learning rate, we chose the default setting of 0.05 in all three models. The optimal epochs for the three models were 5, 10, 10 and the dimensions of embedding are 140, 180, and 150 respectively. The optimal parameters and corresponding performances of three RS models are shown in table 4.

From table 4, in all three models, the pure model-based CF model performed worst in both main metric precision₁₀ metric and complementary metric AUC (0.017067 and 0.746409 respectively). By taking item metadata and user metadata, the performance of the LightFM model improved significantly with precision₁₀ metric increasing from 0.017067 to 0.018649 and AUC metric increasing from 0.746409 to 0.781027. Besides, by considering new features professionalism and service_attitude, the performance of the LightFM model with new extracted features further improved in terms of primary metric precision₁₀ metric with increasing from 0.018649 to 0.018721, while the complementary metric AUC experienced a slight decrease from 0.781027 to 0.780791. Since we cared more about the performance of the top ten items rather than all items, the slight drop in AUC performance would be ignored.

Table 4. Summary of three RS models.

Model	Loss Function	Learning Rate	Epoch	Dimension of Embedding	AUC	Precision_at_10
Pure CF Model	warp	0.05	5	140	0.746409	0.017067
LightFM Model_1	warp	0.05	10	90	0.779783	0.018464
LightFM Model_2	warp	0.05	10	150	0.780791	0.018721

Note: LightFM Model_1 and LightFM Model_2 refer to LightFM Model without and with new extracted features respectively.

Table 5. User metadata of a specific customer

Feature Name	Feature Value
Customer_id	2VABkVdFhHGMo7wAqfofGA
Customer_star	3.78
Frequency_early_morning	246
Frequency_early_morning	2
Frequency_late_morning	0
Frequency_afternoon	1
Frequency_evening	2

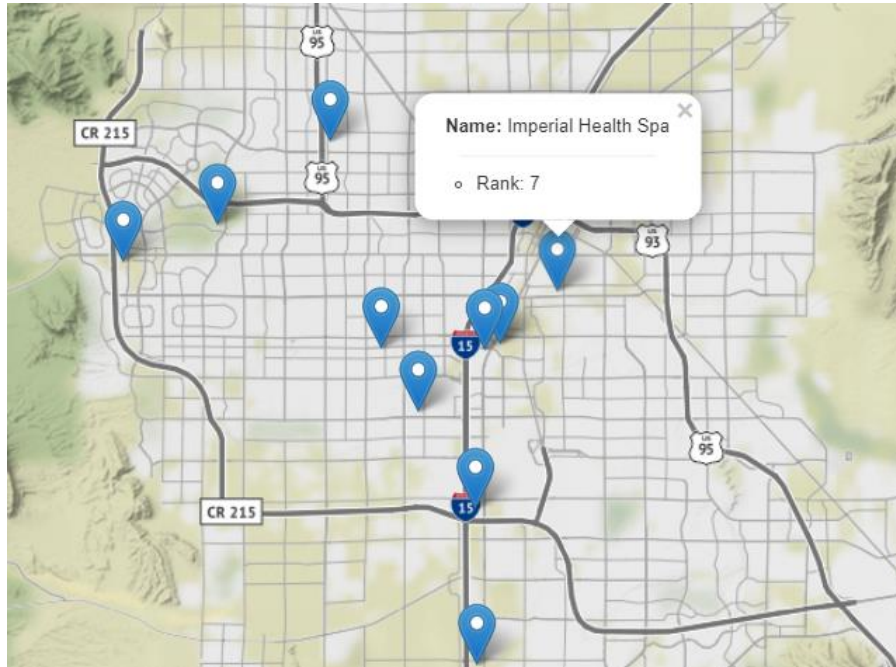


Figure 5. The labelled Las Vegas map of top ten salons for a specific customer.

5. Conclusion and direction

In this paper, we applied three personalized RS models on an underexplored area – beauty & spas salons. The experimental results showed that LightFM models performed much better than the baseline model – a pure model-based CF model by utilizing both item and user metadata. Besides, by introducing the

aspect-based sentiment (ABSA) technique to extract two new features professionalism and service attitude as additional item metadata, the performance of the LightFM model was further improved.

In this paper, limited by computing power, we focused on beauty & spas salons only in one city – Las Vegas. However, theoretically current RS can be generalized to include other cities if necessary. Besides, more metadata would improve the model performance, especially user metadata such as gender and age currently unavailable. Besides, more new features relating to beauty & spas salons may be further explored from customer reviews. At the same time, with the development of NLP, some new techniques such as the deep learning method may be further considered to improve the performance of ABSA. Further, some time-dependent interactions may be factored into the LightFM model to improve model performance.

References

- [1] Gross, B. M. (1964). The managing of organizations: The administrative struggle (Vol. 2). Free Press of Glencoe.
- [2] Kouvaris, P., Pirogova, E., Sanadhya, H., Asuncion, A., & Rajagopal, A. (2018). Text enhanced recommendation system model based on yelp reviews. *SMU Data Science Review*, 1(3), 8.
- [3] Balabanović, M., & Shoham, Y. (1997). Fab: Content-based, collaborative recommendation. *Communications of the ACM*, 40(3), 66-72.
- [4] Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009.
- [5] Kumar, P., & Thakur, R. S. (2018). Recommendation system techniques and related issues: a survey. *International Journal of Information Technology*, 10 (4), 495-501.
- [6] Çano, E. and Morisio, M. (2017) 'Hybrid recommender systems: A systematic literature review', *Intelligent Data Analysis*, 21, pp. 1487–1524. doi: 10.3233/IDA-163209.
- [7] Anwaar, F., Iltaf, N., Afzal, H., & Nawaz, R. (2018). HRS-CE: A hybrid framework to integrate content embeddings in recommender systems for cold start items. *Journal of computational science*, 29, 9-18.
- [8] Walek, B. and Fojtik, V. (2020). A hybrid recommender system for recommending relevant movies using an expert system, *Expert Systems with Applications*. Elsevier Ltd, 158, p. 113452. doi: 10.1016/j.eswa.2020.113452.
- [9] Tahmasebi, F. et al. (2020). A hybrid recommendation system based on profile expansion technique to alleviate cold start problem, *Multimedia Tools and Applications*. Springer, 80(2), pp. 2339–2354. doi: 10.1007/s11042-020-09768-8.
- [10] Lin, W., Li, Y., Feng, S., & Wang, Y. (2014). The optimization of weights in weighted hybrid recommendation algorithm. In *2014 IEEE/ACIS 13th International Conference on Computer and Information Science (ICIS)* (pp. 415-418). IEEE.
- [11] Wang, Y., Wang, M., & Xu, W. (2018). A sentiment-enhanced hybrid recommender system for movie recommendation: a big data analytics framework. *Wireless Communications and Mobile Computing*, 2018.
- [12] Kula, M. (2015). Metadata embeddings for user and item cold-start recommendations. *arXiv preprint arXiv:1507.08439*.
- [13] Hicks, A., Comp, S., Horovitz, J., Hovarter, M., Miki, M., & Bevan, J. L. (2012). Why people use Yelp. com: An exploration of uses and gratifications. *Computers in Human Behavior*, 28(6), 2274-2279.
- [14] Research and Markets Ltd (2021) Spas & Beauty Salons Market Research Report by Type (Beauty Salon, Merchandise Sales, and Spa) - Global Forecast to 2025 - Cumulative Impact of COVID-19. Retrieved from: <https://www.researchandmarkets.com/reports/4968912/spas-and-beauty-salons-market-research-report-by>.
- [15] Pavlopoulos, I. (2014). Aspect based sentiment analysis. Athens University of Economics and Business.
- [16] Madhoushi, Z., Hamdan, A. R., & Zainudin, S. (2019). Aspect-based sentiment analysis methods

- in recent years. *Asia-Pacific J. Inf. Technol. Multimedia*, 8(1), 79-96.
- [17] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. arXiv preprint arXiv:1310.4546.
 - [18] Honnibal, M., & Johnson, M. (2015). An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1373-1378).
 - [19] Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 8, No. 1).
 - [20] Mowlaei, M. E., Abadeh, M. S., & Keshavarz, H. (2020). Aspect-based sentiment analysis using adaptive aspect-based lexicons. *Expert Systems with Applications*, 148, 113234.
 - [21] Python-visualization. (2020). Folium. Retrieved from <https://python-visualization.github.io/folium>.