Analysis of a recommendation system used for predicting medical services

Gjorgji Noveski
gjorgi.noveski@ijs.si
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia

Jakob Valič
jakob.valic@ijs.si
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia

ABSTRACT

Recommendation systems are widely used in prediction of user preferences given a set of data. More often these preferences are part of the domains of the entertainment or general goods industry, striving to improve recommendations for items such as songs, movies, electronics, etc. In this work, we give an overview of using a recommendation system in a electronic and mobile health platform, showcasing the applicability of such a system for recommending healthcare services and keywords relating to user's initial search query.

KEYWORDS

Electronic and mobile health, recommendation systems, Insieme

1 INTRODUCTION

Recommendations systems have been proven to provide solid recommendations in various tasks, such as movie recommendations, song recommendations, general goods recommendations on online shopping platforms, etc. With their help, using online platforms and services has become more enjoyable since they tailor suggestions in respect to each individual user. Focusing on the healthcare domain, we can see that recommender systems are able to offer help in many aspect of a patient's health [3]. Similarly, they are also used by the healthcare professionals, aiding them in decision making scenarios in order to decrease the risk of errors. An interesting approach is proposed by [2], where the entirety of the recommendation system is build as a whole platform. Their platform takes into consideration the user's health status and finds healthcare services which it considers would be of value to the user. Our work is interested in analyzing a recommender system in conjunction with an electronic and mobile health (EMH) platform. We want to recommend relevant healthcare services to the users based on their search queries. Having good recommendations allows the user to find relevant services in a faster and easier way.

Problems with recommendation systems

A recommendation system has the most difficult time achieving good results at the beginning of its work period. This is widely known as the "cold start" problem. This problem is due to the fact that a system can not infer any significant information about its users or items because previous information is scarce. There are several approaches one can take while designing a recommendation system, such as:

Collaborative filtering

Collaborative filtering is based on the idea that users who have similar interest in the past, will be more likely to rate future items the same way. The recommendation is gathered from data from various users, but it is tailored just for the specific user that is doing the query.

Item-item collaborative filtering

Another form of collaborative filtering is the item-item collaborative filtering. Instead of looking at users which are similar based on their previous ratings, it looks at the items that are rated by one user and it suggest a new item which has the highest similarity between the previously rated ones. This form of filtering performs well when the number of users is higher than the numbers of items, which is a good fit for our problem.

Insieme

The environment where the recommendation system will be analyzed is called ISE-EMH (Insieme). It represents an EMH platform which connects various medical institutions and patients. The platform provides information about services which are obtained from medical institutions. An example use case is a patient that requires information about certain illnesses, queries information through keywords on the platform and the platform returns the relevant information.

The users who are using Insieme are not required to have an account to use its services. Because of that, a session based recommendation system is used, meaning the user's queries (searched keywords) are only relevant in only one session. The user may be able to search for different illnesses in different sessions, but we can not assume their previous searches are related to the current one.

2 EXPERIMENT

In order to carry out the analysis of the recommendation model for project Insieme, we decided to simulate a small set of input data due to lack of real data. First, we chose a subset of services and then we simulated some users' choices. The simulation was done by generating use cases which simulate a typical user using the EMH platform and choosing appropriate services. This was done because real user interaction data was unobtainable.

Input data

Insieme services are organized into medical categories. For the purpose of the experiment, we chose the following medical categories: dermatology, oncology and infections. From each of those we chose 3 to 4 services, thus obtaining 10 various medical services. Each service has keywords describing it. Since the services refer to the diseases, the keywords refer to the affected part of the body and describe some additional properties of the disease.

Information Society 2021, 4–8 October 2021, Ljubljana, Slovenia

services \ keywords	dermatology	oncology	infections	skin	brain	lungs	vaccination	cronic illness
pneumonia			+			+		
acne	+			+				
psoriasis	+			+				+
dermatitis	+			+				+
skin cancer		+		+				
lung cancer		+				+		
brain tumor		+			+			
Lyme disease			+	+				+
tick-borne meningoencephalitis			+		+		+	+
COVID-19			+			+	+	

Table 1: A subset of services with corresponding keywords

services \ users	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
pneumonia	+					+							+								
acne				+				+			+				+	+					+
psoriasis		+						+	+	+							+				
dermatitis		+								+									+		
skin cancer			+								+				+			+		+	
lung cancer	+		+									+	+					+		+	
brain tumor			+				+													+	
Lyme disease				+				+					+	+	+						
tick-borne meningoencephalitis							+						+	+					+	+	
COVID-19					+	+						+	+			+	+				

Table 2: Users' choices of services

When choosing the subset of services, we carefully selected those with various keyword intersections. The services and keywords are shown in Table 1. The '+' sign denotes which keywords are associated with a service.

Next, we have simulated the interactions between users and services. The '+' sign denotes that the user has chosen the service. The users' choices are in Table 2. When preparing the data, we bore in mind the possible reasons to some users' choices. E. g., users 1, 7 and 10 might be concerned about some skin problems, while user 11 might be concerned about the lung problems and user 2 might be investigating all information about cancer available. However, in order to make data more realistic, the majority of users' choices don't have agenda.

Recommendation model

Our choice for a recommendation system is LightFM [1]. The reason for using this implementation is the ability to create tag embeddings by supplying user and item features. In our use case, our item features represent tags that further explain the keywords (items), for e.g. : "Acne" item has the corresponding "Dermatology" and "Skin" items associated with it. The benefit of using embeddings is that they capture semantic similarities between the keywords, which in turn will result in better inference of the model and provide an option to choose top N most similar keywords.

With the help of the learned latent vectors improvement is achieved on "cold start" scenarios. If the item features were not supplied, the model would default back to a pure collaborative filtering model. In addition, an implicit feedback model is used that regards the absence of information in the interaction matrix as negative feedback. The motive for this is that a user already made a conscious choice about what kind of services he needs information for, so the keywords that the user didn't search can be regarded as negative interactions in the interaction matrix.

We built the recommendation model using the LightFM library [1]. We compared the outcome of two various models. The first model is trained on the users' choices of the services only. The second model is additionally trained on keywords describing the service. Using these two models, we obtained the suggestions for each user.

Results

The top recommendations for all users are mostly the same. The differences between the recommended services are minimal, e. g. for user 2:

- pneumonia: 0.13
- acne: 0.18
- psoriasis: 0.13
- dermatitis: 0.15
- skin cancer: 0.18
- lung cancer: 0.18
- brain tumor: 0.13
- Lyme disease: 0.17
- tick-borne meningoencephalitis: 0.17
- COVID-19: 0.16

This data is produced by the model with only interactions. The model with item features has the same recommendations, with slightly lower probability. For each user, the top 3 suggested services are acne, skin cancer and lung cancer. These predictions are not satisfactory, since we would like to obtain the suggestions tailored to every user separately. We assume more data would be needed for training of recommendation models.

3 CONCLUSION

In this work we analyzed a recommendation system that is used with an EMH platform. The goal was to see if such a system is applicable on an EMH platform and offer medical service recommendations to users. Because of limited amount of interaction data, the recommendation system faces difficulties in learning meaningful representations. The system requires data which would be gathered during a longer period of time, in order to give more accurate and meaningful suggestions.

ACKNOWLEDGMENTS

The paper was supported by the ISE-EMH project funded by the program: Interreg V-A Italy-Sovenia 2014-2020. We thank IPM Digital, which organized the focus group as part of the HoCare 2.0 project.

REFERENCES

[1] Maciej Kula. 2015. Metadata embeddings for user and item cold-start recommendations. In *Proceedings of the 2nd Workshop on New Trends on Content-Based Recommender Systems co-located with 9th ACM Conference on Recommender Systems (RecSys 2015), Vienna, Austria, September 16-20, 2015.* (CEUR Workshop Proceedings). Toine Bogers and Marijn Koolen, editors. Volume 1448. CEUR-WS.org, 14–21. http://ceur-ws.org/Vol-1448/paper4.pdf.

[2] Choon-oh Lee, Minkyu Lee, Dongsoo Han, Suntae Jung, and Jaegeol Cho. 2008. A framework for personalized healthcare service recommendation. In *HealthCom 2008-10th International Conference on e-health Networking, Applications and Services.* IEEE, 90–95.

[3] Thi Ngoc Trang Tran, Alexander Felfernig, Christoph Trattner, and Andreas Holzinger. 2021. Recommender systems in the healthcare domain: state-of-the-art and research issues. *Journal of Intelligent Information Systems*, 57, 1, 171–201.