Intelligent Match-making between Patients and Family Doctors

Manas Gaur, Mengxin Ji, Inigo Martinez de Rituerto de Troya, Qiwei Han, Laura Szczuczak,
Data Science for Social Good Fellowship,
Nova School of Business and Economic and The University of Chicago,
{mgaur,mji,qhan,lszczuczak}@novasbe.pt

Abstract
In Portugal, patient's appointments with General Practitioners (GPs) or Internal Medicine doctors (IMs) are typically decided randomly based on doctor's availability. This creates a scenario, where patients fail to develop trust with a health care professional. In particular, trust plays a central role in ensuring the sustained relationship between patients and family physicians as a defining characteristic of the quality of primary care. Higher trust in health care professionals results in beneficial health outcomes and higher satisfaction of treatments. However, understanding the development of trust between patients and their family physicians largely relies on the survey-based measures while seldom examining the actual consultation history. In this study, we aim to build the matchmaking system that recommends family physicians to patients and vice versa, in order to achieve the trusting relationship between them.
We partner with a large private health network in a European country and obtain a large-scale data set with over 70 million consultation history recording the interaction of 1.5 million patients and 3500 doctors across Portugal. The large analytical data set provides information on various episodes between patients and doctors, demographics of doctors and patients. We address the problem in three ways. Firstly, we provide a strategic pipeline involving data analysis and pre-processing, feature engineering and model selection, model evaluation and recommendation. Secondly, compare our results with the random procedures used to allocate doctors to patients. Lastly, we manifest the need of matchmaking between patients and doctors.

Keywords : *Recommender System, Content-Based Filtering, Collaborative Filtering, Temporal Cross Validation*

1. Introduction

In general, people typically see doctors based on their availability or personal recommendation. Can we establish a better way to match patients and doctors? In order to provide a precise matching of doctor to a patient, we need to quantify the relationship between the patient and the doctor. For quantifying the relationship between the patient and doctor, we introduce the term "Trust". In this scenario, [1] trust is defined as, how often do patients return to see the same doctor or they choose another doctor? One of the largest European healthcare network partnered with us to provide a large dataset of 72 million interactions involving 1.5 million patients and 3500 doctors generated over a decade. The dataset provides demographic information of patients (age, gender, locality, profession) and doctors (age, gender, education qualification, seniority, location). We define the problem of matchmaking between patients and doctors as a recommendation problem illustrated using the collaborative and content-based filtering approach. The notion of collaborative filtering takes into account the interaction between patients and doctors. This approach recommends doctors to existing patients. Content-based filtering approach recommends doctors to new patients and patients to the new doctor. Hence, we state our system as Hybrid Recommender System.

3. Trust

Customer churn, follower-followee network in social media, wisdom of crowd and dynamic communities in social media are some of the examples where the notion of "Trust" is modeled [2]. Trust can be defined as the ratio of fidelity of one doctor with fidelity of all doctors multiplied by temporal decay parameter defining recency of patient-doctor interaction.

$$T(p, d, t) = \frac{1}{(1+\delta(t))^{\tau}} * \frac{\phi_{pd}(t)}{\sum_{x=1}^{N_D} \phi_{px}(t)}$$

$T(p, d, t)$: Value of trust between patient "p" and doctor "d" at time (year) "t".

$\phi_{pd}(t)$: Number of episodes that took place between patient "p" and doctor "d" at time (year) "t".

$\sum_{x=1}^{N_D} \phi_{px}(t)$: Sum of the number of episodes that took place between patient "p" and all the doctors .

N_D

$\frac{1}{(1+\delta(t))^{\tau}}$: Temporal decay parameter. τ is the year and $\delta(t)$ is the discount rate [5].

2. Methodology

In this section, we briefly explain the procedure undertaken to demonstrate that machine learning has a positive impact and quantification of trust improves recommendation quality [4].

A. Data Analysis and Preprocessing

The dataset obtained from healthcare providers involved 34 relational tables which were complex, inconsistent, less cohesive and redundant. We cleaned the data and created a workable data frame of 701,000 consultation history between 226,000 patients and 179 general practitioners or internal medicine doctors over the year 2012-2017. We analyzed the behavior of the 226K patients across 179 doctors over the span of 5 years using their episodes as the nearest of their visit to the hospital.

B. Feature Engineering and Model Selection

After the cleaning the data and preliminary analysis, we identified features specific to patients, doctors, and their interaction. After identification of features, we create a one hot encoding table containing patients, doctors, demographic features and interaction features. We selected models based on their ability to model continuous label (Trust) [3]. Models were trained and tested using temporal cross-validation. Furthermore, they are compared with the random Gaussian model are the logistic regression (LR), support vector machine (SVM), decision tree (DT), random forest (RF), K-nearest neighbor (KNN), singular value decomposition (SVD), AdaBoost and XGboost.

C. Model Evaluation

We evaluate the machine learning model condition on their representative power, root means square error and precision-recall at K (K: top K recommendation).

D. Recommendation

We recommend top-k doctors to each existing and new patient using a hybrid recommendation system.

References

[1] O'Donovan, John, and Barry Smyth. "Trust in recommender systems." *Proceedings of the 10th international conference on Intelligent user interfaces*. ACM, 2005.
[2] Davoudi, Anahita, and Mainak Chatterjee. "Modeling trust for rating prediction in recommender systems." *SIAM Workshop on Machine Learning Methods for Recommender Systems, SIAM*. 2016.
[3] Zhang, Tong, and Vijay S. Iyengar. "Recommender systems using linear classifiers." *Journal of Machine Learning Research* 2.Feb (2002): 313-334.
[4] Adomavicius, Gediminas, et al. "De-biasing user preference ratings in recommender systems." *RecSys 2014 Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS 2014)(Foster City, CA, USA, 2014)*. 2014.
[5] Beckerman, Wilfred, and Cameron Hepburn. "Ethics of the discount rate in the Stern Review on the economics of climate change." *WORLD ECONOMICS-HENLEY ON THAMES-* 8.1 (2007): 187.