Data-Driven Approach to Patient Flow Management and Resource Utilization in Urban Medical Facilities

Elizaveta S. Prokofyeva
Faculty of Business and Management
National Research University Higher School of Economics
Moscow, Russia
eprokofeva@hse.ru

Svetlana V. Maltseva
Faculty of Business and Management
National Research University Higher School of Economics
Moscow, Russia
smaltseva@hse.ru

Nikita Y. Fomichev
Data Analysis in Disrtibuted Computing Networks
Infogorod
Moscow, Russia
fomichev.n@gmail.com

Alexey G. Kudryashov
Data Analysis in Disrtibuted Computing Networks
Infogorod
Moscow, Russia
gubkinkudr@gmail.com

***Abstract*— Healthcare services are tightly connected with complex data analysis techniques to enable optimal resource allocation in medical institutions. This paper proposes a detailed analysis of incoming patient flow to local polyclinic by integrating clustering techniques, process mining and a concept of self-organizing systems. The study takes into account concepts based on models of managing social networks, the participants of which today can be both people and intelligent software. How could patient flow model be developed using a clinical pathways approach that combines clinical pathways tool, social media analysis, hierarchical agglomerative clustering method and probabilistic topic modeling to investigate the optimal resource utilization of medical facility? The methodology to answer this research question was demonstrated using a time- series clustering (k-medoids, Ward's method, Latent Dirichlet Allocation, Additive Regularization of Topic Models), Naive Bayes classifier based on public real data of 64668 depersonalized patient- doctor of 32 specialties conversions. In this paper, a modeling methodology for heterogeneous patient flow segmentation is proposed. The presented approaches serve as the foundation for the further development of a queuing system model of a medical institution. In addition, the shared economy principles are applied by the development of such service that would reduce the workload of appointments to therapists by matching patients to needed doctors.**

Keywords— patient flow, resource planning, hierarchical clustering, clinical pathways, topic models, social networks

I. Introduction

Optimizing resource utilization in modern healthcare facilities can lead to more effective work organization: data-driven decision about scheduling policies, patient flow characteristics, planning and distribution of resources, reducing queues and other tasks. Modern analytical technologies allow the development of decision-making tools based on empirical data in healthcare domain. For example, depersonalized data on actual movements of patients between medical units and specialists within these institutions allows management to plan the load of resources, ensure a high level of accessibility of services and optimize the work of the organization based on real demand for these services. Based on such data, patient flow management information systems and various services are being actively developed.

In healthcare domain, modern technologies allow one to accumulate massive volume of both structured and unstructured data to make these optimal and informed decisions. However, not only the volume, but also the diversity of these data types provides new possibilities for indepth analysis using advance data analysis technologies in healthcare [1]. Thus, data from social networks, patient chats, electronic prescriptions, pharmacy reviews, news feeds of social platforms, online consultations and many other data sources become resources of new valuable knowledge to learn more about patient flow and real demand for medical services.

Patient community is a self-organizing system in the conditions of opinions sharing on the Internet, so the online platforms enable to measure the result of the existence of such a self-organizing system.

The global trend of *shared economy* is also significantly influencing healthcare systems [2]. A sharing economy is based on collaborative consumption of society members, where products or services are shared with others to reduce waste [3, 4]. This concept of the *sharing economy* appeared as an emerging business model in the beginning of 2000s [4]. The digital technologies, and specifically social networks, played a large role in occurrence of shared services provided by mobile applications and online platforms [5], bringing people together and matching them to relevant and best suitable services and products they are searching for [5, 6, 7].

In healthcare context, the *shared economy* enables doctors provide more flexible medical services in various facilities, polyclinics, private clinics, and other locations as well, similarly to other popular sharing models, such as Uber, Airbnb and others [8]. The application of such shared approaches promises more efficient resource utilization of medical organizations, and ultimately higher quality of services in more personalized manner.

For example, primary medical care in Russia is provided through government-funded polyclinics. The definition of the polyclinic might include the following statement: local polyclinic is a specialized healthcare facility for different medical services, which do not require hospitalization. Local polyclinics offer various services, which are generally considered as general practice. For instance, blood tests, screening, treatment for chronic illnesses, certificates obtaining

2378-1971/20/$31.00 ©2020 IEEE
DOI 10.1109/CBI49978.2020.10060

and other ones. Patients make appointments by online services of information system called EMIAS ("United Medical Information and Analytical System"), which was launched in Moscow in 2011. EMIAS information system includes booking the hospital visits, EHR (Electronic Health Record) management, attachment the patient's personal record to the polyclinic, and electronic prescription services. There is a number of "first- level" doctors available for selfrecording of adult patient in polyclinics:

- Therapist
- Therapist
- Surgeon
- Ophthalmologist
- Urologist
- Obstetrician-gynecologist
- General practitioner (family doctor)
- Otorhinolaryngologist
- Dentist Therapist
- Dental surgeon

Patients with chronic diseases visit the "first- level" specialist to get a referral to narrow-profile specialists, and after the next visit can independently sign up for subsequent appointments.

EMIAS system enables routing the patient flow to polyclinics, which solves the problems of self-registration of patients in medical institutions. The appointment channels include a mobile application, websites (public services, emias.info and others), information kiosks, through receptionists, by phone. Urban outpatient departments as primary and mass types of medical services have data on records and appointments, which are estimated in millions every year, however, the traditional approach to documenting visits does not allow forming a complete picture of all patient trajectories in real time taking into account forecasted flow estimates. Besides, frequently, patients need pre-appointment step in order to understand what specialist would best help with the existing symptoms.

Nowadays, in mentioned shared economy conditions patient often search for extra healthcare related information in social networks, consulting resources and it does affect their decisions [9]. This paper addresses the issue of automatic routing the patient to the right medical specialist based on available public data from social networks, to improve the management of heterogeneous patient flows in urban medical institutions with application of shared economy concept.

II. Related Work

The demand for healthcare support is seasonal in nature, and in the summer the number of patient appeals to polyclinics significantly decreases than in the winter. Therefore, the chief physician of the polyclinic needs to understand, how to most successfully manage all available resources in the next period to maintain a high level of availability. The patient flow model could possible provide some values for the stakeholders in order to plan the campaign more effectively. In summer, when demand is relatively low, most doctors go on vacation. Hence, it allows one to offer the therapeutic help in balance with the demand and use of available resources effectively.

The formal presentation of the patient flow in the form of mathematical models in big data conditions allows not only to detect the main trajectories of movements within the framework of medical facilities [10], but also to identify the bottlenecks of the system to meet the demand of the population for outpatient services for given economic system parameters (restrictions).

One of the successful models of patient flow models is clinical pathway. There are papers [11–15] devoted to the research and analysis of clinical pathways based on initial data of a medical institution. Clinical pathways modeling might be developed on the methodological base of probability theory, mathematical statistics, data mining, graph theory, semantic technologies, process mining, etc.

In a study [11], the authors noted the importance of developing an adaptive approach to modeling clinical pathways due to the high variability of patient trajectories and their individual characteristics. Based on proposed graphs of sequences and data mining methods, patterns of the clinical pathways of stroke patients are distinguished to predict the trajectories of new patients.

The semi-Markov model of individual patient experience in a family practice clinic is presented in [12]. The scheme of the general patient flow in this clinic is represented by an oriented graph, the vertices of which are the rooms and departments of the clinic, and edges correspond to the direction of movement of the analyzed flow. This model allows one to predict the duration of patient care, but such parameters as the waiting time in the queue and the length of the queue are not available for analysis.

A special place among the methods for modeling the clinical pathways of patients is taken by the application of process mining. The development of a process model is based on initial data on the real behavior of patients, their routes and the main characteristics that affect the choice of a particular trajectory. The purpose of process mining is to extract new information about processes from event logs. Thus, process mining as a discipline combines machine learning, data mining, and process modeling techniques. The main ideas of this discipline are described in [16–19].

A number of studies are devoted to the development of process analysis algorithms: for example, the eMotivia algorithm for analyzing the movements of nine patients over 25 weeks [20]. A detailed list of algorithms for process mining in healthcare is given in [21], based on a review of 74 studies in this field. It is important to note that the results of the application of process mining allows one to objectively evaluate the past and current movement of patient flow, however, for a detailed study of the system's behavior and the experimental part for its improvement, it is necessary to develop a simulation model [10, 17].

Besides, models based on data from medical facility, it is highly important to study the public resources, such as social networks, in order to understand the real demand for medical services. Advanced technologies, in particular big data, allow one to collect large volumes of various data types, structured and unstructured. Traditionally, the implementation of big data technologies into healthcare field had rather noticeable lag, when comparing with other industries due to unknown success of the investments [1]. Other obstacles include the privacy issues, resistance to changes from medical staff and the lack of standardization in medical available data.

Another huge barrier for big data technologies implementation is the psychological issue. The responsible managers still are tending to underestimate the power of advanced analytics solutions. Taking into account that these solutions are rather expensive, the managers attempt to escape these expenditures of the unknown efficiency for decision making. Therefore, the difficulty of funding trustful investors is also another challenge for that specific case. Besides, the relevancy of gathered data is becoming a challenge, when implementing advanced analytics solution.

However, the benefits of big data application [22-24] to study additional information about services from social media, are enormous. Worldwide, patients search the Internet for medical information, symptoms and other information related to health conditions (97%) and a visit to the doctor (57%) [25, 26]. In [26] authors mention that an appointment organization is rather complicated process, and there is a need to develop a system that uses clinical information, specifically diagnoses records, available to suggest appropriate medical department. In results, authors mention that their system suggests the correct department with 92.7% of accuracy. Nevertheless, system still needs to incorporate drug and medicine information, images to train the developing solution [26].

Based on related work, an approach to represent the initial patient flow within clustered clinical pathways is considered.

III. Methodology

One of the most widely used principles of system analysis is the black box model (*Fig.1*), where system interacts with the external environment through a set of inputs and outputs [27]. It is important to note that the outputs of this model will satisfy the stated goals, which were mentioned in the previous definition of the system. Accordingly, at the first stage of the analysis, it is advisable to formulate the outputs of the "black box". In relation to modeling the patient flow in polyclinic, the system output would include received medical services and estimated patient flow for optimal decision-making. Patients, resources and ***social networks*** might serve as inputs of analyzed polyclinic. The external environment of polyclinic is characterized by economic, scientific and other significant factors.

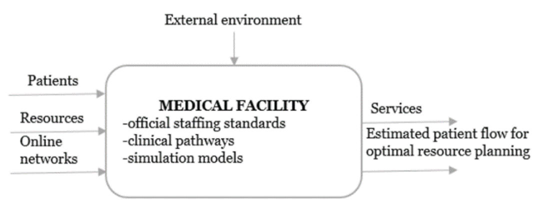

Figure 1. Static model of a medical services management system based on the “black box” principle

Clinical pathways module

As for the model of patient flow in urban polyclinic, healthcare specialists have developed a clinical pathways tool. Generally, clinical pathway means the trajectory of the patient’s movements when receiving medical services in the relevant institutions. In general, clinical pathway of a patient is a trajectory when receiving services in relevant institutions. According to the source [28], clinical pathways have been introduced internationally since the 1980s. This methodology was presented in medical institutions in Sweden in the mid1990s, and in the United States, according to a source [28], approximately 80% of hospitals used clinical pathways to improve the quality of care.

The specific structure of clinical pathways allows the use of cluster analysis of time series. One of the ways to significantly increase the accuracy of the developed model is to segment the initial sample into subgroups of objects similar to each other and to construct separate “personalized” models for each of the selected groups in the future.

Some formal definitions generally accepted for modeling clinical pathways [16-19]. Let E be the set of all real events in the study area that occurred during the medical care process: $E \subseteq A \times T$, where A is the finite set of event identifiers, T is the set of time attributes. Then the event is a pair $e = (a, t)$, where $a \in A$ and $t \in T$. The type of activity and time label of the clinical event is denoted as $e \cdot a$ and $e \cdot t$. It is important to note that when modeling clinical pathways, each event is uniquely determined by a combination of its attributes. Trace σ is the event chain of the patient, a nonempty sequence of patient, a nonempty sequence of events of the clinical path: $\sigma = \langle e_1, e_2, ..., e_n \rangle$, where $e_i \in E$ $(1 \leqslant i \leqslant n)$, $n \in N$ is the length of the patient's route. The set of all routes over E is denoted by E *. The event log L is a non-empty set of patient routes over E *: $L = \{\sigma_1, ..., \sigma_m\}$, where $\sigma_i \in E^*$ $(1 \leqslant i \leqslant m)$, $m \in N$.

A methodology for the formation of patient route groups by the hierarchical agglomerative algorithm with the Ward connection method was performed. For the first time, Additive Regularization of Topic Models (ARTM) was considered to determine patterns of clinical pathways. The obtained results make it possible to conduct a preliminary assessment of the clinical pathways of patients of any event log, to identify bottlenecks in the system and to visualize process maps of the medical institution.

The described approach of clustering clinical pathways to segmentation of the input heterogeneous flow serve as the foundation for the further development of a simulation model of a medical institution and the provision of advisory services to patients, for example, chat bots on web pages of a clinic for consulting services.

Social network analysis module

The module of social network analysis is aimed to provide additional valuable knowledge about patient’s demand for medical specialists in order to build a recommender service in appointment systems. This recommender service is based on text descriptions of the symptoms, which a patient put into a text box while making an appointment. The system offers an optimal choice of different specialists and the best location of medical facility to provide the required service. Therefore, the implementation of such service would reduce the load of appointments to therapists by matching patients to needed doctors, taking into account the schedule, symptoms, and other relevant factors.

In addition, a telemedicine service in medical insitution would reduce the number of physical vitits of patients. Hence, a single input window for patient requests based on such classificator would be automatically processed and route the patient to right specialist.

The Naive Bayes (NB) classifier is probabilistic classifier that was introduced in early 1960s [29] to categorize documents, which assumes that each feature only depends on the class. The idea is that each feature has only the class as a parent.

NB was selected to label and to find the probabilities of classes assigned to texts online chats of patients with doctors for this research, because of its tendency to perform really well for high dimensional data, since each features' probability is estimated independently [30]. Besides, in [31] NB was one of the 10 top data mining algorithms [30]. Observation for class C denotes by X. There is a need to find a class C in which its probability for a given observation would be maximum (1):

$$\mathrm{C} = \arg\max_{C} P(C|X) \tag{1}$$

According to Bayes rule, prediction of the class of X the highest posterior probability should be calculated [30]:

$$P(C|\boldsymbol{X}) = \frac{P(C)P(\boldsymbol{X}|C)}{P(\boldsymbol{X})} \tag{2}$$

As it was mentioned above, in NB classifier features $X_1, X_2, \dots, X_n$ are conditionally independent of defined classes [1]:

$$P(C|\boldsymbol{X}) = \frac{P(C)\prod_{i=1}^{n} P(\boldsymbol{X}_i|C)}{P(\boldsymbol{X})} \tag{3}$$

Machine learning tasks use metrics to evaluate the quality of models and compare various algorithms. An important concept for describing these metrics in terms of classification errors — the confusion matrix (*Fig.2*), where TP- true positive, TN- true negative, FP- false positive and FN- false negative [32].

		Assignment z	
		+	-
Label ℓ	+	TP	FN
	-	FP	TN

Figure 2. Confusion matrix

Based on these matrix, one can calculate the precision (P), recall (R) and F-score metrics:

$$P = \frac{TP}{TP+FP} \tag{4}$$

$$R = \frac{TP}{TP+FN} \tag{5}$$

$$F_\beta = (1+\beta^2)\frac{PR}{R+\beta^2 P} = \frac{(1+\beta^2)\,TP}{(1+\beta^2)\,TP+\beta^2 FN+FP} \tag{6}$$

Another approach to define the topics of patients' questions was performed by topic modelling. One of the most popular topic models is Latent Dirichlet Allocation (LDA), which is the generative hierarchical probabilistic model described in 2003 in a study [33] and originally developed to characterize text documents. The parameters of this model are generated from the Dirichlet a priori distribution, and methods of the Bayesian approach are used for training the model [33]. The document in LDA model is represented by distribution by hidden (latent) topics, each of which is characterized by word distribution.

In probabilistic generative models, available data is considered as the result of a generating process involving hidden variables [33]. Besides, the generating process determined the joint probability distribution over the observed and hidden random variables. As a result, this joint distribution is used to calculate the conditional probability of hidden variables with observables, or a posteriori probability. In the Bayesian approach, one of the options for estimating the parameters of the LDA model is to maximize a posteriori probability.

LDA method is characterized by a priori Dirichlet distribution on the special parameters Φ and Θ. The parameter Φ contains discrete distributions on the set of words w for each topic t: $\varphi wt = p\ (w|t)$. The parameter Θ contains probability distributions on a variety of topics for each document θtd = $p(t|d)$. For the patient's clinical pathway model based on the LDA method, it is necessary to consider the Dirichlet distribution, or a continuous multidimensional distribution on simplex, where $\alpha_k > 0$, $\alpha_k > 0$, $\forall k = 1, \dots, K$ – parameters:

$$Dir(\theta|\alpha) = \frac{\prod_{k=1}^{K}\Gamma(\alpha_k)}{\Gamma(\sum_{k=1}^{K}\alpha_k)}\prod_{k=1}^{K}\theta_k^{\alpha_k-1}, \tag{7}$$

The application of LDA method allows for each patient question in consultation platform to define dominant topic, represented by keywords.

IV. Experiments and Results

For the experiment, the training dataset was collected from public health consultations site, one of the major platforms in Russian, where users chat with doctors of 32 specialties. This source is also containing a catalog of medicines, a catalog of diseases and conditions, a channel to make an appointment in mentioned earlier EMIAS system and other functionalities..

The collected dataset contains 64668 depersonalized patient-doctor conversions with rubric, short header of the problem, question with symptoms, doctor specialization and doctor's answer. There are some questions from patients, which are marked with this special rubric, for instance, cardiology. Accordingly, only marked requests were considered in the experimental part of the study. The collected dataset contains public information necessary for research tasks in the field of optimizing resource utilization by application of shared economy principles.

First, OneHotEncoder from SciKit-Learn library in Python Jupiter Notebook was used to convert categorical data (in context of this study, rubrics of questions) into numbers that predictive models interpret best. Then, the initial dataset was split into test 20% and training 80% of values. By applying CountVectorizer, the input text was converted into a matrix whose values are the number of occurrences of this key (word) in the text. In process vectorizer- transformer- classifier, the transformer step was performed by TfidfTransformer() function.

	precision	recall	f1-score	support
/consultation/list/rubric/proctology/	0.80	0.49	0.61	424
/consultation/list/rubric/psychiatry/	0.71	0.59	0.64	380
/consultation/list/rubric/cardiology/	0.73	0.91	0.81	404
/consultation/list/rubric/vascular/	0.52	0.83	0.64	371
/consultation/list/rubric/endocrinology/	0.73	0.89	0.80	425
/consultation/list/rubric/otolaryngology/	0.60	0.81	0.69	381
/consultation/list/rubric/sexology/	0.88	0.63	0.74	366
/consultation/list/rubric/infection/	0.73	0.80	0.76	436
/consultation/list/rubric/sportdiet/	0.68	0.74	0.71	388
/consultation/list/rubric/onkology/	0.60	0.04	0.07	75
/consultation/list/rubric/trauma/	0.77	0.74	0.76	398
/consultation/list/rubric/estetsurg/	0.65	0.78	0.71	419
/consultation/list/rubric/reproductology/	0.84	0.69	0.76	375
/consultation/list/rubric/patient/	0.96	0.83	0.89	393
/consultation/list/rubric/stomatology/	0.78	0.90	0.84	385
/consultation/list/rubric/surgery/	0.94	0.72	0.81	407
/consultation/list/rubric/allergology/	0.71	0.69	0.70	373
/consultation/list/rubric/oftalmology/	0.93	0.91	0.92	420
/consultation/list/rubric/paediatrics/	0.68	0.46	0.55	412
/consultation/list/rubric/andrology/	0.76	0.77	0.77	403
/consultation/list/rubric/neurology/	0.54	0.34	0.42	369
/consultation/list/rubric/therapy/	0.86	0.73	0.79	386
/consultation/list/rubric/hematology/	0.87	0.84	0.85	395
/consultation/list/rubric/fit/	0.46	0.93	0.62	413
/consultation/list/rubric/somnology/	0.67	0.94	0.78	421
/consultation/list/rubric/dietology/	0.33	0.01	0.02	187
/consultation/list/rubric/gynaecology/	0.65	0.98	0.78	411
/consultation/list/rubric/urology/	0.98	0.70	0.82	311
/consultation/list/rubric/pulmonology/	0.00	0.00	0.00	119
/consultation/list/rubric/gastro/	0.89	0.86	0.87	387
/consultation/list/rubric/logopaedics/	0.62	0.54	0.58	376
/consultation/list/rubric/narcology/	0.37	0.30	0.33	355
/consultation/list/rubric/mammology/	0.70	0.89	0.78	386
/consultation/list/rubric/dermatology/	0.65	0.51	0.57	377
/consultation/list/rubric/genetics/	0.89	0.75	0.81	406
micro avg	0.72	0.72	0.72	12934
macro avg	0.70	0.67	0.66	12934
weighted avg	0.72	0.72	0.70	12934

Figure 3. Results of NB classifier

The most common words from the collected dataset were analyzed by class nltk.probability.FreqDist (Fig.4), which provided the following popular words: *doctor, analysis, child, problem, result, weight, diagnoses etc.* The obtained keywords of online conversations quickly identify the topics and real problems from users in Internet, shaping the understanding of demand for medical services.

Based on precision, recall, f1- score and support metrics NB classifier is estimated, therefore, it can be seen that urology and general patient chats were labelled most effectively, meaning that the question is correctly classified and marked the category to which it belongs. Since the patient's question is an input to the model, the doctor’s response does not affect the result. Besides, the separation of samples into training and validation parts was performed.

Application of LDA modeling to the initial dataset (Fig.5) provided additional tags to topics of patients’ questions, the largest group 1 contains more general words, such as consultation, problem, question, week, photo and etc. The second group is dedicated to blood tests, the third oneto endocrinology and pregnancy, the forth one is mainly about breast cancer. Therefore, these labels of obtained clusters provide additional valuable information about message and improve the quality of a classification task for futher recommender service of patient routing.

V. Conclusion

This paper addresses the issue of automatic routing the patient to the right medical specialist based on available public data from social networks, to improve the management of heterogeneous patient flows in urban medical institutions with application of shared economy concept.

In this paper, the methodology of static model of medical services management was described and modules of clinical pathways approach and social network analysis were studied. For the experiment of the proposed approach the real dataset was collected from public health consultations site in 2020, one of the major platforms in Russian, where users chat with doctors of 32 specialties was considered.

For the subsequent hierarchical cluster analysis of clinical pathways, the Ward method was chosen, for which initially only one object is included in each cluster. After the first clustering approach, the Additive Regularization of Topic Models (ARTM) was applied to the test dataset. All in all, soft clustering method is more flexible for the clinical pathways segmentation, and for the further directions of research the modalities will be added to dataset.

In social network analysis module, Naive Bayes (NB) classifier was applied and evaluated by machine-learning metrics for effectiveness. Then, LDA modelling was performed in order to get additional information about the content of the questions.

The presented approaches serve as the basement for further development of simulation model of urban medical facility and recommendation service for patients. The system will offer an optimal choice of different specialists and the best location of medical facility to provide the required service. Therefore, the shared economy principles are applied by the implementation of such service that would reduce the load of appointments to therapists by matching patients to needed doctors, taking into account the schedule, symptoms, and other relevant factors.

This research is supported by Russian Foundation for Basic Research (RFBR), Investigation into the stability of self-organizing distributed digital systems based on models of the social networks dynamics, # 20-07-00651 A.

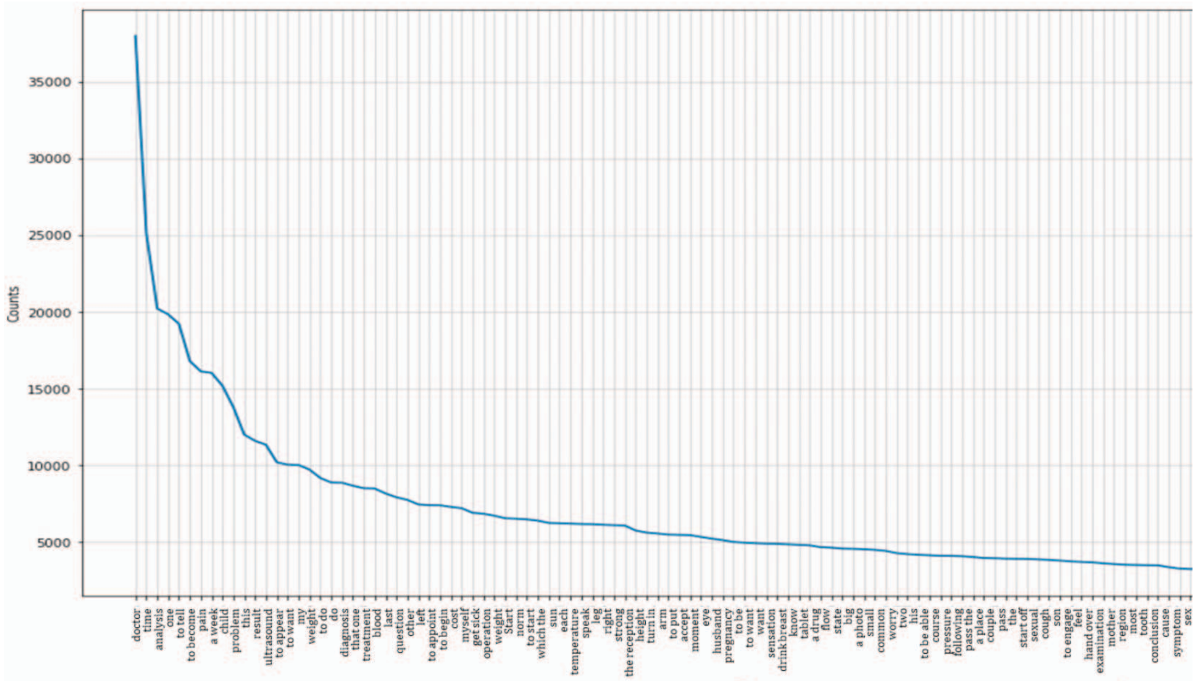

Figure 4. Frequency analysis of words

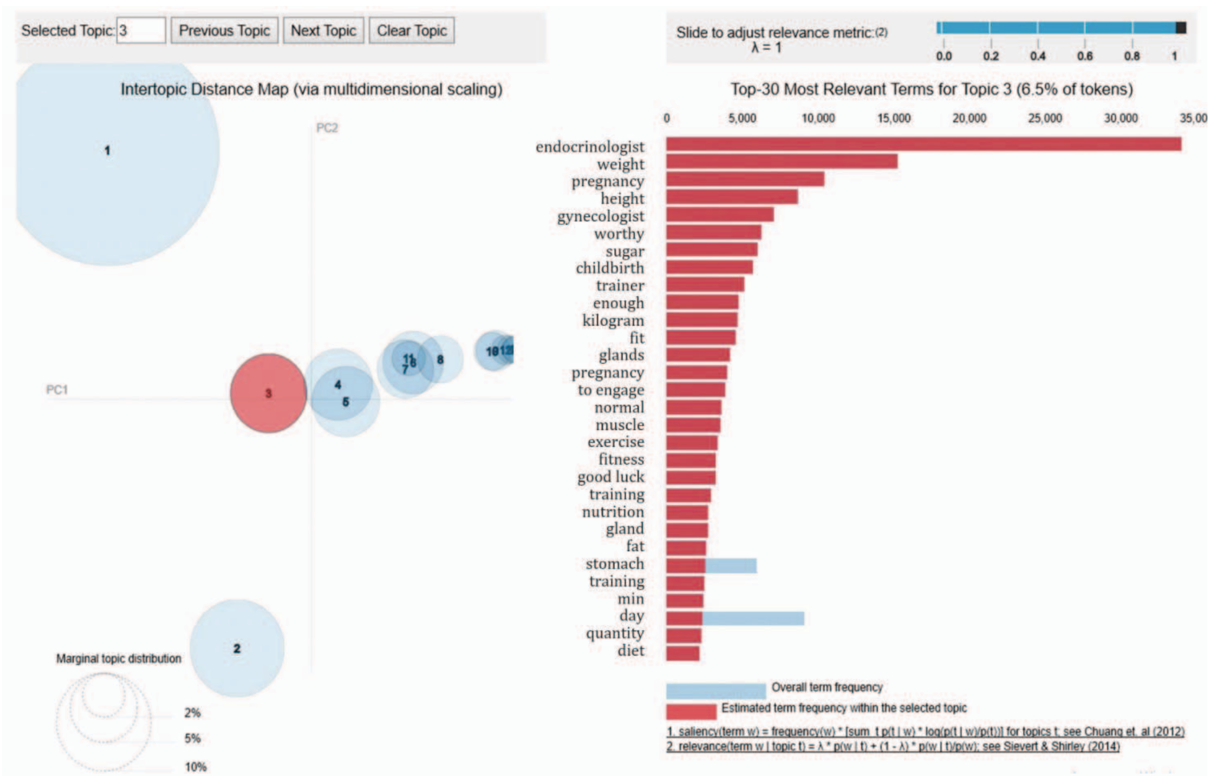

Figure 5. Topics of patients' questions by LDA

REFERENCES

[1] Raghupathi W, Raghupathi V. (2014) Big Data analytics in healthcare: promise and potential. Health Information Science and Systems, vol. 2, no 3, pp.1-10.

[2] Belk, R. (2007) 'Why Not Share Rather Than Own?', The ANNALS of the American Academy of Political and Social Science, 611(1), pp. 126–140. doi: 10.1177/0002716206298483.

[3] Eunsuk, Sung & Kim, Hongbum & Lee, Daeho. (2018). Why Do People Consume and Provide Sharing Economy Accommodation?—A Sustainability Perspective. Sustainability. 10. 2072. 10.3390/su10062072.Martin Prodel. Process discovery, analysis and simulation of clinical pathways using health-care data. Other. Université de Lyon, 2017. English.

[4] Wu H, Chan NK, Zhang CJP, Ming WK. The Role of the Sharing Economy and Artificial Intelligence in Health Care: Opportunities and Challenges. J Med Internet Res. 2019;21(10):e13469. Published 2019 Oct 15. doi:10.2196/13469.

[5] Sutherland, Will & Jarrahi, Mohammad Hossein. (2018). The Sharing Economy and Digital Platforms: A Review and Research Agenda. International Journal of Information Management. 43. 10.1016/j.ijinfomgt.2018.07.004.

[6] May, Spencer & Königsson, Marcus & Holmström, Jonny. (2017). Unlocking the sharing economy: Investigating the barriers for the sharing economy in a city context. First Monday. 22. 10.5210/fm.v22i2.7110.Wakamiya S, Yamauchi K. What are the standard functions of electronic clinical pathways? Int J Med Inform. 2009 Aug;78(8):543–550. doi: 10.1016/j.ijmedinf.2009.03.003.

[7] Botsman, R., and Rogers, R. (2010),What's Mine Is Yours: The Rise of Collaborative consumption. New York, NY, Harpers Collins.

[8] Miller BJ, Moore DW, Schmidt CW. Telemedicine and the sharing economy: the "Uber" for healthcare. Am J Manag Care. 2016 Dec 01;22(12):e420–e422.

[9] Lucinda Roper & Christine Jorm (2017) Preparing medical students for the e-patient: Is a theoretical grounding required?, Medical Teacher, 39:10, 1101, DOI: 10.1080/0142159X.2017.1353072.

[10] Martin Prodel. Process discovery, analysis and simulation of clinical pathways using health-care data. Other. Université de Lyon, 2017. English.

[11] Lin F., Chou S., Pan S., Chen Y. (2001) Mining time dependency patterns in clinical pathways. International Journal of Medical Informatics, vol. 62, no 1, pp. 11–25. DOI: 10.1016/S1386- 5056(01)00126-5.

[12] Cote M.J., Stein W.E. (2007) A stochastic model for a visit to the doctor's office. Mathematical and Computer Modelling, vol. 45, no 3– 4, pp. 309–323. DOI: 10.1016/j.mcm.2006.03.022.

[13] Zhang Y., Padman R., Patel N. (2015) Paving the cowpath: Learning and visualizing clinical pathways from electronic health record data. Journal of Biomedical Informatics, no 58, pp. 186–197.

[14] Blei D.M., Ng A.Y. Jordan M.I. (2003) Latent Dirichlet allocation. The Journal of Machine Learning Research, no 3, pp. 993–1022.

[15] Fernández-Llatas C., Benedi J.-M., García-Gómez J.M., Traver V. (2013) Process mining for individualized behavior modeling using wireless tracking in nursing homes. Sensors (Basel), vol. 13, no 11, pp. 15434–15451. DOI: 10.3390/s131115434.

[16] van der Aalst W.M.P. (2011) Process mining: Discovery, conformance and enhancement of business processes. Springer. DOI: 10.1007/978- 3-642-19345-3.

[17] van der Aalst W.M.P. (2018) Process mining and simulation: A match made in heaven! Proceedings of the 50th Computer Simulation Conference (SummerSim 2018). Bordeaux, France, 9–12 July 2018. DOI: 10.22360/summersim.2018.scsc.005.

[18] van der Aalst W.M.P. (2016) Process mining: Data science in action. Berlin: Springer-Verlag.

[19] van der Aalst W.M.P. (2011) Process mining manifesto. Business Process Management Workshops. Springer, pp. 169–194. DOI: 10.1007/978-3-642-28108-2_19.

[20] Fernández-Llatas C., Meneu T., Benedí J.M., Traver V. (2010) Activity-based process mining for clinical pathways computer aided design. Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. Buenos Aires, Argentina, 31 August – 4 September 2010, pp. 6178–6181. DOI: 10.1109/IEMBS.2010.5627760.

[21] Rojas E., Munoz-Gama J., Sepúlveda M., Capurro D. (2016) Process mining in healthcare: A literature review. Journal of Biomedical Informatics, no 61, pp. 224–236. DOI:10.1016/j.jbi.2016.04.007.

[22] Davenport, T.H. Big Data at Work: Dispelling the Myths, Uncovering the Opportunities; Harvard Business Review Press: Cambridge, MA, USA, 2014.

[23] Aruna Sri, P.S.G.; Anusha, M. Big Data Survey. Indones. J. Electr. Eng. Inform. IJEEI 2016, 74–80.

[24] Schultz, T. Turning healthcare challenges into big data opportunities: A use-case review across the pharmaceutical development lifecycle. Bull. Assoc. Inf. Sci. Technol. 2013.

[25] McMullan, M. Patients using the Internet to obtain health information: How this affects the patient-health professional relationship. Patient Educ. Couns. 2006, 63, 24–28.

[26] Bansal, Veena & Poddar, Abhishek & Ghosh-Roy, R.. (2019). Identifying a Medical Department Based on Unstructured Data: A Big Data Application in Healthcare. Information. 10. 25. 10.3390/info10010025.

[27] CHernyshov, V.N. CH-497 Teoriya sistem i sistemnyj analiz : ucheb. posobie / V.N. CHernyshov, A.V. CHernyshov. – Tambov : Izd-vo Tamb. gos. tekhn. un-ta, 2008. – 96 s. – 150 ekz. – ISBN 978-5-8265- 0766-7.

[28] Kinsman L, Rotter T, James E, Snow P, Willis J. What is a clinical pathway? Development of a definition to inform the debate. BMC Med. 2010;8:31. Published 2010 May 27. doi:10.1186/1741-7015-8-31.

[29] Maron, M. E. (1961). "Automatic Indexing: An Experimental Inquiry". Journal of the ACM. 8 (3): 404–417. doi:10.1145/321075.321084.

[30] Taheri, Sona and Musa A. Mammadov. "Learning the naive Bayes classifier with optimization models." Applied Mathematics and Computer Science (2013).

[31] Wu, X., Vipin Kumar, J., Quinlan, R., Ghosh, J., Yang, Q.,Motoda, H., McLachlan, J., Ng, A., Liu, B., Yu, P. S., Zhou, Z., Steinbach, M., Hand, D. J. and Steinberg, D. (2008). Top 10 algorithms in data mining, Knowledge and Information Systems 14: 1–37.

[32] D.M. Blei, A.Y. Ng, M.I. Jordan. Latent Dirichlet allocation. J Mach Learn Res, 3 (2003).

[33] Goutte, Cyril & Gaussier, Eric. (2005). A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. Lecture Notes in Computer Science. 3408. 345-359. 10.1007/978-3- 540-31865-1_25.