Sentiment Analysis of medical questionnaires to improve adherence to telemonitoring programs

(Discussion Paper)

Chiara Zucco[1], Clarissa Paglia[2], Sonia Graziano[3], Sergio Bella[2] and Mario Cannataro[1]

[1]*Data Analytics Research Center, Department of Medical and Surgical Sciences, University Magna Græcia, Catanzaro, Italy*

[2]*Unit of Cystic Fibrosis, Bambino Gesù Children's Hospital, Rome, Italy*

[3]*Unit of Clinical Psychology, Bambino Gesù Children's Hospital, Rome, Italy*

Abstract

Although it is widely known that high adherence to telemedicine and in particular home telemonitoring programs lead to an improvement in the patient's quality of life, a reduction in hospitalizations, and the containment of healthcare costs, poor adherence is a widespread problem, especially in chronic diseases. Guided by the intuition that the sentiment, i.e., the degrees of positiveness/negativeness, expressed by patients through their responses to ad-hoc designed questionnaires, may be related to their adherence and can be used to predict poor adherence levels, this work describes an integrated software architecture for the online provision, collection, and analysis of questionnaires or surveys and summarizes some preliminary results already presented in our previous work.

Keywords

Text Mining, Sentiment Analysis, Web-based questionnaire, Telemedicine.

1. Introduction

All the set of health services providing medical care in patients' daily living environment are subsumed under the umbrella of telemedicine, which strongly relies on the support of information and telecommunication technologies [2].

Common goals of telemedicine programs range from the increment of patient empowerment to the reduction of healthcare costs [3].

Telemedicine and, specifically, home telemonitoring systems have shown themselves to be cost-effective [4] and to provide an improvement in patients' quality-of-life, in terms, for instance, of significant reduction of both mortality and length of stay of patients in progressive care unit [5], significant improvement of glycemic control for patients with diabetes [6]. However, a telemonitoring program's effectiveness is affected by the extent to which the patient follows medical protocols in terms of frequency and/or dosage [7], namely the adherence levels and, consequently, the degrees of drop-out. In fact, poor adherence is a widespread problem,

CEUR Workshop Proceedings (CEUR-WS.org)

especially in chronic diseases. Therefore, several studies have been carried out to investigate variables related to poor adherence [9],[8],[10].

Adherence levels may be measured in different suitable ways. Here, adherence is intended as the rate of performed monitoring events w.r.t. the ideal number of events, suggested by the telemonitoring protocol, while the degrees of drop-out refer to the percentage of patients the abandon the telemedicine or the telehomecare program they were enrolled in, generally due to poor adherence.

The proposed architecture is intended as a contribution that can help the context of home telemonitoring programs. The basic idea is to integrate within a telehomecare system an online survey instrument that, by collecting textual answers written by patients enrolled in a telemonitoring program, can extract useful features capable of early predict drop-out due to poor adherence.

The system encompasses a novel analysis approach that leverages lexicon-based sentiment analysis techniques [11] and exploits the inferred polarity as a numerical feature to enhance further statistical or machine learning analysis.

To the best of our knowledge, no specific research has been published, nor a system architecture has been proposed that would explicitly monitor changes in patient's opinion across time through the repeated administration of a questionnaire, using the polarity associated with answers to open-ended question as a numerical feature, in a telehomecare system.

Additionally, the paper will present a case study application of the system architecture to discuss whether a predictive relationship, in terms of Granger-causality test modeling, may be assessed between patient adherence in a cystic fibrosis telehomecare program and their opinion about the program they are enrolled in.

The rest of the paper is organized as follows: Section 2 describes the methodology behind the proposed approach and the case-study application. Section 3 provides insights related to collected data and presents the Granger-causality hypothesis tests results and discusses it.

Finally, Section 4 concludes the paper and outlines future works.

2. Materials and methods

In this Section, some preliminary information related to the case study, a description of the experimental protocol used and the analysis pipeline's proposal are presented.

2.1. Experimental protocol and Dataset description

The data analyzed in the present case study application were collected from the Cystic Fibrosis Unit, Bambino Gesù Children's Hospital, Rome, Italy. In this study, 169 online surveys sent by 38 cystic fibrosis patients (F/M=20/18, age= 28.7 ± 9.91, age range= $14 - 49$) recruited among patients already enrolled in a telemedicine program (years of enrollment = 5.9 ± 3.9) were collected and analyzed at five different survey epochs.

The enrollment criteria include patients more than 12 years old with cystic fibrosis who will access the Cystic Fibrosis Unit in ordinary, daytime, or outpatient hospitalization. All patients who have undergone a transplant (liver/lung) were excluded from the study.

The study was formally approved by the local Medical Research Ethics Committee.

2.2. Administration of questionnaire

From June 2019, 38 enrolled patients were asked to complete, every three months, an online questionnaire designed ad-hoc by the clinical team. In the following, each set of surveys submission will be indicated as an epoch.

The Telemedicine Drop-Out (TDO) questionnaire consists of 15 blocks of closed, mixed, and open-ended items with yes/no constraints, and it has been administered through a self-hosted web-based survey instrument built on top of LimeSurvey[1]. The TDO survey was designed as an online, structured version of the interview led by the medical team within the telemedicine program, extended with a series of open-ended questions, whose objective was to infer polarity or, in perspective, to extract emotions from the relative answers [12].

In order to administer surveys to patients, LimeSurvey is set up as a highly customizable, free, and responsive online survey tool. It also provides various API functions through the LimeSurvey RemoteControl 2 (LSRC2). The survey structure and the participants are created through the user interface provided by LimeSurvey. The collection of survey answers is automatized using the Python library Limepy that provides a Python wrapper for the LSRC2 API and the Python library Schedule to automatically update the responses. The DBMS server is MySQL.

As already stated, adherent patients need to transmit the results of the spirometry test at least twice a week. For each survey administration, i.e. survey epoch, the patient's adherence score (Adh-score) to the telemonitoring program was assessed as the total number of spirometry transmissions sent during a three months window starting from the month before until the month subsequent to the survey administration, averaged by twice the total number of weeks following. More in details, let suppose that a survey was carried at month t, then Adh-score$_t = \frac{nS_{t-1}+nS_t+nS_{t+1}}{2(w_{t-1}+w_t+w_{t+1})}$ where nS_{t-1}, nS_t, nS_{t+1} refer to the number of spirometry trasmissions sent at month $t-1, t, t+1$ respectively, while w_{t-1}, w_t, w_{t+1} refer to the number of weeks in the month $t-1, t, t+1$ respectively.

Patients that strictly follow medical advice have a related Adh-score≥ 1. In the following, the percentage of Adh-score, i.e. Adh-score (%) will be considered. Therefore, $0 \leq$ Adh-score (%) = Adh-score $* 100$ and Adh-score (%)> 100 for patients with high rates of adherence. The clinical team provided the number of transmissions per month.

2.3. System architecture

The system architecture encompasses three independent modules, connected in a cascade-fashion. In future works, the modules are supposed to be integrated using a unique user interface. Figure 1 shows the overall architecture for the system, that is organized as three logical levels: i.e. i) data collection, ii) data integration, and iii)data analysis.
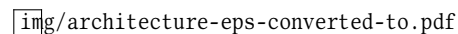
The general pipeline for the analysis of textual data, i.e. answers to open-ended questions, involves: **Text preprocessing :** includes standard NLP (Natual Language Processing) techniques, i.e. tokenization, stop word removal, and lemmatization. The preprocessing step has been executed by using *SpaCy*[2] a popular library for Natural Language Processing in Python,

---

[1]https://www.LimeSurvey.org/
[2]https://spacy.io/

which provides a set of preprocessing algorithms also for the Italian language. **Feature extraction:** to each open-ended free-text answer, a polarity score in the range $[-1, 1]$ has been assigned through the VADER [13] lexicon-based method adapted to the Italian language and considered as a numerical feature. **Statistical hypothesis testing:** data have been sorted by respondents and survey submission date and, for each open-ended question, the sequence of assigned polarity was modeled as a time series, as the sequence of Adh-scores. Augmented Dickey-Fuller Test [14] was used to check for stationarity while Granger-causality hypothesis test model [15] was examined to discuss the existence of directed causal interactions between the polarity score associated with free-text answers and adherence. **Data visualization:** to provide useful insights and summarize patient answers, different visualization techniques have been used. In particular, preprocessed free-text answers have been visualized through word clouds, while a graph shows the time-series of polarity scores at different submission epochs.

Figure 1: The modules of the system architecture, implemented as three independent levels connected to each other in cascade. The architecture is designed to be cyclical, as the system is used for each scheduled administration of the survey.

VADER (Valence Aware Dictionary for sEntiment Reasoning) [13] is a lexicon-based sentiment analysis engine that combines lexicon-based methods with a rule-based modeling consisting of 5 human validated rules.

Starting from a classical lexicon-based approach, the VADER engine's core step is the identification of some general grammatical and syntactic heuristics to identify semantic shifters, i.e. words that increase, decrease or change the polarity orientation of another word. In particular, five heuristics for sentiment polarity shifters have been identified: i.e.punctuation; capitalization; a list of degree modifiers which encompasses noun, adjective, adverbs know to impact sentiment intensity by increasing or decreasing it; contrastive particles and negations.

To extend VADER to the Italian language, Sentix [16], a lexicon that automatically extends the SentiWordNet annotation to the Italian synsets provided in MultiWordNet [17], has been considered.

Among the five heuristics designed in VADER, only three needed to be adapted to the Italian language since the shifter role of capitalization of words and exclamation marks is used as intensifiers for both languages. Words belonging to the VADER set of negation words were translated in the Italian language, and the set was then extended by retrieving MultiWordNet synset terms for each word, while contrastive particle "but" has been simply translated to Italian.

Among the intensifier sets, VADER also considered a few idioms, but, due to discrepancies across different languages, idioms were not considered.

Granger-causality is a statistical hypothesis testing model to determine if there is a directed relationship between two time-series [15]. A time series X is said to Granger-cause Y if it can be shown that there is a statistically significant improvement in predicting future values of Y by using past values of X (i.e. lagged values of X) and Y, compared to predictions based only on past values of Y.

The possibility to relate past values of X to Y's actual values is in virtue of a lag factor. Here, the Granger-causality test was computed for X's lagged values. All the lags ranging from one

to four were tested, where four is the number of considered submission epochs minus one.

Here, the considered alternative hypothesis is that the polarity-score time-series associated with each considered open-ended question Granger-cause the time-series of adherence. The level of significance was set at 5%, i.e. *p-value* < 0.05. The Granger-causality test assumes the hypothesis that the investigated time-series are stationary. Therefore the augmented Dickey-Fuller method was exploited to check stationarity conditions [14].

3. Results and discussion

A comprehensive analysis of the responses to the TDO survey is beyond the scope of this study. Instead, only answers to two open-ended questions collected from the TDO survey will be discussed, namely: i) **Q1** - "What do you think about telemedicine?", and **Q2** - "Since you joined the telemonitoring program, what has improved the quality of your life?". A polarity score ranging in $[-1, 1]$ has been inferred by adapting the VADER framework to the Italian language and considered as a numerical feature for each set of answers.

3.1. Testing Granger-causality

Three time-series have been considered, i.e. polarity score related to Q1 and Q2 respectively and the time-series of Adh-scores. The augmented Dickey-Fuller test showed that for all the three considered time-series the stationarity condition holds (p-value= 4.6124e-18, p-value= 3.2185e-07, p-value= 0.0035 respectively).

Two Granger-causality tests were performed to check whether Q1 Granger-causes Adh-score and whether Q2 Granger-causes Adh-score. Moreover, since all the three series are considered contemporaneously, we also need to check whether Adh-score Granger-causes Q1 and whether Adh-score Granger-causes Q2. Three different test-statistics, i.e. F-test, chi2, and likelihood-ratio were considered, with the number of lags varying from one to four. Table 1 and Table 2 show the results in terms of p-values. It can be seen that both Q1 and Q2 Granger-causes Adh-score for lag=1. On the other hand, Adh-score appears to not Granger-causes Q1 or Q2.

Therefore, the results suggest the existence of a predictive relationship between the polarity scores series associated with Q1 and the polarity score series associated with Q2 w.r.t. the Adh-score.

Table 1

Q1 and Adh-score: p-value of Granger-causality test performed with three different statistics and four different lags.

	Q1$\longrightarrow$ Adh-score			Adh-score$\longrightarrow$ Q1		
	F-test	Chi2 test	Likelihood-ratio	F-test	Chi2 test	Likelihood-ratio
1	**0.0368**	**0.0339**	**0.0350**	0.4076	0.4027	0.4031
2	0.0998	0.0908	0.0936	0.6667	0.6585	0.6592
3	0.1063	0.0917	0.0963	0.6628	0.6479	0.6496
4	0.2032	0.1760	0.1833	0.8200	0.8061	0.8064

Table 2
Q2 and Adh-score: p-value of Granger-causality test performed with three different statistics and four different lags.

	Q2$\longrightarrow$ Adh-score			Adh-score$\longrightarrow$ Q2		
	F-test	Chi2 test	Likelihood-ratio	F-test	Chi2 test	Likelihood-ratio
1	**0.0020**	**0.0016**	**0.0028**	0.9970	0.9969	0.9969
2	**0.0174**	**0.0141**	**0.0155**	0.7455	0.7390	0.7394
3	0.0547	0.0446	0.0482	0.8235	0.8148	0.9154
4	0.0865	0.0685	0.0744	0.7566	0.7387	0.7406

The results are consistent with the hypothesis that the polarities extracted from patients' opinion on telemedicine may help predict their average adherence one epoch after the survey administration.

4. Conclusions

In the present paper, a system architecture for the extraction of emotional states from textual contents, designed to support the monitoring of patients with chronic disease has been presented. The main goal of the proposed system is basically to implement a methodology to capture the underlying opinions that chronic patients have about the program they are enrolled in and to investigate whether these features may help in the early prediction of patients drop-out from the telemedicine program.

The proposed system is designed in an end-to-end fashion to provide support through the whole process, including the implementation of the questionnaire, the survey administration at scheduled intervals, as well as the analysis. Specific contributions are: i)the design of a self-hosted web-based survey instrument built on top of Limesurvey for the management of online inquiries over time; ii)an analysis pipeline that exploits sentiment analysis techniques to infer a sentiment polarity score for each open-ended answer and uses it as a numerical feature (to the best of our knowledge, this is the first time that this kind of approach has been proposed); iii)the validation of both the survey instrument and the analysis pipeline, that have been applied to collect and analyze 169 TDO survey responses sent by 38 patients enrolled in a home telemonitoring program provided by the Cystic Fibrosis Unit at the "Bambino Gesu" Children Hospital in Rome, Italy.

Limitations to the present analysis may be found in the small amount of data collected up-to-date, which does not consent the investigation of changes for a single patient through time.

Nevertheless, the promising results encourage us to further investigate the potentiality of the proposed architecture and the analysis pipeline with the aim to develop, as future work, a predictive system for the early detection of poorly-adherent patients that may also alert doctors to contact patients and eventually update/personalize their telemedicine program (e.g. in terms of timing, technological equipment, psychological counseling, etc.)

References

[1] Zucco, C., Paglia, C., Graziano, S., Bella, S., & Cannataro, M.: Sentiment Analysis and Text Mining of Questionnaires to Support Telemonitoring Programs. Information, 11(12), 550 (2020)

[2] Ryu, S. Telemedicine: Opportunities and Developments in Member States: Report on the Second Global Survey on eHealth 2009 (Global Observatory for eHealth Series, Volume 2); *Healthc. inf. res.*; *18*, 153 (2012).

[3] Nielsen, M.K.; Johannessen, H. Patient empowerment and involvement in telemedicine. *J. Nurs. Educ. Pract.*, *9*, 54–58 (2019).

[4] Delgoshaei, B.; Mobinizadeh, M.; Mojdekar, R.; Afzal, E.; Arabloo, J.; Mohamadi, E. Telemedicine: A systematic review of economic evaluations. *Med J. Islam. Repub. Iran (MJIRI)*, *31*, 754–761 (2017).

[5] Armaignac, D.L.; Saxena, A.; Rubens, M.; Valle, C.A.; Williams, L.M.S.; Veledar, E.; Gidel, L.T. Impact of Telemedicine on Mortality, Length of Stay, and Cost Among Patients in Progressive Care Units: Experience From a Large Healthcare System. *Critical care medicine*, *46*, 728 (2018).

[6] Polisena, J.; Tran, K.; Cimon, K.; Hutton, B.; McGill, S.; Palmer, K. Home telehealth for diabetes management: a systematic review and meta-analysis. *Diabetes, Obesity and Metabolism*, *11*, 913–930 (2009).

[7] Osterberg, L., & Blaschke, T.: Medication Adherence. New England Journal of Medicine, 353, 487-497 (2005)

[8] Gorst, S.L.; Armitage, C.J.; Brownsell, S.; Hawley, M.S. Home telehealth uptake and continued use among heart failure and chronic obstructive pulmonary disease patients: a systematic review. *Annals of Behavioral Medicine*, *48*, 323–336 (2014).

[9] Mathes, T., Jaschinski, T., & Pieper, D.: Adherence influencing factors–a systematic review of systematic reviews. Archives of Public Health, 72(1), 1–9 (2014)

[10] Tagliente, I.; Solvoll, T.; Murgia, F.; Bella, S. Telemonitoring in cystic fibrosis: A 4-year assessment and simulation for the next 6 years. *Interact. J. Med. Res.*, *5*, (2016).

[11] Zucco, C.; Calabrese, B.; Agapito, G.; Guzzi, P.H.; Cannataro, M. Sentiment analysis for mining texts and social networks data: Methods and tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, p. e1333 (2019).

[12] Zucco, C.; Bella, S.; Paglia, C.; Tabarini, P.; Cannataro, M. Predicting abandonment in telehomecare programs using Sentiment Analysis: a system proposal. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 1734–1739 (2018).

[13] Hutto, C.; Vader, G.E. A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the Eighth International AAAI Conference on Weblogs and Social, Ann Arbor, MI, USA, 1–4 (2014).

[14] Dickey, D.A.; Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.*, *74*, 427–431 (1979).

[15] Granger, C.W. Testing for causality: A personal viewpoint. *J. Econ. Dyn. Control*, *2*, 329–352 (1980).

[16] Basile, V.; Nissim, M. Sentiment analysis on Italian tweets. In *Proceedings of the 4th*

Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Atlanta, Georgia, 14 June 2013; Association for Computational Linguistics: Atlanta, Georgia; 100–107 (2013).

[17] Pianta, E.; Bentivogli, L.; Girardi, C. *MultiWordNet: Developing an Aligned Multilingual Database*; In Proceedings of the First International Conference on Global WordNet, Mysore, India, 293–302 (2002).