Lahti, Lauri; Tenhunen, Henni; Nieminen, Marko

How patients talk about care? identifying patient experience expressions from online discussions

Published in:
Building Continents of Knowledge in Oceans of Data

DOI:
10.3233/978-1-61499-852-5-116

Published: 01/01/2018

Document Version
Publisher's PDF, also known as Version of record

Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth
A. Ugon et al. (Eds.)

doi:10.3233/978-1-61499-852-5-116

How Patients Talk About Care? Identifying Patient Experience Expressions from Online Discussions

Lauri LAHTI[a1], Henni TENHUNEN[b] and Marko NIEMINEN[a]
[a] *Department of Computer Science, Aalto University, Finland*
[b] *Institute of Healthcare Engineering, Management and Architecture (HEMA), Department of Industrial Engineering and Management, Aalto University, Finland*

Abstract. Patient experience is an emerging concept that supports the improvement of healthcare services through identified patient expectations and experiences. In addition to structured feedback through official channels, experiences about healthcare appear increasingly in digital services and social media. We explore a new patient experience harvesting process based on linguistic patterns to identify relevant expressions in online discussions about children's health. Our results from the analysis of 98 229 unique sentences suggests that the 7-step process can be useful in discovering patients' evaluations of their care experiences. We propose ways to extend the process to other care contexts by adjusting the semantic reference models.

Keywords. patient experience, word collocations, online discussions

1. Introduction

Patient experience (PX) reflects events across the continuum of care including focus on personalized care, patient-centered care and patient satisfaction [11]. However, as a concept, PX is still fragmented and flexible: It covers a broad range of topics including information and communication, patient's emotions, knowledge, opinions and decision-making. Conceptually, PX connects to physical conditions and health behavior [3, 5, 9]. A known instrument for gathering patient feedback is the HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems satisfaction survey) [12]. HCAHPS is implemented as a questionnaire, which is suitable to handle pre-categorized and structured data. Real life, however, often escapes the rigid formats of such measures.

Standardized surveys enable longitudinal monitoring of patient satisfaction. Reasons for changes in measures with related experiences, however, are not easily identifiable so that they would support detailed improvement actions in healthcare. At the same time, patients increasingly voice their views in the "digital world" at online forums in order to get peer support and to compare experiences. To certain extent, such online discussions can form a valuable database of organic unstructured expressions describing healthcare experiences. As digital communication in healthcare increases, new approaches are needed to answer this call. Currently, the potential of large text corpuses is largely

[1] Corresponding Author, Department of Computer Science, Aalto University School of Science, Espoo, Finland; E-mail: lauri.lahti@aalto.fi.

unrealized due to the limitations of qualitative methods and traditional statistical approaches to handle masses of online discussion information. Recent studies address how identification of patterns of linguistic health data can help to support care [2, 7].

In this article, we present a mixed method process combining qualitative and computational estimations. We harvest patient experience expressions from online discussions concerning children's health. The source of the online discussions is the Finnish Suomi24 discussion forum [10], providing a rich set of unmoderated expressions. Our exploration can serve as the basis for automated harvesting of patient experience in emerging digital communication platforms in healthcare.

2. Objectives and research questions

Our objective is to study how online discussions reveal patients' experiences. What is the organically formed, unprompted "online language" about patients' experiences? Due to the nature of the data, we explore what could be a recommended process to efficiently find focused expressions from masses of data. Our research questions are following: Q1: How do patients express their experiences on care in digital/online discussions? What words and expressions do they use to convey PX? Q2: How can we find the relevant PX expressions from masses of text? What process can be used to extract the valuable expressions? The practical aim is to define an effective method for finding and identifying relevant expressions in online discussion for improving patient experience.

3. Data and methods

The source of the online discussions is the Finnish Suomi24 discussion forum in a time frame 2001-2016. It is available as an annotated data set for linguistic research purposes at the Language Bank of Finland [10]. The data is supplied with a morphosyntactic annotation that is based on Turku Dependency Parser. Our set of 98 229 unique sentences from discussion category Children's health ('Lasten terveys' in Finnish) have been extracted and further computationally analyzed with various R language scripts by author LL. Analysis of the sentences in the data set is facilitated by having the part-of-speech tags indicated for each word in the sentences.

In our procedure, we combine qualitative rating and thematic analysis with quantitative methods. Many common statistical tests are designed for independent and identically distributed data and do not suit well for measuring all kinds of associative patterns of data, such as collocation of words in texts. The frequency of a word pair collocation can indicate the popularity of a certain expression, but many distinctive patterns can remain unnoticed due to the noise caused by dominant but trivial words. Log-likelihood has been proposed as a convenient measure for analyzing associative patterns of data [1, 4]. Likelihood ratio tests rely on the ratio computed between the maximum likelihood of the observed data in respect to the null hypothesis and its unconstrained maximum likelihood. Computation of log-likelihood values can be carried out with a multinomial sampling distribution.

4. The 7-step PX harvesting process (PX-HARV)

Author LL calculated frequency and log-likelihood values for all word pair collocations in 98 229 sentences (**Step 1**). Our extraction of the PX expressions started by identifying word pair collocations that define a relationship between a PX relevant noun and adjective (e.g. "sick child"). The authors formulated 72 Finnish key terms[2] to be matched with at least one word in adjective-noun or noun-adjective collocations occurring in 98 229 sentences. These key terms were formed based on the conceptual themes of two patient experience instruments: the HCAHPS [12] and a newly developed survey for the parents of pediatric patients (PPX) [6] (**Step 2**). Among collocations matching with the key terms author LL extracted the 16 highest-ranking word pairs[3] in respect to frequency and log-likelihood to be used for further analysis, 20 word pairs when overlaps removed (**Step 3**). These 16 highest-ranking word pairs occurred in 869 unique sentences. The three authors independently reviewed each of these 869 sentences and evaluated whether the sentence contained a description of patient experience (yes/no) (**Step 4**). Based on this identification task, we formed a ranking of word pairs: How likely is it that certain word pairs can reveal sentences describing patient experience from the data set? The ranking was based on relative frequency indicating how many of the word-pair sentences were identified as "contains a PX expression" by all three authors (**Step 5**) (Table 1).

Table 1. Highest-ranking word pairs for which all the three authors agreed that the sentences represent PX

Word pair		Frequency of sentences for the word-pair category	Proportion of sentences about PX in the word-pair category agreed by all three authors
Adjective	**Noun**		
private	doctor	35	0.80
good	care	29	0.76
private	healthcare provider	21	0.71
own	doctor	27	0.63

Further analysis was done with 50 sentences for the word pairs "private doctor" ("yksityinen lääkäri", PD) and "good care" ("hyvä hoito", GC), since these word pairs reached the highest proportion of sentences about PX agreed by all three authors, and they also represent a general theme in common vocabulary in respect to healthcare services (**Step 6**). The three authors independently evaluated 28 "private doctor" sentences (0.80×35=28) and 22 "good care" sentences (0.76×29≈22) in respect to their semantic fit with HCAHPS and PPX categories (Table 2) (**Step 7**).

Table 2. HCAHPS and PPX categories used in the classification of the 50 PX sentences

HCAHPS categories [11]	PPX categories [6]
C1. Communication with nurses and doctors **C2.** The responsiveness of hospital staff **C3.** The cleanliness and quietness of the hospital environment **C4.** Pain management **C5**. Communication about medicines **C6.** Discharge information **C7.** Overall rating of hospital **C8.** Would they recommend the hospital	**C9.** Success of the treatment **C10.** Arrangements for the treatment **C11.** Personnel **C12.** Relation/attitude to the illness **C13.** Support and arrangements for everyday life

[2] **72 key terms** and other details of the analysis are available in our open data supplement (see [8]).

[3] **20 word pairs** (overlaps removed): sick child (frequency 102; log-likelihood 535.09), other symptom (98; 345.01), own experience (97; 489.80), other experience (78; 311.53), domestic grain (49; 735.68), allergic reaction (46; 529.71), positive feedback (40; 571.39), neurologic illness (37; 329.22), private doctor (33; 158.10), corresponding experience (32; 276.32), effective substance (30; 448.94), personal assistant (30; 423.40), good care (29; 97.32), personal experience (29; 357.29), neurologic examination (28; 226.13), own doctor (28; 25.83), appropriate answer (23; 267.94), academic hospital (21; 270.47), private healthcare provider (21; 272.46), negative feedback (17; 231.74)

5. Results

28 PD sentences and 22 GC sentences were labeled with HCAHPS and PPX categories by all three authors (Table 3). HCAHPS categories received labels less evenly than PPX. The HCAHPS categories C3 (the cleanliness and quietness of the hospital environment) and C4 (pain management) did not receive any labels from the authors. Categories C5 (communication about medicines) and C6 (discharge information) received only few labels. For the sentences to which all three authors gave the same HCAHPS category label, C7 (overall rating of hospital) received the highest number of labels (11). C8 (would they recommend the hospital) received 6 labels, C1 (communication with nurses and doctors) 5 labels, and C5 (communication about medicines) 1 label. For the sentences to which all three authors gave the same PPX category label, C9 (success of the treatment) received the highest number of labels (5). Both HCAHPS and PPX received labels relatively equally on PD sentences. PPX, however, received more labels on GC sentences than HCAHPS . Among 50 sentences on PD or GC, 8 sentences were given the same HCAHPS and PPX category labeling by all three authors. Examples of such sentences are “Good care and nice personnel” and “We visited an expensive private doctor, said that it was just some rash”.

Table 3. Amount of sentences getting the same HCAHPS [12] or PPX [6] category label by the authors (28 "private doctor" (PD) sentences and 22 "good care" (GC) sentences)

Same HCAHPS category label for the sentence			Same PPX category label for the sentence		
Category	**Same label from all three authors**	**Same label from at least 2/3 authors**	**Category**	**Same label from all three authors**	**Same label from at least 2/3 authors**
C1	5 (PD 5; GC 0)	8 (PD 6; GC 2)	C9	5 (PD 3; GC 2)	13 (PD 4; GC 9)
C2	0 (PD 0; GC 0)	6 (PD 3; GC 3)	C10	0 (PD 0; GC 0)	10 (PD 6; GC 4)
C3	0 (PD 0; GC 0)	0 (PD 0; GC 0)	C11	0 (PD 0; GC 0)	6 (PD 5; GC 1)
C4	0 (PD 0; GC 0)	0 (PD 0; GC 0)	C12	0 (PD 0; GC 0)	11 (PD 1;GC 10)
C5	1 (PD 1; GC 0)	1 (PD 1; GC 0)	C13	0 (PD 0; GC 0)	5 (PD 1; GC 4)
C6	0 (PD 0; GC 0)	0 (PD 0; GC 0)			
C7	11 (PD 1;GC 10)	11 (PD 3; GC 8)			
C8	6 (PD 4; GC 2)	6 (PD 4; GC 2)			

6. Discussion and conclusions

Our results suggest that the 7-step PX harvesting process can be used in discovering patients’ evaluations and descriptions of their care experiences. The method mixes quantitative and qualitative methods enabling the use of big data sources for finding practical insights supporting healthcare service improvement. Computational methods enable the use and effective filtering of large data sources and human interpretation adds significance and semantic certainty on the revealed expressions.

Our process focused on statements of patients’ expressions about healthcare in Finnish online discussions. For children’s health, word pairs “good care” and “private doctor” were expressions that led us to PX focused sentences that fit conceptual categories in patient feedback and experience models (HCAHPS [12], PPX [6]). The expressions relating to HCAHPS addressed overall rating of hospital, recommendations of hospitals, and communication with nurses and doctors. The PPX related expressions addressed success of the treatment, relation/attitude to the illness, and arrangements for

the treatment. Overall, it appeared that the continuum of care in respect to PPX, including patient-centered care, was rather well covered by these expressions.

For systematic semantic focus, our "PX-HARV" method requires the usage of a conceptual reference model such as HCAHPS or PPX. By changing the conceptual reference models, the process may be used in other contexts of care. With proper conceptual framing, we anticipate that our method can be applied, for instance, to analyze online discussions regarding pregnancy with a pregnancy-related PX survey or prostate cancer discussions with a prostate cancer specific content.

The increasing amount of healthcare communication in a digital form provides interesting possibilities for computer-supported semi-automatic analysis of PX. Our 7-step process can serve as the basis for automated harvesting of patient experiences in emerging digital communication platforms in healthcare, such as chat-based healthcare services. For practical purposes, our proposed process may still be too labor-intensive requiring several persons to manually classify the sentences. Therefore, further work on the PX harvesting process would benefit from focus on increased efficiency in spotting relevant experience statements. Future work should focus on automating some of the qualitative steps in the process and explore longer phrases with machine learning.

References

[1] A. Agresti, *Categorical Data Analysis* (1990), John Wiley & Sons, New York.

[2] J. Ashutosh., D. Andrews, A. Fiksdal, A. Kumbamu, J. McCormick, A. Misitano, L. Nelsen, E. Ryu, A. Sheth, S. Wu & J. Pathak, Comparative Analysis of Online Health Queries Originating From Personal Computers and Smart Devices on a Consumer Health Information Portal. *Journal of Medical Internet Research* 16(7) (2014).

[3] A. Chen, H. Tenhunen., P. Torkki, S. Heinonen, P. Lillrank & V. Stefanovic, Considering medical risk information and communicating values: A mixed-method study of women's choice in prenatal testing, *PLOS ONE* 12(3) (2017), e0173669.

[4] T. E. Dunning, Accurate methods for the statistics of surprise and coincidence, *Computational Linguistics*, 19(1) (1993), 61–74.

[5] S. Greenfield, S. Kaplan, J.E. Jr. Ware. Expanding patient involvement in care: effects on patient outcomes, *Ann Intern Med* 102 (1985), 520–8.

[6] J. Kaipio, P. Mannonen, H. Stenhammar, N. Karisalmi, P. Lahdenne, K. Hiekkanen, M. Nieminen, Potilaskokemuskyselyn kehittäminen lastensairaalan potilaiden vanhemmille, *FinJeHeW* (2017), 9(2–3).

[7] L. Lahti, Interpretation of health-related expressions and dialogues: enabling personalized care with contextual measuring and machine learning. *International Journal of New Technology and Research* (IJNTR), 3(11)(Nov 2017), 171-179, ISSN 2454-4116. https://www.ijntr.org/download_data/IJNTR03110081.pdf http://urn.fi/URN:NBN:fi:aalto-201712298340

[8] L. Lahti, H. Tenhunen, & M. Nieminen, Data supplement to the research article Lahti, Lauri, Tenhunen, Henni, & Nieminen, Marko, "How patients talk about care? Identifying patient experience expressions from online discussions", Aalto University School of Science, Finland, (2018). (Open access: http://urn.fi/URN:NBN:fi:aalto-201802021206).

[9] T.D. Sequist, E.C. Schneider, M. Anastario, E.G. Odigie, R. Marshall, W.H. Rogers, et al. Quality monitoring of physicians: linking patients' experiences of care to clinical quality and outcomes. *J Gen Intern Med.* 23(11) (2008), 1784–90.

[10] The Suomi 24 Corpus (2016). The corpus of discussion forums of the Suomi24 online social networking website. A linguistic resource provided by the Language Bank of Finland and CSC - IT Center for Science. http://urn.fi/urn:nbn:fi:lb-2017021630

[11] J.A. Wolf, V. Niederhauser, D. Marshburn & S. LaVela S. Defining Patient Experience. *Patient Experience Journal* 1(1) (2014).

[12] E.E. Zusman, HCAHPS replaces Press Ganey survey as quality measure for patient hospital. *Neurosurgery* 71(2) (2012), N21–N24.