Patient Experience in Online Support Forums: Modeling Interpersonal Interactions and Medication Use

Annie T. Chen
University of North Carolina, Chapel Hill
School of Information and Library Science
atchen@email.unc.edu

Abstract

Though there has been substantial research concerning the extraction of information from clinical notes, to date there has been less work concerning the extraction of useful information from patient-generated content. Using a dataset comprised of online support group discussion content, this paper investigates two dimensions that may be important in the extraction of patient-generated experiences from text; significant individuals/groups and medication use. With regard to the former, the paper describes an approach involving the pairing of important figures (e.g. family, husbands, doctors, etc.) and affect, and suggests possible applications of such techniques to research concerning online social support, as well as integration into search interfaces for patients. Additionally, the paper demonstrates the extraction of side effects and sentiment at different phases in patient medication use, e.g. adoption, current use, discontinuation and switching, and demonstrates the utility of such an application for drug safety monitoring in online discussion forums.

1 Introduction

Online support groups are a rich source of information concerning patient experiences, but they are far different from clinical content. Instead of "The patient presents with…" and "denies vomiting," patients may speak of their "doc" and "rheumy." There may be utterances like "LOL" (laugh out loud) and "Hugs." Patients may raise issues that they may be reticent to speak with health care practitioners about, day-to-day condition management issues, or personal strategies that they have for taking medicine.

In recent years, it has been observed that patients may be a valuable source of expertise to other patients, and that they provide information that is different from the expertise of clinicians (Civan & Pratt, 2007; Hartzler & Pratt, 2011). This may include: action strategies, recommended knowledge, suggested approaches, and information resources for dealing with problems. This content can be extremely valuable to clinicians and patients alike; however, to date most interfaces for patient-generated content offer few features tailored to the unique nature of the content in these support forums.

Thus, the objective of the current paper was to explore techniques for extracting and visualizing dimensions of patient experience. For this preliminary work, two specific dimensions were selected: interpersonal interactions and medication use.

Interpersonal interactions are an important dimension to consider because others have such a profound impact on patient experience. For example, social support from family, friends and even practitioners can be invaluable to patients; and understanding, (or the lack of it), from practitioners and other people in one's life can be enormously difficult for people, especially those dealing with a stigmatized condition such as fibromyalgia (Barker, 2008). Thus, automatic identification of patient experiences with others, e.g. family, husband, wife, son, daughter, doctor, etc., and ways of highlighting similar types of experiences across patients, might serve various uses. Scientists could use this to study social support, physician-patient communication and other types of interpersonal interactions. Integrated into a search interface, patients could search for others with similar experiences, and see if there are strategies that they could use to address their own problem.

Medication use is another important dimension of patient experience. There has also been an increased interest in the use of online discussion content to predict adverse events and monitor off-label prescription practices (e.g. Wang et al., 2011; Leaman et al., 2010; Chee et al., 2011). This work differs from previous literature in that the method identifies and visually contextualizes patient medication experiences, particularly in terms of stages of use and affect.

Proceedings of the ACL Student Research Workshop, pages 16–22,
Sofia, Bulgaria, August 4-9 2013.

2 Background

This work draws primarily upon two streams of literature: automated analyses of health-related discussion content, and extraction of medication-related information from text. With regard to the former, a large number of studies have employed the software, Linguistic Inquiry and Word Count (LIWC), to compare emotional expression in communities or associations between emotional expression and health outcomes (e.g. Siriaraya et al., 2011; Han et al., 2008).

Other studies of online support groups have focused on social support. Wang, Kraut and Levine (2012) used machine learning with features generated using LIWC and Latent Dirichlet Allocation to investigate whether different types and amounts of social support are associated with length of membership. Namkoong et al. (2010) examined the effects of exchanging treatment information within computer-mediated breast cancer support groups on emotional well-being. Treatment information exchange was assessed using InfoTrend, a software program that employs a rule-based system for computer-aided coding of key ideas.

The task of extraction of medication-related information has often been explored in past literature. For example, the 2009 i2b2 medication challenge focused on the extraction of medication-related information from discharge summaries including: medication name, dosage, mode, frequency, duration and reason (Spasic et al., 2010). This study differs from previous work in that, the focus is not on the time of day or the frequency of medication use, but rather, the stage in the adoption/discontinuation of a medication an individual is at.

3 Method

Discussion content was downloaded through a series of focused crawls of a health-related social networking site (SNS), DailyStrength (`http://www.dailystrength.org`). The content from the corpus encompasses a span of time of approximately 3.5 years, from the site's inception in 2006, to early 2010.

Text pre-processing was done to strip code and extract post metadata. The text was parsed and tagged using the Stanford Parts-of-Speech Tagger (Toutanova et al., 2003). An affective lexicon, WordNet-Affect was used to identify words with emotional content in the text (Strapparava & Valitutti, 2004). There are many specialized resources that could be used to extract medical terminology. However, forum participants wrote in ways that often departed from medical terminology; thus, it was decided that manually constructed lexicons of medication names, side effects and people would be more effective.

4 Results

The results are reported in three parts: descriptive statistics for the corpus, interaction with others and medication information.

4.1 Corpus

The corpus is comprised of discussion posts for three conditions. Since the first part of this study examines interpersonal interactions, three conditions were selected in which key support interactions and level of affect were expected to differ.

Unit/Condition	Breast cancer	Type 1 diabetes	Fibromyalgia
Threads	614	514	763
Posts	2,847	3,259	6,095
Tokens	366,121	389,392	541,233
Types	18,181	18,755	25,942

Table 1: Corpus Statistics

4.2 Modeling Interpersonal Interactions

This work first addresses the challenge of modeling interpersonal interactions. There are two methods of visualization that are explored: the coupling of people and affect, and that of people and actions.

The first step in the pairing of people and affect, was to identify and estimate rates of occurrence of important figures appearing in the text such as: family, husband, wife, mother, friend, and doctor. In order to extract these relations, the researcher manually compiled a list of terms indicating such relations through review of social support literature pertaining to the focal conditions and manual analysis of the text. Alternative names such as "hubby" for husband, "doc" and "dr" for doctor, and "rheumy" for rheumatologist, were included. Many of the instances of the word "family" appeared were references to family doctors or family history; these references were excluded from the estimates (Table 2).

These results show that certain types of individuals tend to appear in forum conversations for certain conditions over others: mothers, family and friends in breast cancer; sons, daughters, and "people" in Type 1 diabetes; and doctors in fibromyalgia. As can be seen, many posts do not include references to other people, but rather, focus on other areas such as patients' own experiences.

Term	Breast cancer	Diabetes	Fibromyalgia
Family	5.09	3.95	1.75
Husband	4.03	3.4	3.02
Wife	0.57	0.92	0.41
Mother	8.62	3.86	1.54
Son	1	3.86	0.61
Daughter	2.23	4.8	1
Friend	3.13	2.23	1.85
Doctor	13.28	13.49	16.42
People	8.42	13.37	8.96

Table 2: Percent of Posts Mentioning Important Roles

Next, the degree of affect expressed in proximity with these roles was investigated. Various sentiment lexicons are available, e.g. SentiWordNet, WordNet-Affect and the LIWC lexicon. Many lexicons classify words as positive or negative or by a limited set of emotions; however, with complex issues like health and interpersonal interactions, there may be multiple dimensions. WordNet-Affect was selected for its diverse set of emotion categories. Following a review of emotion research and preliminary content analysis of the corpus, seven emotions were selected on the basis of frequency and relevance to the conditions (Table 3).

Emotion	Example
Anger	I'm happy but also **mad** cuz I've been suffering and no doctor bothered to tell me about this.
Fear	I have the prescription right here and **afraid** to try it.
Frustration	It is extremely **discouraging** to hear that repeatedly...
Sadness	It's **sad** that I am so excited about getting some sleep.
Anxiety	You may be **worrying** over nothing.
Happiness	I'm so **happy** Lyrica is working for you.
Hope	I really **hope** this works for you.

Table 3: Examples of Emotional Expression

The percentage of posts expressing various affect types was calculated (Table 4). Across all conditions, fear and hope were most common. The highest proportion of fear, happiness and hope were seen in the breast cancer forum. Though anger and frustration were not as common as other emotions, higher levels of these emotions in diabetes are perhaps worthy of note.

Emotion	Breast cancer	Diabetes	Fibromyalgia
Hope	15.55	12.2	13.36
Anger	1.5	2.82	1.38
Frustration	0.78	3.49	1.75
Fear	16.05	11.11	6.32
Happiness	10.31	6.41	5.98
Sadness	10.26	9.91	8.21
Anxiety	9.85	8.29	4.81

Table 4: Percentage of Posts Expressing Emotions

Radar graphs illustrating the extent of emotional expression for the various roles were generated to facilitate comparison (Fig. 1). The light blue line, representing "doctor," is the innermost ring in all cases, demonstrating that emotional expression occurs least often in posts that mention doctors. Posts mentioning family (dark blue lines) were generally associated with higher degrees of emotional expression. Moreover, it is interesting that the patterns of emotional expression are quite different across conditions.

This paper also explores the visualization of discussion content by combining important figures and their actions. For this preliminary work, this action was undertaken with the fibromyalgia forum. The approach taken here was to extract high frequency verbs co-occurring in the same sentence with the target role. High frequency verbs that co-occurred with "doctor" included: "said," "told," "gave," "prescribed," "started," and "diagnosed." The verb "asked" also occurred frequently because forum participants often discussed or suggested questions to ask of doctors. Arranging person-verb pairings together in an interface could be a convenient way to acquire a sense of what patients are being told, and what medications doctors are prescribing. One might even add additional search constraints. For example, with regard to patient experiences with doctors prescribing Lyrica, the system might retrieve: "Recently my doctor put me on Lyrica which did help but had me…," "My doctor gave me a 'taper down' schedule," "My doctor prescribed Lyrica which I refused to take." Patients could use such a system to acquire a sense of the range of experiences that others have had with the drug.

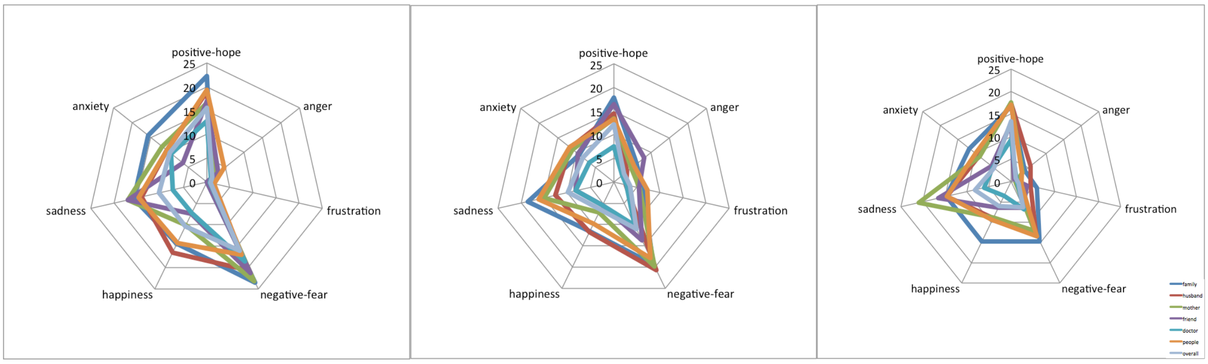

Figure 1: Percent of Posts Containing Affect for Breast Cancer (left), Diabetes (center), and Fibromyalgia (right)

4.3 Modeling Medication Use

Extant literature has found that though some categories of discussion content, e.g. self-introductions, research results and study invites, are common across conditions, other types of discussion content differ, e.g., breast cancer discussion more commonly focuses on treatments, and fibromyalgia discussions tend to focus more on medications (Chen, 2012). Thus, in this study, the researcher selected fibromyalgia as a case study for modeling discussion or comments about medication use.

The researcher created a lexicon of drug names for use in this study, drawn from a review of fibromyalgia literature and information resources, as well as manual review of corpus content. The most common medications are listed in Table 3.

Medication Name	# Posts	% Posts
Lyrica	670	11.36
Cymbalta	329	5.58
Savella	215	3.65
Neurontin	175	2.97
Tramadol	137	2.32
Ultram	79	1.34

Table 5: Common Medications

In order to model temporal differences in patient experience with medications, this study implemented a rule-based system for extraction at five phases in the adoption and use of a medication: adoption, current use, transition, switching, and discontinuation (Table 4). Adoption referred to when an individual began taking a medication. Current use referred to the period in which a person is taking a medication, and has no plans (that he or she reveals at least) to discontinue it. If an individual said that they first had a certain kind of experience with a medication, but that later on it changed, this was referred to as "transition." Discontinuation referred to when an individual stopped using one medication, and switching to when an individual changed from using one medication to another. Information such as whether side effects were temporary, withdrawal symptoms and interactions/contraindications was also extracted. These rules were implemented at the sentence level to prevent misattributions of side effects when multiple medications are mentioned in the same post.

Phase	Rule
Start	"start", "began"
Current Use	"I take", "is working", "currently", "been on" etc.
Transition (A & B)	A: "initially", "at first", "in the beginning" B: "after", "but", "then"
Switching	Fulfills both start and stop criterion or contains "switch".
Discontinuation	"stop", "off" or "quit"

Table 6: Medication Phase Extraction Rules

Using an interface designed for this study, the researcher investigated the reporting of side effects during each phase. Though the most common side effects for a drug were generally reported in multiple phases, certain side effects were reported in a given phase but not another. The last column, "no stage," depicts posts that did not contain explicit references to a specific phase of medication use.

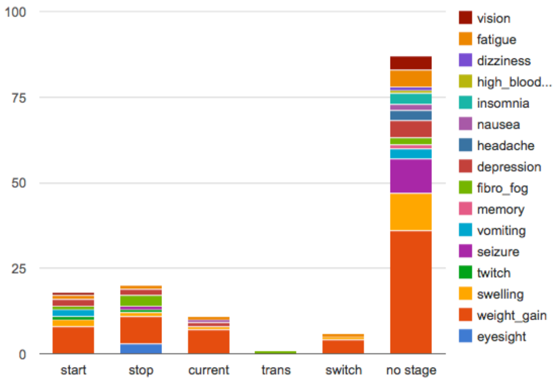

Figure 2: Mentions of Side Effects for Lyrica, Distinguished by Phase

Figure 2 shows that, for Lyrica, the predominant symptom that was reported by patients was weight gain, which appeared in almost all phases. Those who took Savella reported symptoms such as nausea, high blood pressure and dizziness, but there were also a number of reports that these disappear over time (Fig. 3).

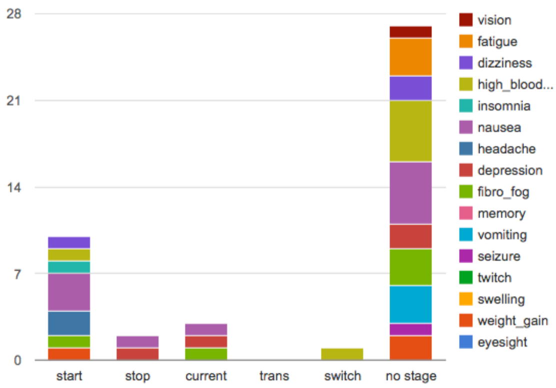

Figure 3: Mentions of Side Effects for Savella, Distinguished by Phase

Another important aspect of patients' experience with certain medications is their attitude towards it. In the previous section, the focus was on emotions because they are important for understanding interpersonal interaction. In the case of medications, rather than tracking the appearance of emotion, it may be useful to consider positive/negative polarity, whether the medication works or not, and side effects.

Thus, in addition to side effects, sentences with positive and negative sentiment were extracted using WordNet-Affect. Words from the "happiness" and "hope" categories of Word-Net Affect were used for positive sentiment, and the "fear," "anger" and "sadness" categories of WordNet-Affect were used for negative sentiment. A lexicon constructed by examining the corpus supplemented the words from WordNet-Affect.

A rule-based system was implemented to identify instances in which participants mentioned whether a medication worked or not. This was implemented using keywords such as "effective," "work" or "help," and recognizing negation. Table 7 lists the number of sentiment and perceptions of efficacy mentions. These do not add up to the number of medication mentions, as many times when medications appear, sentiment is neutral or ambiguous, and perception of efficacy is not the topic of the post. For example, the text might say, "The doctor started me on Savella yesterday."

These results illustrate the utility of extracting multiple facets of patient medication experience, e.g. positive/negative valence, efficacy and side effects, in order to better understand these experiences. Of particular note in these findings are that the estimates of one dimension may appear to conflict with another. For example, overall sentiment towards many medications is negative, but they are reported as working more often than not. The side effects tell yet another story; in many cases, the side effects are different in different phases. Reading the content, one comes to understand that, in an overwhelming number of cases, it is not that patients have found medications that solve all their problems, but that they are selecting ones that work and weighing the costs of the side effects. Thus, an interface that enables users to view all these nuances could be an invaluable asset.

Medication (# mentions)	**Polarity**		**Works**	
	Pos	**Neg**	**Yes**	**No**
Lyrica (934)	42	96	72	21
Cymbalta (413)	11	49	43	20
Savella (338)	21	23	23	2
Neurontin (235)	6	15	13	3
Tramadol (178)	6	3	16	4

Table 7: Sentiment and Efficacy of Medications

The last facet of medication use that was modeled was suggestions and/or recommendations from forum participants. One rule for doing this was by extracting sentences that began with verbs such as: "try," "take," "ask," "tell," and "go." Another was to extract sentences with "suggest" or "recommend." Doing so would retrieve advice such as: "Ask ur doctor about Elavil and Lyrica combination," "She suggested staying on the Lyrica…. while… doing the Vitamin D treatment," and "Word of advice: stop taking SSRI 's at least one week prior to start of Savella."

Forum posts are valuable because they are rich troves of patient experience; however, their richness means that it is also possible to get lost in the story. An interface that organizes the advice, but also allows one to link to the full text, can help users to orient themselves.

5 Discussion and Implications

This study employed NLP techniques in order to model two dimensions of patient experience in online support forums: interpersonal interactions and medication use. With regard to interactions with others, the prominence of different individuals and associated affect differed depending on condition. With regard to medication use, patients' experiences of medication use differed along phase of adoption.

These results may have important implications for the design of support forums. For example, in posts about family that contained fear and anxiety, certain topics tended to occur often: family history, families being supportive or non-supportive, and concerns of worrying the family. Forum participants presented various perspectives and suggestions concerning these issues. Thus, one recommendation is that systems could be designed to organize these various perspectives and suggestions in a form that is easier for the viewer to understand.

The results of this study also yielded various insights concerning fibromyalgia. In particular, the prominence of doctors, and relatively infrequent mention of family and friends was worthy of note. Previous research has found that fibromyalgia patients report a lack of understanding from medical practitioners and others around them (e.g. Madden & Sim, 2006; Sim & Madden, 2008). These reports of interactions with medical practitioners could help researchers to understand where gaps in knowledge and communication exist in both parties, and attempt to rectify them. The content from online support forums may also be helpful for researchers seeking to understand patients' patterns of interpersonal interaction.

The framework presented here for modeling medication use could be useful in many settings. Visualizing side effects at various points in the adoption, use and perhaps discontinuation of a medication could avert potential misunderstandings. For example, sentiment analysis on a medication X might be favorable overall; however, decomposing the posts by phase might show that users initially react favorably, but develop problems with it over time. Of course, the converse, that individuals experience certain side effects initially, but that these disappear over time, could also be true. Such information could be useful to a wide audience, including patients, clinicians, researchers and the pharmaceutical industry.

5.1 Limitations and Future Directions

There are many directions in which the current work could be improved. First, in the case of interpersonal interactions, affect was modeled as dichotomous variable indicating presence or absence. However, the level of emotional expression in a post could vary substantially. Thus, it may be useful to employ a lexicon that provides word rankings, such as SentiWordNet (Esuli & Sebastiani, 2006).

In the case of medication use, extraction of relevant sentences was based on presence of the medication name; thus, the system would not have identified sentences in which pronouns were used. A system that performed coreference resolution might identify significantly more references to medications.

Because previous research has indicated that medications are a common topic in fibromyalgia-related discussion, medication use was a natural target for modeling discussion content. However, it would also be useful to extend the modeling to include treatment experiences. Treatments such as massage and aqua therapy are often used in fibromyalgia, and treatments are the foci for many other conditions, such as breast cancer. Rather than considering phases of medication use, one might consider psychological state and expectations prior to, during and after treatment. Lastly, the interface that was developed for exploring medication use was specific to fibromyalgia; moving forward, it would be useful to expand the interface to other conditions.

Acknowledgments

The author would like to thank Dr. Stephanie W. Haas at the University of North Carolina, Chapel Hill, and the anonymous reviewers for their helpful suggestions in the preparation of this manuscript.

References

Kristin K. Barker. 2008. Electronic support groups, patient-consumers, and medicalization: The case of contested illness. *Journal of Health and Social Behavior, 49*(1):20-36.

Brant W. Chee, Richard Berlin, and Bruce Schatz. 2011. Predicting adverse drug events from personal health messages. In *AMIA Annu Symp Proc. 2011*: 217–226.

Annie T. Chen. 2012. Exploring online support spaces: Using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. *Patient Education and Counseling, 87*(2): 250-257.

Andrea Civan and Wanda Pratt. 2007. Threading together patient expertise. In *AMIA Annu Symp Proc. 2007:* 140–144.

Andrea Esuli and Fabrizio Sebastiani. 2006. SENTI-WORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of LREC-06, 5th Conference on Language Resources and Evaluation*: 417–422.

Jeong Yeob Han, Bret R. Shaw, Robert P. Hawkins, Suzanne Pingree, Fiona McTavish, and David H. Gustafson. 2008. Expressing positive emotions within online support groups by women with breast cancer. *Journal of Health Psychology, 13*(8):1002-1007.

Andrea Hartzler and Wanda Pratt. 2011. Managing the personal side of health: How patient expertise differs from the expertise of clinicians. *J Med Internet Res, 13(3):*e62.

Robert Leaman, Laura Wojtulewicz, Ryan Sullivan, Annie Skariah, Jian Yang, and Graciela Gonzalez. 2010. Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*: 117–125.

Kang Namkoong, Dhavan V. Shah, Jeong Yeob Han, Sojung C. Kim, Woohyun Yoo, David Fan, . . . David H. Gustafson. 2010. Expression and reception of treatment information in breast cancer support groups: How health self-efficacy moderates effects on emotional well-being. *Patient Education and Counseling, 81*(Supp1):S41-S47.

Sue Madden and Julius Sim. 2006. Creating meaning in fibromyalgia syndrome. *Social Science & Medicine, 63*: 2962–73.

Panote Siriaraya, Caleb Tang, Chee Siang Ang, Ulrike Pfeil, and Panayiotis Zaphiris. 2011. A comparison of empathic communication pattern for teenagers and older people in online support communities. *Behaviour & Information Technology, 30*(5): 617-628.

Irena Spasic, Farzaneh Sarafraz, John A Keane, and Goran Nenadic. 2010. Medication information extraction with linguistic pattern matching and semantic rules. *J Am Med Inform Assoc*;17:532-535.

Carlo Strapparava and Alessandro Valitutti. 2004. WordNet-Affect: an Affective Extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*:1083-1086.

Julius Sim and Sue Madden. 2008. Illness experience in fibromyalgia syndrome: A metasynthesis of qualitative studies. *Social Science & Medicine, 67*: 57–67.

Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL 2003*: 252-259.

Wei Wang, Krystl Haerian, Hojjat Salmasian, Rave Harpaz, Herbert Chase, and Carol Friedman. 2011. A Drug-Adverse Event Extraction Algorithm to Support Pharmacovigilance Knowledge Mining from PubMed Citations. In *AMIA Annu Symp Proc. 2011:* 1464–1470.

Yi-Chia Wang, Robert Kraut, John M. Levine. 2012. To stay or leave? The relationship of emotional and informational support to commitment in online health support groups. Presented at *CSCW 2012*. Seattle, WA.