Learning Relationship between Authors' Activity and Sentiments: A case study of online medical forums

Marina Sokolova

IBDA and University of Ottawa sokolova@uottawa.ca

Victoria Bobicev

Technical University of Moldova vika@rol.md

Abstract

Our current work analyses relations between sentiments and activity of authors of online In-Vitro Fertilization forums. We focus on two types of active authors: those who start new discussions and those who post significantly more messages than other authors. By incorporating authors' activity information into a domain-specific lexical representation of messages, we were able to improve multi-class classification of sentiments by 9% for Support Vector Machines and by 15.3 % for Conditional Random Fields.

1 Introduction

User-friendly information and communications technologies and easily available access to the Internet were critical in development of Social Web, a socio-technical phenomenon that enables people to connect, support and learn from each other (Ho et al, 2014). The world-wide social media helped to create a digital resource of texts written by the general public. Those texts aggregate sentiments expressed by millions of people in relations to consumer goods, political campaigns, climate change and other matters of social importance. However, not all participants in online communities contribute equally to that resource: there are visitors who only read the posted texts, authors posting occasional messages and a small group of active authors whose online contributions significantly overweigh contributions of other authors. Those most active participants significantly influence online discussions (Tan et al., 2011: Zafarani et al., 2010).

Our current work studies relations between authors' activity levels and expressed sentiments in an online IVF forum. The forum is a public platform for discussion of In-Vitro Fertilization (IVF) treatment. It has been shown that sentiments on forums dedicated to specific health conditions dependent on the topic of discussion

(Ali et al, 2013). We use a set of sentiment and factual categories tailored for on-line IVF discussions: *encouragement, gratitude, confusion, endorsement* and *facts*.

We are interested in two types of active authors: a) those who start discussions (a.k.a. first authors), b) those who post significantly more messages than other authors (a.k.a. prolific authors). The remaining authors usually post one-two messages, and their contributions are rather sporadic. We have found that distribution of sentiments appeared in text written by different types of authors differs considerably. For example, the authors who start new topics and actively post in the following discussion usually express more *gratitude*: 26% of messages posted by the first authors vs. 9% of messages for all the authors.

We wanted to confirm that information about the author's activity has practical implication and can enhance sentiment and subjectivity lexicons. We used automated prediction of sentiments, where messages were first represented through a domain-specific subjectivity lexicon and then authors' activity information was added to the representation. This enhancement helped to improve the sentiment classification up to 9 % for Support Vector Machines and up to 15.3% for Conditional Random Fields.

2 Related Work

Subjectivity, opinion and attitude classification, mood summarization, emotion and affect detection exemplify Sentiment Analysis and Opinion Mining research (Banea et al, 2012). Those studies increasingly apply to health-related issues, with drug-related sentiment studies emerging as a new sub-topic (Nikfarjam and Gonzalez, 2011). Sentiment dynamics in a health-related online community was studied by Qiu et al. (2011). The authors collected the data from the American Cancer Society Cancer Survivors Network; the data represented a 10-year time span from July

2000 to October 2010. The authors applied binary classification of positive and negative sentiments; e.g., My mom became resistant to carbo after 7 treatments and now the trial drug is no longer working: (, *Negative*; ID-x, I love the way you think, ..., hope is crucial and no one can deny that a cure may be right around the corner!!! *Positive*.

The results demonstrated that the initial negative posts were often followed by positive posts of the same participant. The change was attributed to interaction with other participants of the same thread. The authors hypothesized that the use of multiple categories of sentiments can improve sentiment analysis of the same data.

A predefined set of general sentiment labels may not be sufficient for emotionally charged discussions. Sentiment transition and topic influence on Twitter were studied in (Kim et al, 2011). The results showed that an extensive set of sentiment categories including teasing, complaining, sympathy, and apology provided for a more accurate sentiment prediction and classification than positive, negative and neutral sentiments or six basic emotions: *anger*, *disgust*, *joy*, *fear*, *sadness*, *surprise*. The authors concluded that the 'social' sentiments *sympathy*, *apology*, and *complaining* were influential in sentiment change.

Celli and Zaga (2013) demonstrated that personality traits help in a sentiment analysis task. The authors used the Big5 model (Costa, and MacCrae, 1992) which describes personality along five traits formalized as bipolar scales: extroversion (sociable or shy), neuroticism (calm or neurotic), agreeableness (friendly or uncooperative), conscientiousness (organized or careless) and openness to experience (insightful or unimaginative). Life cycles of online groups had been studied by Patil et al. (2013). The authors determined that 'prolific' members play an important role in maintaining the group stability.

Not all subjective statements are perceived equal: messages posted by frequent contributors may trigger a bigger effect than those posted by occasional authors. At the same time, few sentiment analysis studies of online health-related forums connect activity levels of authors with sentiments and opinions expressed in their messages. In the current work, we study activity characteristics of the forum authors, such as their message productivity, willingness to start new topics and maintain dialogue started by others.

3 The IVF Data Set

In this research, we have used the IVF data set introduced in (Sokolova & Bobicev, 2013). The data is available for research purposes upon request. All the messages were collected from online medical forums dedicated to infertility issues and reproductive technologies. The data set consists of 1321 messages written by 359 female authors and posted on 80 discussions. The average length of the discussion - 16.5 posts (s.t.d. = 9.6). The average number of participants in one topic - 9.5 persons (s.t.d. = 4.2). The average post had 750 characters and 5-10 sentences.

Each post was annotated by two independent annotators. They categorized a post into one category selected among three sentiment categories (encouragement, gratitude, confusion) and two factual categories (facts, endorsement). For example,

post id 300078 "I am so so sorry for your loss, but I want to give you some hope. EXACTLY the same happened to me, only this past May gone. I was ready to give up; I didn't think I could try again. We ended up doing IVF in August and I am now 20 weeks pregnant. Take the time to take good care of yourself over the holidays and enjoy some wine... All the best to you and your dear wee family.RG" endorsmentpost_id_300144 Candis I am sorry about your lossHope you get well soon and have a successful cycle next yearIn the mean time take good care of your selfSam encouragement post_id 300160 Thanks so much everyone ..all your kind words have truly made my day gratitude

The annotators achieved a high Fleiss Kappa = 0.791 that indicates a near-strong agreement¹. There were 433 posts marked as *facts*, 310 obtained the label *encouragement*, 162 posts marked with *endorsement* and 124 as *gratitude*. 176 posts were left ambiguous as the annotators did not agree on their sentiment label.

The analyzed forum discussions are intrinsically heterogeneous. We identify three main factors contributing for the diversity:

• The authors go through different experience (successful IVF treatment vs. complications and uncertainty), exhibit conflicting personal traits (reserve vs. openness) and vary in contributing to the forum. (e.g., many authors add one or two messages per discussion).

1

¹ A *strong* agreement is indicated by Kappa = 0.80.

- Time delay (the last message can be posted weeks or even months after the first one) might weaken relations between messages and the expressed sentiments.
- New participants were bringing new ideas and emotions in already established discussions.

We observed that despite those diversifying factors discussions exhibited a common content flow: it could start by a participant by expressing her doubts and concerns, continued by describing a treatment and concluded by posting the results. Within discussions, the messages were coherent and related, i.e., every posted message answered to one or several previous messages, and in most cases did not diverge from the discussed topic.

To estimate divergence of sentiment categories due to discussion progress, we computed the sentiment categories in the first messages of discussions, last messages of the discussions and all the messages. The first posts of the discussions express the author's confusion more often than not (56% of the post) or describe the author's situation in more objective manner (facts - 17%). With the progress of discussions, confusion decreased to 9% of all the posts, and facts increased to 33% of all the posts. There were no discussions that started with positive sentiments, i.e. gratitude and encouragement, and only one stated with endorsement. Those three categories appeared in the following posts as responses to the *confusion* posts.

They eventually formed 44% of the all messages (encouragement – 24%; endorsement – 12%, gratitude – 9%). The first posts were more difficult for annotation than others, as 26% of the first posts were ambiguous whereas only 13% of all the messages were ambiguous.

We gathered posts from discussions marked "inactive" by the forum. Thus, we considered that the discussions have the "last" post. In most cases, discussion was perceived as completed and became inactive when participants posted a post conveying necessary information (facts – 39%, endorsement – 8%), or a moral support (encouragement – 25%, gratitude - 11%). Only one of the analyzed threads became inactive after a post labeled as confusion. The reported results support our hypothesis that the position of the post in discussion provides additional insight about the sentiments it could contain. We used the po-

sition information in Machine Learning classification of sentiments. Figures 1- 3 visualize those results.

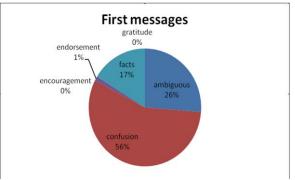


Figure 1. Sentiment distribution in the first messages.

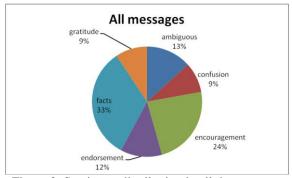


Figure 2. Sentiment distribution in all the messages.

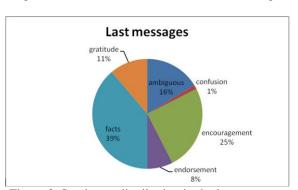


Figure 3. Sentiment distribution in the last messages.

4 Authors' Activity on the Forum

We focused on how information about the authors and their activity on the forum can help in prediction of expressed sentiments. We looked at

- 1) the total number of messages posted by an author;
- 2) initiation of new discussions; and
- 3) contribution to discussions initiated by other authors.

The authors who start discussions (a.k.a. first authors) actively participate in the initiated discussion and guide it in the direction they need. In only 10% of cases they posted only the first mes-

² This discussion has not been re-activated on the time of this paper submission.

sage in the discussion and did not respond on messages posted afterwards.

On average, 25% of messages in discussions were posted by the author of the first post. Figure 4 shows distribution of sentiments in messages of the first authors. Per cent with the posts with *confusion* is larger in comparison with the other authors. Recall that 56% of discussions started with posts marked *confusion*. However, *confusion* posts for these participants decrease considerably as discussions progress and result in 17% of all their posts. The first authors post many messages with *facts* and 26% of their posts express *gratitude* to the participants who helped them with information or moral support.

Our results support those obtained by Qiu et al. (2012) although both studies were conducted on data gathered from different health-related forums.



Figure 4. Distribution of sentiment categories in messages of the first authors.

We intended to explore whether the active participants have any specific characteristics regarding sentiments expressed in their posts and whether sentiments the threads they actively participated in are more predictable. We call the most active authors "prolific". We estimated "prolificness" of the authors as the ratio of the total number of author's posts to the total number of posts of the most prolific author in the studied topics (Patil et al, 2013). Thus, prolificness ranges between [0, 1] and the participant with the greatest number of posts has prolificness equal to 1. In our data, the average prolificness of the prolific authors is 0.44, while the overall prolificness is 0.06. More detailed analysis of these authors' activity can be found in (Bobicev et al., 2015b).

The prolific authors mostly conveyed *facts* and *encouragement*: 39.1% and 24.1% of their messages. In messages posted by the prolific authors, *confusion* appeared less than other sentiments: 22

posts in total, or 5.7% of their messages. *Gratitude* was the second least frequent sentiment among the authors: 6.7% of their posts marked as *gratitude*. The prolific authors showed considerably more confidence and assurance than the authors who posted only 1-2 messages on the forum. Figures 5 and 6 compare sentiments in messages of the prolific authors with sentiments of messages written by the infrequent authors.

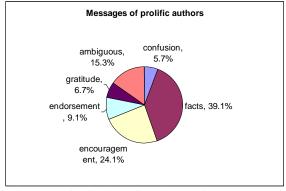


Figure 5. Distribution of sentiments in messages of the prolific authors.

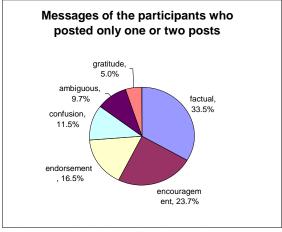


Figure 6. Distribution of sentiments in messages of the infrequent authors.

Comparing the group of the prolific participants and the group of first authors, we observed that 14 of 80 discussions were started by the prolific authors. 10 prolific authors started at least one topic, two of them started two and one started three topics. Thus, they were active not only in participating in various discussions but also in starting the new ones. On the other hand, the average prolificness of the first authors is 0.15 which means that the participants who start new topics are more active in general than in the average participant whose prolificness is 0.06.

It was much easier to predict the characteristics of the message posted by an interlocutor already involved in discussion while a message posted by a person who decided to join this thread was rather unpredictable. Thus, we pooled together messages posted by authors joining discussion for the first time (a.k.a. discussion newcomers) (Figure 7). In comparison with the sentiment distribution of all the authors, there were fewer messages with gratitude and more with confusion as many participants post the first message describing their problems. 75% of the discussion newcomer's posts contained facts or/and encouragement addressed the previous thread participants; thus they were not as much unpredictable as we expected.

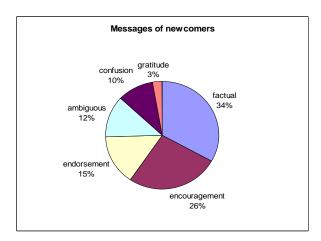


Figure 7. Distribution of sentiments in messages of discussion newcomers.

5 Sentiment Classification

Sentiment analysis of the IVF forum demonstrated that a domain-specific HealthAffect lexicon is effective in prediction of expressed sentiments. HealthAffect (HA) is built by applying Pointwise Mutual Information on a small number of training examples and candidates (unigrams, bigrams and trigrams) with occurrence > 5 in the training data. The detailed description of the HA lexicon creation can be found in (Sokolova & Bobicev, 2013). To represent the data, we used the top frequent 207 terms that appear in Health Affect (HA 207 terms).

We applied 6-class classification to classify 1321 posts into *confusion*, *encouragement*, *endorsement*, *gratitude*, *facts*, and *ambiguous*.

We used Support Vector Machines (SVM) from WEKA toolkit and Conditional Random Fields (CRF) from Mallet toolkit. SVM used the logistic model and normalized poly kernel; CRF had default settings. The best classifier was selected by 10-fold cross-validation.

We obtained the baseline classification accuracy by represented the messages through the

HA 207 terms. We then reinforced the HA representation by adding information about positioning of the post in discussion and information about the author activities. We used two categorical features to represent the position of the post in discussion:

- an indicator showing that the current post holds the first, last or mid position in discussion.
- an indicator showing that the previous post holds the first, last or mid position in discussion:

We used three binary features describing author's activity:

- an indicator that the author of the post started the discussion from which the post was collected:
- an indicator whether the author of the post is a prolific author;
- an indicator that the author of the post joined the discussion from which the post was collected.

Tables 1 and 2 report the classification results for SVM and CRF respectively. For both algorithms, the access to the author information has shown to be beneficial: F-score improved up to 9% for SVM and up to 15.3% for CRF.

The aim of the next set of the sentiment classification experiments was to study what group of authors expressed sentiments in a way more predictable for automated classification. We built three sets:

- First authors: we collected 269 posts from 10 discussions which had the largest number of posts posted by the initial author;
- Prolific authors: we gathered 224 posts from 10 discussions which the largest number of posts posted by prolific authors among all the discussions;
- Discussion newcomers: we collected 130 posts from 10 discussions which had the largest number of authors joining the discussion.

The posts were represented by 207 HA terms. The results of 6-class sentiment classification in Table 3 show that SVM classifies sentiments more accurately when the initiators of discussions actively participate in following message exchange. CRF better recognizes sentiments if many new authors join the discussion.

Features	SVM			
	P	R	F	
HA 207	0.40	0.43	0.41	
HA 207 + pos of current	0.42	0.45	0.43	
post				
HA 207 + pos of current	0.41	0.44	0.42	
and prev. post				
HA 207 + pos of current	0.43	0.46	0.44	
post + if the author is the				
first				
HA 207 + pos of current	0.42	0.45	0.43	
post + if the author is the				
new one				
HA 207 + pos of current	0.41	0.45	0.42	
post + if the author is				
prolific				
HA 207 + pos of current	0.44	0.47	0.45	
post + if the author is the				
first, new or prolific				
HA 207 + if the author is	0.43	0.45	0.43	
the first, new or prolific				

Table 1: 6-class sentiment classification by SVM. Baseline results are in *italic*. The best F-score is in **bold**, the 2nd best is in *that font*.

Features	CRF			
	P	R	F	
HA 207	0.31	0.29	0.30	
HA 207 + pos of current	0.32	0.30	0.31	
post				
HA 207 + pos of current	0.34	0.32	0.33	
and prev. post				
HA 207 + pos of current	0.36	0.33	0.34	
post + if the author is the				
first				
HA 207 + pos of current	0.33	0.31	0.32	
post + if the author is the				
new one				
HA 207 + pos of current	0.35	0.31	0.33	
post + if the author is				
prolific				
HA 207 + pos of current	0.35	0.31	0.33	
post + if the author is the				
first, new or prolific				
HA 207 + if the author is	0.36	0.34	0.35	
the first, new or prolific				

Table 2: 6-class sentiment classification by CRF. The best F-score is in **bold**, the 2nd best is in *that font*.

Discussion	SVM			CRF		
sets	P	R	F	P	R	F
First authors	0.44	0.46	0.44	0.33	0.39	0.35
Prolific au-	0.38	0.35	0.36	0.33	0.33	0.33
thors						
Discussion	0.39	0.39	0.39	0.40	0.33	0.37
newcomers						

Table 3: Sentiment classification related to three types of the author activities.

6 Discussion and Future Work

Currently 19%-28% of Internet users participate in online health discussions (Balicco, Paganelli, 2011). Analysis of sentiments and opinions posted online can help in understanding of sentiments and opinions of the public at large. Such understanding is especially important for the development of public policies whose success greatly depends on public support, including health care (Atkinson, 2009; Eysenbach, 2009).

In this work, we have focused on relations between sentiments and authors' activity on online health-related forums. We worked with 6 sentiment and factual categories: *encouragement*, *gratitude*, *confusion*, *endorsement* and *facts*.

We have identified three groups of the forum authors: the most prolific authors, the authors who start new discussions, and the authors who join discussions started by other authors. We have shown that distribution of sentiments differs considerably for those categories of the authors. Annotation agreement is the strongest (Kappa = 0.806) on messages with the greatest presence of the new authors, as well as ability of CRF to identify the six sentiments (F-score = 0.37). At the same time, SVM achieved the most accurate classification on messages with the greatest contribution from the first authors (F-score = 0.44 in six-class classification). We have shown that adding the author information to a semantic representation of the messages can significantly improve sentiment recognition (up to 15.3%).

As a future work we intend to study participants' interaction in more details. In (Bobicev et al., 2015a) we analyzed message sequences and found patterns of sentiments in the consecutive posts. However, many posts were addressed to the one specific interlocutor by her name. We plan to analyze these direct communications and interaction of sentiments expressed in these sequences of posts.

Also, we plan to investigate the ambiguous messages and find a suitable solution for their sentiment annotation. One of the solutions would be to allow multiple annotations for one post. In this case we can use both labels assigned by the annotators to the ambiguous post and find a way to automate learning of multiple annotations. Taking into consideration that the messages are comparatively long (5 to 10 sentences) the other possible solution is to annotate some parts of one message with different labels. This could be done by automatically applying a sentiment lexicon.

References

- Ali, T., D. Schramm, M. Sokolova and D. Inkpen. 2013. *Can I hear you? Sentiment Analysis on Medical Forums*, International Joint Conference on Natural Language Processing, pp. 667-673.
- Atkinson N.L., Saperstein S.L., Pleis J. 2009. *Using the internet for health-related activities: findings from a national probability sample*. J Med Internet Res. 2009 Feb 20;11(1):e4.
- L. Balicco, C. Paganelli. 2011. Access to health information: going from professional to public practices. SIIE'2011.
- Banea, C., R. Mihalcea, and J. Wiebe. 2012. *Multilingual sentiment and subjectivity analysis, in Multilingual Natural Language Applications: From Theory to Practice*. D. Bikel and I. Zitouni (eds). Prentice-Hall. 2012.
- Bobicev, V., Sokolova, M., Oakes. M. 2015a. What Goes Around Comes Around: Learning Sentiments in Online Medical Forums, *Journal of Cognitive Computation*, 2015.
- Bobicev, V., Sokolova, M., Oakes. M. 2015b. Sentiment and Factual Transitions in Online Medical Forums. Proceedings of Canadian AI 2015 conference.
- Celli, F., C. Zaga. 2013. Be *Conscientious, Express your Sentiment!* ESSEM@AI*IA 2013: 140-147.
- Costa, P. T. and MacCrae, R. R. 1992. Normal personality assessment in clinical practice: The neo personality inventory. Psychological assessment, 4(1):5.
- Eysenbach, G. 2009. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behaviour on the Internet. *Journal of Medical Internet Research*, 11(1).
- Ho, K., Peter Wall Workshop Participants. 2014. Harnessing the Social Web for Health and Wellness: Issues for Research and Knowledge Translation. *Journal of medical Internet research* 16.2.
- Kim, S., J. Yeong Bak, Y. Jo, Alice Oh. 2011. *Do you feel what I feel? Social Aspects of Emotions in Twitter Conversations*, Workshop on Computational Social Science and the Wisdom of Crowds. NIPS.
- Nikfarjam, A., and Gonzalez, G. H. 2011. *Pattern mining for extraction of mentions of adverse drug reactions from user comments*. In AMIA Annual Symposium Proceedings (Vol. 2011, p. 1019).
- Qiu, B., K. Zhao, P. Mitra, D. Wu, C. Caragea, J. Yen, Greta E. Greer, K. Portier. 2011. Get online support, feel better sentiment analysis and dy-

- namics in an online cancer survivor community. Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing.
- Patil, A., J. Liu, J. Gao. 2013. *Predicting Group Stability in Online Social Networks*. Proceedings of the 22nd international conference on World Wide Web, , pages 1021-1030.
- Sokolova, M., V. Bobicev. 2013. What Sentiments Can Be Found in Medical Forums? RANLP 2013: 633-639.
- Tan, C., L. Lee, J. Tang, L. Jiang, M. Zhou, P. Li. 2011. User-level sentiment analysis incorporating social networks. The 17th ACM SIGKDD international conference on Knowledge Discovery and Data Mining, pp. 1397-1405.
- Zafarani, R., W. Cole, and H. Liu. 2010 Sentiment Propagation in Social Networks: A Case Study in LiveJournal. Advances in Social Computing (SBP 2010), pp. 413–420.