Analyzing Perceived Intentions of Public Health-Related Communication on Twitter Preprint

Elena Viorica Epure, Rebecca Deneckere, Camille Salinesi

Analyzing Perceived Intentions of Public Health-Related Communication on Twitter
Preprint

Elena Viorica Epure, Rébecca Deneckere, and Camille Salinesi

Université Paris 1 Panthéon-Sorbonne, CRI, 90 Rue de Tolbiac, 75013 Paris, France,
{Elena.Epure@malix.univ-paris1.fr,Rebecca.Deneckere@univ-paris1.fr,
Camille.Salinesi@univ-paris1.fr

Abstract. The increasing population with chronic diseases and highly engaged in online communication has triggered an urge in healthcare to understand this phenomenon. We propose an automatic approach to analyze the perceived intentions behind public tweets. Our long-term goal is to create high-level, behavioral models of the health information consumers and disseminators, relevant to studies in narrative medicine and health information dissemination. The contributions of this paper are: 1) a validated intention taxonomy, derived from pragmatics and empirically adjusted to Twitter public communication; 2) a tagged health-related corpus of 1100 tweets; 3) an effective approach to automatically discover intentions from text, using supervised machine learning with discourse features only, independent of domain vocabulary. Reasoning on the results, we claim the transferability of our solution to other healthcare corpora, enabling thus more extensive studies in the concerned domains.

Keywords: intention mining, text mining, natural language processing, classification, machine learning, Twitter, speech acts, linguistics

1 Introduction

The Internet has nurtured a highly available and accessible environment for disseminating health information. While in 2001, 70 000 websites contained health-related information [4], the order of magnitude for *healthcare*-verbatim websites only, has increased by three by 2013 in the United States [7]. Correlated to the massive production of online health-related content, the number of online health information seekers has doubled in this period, reaching 100 million [4, 7]. Nowadays, the dissemination and consumption of health information have been also impacted by the tremendous adoption of social media [18]. This has led to the creation of communities, fostered by the interpersonal interactions with acknowledged advantages such as anonymity and 24-hour availability [4].

Social media allows to disseminate health information, express beliefs, feelings about health matters and react to existing content. One's beliefs are built on or altered by the information to which he or she is exposed [4]. This could be further reflected in new behavioral intentions and eventually new behaviors [1].

Social media has even a stronger influence on consumers because of the social norms: the attitudes and behaviors of the community towards health matters are transparent in this online environment. This aspect could be exploited to promote healthy behavior such as quitting smoking. However, inaccurate information, available to a very large and generally vulnerable target–people impacted directly or indirectly by chronic diseases, could become harmful and have mass consequences [4]. Therefore, there is an urge in healthcare *to understand the effects* of the exposure to health information disseminated through social media, on consumers. The reason is *to predict these effects* as potential immediate or long-term behaviors [4, 3]. However, suitable methods are necessary for this.

In the current paper, we propose a means *to discover automatically the perceived intentions of the Twitter public posts.* We call them *perceived* because they are interpreted from the stance of the information consumer. The link between perceived intentions and revealing or predicting behavior is as follows. First, the perceived intentions are a behavioral component of health disseminators. Second, the perceived intentions allow to create automatic techniques to measure the impact of various message formulation on health information consumers, similar to message framing [16]. For instance, is a consumer more likely to read an online, health-related article if its link is tweeted in a rhetorical question or in an informative fact? Third, we want to support the automatic discovery of collective narratives concerning healthcare, as they can be strong drivers for behaviors [11]. Automatic techniques already exist for identifying components of the crowd narratives such as topics or events [19]. However, disseminators' disposition towards presented events is necessary for a thorough narrative's representation [14].

2 Related Work

Related works to analyze the exposure to health information disseminated on social media rely on traditional research methods [3, 4, 16]. First, data is gathered through interviews and questionnaires. In order to reach people consuming online health information, clinics, hospitals or online communities are targeted. Then, the data is analyzed using statistics or qualitative methods. Although, these methods empowered the healthcare community to gain valuable insights, there are several limitations. The research results are mainly descriptive. However, for acting on the existing knowledge, *predictive models* are required too. Further, an extensive study is rarely possible, the reported samples being at best of several thousand subjects. These studies also come with localization and time-span constraints. In reality an online community could include worldwide members. Moreover, the questions of how certain results change over time, or how the discovered knowledge depends on current temporal trends, or how predictions could be made in real-time are challenging to answer.

Computer science methods to complement existing studies, by exploiting automatically social media in creating predictive models, exist [5, 10, 23]. However, predicting behavior in social media requires deep consideration of underlying processes, of how the cause–the disseminated health information–leads to the effects–the behavior changes. Hence, a solution should go beyond the black-box approach often employed in data mining and incorporate also theoretical knowledge from humanities. In conflict resolution, a model for predicting behavior changes from Twitter, based on narrative theory, is proposed [11]. Focused on perception, Myslin *et al.* [15] discover narrative-related elements from tweets about smoking: genres (e.g. first-hand experience, opinion) and themes (e.g. cessation, pleasure). Priesto *et al.* [19] identify health topics from tweets, similar to narrative's themes (e.g. depression, flue). Though, a solid start for our long-term goal, these works can be augmented with more high-level behavioral cues.

3 Research Design

Our objective is to analyze the perceived intentions of the publicly disseminated tweets. Specifically, several research questions are identified:

1. Could a *valid taxonomy of perceived intentions*, *representative* for Twitter public communication be defined?
2. How do various *supervised machine learning algorithms* with various configurations of features compare for discovering perceived intentions?
3. What are *the most predictive features* for each type of intention?

Data Collection. We collected two sets of tweets via Twitter Streaming API. The first consists of 2714 tweets and the second of 43153 tweets. The first set was fetched by the keyword *autoimmune*. The second set was collected with commonly used medical terms and jargon, proposed by Ridpath et al. [20]. Even though the communication on autoimmune diseases was initially targetted, we included the second set for data diversity in evaluation, thus aiming at the input generalizability. Further, we sampled randomly 600 tweets from the first set and 500 tweets from the second set, with no message format duplicates. The selected 1100 tweets were used in the intention taxonomy's validation and in creating the ground-truth corpus for the machine learning experiment.

Research Method. For the first research question, we considered as *valid* an intention taxonomy that is consistently applied by raters, showing thus an alignment in the tweets' perception and ensuring experimental reproducibility. For this, the selected corpus was tagged in parallel by two raters: one expert involved in defining the taxonomy, the other seeing the taxonomy for the first time (researcher in computer science). The taxonomy was briefly presented as the inexperienced rater was expected to rather rely on its intuition when tagging. At least one intention had to be chosen per tweet. More intentions were allowed when the tweet had multiple sentences or one intention could not be clearly conveyed. After the tagging, the Fleiss' Kappa statistical test [6] was used for evaluating the validity. When multiple tags were used, an alignment of the sets was necessary. For instance, *[t_1, t_2]* and *[t_2, t_1, t_3]* would align as: t_1-t_1, t_2-t_2, t_3-*unknown*. The test considers the degree of agreement for each such pair and the probability of agreeing by chance. A score of the Fleiss' Kappa test over 0.6 is considered a good result, specifically between 0.61 and 0.8 *substantial* and between 0.81 and 0.99 *almost perfect* [6]. Further, for assessing if the taxonomy was *representative*, a tag *other* was created to be used when none of the proposed intentions was a suitable choice and its frequency was computed.

For the next research questions, several steps were required. First, the truth set was created based on the tagged corpus. For each non-agreement, a discussion took place between the raters and a collective final decision was made. In the final corpus[1], 89% of the tweets were single-tag and the rest had associated 2-3 intentions. Second, text processing and Tweet NLP [17] were applied for feature extraction. Two types of features were defined: *Content* and *Discourse* features. *Content* features consisted of standard text mining features: *BagOfWords* and *OpinionKeywords*. These were computed after the corpus' lower case conversion and lemmatisation. *BagOfWords* was a dictionary of tokens and their frequencies. The tokens were extracted from the pre-processed Twitter corpus before the classification. *OpinionKeywords* included the frequencies and ratios of negative and positive opinion words from a predefined lexicon [13]: *freqPositiveWords*, *ratioPositiveWords*, *freqNegativeWords*, *ratioNegativeWords*. *Discourse* features are novel and defined by considering linguistic means to express intentions. They are described in Section 5, after introducing the proposed taxonomy.

Logistic Regression, *Linear SVM*, *Random Forest* and *Multinomial Naive Bayes* were the selected classifiers. The *scikit-sklearn*[2] implementations were used with default parameters. *Multinomial Naive Bayes* was selected as its library's implementation allowed continuous features too. *Naive Bayes* and *Random Forest* handled inherently multiple classes while *Logistic Regression* and *Linear SVM* in an on-versus-all strategy. Various configurations of features were evaluated: discourse features only (*DiscF*), content features only (*ContF*), all features (*AllF*). Also, the features were scaled beforehand. The metrics for performance evaluation of the *single-tag* corpus were *precision*, *recall* and *f-score*. These were computed as *macro scores*, weighted by the support of each intention and averaged over *10 folds*. Same evaluation decisions were applied within each fold in cross-validation. The *hamming loss* was used to evaluate the accurate prediction of *multi-tag* corpus. For this, the single-tag corpus was the train set and the multi-tag corpus the test set. The *hamming loss* is the ratio of intentions in average that are incorrectly predicted. Finally, with no interaction in models, the most predictive features per intention were found by analyzing the weights of the best classifier, trained on standardized features' values [8].

4 An Intention Taxonomy for Public Tweets

Human behavior is intrinsically intentional as thoroughly discussed in philosophy [2] and psychology [1]. Though, behavior is not necessarily linked to only physical human acts but also to language. Generally, people communicate with various intentions. Utterances are considered thus as acting through words while their leading intentions are called *speech acts* [21]. For example, *This hospital has*

[1] http://tinyurl.com/hk9t83y
[2] http://scikit-learn.org/stable/supervised_learning.html

a nonstop emergency service asserts the speaker' belief about the world while *Could you please give me a painkiller?* requires the listener to act. Searle [21] proposed five *classes* of speech acts. *An assertive* is used for stating information being true or false about the state of affairs in the speaker's world. *A commissive* denotes the engagement of the speaker to a future course of action. *A directive* implies the listener carrying out an action as a result of the speaker's utterance. *An expressive* is used to express the speaker's feelings towards the state of affairs in the world. Finally, *a declarative* is the type of utterance, changing the world' state such as firing someone. *The speech act theory* emerged over time, as a highly-adopted framework to extract or predict behavior from text.

However, we aimed at more granular intentions than Searle's classes [21]. This was enabled by the work of Vanderveken [22] who proposed a lexicalization of the intentions through 300 English verbs. These verbs are organized in five hierarchies where the roots are the speech act classes and each level is a specialization of the parents. Our taxonomy emerged from the Vanderveken's theoretical work [22] but its refinement was based on manual corpus analysis. Thus, we empirically discovered that *assertive* and *directive* classes cover most of Twitter public communication (see Table 1). These findings are not surprising considering that public tweets rarely contain personal feelings (*expressive*) or personal goals (*commissive*). Such communication takes place rather privately. Moreover, declarative speech acts are very rare even in live settings.

Table 1. Identified intentions from Twitter public communication.

Class	**Intention**	**Tweet Example**
Assertive	**assert**	New study reveals autoimmune/inflammatory syndrome triggered by HPV vaccine URL
Assertive	**hypothesize**	Vitamin B1 may help relieve fatigue in Hashimoto's thyroid patients URL
Directive	**propose**	The gut microbiota and inflammatory bowel disease #microbiology #autoimmune URL #gutmicrobiota
Directive	**direct**	Are you a Cure Champion? Sign up for the Walk to Cure Psoriasis in a city near you...URL
Directive	**advise**	How To Avoid Holiday Autoimmune Flares URL
Directive	**warn**	Why you shouldn't be going from competition to competition URL #thyroid #metabolism #autoimmune

The rational behind the association of intentions to tweets is presented further. An *assert* is a tweet that clearly conveys the message such as news or personal public declarations. Though often it has a url, the linked resource appears with the role to sustain or detail the message. A *hypothesize* is a tweet containing a weak assertion such as probable statements or hypothetical questions. Compared to the assertives, a *propose* is a tweet that always references an external resource, which must be accessed in order to consume the message. The *propose* tweets usually provide key or opinion words about the resource's content and could be considered a weak attempt to make the reader access the url. In contrast, a *direct* is a strong attempt to make the reader act to demands, requests, invitations, encouragements or questions. Finally, *advise* and *warn*, which are very similar, are directives to a future action or resource consumption that is supposed to be good or bad for the reader.

5 Discourse Features for Intention Discovery

A person could use Twitter for addressing utterances to the community or to specific users. In a live communication, the utterance's intention is implicitly understood from the utterance's content, speaker's gestures and voice. Though not as rich as this case, the written tweets have also characteristics that convey their intentions. We relate them to *Discourse* features, which are further presented.

PronominalKeywords are frequencies of various pronominal forms. The first person singular (*freq1stPersonSg*) is chosen because it could be a sign of personal declarations specific to assert tweets (e.g. *I'm for vaccination.*). Similarly, the third person (*freq3rdPerson*) could be linked to assert tweets when reporting. By contrary, the first person plural form (*freq1stPersonPl*) or second person (*freq2ndPerson*) could be cues of direct tweets (e.g. *We must vaccinate our kids! You should too*). Further, *PunctuationMarks* are indicators of the discourse' functions: the presence of exclamation (*hasExclamation*), interrogation (*hasQuestion*), ellipsis (*hasEllipsis*; for omissions, hesitations); colon (*hasColon*; for titles, explanations) or quotes (*hasQuotes*). *QuestionKeywords* complements the punctuations for revealing discourse functions (e.g. questions). We separate the frequency of *what, when, where, why, who* (*freq5W*), from that of *how* (*freq1H*) because *how* is also used in advice or proposals (e.g. *How I Gave Up Smoking*).

The frequency of *EmoticonCues* (*freqEmoticons*) could be inversely correlated with the impersonal reporting; hence possibly linked to assert and propose tweets. Both ASCII and Unicode emoticons were checked. *TitleCues* (*hasTitle*) seems to be an often marker for news, being thus a potential discriminator for the assert and propose tweets (e.g: *New Releases in Science*). *VerbPhrases* (*hasVerb*, present and past participles not considered) could be an indicator of weak directives when false. These tweets often lack the subject-predicate form. *VerbMoods* contains *hasImperative*, which is frequent in requests, demands; thus being an indicator of direct tweets. For identifying imperatives, we created a rule-based algorithm, having as input the part-of-speech (POS) tags. *VerbKeywords* encompasses features regarding the modals (*hasCan*, *hasCould*, *hasMust*, *hasMay*, *hasMight*, *hasShould*). Modals could show various intentions: hypothesize, advise, direct etc. For verb features, negative forms were identified too.

SyntaticConstructs features are created with the goal of incorporating syntactic characteristics of the discourse. The assumption was that tweets with same intention might share similar discourse form. We used the POS tags in order to dynamically discover representative POS-related features. The way we proceeded was: POS-tag the corpus using a dedicated tweet NLP parser [17]; compute the normalized frequencies of each two consecutive POS tags from the output; select those with a score of at least *0.5* per intention. The final syntactic features are: *hasNV*, *hasNN*, *hasAN*, *hasNComma*, *hasPN*, *hasDN*, *hasVN*, *hasCommaU*, *hasNP*. The encoding of these features is: *N* nouns; *V* verbs, *A* adjectives, *Comma* punctuation, *P* pre-, post-position or subordinating conjunctions, *D* determiners, and *U* urls. Compared to the original output of the parser [16], two changes were made. *N* incorporates also the proper nouns (the symbol ˆ) and pronouns (the symbol *O*). *Comma* replaces ,.

6 Results and Discussion

The first research question sought to answer if the proposed intention taxonomy was *representative for Twitter public communication* and *valid*. The condition of being *representative* could be considered fulfilled through artifact design, by being both theory- and corpus-driven. Relying on theory, we ensured that the taxonomy was linguistically representative for written utterances. Relying on corpus analysis, we ensured that the taxonomy was mapped on the actual Twitter public communication. Further, we agreed that a high frequency of the tag *other* would denote a lack of representativeness. However, *other* was used only in 0.005 of the cases by both raters, supporting thus our taxonomy's design. The tag *other* replaced *expressives* such as greetings or *commissive* such as public promises. The Fleiss' Kappa test was performed to assess the taxonomy's *validity* (see Table 2). The overall *intention-wise* Fleiss' Kappa score was 71.5% (z=40.9, p=0.001) being considered a *substantial* result. All intentions apart from *advise* and *other* achieved substantial scores. The result for *other* is not surprising given its very low frequency. The overall *class-wise* Fleiss' Kappa score was 78.3% (z=28.1, p=0.001), showing thus that mismatches occurred sometimes between intentions of the same class. Related works of speech act tagging for tweets reported similar results (Kappa scores between 0.6 and 0.85) [5, 10, 23].

Table 2. Fleiss' Kappa reported as *result, z-score* for intentions and classes.

Assertive: **0.8**,27.2		**Directive**: **0.8**,27.1				**Other**:0.41,13.9
assert	**hypothesize**	**propose**	**advise**	**direct**	**warn**	**other**
0.77, 26.4	**0.78**, 26.7	**0.70**, 24.1	0.49, 16.7	**0.68**, 23.3	**0.76**, 25.9	0.35, 12.1

Several decisions were made for the final corpus creation. 82% of mismatches between raters concerned the *advise* tweets, specifically with propose (55%), direct (25%), assert (15%), warn (5%). By further manual analysis, we concluded a tweet was an advice either because it redirected the reader to an external source that actually contained an advice or it directly contained the advice. The first case corresponded to mismatches involving *propose* (e.g. *4 Steps to Heal Leaky Gut and Autoimmune Disease URL*), while the second case to mismatches regarding *direct* (e.g. *#TipTuesday: Vitamin D deficiency is linked to autoimmune diseases. Add mushrooms to Thanksgiving!*) and *assert* (*The quicker you receive* *treatment, the better your chances for a good recovery from #Stroke are*). It might be surprising an advice is stated as an assert but this is an example of *indirect* speech acts. Nevertheless, what emerges is that an *advise* tweet could be ultimately a *propose*, *direct* or *assert* too. Considering its low Fless' Kappa score, we decided for now to transform the *advise* tweets in their secondary speech acts. Moreover, we transformed the *warn* tweets too because of insufficient instances (0.02% of the corpus) and for maintaining consistency with *advise*.

The next research question looked into the comparison of classifiers with different features' configuration. The results are summarized in Table 3. Statistically, Logistic Regression and Linear SVM are comparable ($p > 0.5$, two-tailed t-test) and both outperformed Random Forest. Multinomial Naive Bayes is not reported as yielded results similar to Random Forest for *DiscF* and *ContF*, and worse for *AllF*. We can notice that *DiscF* systemically leads to similar results as *AllF* (apart from *hypothesize* with Logistic Regression) and improves results over *ContF* up to 5 times with Random Forest. For multi-label classification, the minimum *hamming loss* scores are obtained when using *DiscF* with Logistic Regression (0.287) and Linear SVM (0.264).

The last research question aimed at the evaluation of the features' predictive power in relation to each intention. This was assessed based on the features' weights estimated for Linear SVM. As expected, for *assert*, the most predictive discourse features are related to reporting information or making public declarations: hasVerb, hasNV, hasQuotes, freq3rdPerson, freq1stPersonPl, freq1stPersonSg. The most representative content features seem to be related to news and particularly to scientific ones ("implications", "future", "bacteria", "influence"). For *hypothesize*, both the discourse and content features reveal the importance of modals ("might", "may" and "could"). Apart from these, the most important discourse feature is hasQuestion. For *direct*, hasImperative is the most discriminatory feature. This also emerges from top predictive content features that contain multiple verbs ("know", "learn", "read", "check", "enjoy"). Then, features linked to requests, encouragements and questions follow in importance in the classification of *direct*: for discourse features–hasQuestion, hasExclamation, freq2ndPerson, freq1H, freq5W; for content features –"what", "please", "let". For *propose*, the top most important discourse features are related to news summaries (hasTitle, freq5W, hasColon) and advice or warnings (freq1H, hasMust, hasShould, freq2ndPerson). These features are correlated to the content ones, which incorporate interrogation cues ("why", "how"), impersonal scientific words ("epidemic", "lupus") and advice/warning words ("step", "recipes", "good"). Finally, the POS-related features appear highly predictive for all intentions, in particular for *assert*, *direct* and *hypothesize*.

In conclusion, the proposed *discourse features* improved significantly the discovery of intentions and we often observed that they were also correlated to the top most important content features. However, the discourse features benefit of being much fewer (30 vs. 4608) and corpus-independent, allowing thus reproducibility on other medical English corpora. In the analyzed corpus, the most popular intentions are *assert* (46%) and *propose* (41%) revealing thus that Twit-

Table 3. Results of the classification experiment using various feature sets.

		LogisticRegression			**RandomForest**			**LinearSVM**		
Intention	**Metric**	DiscF	ContF	AllF	DiscF	ContF	AllF	DiscF	ContF	AllF
assert sup.=416	*precision*	0.75	0.67	0.77	0.70	0.57	0.66	0.77	0.69	0.79
	recall	0.83	0.71	0.84	0.79	0.64	0.86	0.82	0.66	0.83
	f-score	0.79	0.69	<u>**0.80**</u>	**0.74**	0.60	**0.74**	**0.80**	0.68	**0.81**
hypothesize sup.=49	*precision*	0.62	0.62	0.77	0.65	0.50	0.60	0.71	0.55	0.75
	recall	0.41	0.10	0.47	0.49	0.06	0.12	0.76	0.33	0.55
	f-score	0.49	0.18	**0.58**	**0.56**	0.11	0.20	<u>**0.73**</u>	0.41	0.64
direct sup.=119	*precision*	0.71	0.66	0.78	0.70	0.59	0.67	0.75	0.50	0.77
	recall	0.79	0.18	0.75	0.70	0.11	0.35	0.82	0.27	0.79
	f-score	0.75	0.28	**0.76**	**0.70**	0.18	0.46	<u>**0.78**</u>	0.35	<u>**0.78**</u>
propose sup.=391	*precision*	0.80	0.63	0.79	0.74	0.57	0.75	0.80	0.62	0.78
	recall	0.72	0.79	0.75	0.66	0.71	0.70	0.72	0.76	0.76
	f-score	0.76	0.70	<u>**0.77**</u>	0.70	0.64	**0.72**	0.76	0.68	<u>**0.77**</u>

ter is publicly used for information dissemination. However, it is quite interesting that the strategies are different, half of the tweets' messages being self-standing (*assert*) while the other half requiring external redirection (*propose*). The ratio of *direct* tweets (18%) shows also a significant expected reaction from consumers, by replying or following advice, warnings, requests or invitations. Similar to us, Godea *et al.* [9] identify tweets' purposes in healthcare (advertising, informational, positive or negative opinions). However, we focus on intentions as established by pragmatics, ensuring thus a domain-independent, general approach.

7 Conclusion and Future Work

An approach for analyzing the perceived intentions in the Twitter public communication was proposed. An *intention taxonomy* for public tweets was defined and validated. Its automatic discovery proved effective, with f-scores between 0.73 and 0.8 using *Linear SVM* and *discourse features* only. The most predictive content features were often linked to the discourse ones and intentions, acting thus as a positive feedback loop to the proposed taxonomy and features' decisions.

Future work must address several limits. As *advise* and *warn* had low scores, the taxonomy must be revised and experiments deployed with more raters. Twitter private communication should be researched too, including intentions from *expressive* and *commissive*. The impact of parameters' values on classifiers' performance must be assessed. Finally, as the proposed approach is domain-independent, we envision a large-scale analysis of public, health-related tweets. Information dissemination in different communities (e.g. diseases) can be compared. The public's reactions to the same, but differently formulated messages can be analyzed (e.g. which formulation is most re-tweeted?). Then, within a community, conflicting narratives could emerge. The work of Houghton *et al.* [11] can be extended to identify them while also considering disseminators' disposition as perceived intentions. The current work is our first attempt in joining narrative medicine, for *bringing the patient as a subject back into medicine* [12].

References

1. Ajzen, I.: The theory of planned behavior. Organizational Behavior and Human Decision Processes J. 50(2), 179–211 (1991)
2. Bratman, M.: Intention, plans, and practical reason. Harvard Univ. Press (1987)
3. Calvert, J.K.: An ecological view of internet health information seeking behavior predictors: Findings from the chain study. Open AIDS J. 7(1), 42–46 (2013)
4. Cline, R.J.W.: Consumer health information seeking on the internet: the state of the art. Health Education Research 16(6), 671–692 (2001)
5. Ding, X., Liu, T., Duan, J., Nie, J.Y.: Mining user consumption intention from social media using domain adaptive convolutional neural network. In: the 29th AAAI Conf. pp. 2389–2395. AAAI Press (2015)
6. Fleiss, J.L., Cohen, J., Everitt, B.S.: Large sample standard errors of kappa and weighted kappa. Psychological Bulletin 72(5), 323–327 (1969)
7. Fox, S., Duggan, M.: Health online 2013. Tech. rep., PEW (2017), http://www.pewinternet.org/2013/01/15/health-online-2013/
8. Gelman, A.: Scaling regression inputs by dividing by two standard deviations. Statistics in Medicine 27(15), 2865–2873 (2008)
9. Godea, A.K., Caragea, C., Bulgarov, F.A., Ramisetty-Mikler, S.: An analysis of twitter data on e-cigarette sentiments and promotion. In: Holmes, J.H., Bellazzi, R., Sacchi, L., Peek, N. (eds.) AIME 2015. pp. 205–215. Springer, Cham (2015)
10. Hemphill, L., Roback, A.J.: Tweet acts: How constituents lobby congress via twitter. In: CSCW'14. pp. 1200–1210. ACM, NY, USA (2014)
11. Houghton, J., Siegel, M., Goldsmith, D.: Modeling the influence of narratives on collective behavior case study. In: Int. System Dynamics Conf. (2013)
12. Kalitzkus, V.: Narrative-based medicine: Potential, pitfalls, and practice. The Permanente Journal 13(1) (2009)
13. Liu, B., Hu, M., Cheng, J.: Opinion observer: Analyzing and comparing opinions on the web. In: the 14th Int. Conf. on WWW. pp. 342–351. ACM, NY, USA (2005)
14. Murphy, J.: The role of clinical records in narrative medicine: A discourse of message. The Permanente Journal (2016)
15. Myslin, M., Zhu, S.H., Chapman, W., Conway, M.: Using twitter to examine smoking behavior and perceptions of emerging tobacco products. Medical Internet Research 15(8), e174 (2013)
16. O'Keefe, D., Jensen, J., Jakob, D.: The relative persuasiveness of gain-framed loss-framed messages for encouraging disease prevention behaviors. Health Communication J. 12(7), 623–644 (2007)
17. Owoputi, O., Dyer, C., Gimpel, K., Schneider, N., Smith, N.: Improved part-of-speech tagging for online conversational text with word clusters. In: NAACL (2013)
18. Perrin, A., Duggan, M., Greenwood, S.: Social media update 2016. Tech. rep., PEW (2017), http://www.pewinternet.org/2016/11/11/social-media-update-2016/
19. Prieto, V.M., Matos, S., Alvarez, M., Cacheda, F., Oliveira, J.L.: Twitter: A good place to detect health conditions. PLOS ONE 9(1), 1–11 (01 2014)
20. Ridpath, J.R., Wiese, C.J., Greene, S.M.: Looking at research consent forms through a participant-centered lens. J. of Health Promotion 23(6), 371–375 (2009)
21. Searle, J.R.: Speech acts. Cambridge University Press, 1 edn. (1969)

22. Vanderveken, D.: Meaning and speech acts. Cambridge University Press (1990)
23. Wang, J., Cong, G., Zhao, W.X., Li, X.: Mining user intents in twitter. In: the 29th AAAI Conf. pp. 318–324. AAAI Press (2015)