Social Media and Patient Health Outcomes
Findings from the Yearbook 2014 Section on Consumer Health Informatics

P. Staccini[1], N. Douali[2], Section Editors for the IMIA Yearbook Section on Consumer Health Informatics
[1] INSERM UMR 912 SESSTIM, IRIS, UFR Médecine, Université Nice-Sophia Antipolis, Nice, France
[2] INSERM UMRS 1142 LIMICS, Université Pierre et Marie Curie, Paris, France

Summary

Objectives: To provide a review of the current excellent research published in the field of Consumer Health Informatics.

Method: We searched MEDLINE® and WEB OF SCIENCE® databases for papers published in 2013 in relation with Consumer Health Informatics. The authors identified 16 candidate best papers, which were then reviewed by four reviewers.

Results: Five out of the 16 candidate papers were selected as best papers. One paper presents the key features of a system to automate the collection of web-based social media content for subsequent semantic annotation. This paper emphasizes the importance of mining social media to collect novel data from which new findings in drug abuse research were uncovered. The second paper presents a practical method to predict how a community structure would impact the spreading of information within the community. The third paper presents a method for improving the quality of online health communities. The fourth presents a new social network to allow the monitoring of the evolution of individuals' health status and diagnostic deficiencies, difficulties or barriers in rehabilitation. The last paper reports on teenage patients' perception on privacy and social media.

Conclusion: Selected papers not only show the value of using social media in the medical field but how to use these media to detect emergent diseases or risks, inform patients, promote disease prevention, and follow patients' opinion on healthcare resources.

Keywords

Social media, epidemiology, substance-related disorders, patients rights, health communities

1 Introduction

Consumer Health Informatics (CHI) is a subspecialty of medical informatics which studies consumers' needs for information to improve healthcare [1]. Consumer Health Informatics includes various health informatics research areas at the intersection of telemedicine, social media, serious games, Public Health, etc. The success of social networks and their development in all age categories of the population makes them interesting for researchers and officials in public health [2]. Several studies have examined the use of social media not only in disease prevention but also to study populations by extracting information about lifestyle and its relationship with health [3].

Enterprises have created healthcare social media analytics databases [4] by studying conversations within patient communities. They provide their clients (public health organizations, governmental agencies, communication agencies, pharmaceutical companies, healthcare researchers, and academics actors) with tailored data across variables such as diseases and drugs, professions, physicians, pharmacies and hospitals, time, relevancy, influence, and trust. Folksonomies are now available based on hashtags. From a scientific point of view, questions are raised about the level of relevance of the knowledge extracted from these databases in comparison with "professional" databases (DRG-based databases for example). If the impact of social media in helping patients to manage their disease has to be measured in terms of credibility and efficiency, relationships between privacy and massive data analysis still remain a hot topic. In this paper we attempted to illustrate these questions across the presentation of the five best papers published in 2013.

2 Method

We queried PubMed and Web of Science to conduct our review. We searched for articles on CHI and Big Data, the latter being the special topic for the 2014 IMIA Yearbook. Our approach was based on a customized query.

For the PubMed search, we first identified two queries: "consumer health informatics [Text Word]" and "social media"Text Word]. Since "social media" is not a MesH keyword, we did not restrict our query to only MesH words. From the retrieved papers, we analyzed the MeSH terms used to index the latest articles. The main related MeSH terms found were: Internet, Consumer Health Information, Medical Informatics, Health Education, Consumer Participation, Information services. We added other terms to specify usage topics or tools: Community Networks, Consumer Product Safety, Consumer Satisfaction, mHealth, Patient Portals, Patient Safety, Personal Health, Personal Health Records, Personalized Medicine, Social Media, Big Data. We identified 12 papers and only 2 papers were related to both the field of Consumer Health Informatics and "Big Data". To date, "Big Data" is not considered a scientific keyword. It was not mentioned in

all retrieved papers. We reviewed the initial query and created a new query including all journals of Health Informatics. To mention "Big Data" was not a required condition. We found 333 papers. As the journal "Studies in Health Technology and Informatics" concerns only health informatics conference papers (MIE, MEDINFO, etc.), we excluded it, and we finally selected 130 papers with the last query.

For the Web of Science search, we originally used a simple query ("Social Media" and "Consumer Health Informatics"). 496 citations were found with a very high number of papers that were irrelevant. We optimized the query by limiting the search to articles in the following categories: Computer Science Software Engineering, Computer Science Information Systems, Computer Science Theory Methods, Computer Science Artificial Intelligence. Many of the retrieved articles were published in statistics and mathematics resulting in the design of a filter to get only the papers from the field of "medical informatics".

Each section editor independently analyzed the list of the papers resulting from the union of the two sets of retrieved papers to select the candidate best papers. After a consensual step, 16 candidate best papers were finally selected. Then, candidate best papers underwent a classical reviewing process by external reviewers.

3 Results

The reviewing process highlighted five best papers whose content overlapped what we intended to address: relationships between massive data analysis and consumer health information. The first one is related to the design, the development, and the evaluation of a new semantic web platform aimed to automate the extraction of semantic information from web-forum content to facilitate drug abuse research using social media. Beyond the fact that this tool is dedicated to drug abuse, this paper gives a consistent and valid model to extract and analyze social media content. This model combines lexical, pattern-based, and rule-based approaches. Furthermore, the results

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2014 in the section 'Consumer Health Informatics'. The articles are listed in alphabetical order of the first author's surname.

Section
Consumer Health Informatics

- Cameron AM, Massie AB, Alexander CE, Stewart B, Montgomery RA, Benavides NR, Fleming GD, Segev DL PREDOSE: A semantic web platform for drug abuse epidemiology using social media. J Biomed Inform 2013;46(6):985-97.
- Huh J, Yetisgen-Yildiz M, Pratt W. Text classification for assisting moderators in online health communities. J Biomed Inform 2013;46(6):998-1005.
- Subirats L, Ceccaroni L, Lopez-Blazquez R, Miralles F, García-Rudolph A, Tormos JM. Circles of health: Towards an advanced social network about disabilities of neurological origin. J Biomed Inform 2013;46(6):1006-29.
- Van Der Velden M, El Emam K. "Not all my friends need to know": a qualitative study of teenage patients, privacy, and social media. J Am Med Inform Assoc 2013;20(1):16-24.
- Weng L, Menczer F, Ahn, YY. Virality prediction and community structure in social networks. Scientific reports 2013;3:2522.

---

propose an answer to an epidemiological question, in this case drug abuse epidemiology. Authors showed how they are able to explore trends of either a specific drug (e.g. loperamide) by week, month, quarter or year, or to detect patterns of co-occurrent drugs (e.g. Buprenorphine–Benzodiazepine). In future, they plan to implement a module for entity disambiguation and enhance the existing modules for relationship, triple patterns extraction, and sentiment extraction. Prevalence of risky situations, prevalence of diseases among sub-populations, and incidence of emerging healthcare problems are some of the most popular epidemiology questions that can be addressed by such a tool.

The second paper addresses the dynamics of the spreading of information units (called memes). As recent studies demonstrate that diseases and behaviors spread differently (simple versus complex contagions), the authors wished to determine the success of a meme and how a meme spreading is related with the underlying community network structure. Authors provided a direct approach for translating data about community structure into predictive knowledge about what information will spread virally. They identified five models of community structure according to the spreading of social contagions within or across communities: structural trapping, social reinforcement, homophily, retweet network, follower network. The authors declare that their method does not exploit message content and thus can be easily ap-

plied to any socio-technical network from a small sample of data.

The third paper was selected to complete the previous approaches of forum content analysis. It discusses the role of community managers, especially their ability to moderate online health communities. As a lot of forums are used by patients to post questions and wait for answers, some conversations need expertise and it becomes rapidly impossible for a single community manager to handle and treat all conversations. The purpose of this study was to define and implement a methodology aimed at detecting the threads that need moderator's help. Different dictionaries have been validated with categories such as social, health, bio, negative emotion, positive emotion, for instance. While using these dictionaries, authors determined a classification method to prioritize which posts they need to respond among the thousands of messages being posted each day. The performances of this method were stratified according to the time moderators could spend to read posts. In order to apply the method to medical forums other than diabetes, the authors suggested exploring applying keywords or medical terms, using built-in tools such as Metamap, or consumer health vocabularies as additional features to discriminate posts with medically-oriented topics.

With the selection of the fourth paper, we intended to highlight how simple social media tools could inspire researchers to create new social-network platforms

for people with specific diseases. The authors explored how basic features of circles of friends can affect a population of patients with disabilities of neurological origin. The objectives of the research were: 1) to provide stakeholders (people with disabilities, their family, health professionals and therapists, care givers, and institutions) with a new social-network platform called "circles of health" aimed at supporting multicenter studies using standard indicators; 2) to promote knowledge democratization, user empowerment, and informed decision making in order to facilitate the e-inclusion of people. The authors described the data model used to implement the "circles of health" and the dedicated management functions. The main result obtained when evaluating this new social-network was that people with disabilities found it helpful to enrich their personal knowledge with the experience of other users and to perform online follow up questionnaires after clinical discharge.

The last paper selected is a qualitative study on the ways teenagers manage personal health information and their privacy on social media sites. The questions asked by the authors were: 1) do teenage patients use social media and to they share health information?, and 2) do privacy concerns explain the fact that teenage patients choose not to share personal health information on social media? The first finding was that despite teenage patients spend many hours a day on the internet, they don't define themselves as patients and they seem to be neither interested in sharing information about their diagnosis, nor to find peers with a similar diagnosis, or to meet them online or offline. Social and psychological privacy they seek is not so much expressed in terms of controlling the access to their personal health information for people they know, but about being in control of defining who they are, and how they want to be perceived. Another interesting observation from the interviews was that teenage patients do not use email. The study found that the disclosure of personal health information in social media is best explained by the control paradox and the privacy dilemma, rather than the privacy paradox.

Conclusion

If we recall internet groups maintained by former access providers such as Compuserve for instance, current social media do not differ by their content but are more democratized and widely used by every one either with classical internet access or mobile phone access. The questions of research include three categories: Who is using social media, and for patients with disease for what purpose? How can we use content to improve patients' information literacy, disease awareness, and level of treatment compliance? Is social media content useful for public health purposes such as to detect epidemiological trends or emerging diseases? This year, we selected five best papers, which addressed part of these questions and explored links between data provided by users and patients, data mining methods, and qualification of use by sub-groups of users and patients. Analysis of massive data (such as social media content for instance) has adapted text classification methods and could rapidly provide for regional institutions and hospitals smart tools to describe and follow patients' satisfaction as well as epidemiological variations.

References

1. Eysenbach G, Jimison H, Kukafka R, Lewis D, Stavri PZ, editors. Consumer health informatics. Springer Science; 2005.
2. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social Media: A Review and Tutorial of Applications in Medicine and Health Care. J Med Internet Res 2014;16(2):e13.
3. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85.
4. Symplur. Healthcare Social Media Analytics. http://www.symplur.com/healthcare-social-media-analytics/ (last visit: 07/06/2014).

Correspondence to:
Pr Pascal Staccini
INSERM UMR 912 SESSTIM, IRIS, UFR Médecine,
Université Nice-Sophia Antipolis
28 avenue de Valombrose
06107 Nice cedex 2, France
E-mail: pascal.staccini@unice.fr

Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2014

Cameron D, Smith GA, Daniulaityte R, Sheth AP, Dave D, Chen L, Anand G, Carlson R, Watkins KZ, Falck R

PREDOSE: A semantic web platform for drug abuse epidemiology using social media

This paper describes the development of a semantic web technologies platform called PREDOSE (Online Prescription Drug Abuse Surveillance and Epidemiology). PREDOSE is designed to facilitate the epidemiologic study of prescription (and related) drug abuse practices using social media. The study aims to automate the extraction of semantic information from web-forum content to facilitate drug abuse research using social media. The PREDOSE platform provides three functions: a data collection tool; an automatic coding tool and a data analysis and interpretation tool. The use of this platform allows to identify trends for a concept (single medication) or a combination of drugs.

Weng L, Menczer F, Ahn, YY

Virality prediction and community structure in social networks

This paper presents a practical method to translate information on the structure of a community into a predictive knowledge about what information will spread virally in that community. The authors show that most memes spread like complex contagions. However, a few viral memes spread across many communities, similar to a highly infectious disease. The future popularity of a meme can be predicted by quantifying its early spreading pattern in terms of community concentration. Authors conclude that their approach could improve the understanding in computational social science, social media

analytics, and marketing applications, while using some indicators such as "usage and adoption entropy" or "fraction of intra-community user interactions". Social complex processes such as ones existing in health education between professionals, patients, and institutions could be better understood by this kind of methodology in order to spread memes connected with behavior changes.

Huh J, Yetisgen-Yildiz M, Pratt W
Text classification for assisting moderators in online health communities
J Biomed Inform 2013;46(6):998-1005

Social media environments increasingly offer unique venues through which patients can gain health-related information. All posts in online community forums usually require a response. Whereas some responses need to be moderated, others, such as recipes for healthier diet, could be supported purely by patients' responses. The authors present an exploration of text classification methods for prioritizing online posts that require a moderator's help. The method was to employ a binary classifier on WebMD's online diabetes community data. They used a dictionary-based approach (word unigrams and features derived from a sentiment analysis tool called Linguistic Inquiry Word Count)

coupled with a naïve Bayes classifier to define a "low-cost" text classification method, and implemented an algorithm to assist moderators in prioritizing posts that need help. Despite the method applies to large-scale social media environments, it uncovered social, legal, and ethical challenges in providing moderators' expertise. Performance must be improved to discriminate posts with various medically-oriented topics.

Subirats L, Ceccaroni L, Lopez-Blazquez R, Miralles F, García-Rudolph A, Tormos JM
Circles of health: Towards an advanced social network about disabilities of neurological origin
J Biomed Inform 2013;46(6):1006-29

In this paper, the authors present a multi-center study that uses a periodically collected data set forming the basis of rehabilitation processes. Standardized, multidimensional indicators are automatically and dynamically collected, transformed, represented, and shared. The bio-psycho-social data is standardized into multidimensional indicators (ICF). The aim is to promote knowledge democratization, user empowerment, and informed decision making, with *Circles of Health* (some kind of Circles of Friends or Google + network) that allow individuals to

monitor their rehabilitation progress, and compare it with others including people without disabilities. Graphical representations are useful to diagnose and compare deficiencies quickly. Users can find information to generate knowledge and to promote knowledge democratization and informed opinions. The ease at which knowledge can be gained facilitates the e-inclusion of people with disabilities of neurological origin.

Van Der Velden M, El Emam K
"Not all my friends need to know": a qualitative study of teenage patients, privacy, and social media
J Am Med Inform Assoc 2013;20(1):16-24

This paper addresses the problem of confidentiality and the use of social media by teenage patients. It's a qualitative study based on a content analysis of semi-structured interviews with 20 hospital patients. The majority of teenage patients do not use social media to build a connection with others with similar conditions. They do not use the Internet to find health information about their diagnosis. This study concludes that teenage patient's online privacy behavior is an expression of their need for self-definition and self-protection. Another finding is that teenage patients do not use emails.