Appendix: Content Summaries of Best Papers for the Education and Consumer Health Informatics Section of the 2019 IMIA Yearbook

Abdellaoui R, Foulquié P, Texier N, Faviez C, Burgun A, Schück S

Detection of Cases of Noncompliance to Drug Treatment in Patient Forum Posts: Topic Model Approach

Non-compliance (or non-adherence) to long-term treatment is a worldwide problem detrimental to the overall effectiveness of the health system. Social media holds a lot of promise in improving communication and patient engagement. The example of benfluorex illustrates how social media could be valuable sources for experts. Methods to identify messages with adverse events mentions have been developed and it has been showed that social media may even impact treatment adherence. The objective of this study was to evaluate a topic model approach to detect patient non-compliant behaviours (dose change and treatment cessation) associated with antidepressant drug (escitalopram) and antipsychotic drug (aripiprazole) in online forums. Authors implemented a probabilistic topic model to identify the topics that occurred in a corpus of messages mentioning these drugs, posted from 2004 to 2013 on three of the most popular French forums. Around 6% (154/2691) of online posts were detected on escitalopram non-compliance and 7% (122/1778) on aripiprazole. The topic models approach detected cases of non-compliance behaviours with average recall and precision scores of 98.5% (272/276) and 32.6% (272/844), respectively. Authors concluded topic modelling was a valuable sensitive method to detect non-compliance. However, it lacks specificity and manual review was required to distinguish between true and false positives in each dataset. They suggested that syntactic and semantic methods could be developed to recognize the experience, the temporal features, and the object concerned by the action in the sentences.

Jones J, Pradhan M, Hosseini M, Kulanthaivel A, Hosseini M

Novel Approach to Cluster Patient-Generated Data Into Actionable Topics: Case Study of a Web-Based Breast Cancer Forum

Despite the proliferation of social media use, such as blogs and forums, little is known about the scope and quality of information shared, or the purposes that social media sites serve for consumer decisional and support needs. This study explores approaches for analysing the free-text social media data to discover hidden, less obvious, aspects of health consumers' lives and extract potential valuable information on managing health and well being beyond the context of health care. This was applied to breast cancer management and recovery in five online breast cancer forums (mainly breastcancer.org community). Natural language processing and statistical modelling approach were used to cluster >4 million postings into manageable topics. Topic modelling (cluster of words that frequently occur together) was performed with the machine learning language toolkit open source tool. It was followed by multiple linear regression analysis to detect highly correlated topics among the different website forums. Quantitative content analysis of the forums resulted in 20 categories of user discussion. Topic model organized posts into 30 topics which were grouped into four distinct clusters of highly correlated computationally modelled topics. These clusters were labelled "symptoms and diagnosis", "treatment", "financial", "friends and family". Multiple regression analysis was performed to identify the most significant topics discussed among the forum participants. They were arranged in a descending order based on the Akaike information criterion value: 1) lingering side effects while in remission, 2) chemotherapy side effects and change of treatment, 3) radiation and side effects, 4) genetic risk and testing, 5) support from caregiver and medical team for long term recovery, and 6) looking for support from people in similar circumstances.

Park A, Conway M, Chen AT

Examining Thematic Similarity, Difference, and Membership in Three Online Mental Health Communities from Reddit: A Text Mining and Visualization Approach

Studies have consistently shown individuals can gain positive effects from interacting with other individuals in similar circumstances. Online interactions have been shown to improve depression, anxiety, stress, and negative mood, as well as to facilitate coping and empowerment. Moreover, members of online health communities consistently emphasize the benefits of participation with respect to their treatment decisions, symptom management, clinical management, and outcomes. In this study, authors examine the nature of online discussion (main themes expressed in the communities) and compare issues (thematic overlap, similarity and differences among the communities) pertaining to three mental health conditions: anxiety, depression, and post-traumatic stress disorder (PTSD). The corpus was based on Reddit (http://www.reddit.com), a popular social networking, online gathering, and news exchanging platform. Between the months of Oct 2015 to Dec 2015, a total of 7,410 posts and 132,599 associated comments made by 41,967 unique members were downloaded. Discussion themes were identified using knowledge resources like Unified Medical Language System or clusters analysis. Similarity among clusters in the network visualization used Louvain modularity algorithm. For each of the three main themes (anxiety, depression, and PTSD), 15 clusters had been generated. Using r/Anxiety subreddit discussion content, clusters including "social anxiety", "medication", "school", "panic attack", and "therapy/therapist" contained terms and labels which clearly differentiated the clusters from one another. A few clusters, such as "positive emotion" and "gratitude" shared terms. For the r/Depression subreddit, clusters including "birthday", "school",

"sleep", "work", and "gratitude" were clearly differentiated from one another. Clusters such as "talking to friends" and "friends and family" shared identical or semantically similar terms. For the r/PTSD subreddit, many clusters including "trauma therapy", "work", "sleep", "trauma trigger", "EMDR therapy", "nightmare", "animal", "research", were clearly distinguishable. A few clusters, such as "sleep" and "nightmare" shared similar terms but also had distinctive and non-overlapping terms. Venn diagrams were built to summarize and highlight common themes: "school" and "social related" between Anxiety Disorder and Depression, "living with" between Anxiety Disorder and PTSD. The global intersection between the three communities shared overlapping concerns and discussion patterns such as: "gratitude", "sleep", "work" and "positive emotion". However, Depression clusters focused on self-expressed concerns (*e.g.*, events associated with depressed moods), whereas Anxiety Disorders and Post-Traumatic Stress Disorder clusters focused around treatment- and medication-related issues.