

HARVESTING PATIENT OPINIONS ABOUT IMPLANTS FROM SOCIAL MEDIA SOURCES

Yves Noirjean, Markus Degen

Institute for Medical and Analytical Technologies, University of Applied Sciences and Arts
Northwestern Switzerland, Switzerland

yves.noirjean@fhnw.ch

Abstract: The increasing use of social media (e.g. Facebook, Twitter, blogs, discussion boards) provides a new valuable information source for implant manufacturers about the public reception of a specific implant. On the other hand, patients and caregivers can get first hand experiences from other patients. However, because of the overwhelming amount of data available, tools are required to locate useful information. Often it is more important to get the "Wisdom of the crowd" instead of a single opinion. This paper describes a system which can be used to harvest data from social media and use different filters and analysis tools to aggregate search results.

Keywords: social media, sentiment analysis, cochlear implant, knee replacement, hip resurfacing

Introduction

Implant manufacturers are very interested in getting first hand and subjective information about the reception of their products at the patient side. The use of social media by users of medical products (e.g. patients who carry an implant) hold significant information for designers of the respective products [1]. There are common text analysis tools which can be used to get statements about a text without having to read the whole text. While most tools can be applied independent of the domain, three groups of medical implants were chosen as a showcase in this project: cochlear implants, knee replacements and hip resurfacing.

Methods

Figure 1 shows the architecture of the software system built.

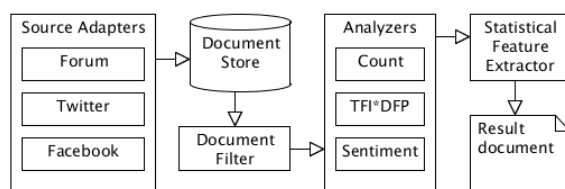


Figure 1: System overview

Sources were manually chosen, and then continuously harvested, converted to a unified data model and written to a document store.

For the Twitter source, a Naive Bayes classifier [2] was trained with 1196 tweets. The tweets were manually tagged

as *wanted* or *unwanted* where the *wanted* tweets contained personal experiences of a patient with an implant.

After filtering with the classifier, the texts were grouped (e.g. by implant model or manufacturer) and analyzed. The remaining texts were used to form a control group.

Metrics

The following metrics were applied to each document:

TF*IDF: Calculates the Term Frequency tf (Eq. 1) multiplied by the Inverse Document Frequency idf (Eq. 2) for a given set of terms. tc represents the term occurrence count, dc the count of documents containing the term and tdc is the total document count.

$$tf(tc) = \begin{cases} \log_{10}(1 + tc), & \text{if } tc \text{ is } >0 \\ 0, & \text{if } tc \text{ is } 0 \end{cases} \quad (1)$$

$$idf(dc, tdc) = \log_{10}\left(\frac{tdc + 1}{dc + 1}\right) \quad (2)$$

$tf * idf$ equals to the prominence of the analyzed terms [3].

Sentiment Analysis: SentiStrength [4] was used to estimate positive and negative sentiments expressed.

A t-test resulting in a p -value was done for each pair of groups to determine the observed significance level of each comparison. Those results, along with the mean and the standard deviation for each group, are stored in the result document.

Results

The project is currently in its last stage (evaluation). Therefore the results shown here are preliminary.

Source evaluation

Figure 2 shows the quantity of hits for a selection of keywords in the domain of cochlear implants and hearing aids for different sources. The result demonstrates how unspecific queries lead to many hits which can be reduced by using more specific ones. For the alld deaf internet forum [5] this is true to a lesser extent because it is specific to the domain of hearing impairment.

Tweet classification quality

143 of the 1196 manually tagged Tweets were *wanted*. The classifier trained with half of the tagged set had a sensitivity

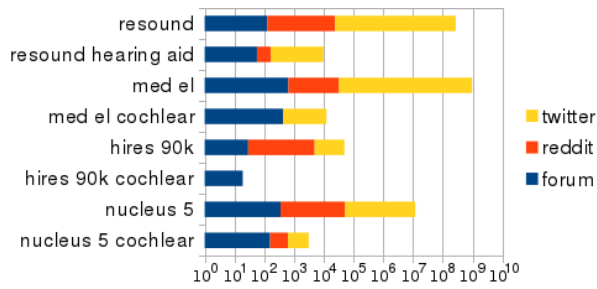


Figure 2: Hit quantity for keywords in terms of manufacturer MED-EL (cochlear implants) and ReSound (hearing aids) and products HiRes 90K (cochlear implant) and Nucleus 5 (sound processor). One can see that the precision of the queries has different impacts depending on the source: While the hit reduction on the domain specific forum is minor, it is rather big on Twitter and reddit [6]).

of 0.97 and a specificity of 0.72 when tested on the other half. During four months, 329666 Tweets were collected. Of those, 251022 (76%) held original content (no retweets) and 25149 (7.6%) were classified as *wanted*.

Statistical results of text analysis

One of the words yielding significant results when applying the TF*IDF method on the alldeaf forum data was "adjustment". Figure 1 shows p -values for the comparison between three different hearing aid manufacturers (A, B, C) and the control group. The low p -values of comparisons with group B indicate that it is different.

Table 1: p -values for TF*IDF of word "adjustment".

	control	A	B	C
control	1.0000	0.3884	0.0000	0.5400
A	0.3884	1.0000	0.0172	0.6939
B	0.0000	0.0172	1.0000	0.0002
C	0.5400	0.6939	0.0002	1.0000

The mean and standard deviation values indicate a group of outliers within group B. Manual analysis of the texts for investigation confirmed the results and revealed reports of people having frequent visits to an audiologist for readjustments.

Table 2: mean and sd for TF*IDF of word "adjustment".

	control	A	B	C
mean	0.0028	0.0074	0.0222	0.0049
sd	0.0469	0.0849	0.1368	0.0592

Discussion

We have built a system which can be used to harvest patient opinions from several online sources. The system is fully configurable and easily extendable for further online sources and analysis methods. First results from analyzing the harvested data look promising.

Of the 445267 documents collected so far, 115601 are from the alldeaf forum and 329666 are from Twitter. The content of those documents totals to 106 MB of raw text data. With the planned addition of Facebook as a source and continuous harvesting of documents, these numbers will increase. The classifier based filtering of Twitter content improves the quality of the results at the cost of quantity. For the already analyzed search terms, the most valuable data is found in the domain specific, manually selected forum.

The statistical results show that it is possible to find relevant aspects of implant product groups. Queries done while testing the system showed that it is important to carefully choose meaningful search terms - on one hand they have to be specific and on the other hand they have to be generic enough to allow a certain degree of fuzziness.

Outlook

This project is now in its final stage and the main focus is now on the analysis of the harvested data. The final results will be available at the end of the project by mid 2013. Extensions to the system are conceivable in several areas:

- Improve classification methods, i.e. according to [7], which would probably lead to a better specificity
- Weight posts by analyzing the expertise of authors
- Add domain specific knowledge to the analyzer (i.e. extract "range of motion" after knee replacement)
- Add more data sources, i.e. use an aggregating search engine

Bibliography

- [1] C. Hawn, "Take two aspirin and tweet me in the morning: how twitter, facebook, and other social media are reshaping health care," *Health Affairs*, vol. 28, no. 2, pp. 361–368, 2009.
- [2] A. McCallum, K. Nigam, *et al.*, "A comparison of event models for naive bayes text classification," in *AAAI-98 workshop on learning for text categorization*, vol. 752, pp. 41–48, Citeseer, 1998.
- [3] J. Ramos, "Using tf-idf to determine word relevance in document queries," in *Proceedings of the First Instructional Conference on Machine Learning*, 2003.
- [4] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment strength detection in short informal text," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2544–2558, 2010.
- [5] "alldeaf internet forum." <http://www.alldeaf.com/>. Accessed: 27/03/2013.
- [6] "reddit social discussion site." <http://www.reddit.com/>. Accessed: 27/03/2013.
- [7] F. Peng and D. Schuurmans, "Combining naive bayes and n-gram language models for text classification," *Advances in Information Retrieval*, pp. 547–547, 2003.