

A Domain-Based Approach for Retrieving Trustworthy Health Videos from YouTube

Randi Karlsen^a, Jose Enrique Borrás Morell^a, Luis Fernández Luque^b, Vicente Traver Salcedo^c,

^aDepartment of Computer Science, University of Tromsø, Norway

^bNorthern Research Institute, Tromsø, Norway

^cITACA - Health and Wellbeing Technologies, Universidad Politécnica de Valencia, Spain

Abstract

Health information retrieval and YouTube can be used as powerful tools to improve user's health knowledge. However, YouTube videos must be carefully analysed in order to avoid misleading, inaccurate, obsolete and incorrect health content.

We present an approach for re-ranking health videos obtained from YouTube, called Domain-based ranking. Our system automatically identifies videos coming from trusted sources (channels), such as hospitals and health organizations, and re-ranks YouTube results so that such videos are presented first in the ranking list. Video and channel metadata are used to automatically determine if a video is provided by a trusted source. The approach is tested and results show that the amount of relevant and reliable videos ranked within top-10 increase when using Domain-based ranking, compared with the original YouTube ranking.

Keywords:

Health video retrieval, information retrieval, social media, medical informatics.

Introduction

YouTube is increasingly being used to share health information offered by hospitals, organisations, government, companies and users [1]. However, YouTube video ranking favours content from popular channels, causing for instance hospital videos, where social interaction through likes/dislikes and comments are less common, to appear low in the ranked list. Also, misleading and incorrect videos may well be popular and therefore given a high ranking.

User studies have shown that the credibility of an information source is one of the most powerful factors affecting information trustworthiness [2]. Users are for example more likely to trust health information published or authored by physicians or major health institutions. We therefore assume that videos from hospitals and health organizations are in general more trustworthy than the average health video on YouTube, and aim at identify videos from such trusted sources and improve their ranking so that they are more easily available to users.

Methods

Domain-based ranking is based on automatically classifying videos as belonging to four different domains: hospitals, organizations, active users and others. For each of the three first domains, our system automatically generates a white list of channels by analyzing channel properties and social features, such as channel description, number of shared videos, number of views per day, comments, likes/dislikes and favourites.

Using the white lists, a new re-ranked list of videos is generated. Videos from hospitals and health organizations are highest ranked, followed by videos from active users on the specific disease, and finally videos from other users.

In a test, the Domain-based ranking system retrieved and re-ranked the top-100 YouTube results for 20 different queries, all with diabetes as topic. Domain-based and YouTube ranking were compared based on two components; relevance to the query and perceived reliability. Video reliability was analysed by i) checking video comments for any comments on wrong or misleading information in the video, and ii) watching and evaluating the content.

Results

When comparing the top-10 results from Domain-based ranking with top-10 from the original YouTube ranking, we found a significant increase of videos from trusted sources (see Table 1).

Table 1 – Average amount of videos from hospitals, health organizations and active users within top-10 ranking.

| | hospitals | organizations | active users |
|--------------|-----------|---------------|--------------|
| YouTube | 3% | 6% | 4% |
| Domain-based | 12% | 36% | 13% |

Precision scores for relevant and reliable videos were calculated for each query and ranking method. The average precision scores show that Domain-based ranking increased precision from 0.59 in the original YouTube ranking to 0.81. These results indicate that our approach can increase the proportion of reliable and relevant videos among top ranked videos.

Conclusion

We have in this work focused on retrieval of trustworthy health videos, and have developed a novel method for classifying and ranking YouTube videos so that videos from trustworthy sources are presented early in the ranking list.

References

- [1] Bennett, E., Social media and hospitals: from trendy to essential, Futurescan 2011, Health Administration Press.
- [2] Freeman, K.S., Spyridakis, J.H., Effect of Contact Information on the Credibility of Online Health Information, IEEE Transactions on Professional Communication, Vol. 52, No. 2, June 2009.