Soc. Netw. Anal. Min. (2017) 7:31
https://doi.org/10.1007/s13278-017-0452-3

ORIGINAL ARTICLE

Advanced searching framework for open online educational video lectures

Dimitrios Kravvaris[1] · Katia Lida Kermanidis[1]

Received: 15 December 2016 / Revised: 16 May 2017 / Accepted: 14 July 2017 / Published online: 26 July 2017

Abstract The appearance of massive open online courses has caused an increase in the volume of open online educational videos on the web. Therefore, there is a vast amount of information to be managed by the Internet users. The present work aims to optimize the educational video lecture searching in social networks. The research presents a novel ranking procedure for the educational video lectures that takes into account their popularity with content-based social media communities. The popularity formula combines quantitative and qualitative characteristics, taking into account not only the positive and the negative elements of the web page containing the video, but also the opinion of the users based on their comments. Thus, a novel social parameter is proposed which is embedded in the content-based ranking process. Furthermore, a user evaluation procedure is carried out, and initial results indicate that this integration produces a ranking output that better matches the user's preferences.

Keywords Educational video lecture · Popularity · Opinion mining · Ranking · Cosine similarity · Social parameters

✉ Dimitrios Kravvaris
jkravv@gmail.com

Katia Lida Kermanidis
kerman@ionio.gr

[1] Department of Informatics, Ionian University, 7 Tsirigoti Square, 49100 Corfu, Greece

1 Introduction

Technological changes have always affected education and learning styles (Balakrishnan and Gan 2016). From the blackboard to the modern era of interactive whiteboard, to long distance learning through correspondence, and finally to the current Massive Open Online Courses (MOOCs; Siemens 2013), technology has been a powerful tool in the struggle for the conquest of knowledge. But, as is the case with all changes, challenges arise, the most important issue of which is the management of the large amount of educational content available mainly in video format. The oversupply of open educational videos on the web has increased with the appearance of MOOC and the participation of many prestigious universities (Klobas et al. 2014), making it harder for the students to choose the video lecture that is suitable for their needs. The search for the selection of the most appropriate video lecture by the user can be very time consuming. Due to the number of video lectures offered, the query processing time complexity is quite challenging and continues to increase as new video lectures are uploaded on a daily basis.

This paper presents the design, the implementation and the evaluation of an innovative advanced searching framework for educational video lectures on real data. The framework involves searching for video lectures in the content-based social media communities. The study aims to exploit the social characteristics of a content-based social network such as YouTube. This creates an additional value for each educational video lecture based on these characteristics. This additional value is then incorporated in the ranking of educational video lectures, optimizing the original ranking and, thus, presenting to the user a set of results that better match his needs, in order to choose the one that he considers more appropriate. The present paper's contribution is twofold. Firstly, it presents a novel advanced searching framework for educational videos at the video selection level. The framework aims to reduce the search time for the user in order to find the suitable video for his search. Secondly, it applies state-of-the-art algorithms and methodologies a) to web pages that contain and offer educational videos, and b) to their verbal content which provides the information sought. A study of the parameters of the algorithms is conducted, and significant results are exported for their application to real data of educational video lectures, used both in media sharing networks and in complete courses either in the form of MOOCs or as traditional online courses. An important part of the present research is the design of a formula for the estimation of the popularity of educational videos in social media, which can also be applicable to other fields of study.

The first part of the paper presents work related to video lecture search in content-based social networks. Section 3 defines the concept of popularity; Sect. 4 describes the proposed ranking methodology, the experimental procedure and its results. The methodology, the experimental setup, and the results of the user evaluation process are presented in Sect. 5. Finally, Sect. 6 extensively discusses the findings, provides aggregated results and proposes future research work.

1.1 Research objectives

The goal of the first part of the present work is to design an educational video lecture ranking procedure that incorporates the opinion of users, extracted from the lecture's view history. Thereby, future users will save time when searching for a lecture that is suitable for them. For this purpose, the quantitative parameters (videos' number of likes, dislikes and views) and qualitative parameters (users' comments and videos' transcripts) of the web page containing the video lectures are first recorded. The qualitative parameters of users' feedback are then studied in order to extract their views on the video lecture examined. Both the quantitative and qualitative parameters are utilized to create a formula in order to estimate the popularity of each video lecture, which takes into account both the positive and the negative opinion features posted by users. The popularity of each video is incorporated in the form of social weight in the classification of the video lectures when searching takes place, and, thus, a new ranking method is generated, which is both content- and social-based. This new search strategy is studied with respect to the changes introduced in the original ranking of video lectures, and a user evaluation procedure is performed to examine its effects.

The volume of online educational video lectures, both in number and in time length, increases rapidly. The changes proposed in this paper on how to locate the appropriate video lecture are necessary and imperative. Through the application of the proposed method, the time the users need to search for the most appropriate video lecture can be reduced significantly by taking into account the opinion of users who have already watched the video.

2 Related work

The correlation of the contents of the video and the opinion of the users to rank and search for information has preoccupied the scientific community. The use of content similarity in many studies (Zachara and Pałka 2016; Wang et al. 2016; Deldjoo et al. 2016) and the use of social media in education (Blankenship 2011; Fernandez et al. 2011; Manca and Ranieri 2016; Krutka and Carpenter 2016) have led to the following interesting studies.

The study by Chelaru et al. (2012) focuses on the keyword-based video search and investigates the impact of the social features (views, likes, dislikes and comments) on the retrieval performance. The authors show that social features can help retrieving more relevant videos for the queries when they are combined with the basic video features (title and tags). The work by Telang et al. (2012) suggests the correlation between content similarity and user's similarity. Thus, ranking of the web content is enriched with the user-similarity parameter, which, for a given search, shows the selections performed by previous users in the history of the same query. A social-textual approach is suggested by Khodaei and Shahabi (2012) for search queries on the web. The relevance of search results is based on the social parameters derived from the user's connections with others in a social network. A new ranking framework is suggested in the research by Gou et al. (2010). This framework incorporates a content value and a social value to get the final ranking position of a video. The content value is based on the tf-idf score of the video's metadata, and the social value is the similarity between the user searching for a video and the owner of the video. A case study on recommending YouTube videos to Facebook users based on their social interactions is presented in the study by Nie et al. (2014). The authors developed a new recommendation algorithm that exploits user social interactions on Facebook to improve the recommendation accuracy for YouTube videos. Through experiments, authors found that social interaction information can be added on top of the existing similarity-based collaborative filtering algorithm to achieve significantly higher accuracy. The above studies have shown that the inclusion of the social parameters in the search process has a positive effect on the final result.

The present research confirms this positive effect of the social parameter, adding three important elements compared to previous research. First, our research is applied to the largest database (i.e., YouTube), and thus, there is richer information database concerning the social characteristics of an educational video. Secondly, the application and its evaluation results are not limited to the registered users of the social network, but are accessible to everyone. Thirdly, the present study addresses also the opinion of the users, extracting not only the measurable quantitative data, but also the qualitative data based on the users' comments on each video.

3 Video popularity

3.1 Characteristics analysis

The popularity of a video on YouTube has been investigated (Chatzopoulou et al. 2010; Figueiredo et al. 2014), and the importance of characteristics, such as the number of views, likes, dislikes, the users' comments, the number of users who have chosen it as a favorite and, finally, the number of video responses, has been confirmed. A specific formula for the estimation of video popularity, however, has not been validated so far, and such an evaluation schema is presented herein.

The most important features derive from the registered YouTube users, and they include the number of likes and dislikes, as well as the comments, which clearly give the user's view on a video. The number of views also plays an important positive role (Zhou et al. 2016; Szabo and Huberman 2010) since it shows how many times a video has been watched. YouTube uses an algorithm to detect and eliminate fraudulent views, ensuring that the video has been really watched. While the number of likes, dislikes and views constitute numeric features that are easily measurable, the comments feature requires special processing in order for the opinion of the users to be extracted (Liu 2012; Chen and Zimbra 2010).

3.2 Opinion mining

The opinion of the users on a video may vary but it is recorded in the comments and can even cause discussions among users. The comments may concern part of the video or the video as a whole. They can even concern the speaker's words or other characteristics such as the presentation, the colors or the way the video was filmed. Aiming at an overall assessment of a video lecture, all users' comments on a video lecture are utilized and are automatically classified into two categories as positive and negative, following that way YouTube's binary characterization of the video as liked or disliked.

Initially, a training model based on the support vector machines (SVM) linear formula is created using Rapidminer v5.3 (Hofmann and Klinkenberg 2013), using two dictionaries, one of positive and one of negative words (Hu and Liu 2004). The SVM linear formula is suitable for binary classification (Yu and Kim 2012; Park and Cardie 2014) of users' comments into either the positive or the negative class. Then, in the second stage, the comments are filtered and only the words that are of value to sentiment classification have been kept. Filtering consists of the following procedures: (a) tokenization, (b) stop words removal (Ali and Ibrahim 2012), (c) transformation of characters into lower case and (d) filtering out tokens shorter than 3 and longer than 25 characters, removing that way 2-character emotion symbols and multiple-character strings from users' comments. In the third stage, each one of the previously processed comments of the users is evaluated through the training model. In the end, the positive and the negative feedback is combined and transformed into a metric that represents the opinion of the users on the video lecture under examination.

3.3 Formula

Unlike a previous research approach to educational video ranking (Kravvaris et al. 2015), which takes into consideration only the positive attributes, or only some of the positive and negative social markers (Kravvaris and Kermanindis 2015), the present work takes into account a wider set of quantitative and qualitative markers pertaining to the positive and the negative social parameters, in order to estimate video popularity. Positive parameters include the number of likes, the number of views and the number of positive comments, while the negative ones include the number of the dislikes and the number of negative comments. Both the positive and the negative parameters result from actions on the part of the users that watch the video.

We define as popularity p of a video i, which belongs to a group of videos that have resulted from a particular search sr, the value of positive parameters minus the value of negative parameters. The value of positive parameters is the sum of the normalized values of likes l, views v and positive comments pc. The value of negative parameters is the sum of the normalized values of dislikes dl and negative comments nc according to the formula:

$$P_{i,\mathrm{sr}} = \left[\left(\frac{l_i}{\mathrm{maxl}_{\mathrm{sr}}} + \frac{v_i}{\mathrm{maxv}_{\mathrm{sr}}} + \frac{\mathrm{pc}_i}{\mathrm{maxpc}_{\mathrm{sr}}}\right) - \left(\frac{\mathrm{dl}_i}{\mathrm{maxdl}_{\mathrm{sr}}} + \frac{\mathrm{nc}_i}{\mathrm{maxnc}_{\mathrm{sr}}}\right)\right] \quad (1)$$

where maxl, maxv, maxpc, maxdl and maxnc are the maximum values of likes, views, positive comments, dislikes and negative comments correspondingly, which are observed in the search results of the videos under examination.

3.4 Social and content similarity

One of the novel aspects of the methodology proposed herein is the incorporation of the popularity of each video lecture in this formula. Therefore, a new parameter is proposed, henceforth called social weight (sw) that is the popularity (p) defined in the subsection above. The new similarity metric form is called social-content similarity (scs) and combines the cosine similarity value (s), between the transcript of the video lecture and a search query, with the popularity value for each video lecture as shown in the following formula.

$$\mathrm{scs} = s(2 + \mathrm{sw}) \tag{2}$$

Only the video transcripts that are related are studied, so the angle between the transcript of the video lecture and the search query vectors is between 0 and 90 degrees and, therefore, the value of parameter s ranges between 0 and 1. The sw parameter theoretically takes values between -2 and 3, since 2 negative and 3 positive characteristics define the parameters of the popularity, respectively. In order to avoid negative values, the parameter $2 + \mathrm{sw}$ was used. Combining the above, it is found that the proposed scs measure values range between 0 and 5.

4 Ranking

4.1 Data

Of particular importance to our research are the data that have been collected from video lectures on YouTube in the category of education. The video lectures vary in their subject matter and the object they present. Their searching subject areas are presented in Table 1.

Out of a total of 20.830 video lectures collected searching the subjects above, only 1.116 that are accompanied by English transcripts have been utilized in the present research. Subtitles are an important element in the proposed searching framework, since they provide immediate access to the content of the lecture. Apart from the transcripts, users' comments have also been collected, and information including the number of likes, dislikes and views for each video lecture, using the YouTube Data API v3.[1] All these constitute important parameters to determine the opinion of the users on a video.

[1] https://developers.google.com/youtube/v3.

Table 1 Educational video subject areas

Subjects			
Computer science	Data mining	Machine learning	Biology
Medicine	Web	C++	Statistics
Theory	Art	Php	Social
Physics	Health	Java	Analysis
Space	Classification	Network	Geography
Human	Programming	Public	Mathematics
Teaching	Economics	Ted	Internet
Philosophy	Financial	Media	Laboratory
Chemistry	Learn	Algorithms	Experiment
Database	Lecture	History	Energy

4.2 Methodology

For the implementation of the new ranking in the video lectures search results, which are based on the social-content similarity formula, the following procedure is adopted. In the first phase, the transcript of the video lectures is processed in order to keep the words that have some meaning for the content. In the second phase, the content similarity value of each video lecture is calculated in relation to the corresponding search query and thus the first ranking is performed. In the third phase, the popularity of each video lecture is calculated and then the social-content similarity formula is applied, calculating the value for each video. The phase ends with the final ranking procedure of the video lectures.

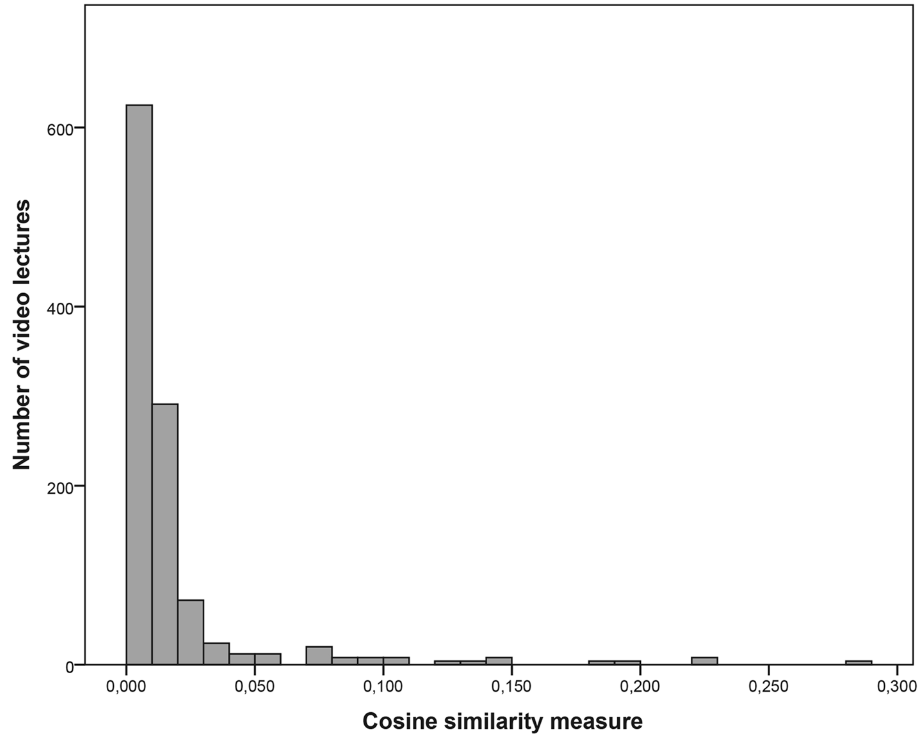
4.2.1 Transcript processing

In order to keep the words that have some meaning for the content of each transcript, Rapidminer v5.3 (Jungermann 2009) is used. More analytically, the following four procedures are followed in the order described next. The first procedure splits the text of a transcript into a series of single word tokens. The second procedure removes stop words that do not add meaning to the transcript. The third procedure removes the most common morphological and inflectional endings from English words (Willett 2006) according to the Porter stemming algorithm. Finally, the fourth procedure transforms all characters into lower case.

4.2.2 Content similarity

Forty groups of video lectures are used originally extracted from the corresponding forty search queries posed on YouTube. From each group, the corresponding processed transcripts of the video lectures are collected and the text of the search query is added as a new record. Then the similarity among texts is calculated for each group separately.

Fig. 1 Distribution of the cosine similarity values



The result of this process is a table showing the value of the cosine similarity for each pair of all possible combinations of the texts. From these combinations, the results of the pairs' transcript and search query are chosen. So, by ranking them, the similarity rank of the transcripts in relation to the search query can be seen based exclusively on the video lectures content.

4.2.3 Social-content similarity

In order to calculate popularity to make it more meaningful, all irrelevant comments have been removed from the process. More specifically, only those comments that contain words shared with the processed transcript of the video lecture are chosen, so they are related to the speaker's speech object. Taking into account the previously estimated cosine similarity, the social-content similarity can now be calculated. Using the proposed formula, new correlation values are obtained. After this calculation, the order of appearance of the video lectures is re-ranked. The study of the differences and the similarities between the initial and the final ranking procedure provides interesting results on the qualitative addition of the users' opinion.

4.3 Experimental procedure

The procedure is applied to forty groups of video lectures. Figure 1 shows the distribution of the cosine similarity measure of each video lecture, as described above. Each bar on the graph represents the number of video lectures grouped in the corresponding range of cosine similarity. The best value is close to 0.3, while there are 21 video lectures returned by YouTube that present a cosine similarity value of zero, i.e., the transcript contains no word from the search query. Figure 2 shows the distribution parameter $2 + \mathrm{sw}$, which constitutes the social addition to the calculation of social-content similarity. Each bar on the graph represents the number of video lectures grouped into the corresponding social parameter $2 + \mathrm{sw}$. Most video lectures have sw values greater than 2, which means that sw has a positive effect on these. There are, however, a few video lectures that have more negative elements (dislikes and negative comments) than positive ones, and their sw value is lower than 2. The cases that have only positive elements are far more and reach 9% of the video lectures under examination.

The new ranking, based on the social-content similarity method, changes the ranking order, compared to the ranking estimated only with the cosine similarity method. In Fig. 3, the rates of rank position change are shown. 45% of the video lectures are ranked in a higher position, 19% shift to a lower position, while only 36% remain in their original position. The position changes vary, and their distribution is shown in Fig. 4. As can be seen, the changes in the ranking of the video lectures are usually between 1 to 4 positions up or down, which cover about 84% of the cases. But also extreme cases arise, like the one where a lecture is accompanied by only negative elements and is therefore placed 23 positions lower.

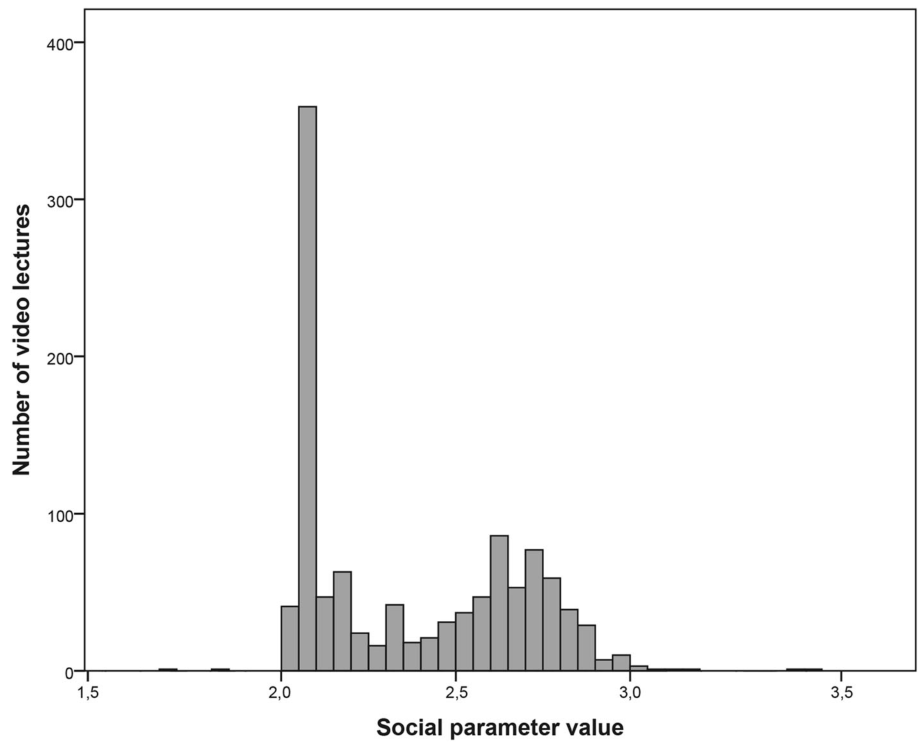


Fig. 2 Distribution of the social parameter (2 + sw) values

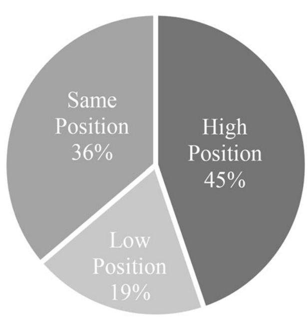


Fig. 3 Rates of the position change

4.4 Results

The ranking of the educational video lectures on real data reveals many interesting issues. The incorporation of the users' opinion to the process changes the order of 64% of the video lectures compared to their initial positions, which take into account only their content, showing that this incorporation is an important factor that highly affects the ranking process. The new ranking places 84% of the video lectures between 1 and 4 positions higher or lower than their initial positions. Thus, there is a special emphasis on the importance of the cosine similarity measure in the present proposal, which prevents a great change in the positions of the video lectures except for some few extreme cases.

Compared to previous work (Kravvaris et al. 2015), which took into account only the positive opinion of the users, it is found that the inclusion of negative parameters in the process increases the number of videos that change positions. Thus, the new social-cosine similarity method affects 7% more video lectures. Moreover, it limits the range of position change (whether higher or lower), thus discouraging extreme changes. Finally, it is found that there is an increase by 4% in the ranking of video lectures to higher positions. Overall, the new ranking results are more comprehensive, they include qualitative parameters related to the users' opinion (positive or negative) about the video content, and they derive from the recalculation of the popularity of every video according to the search query.

5 User evaluation

An important point of the present research is the investigation of (a) whether the proposed ranking correlates to the ranking results selected by the users, and of (b) whether it eventually improves the ranking with respect to the simple ranking based on the content only. Thus, using the same videos the ranking results obtained by real users are compared with those returned by the cosine similarity measure and the content-social similarity measure.

5.1 Data

Thirty users, who are students at different university departments, take part in the evaluation procedure. Six

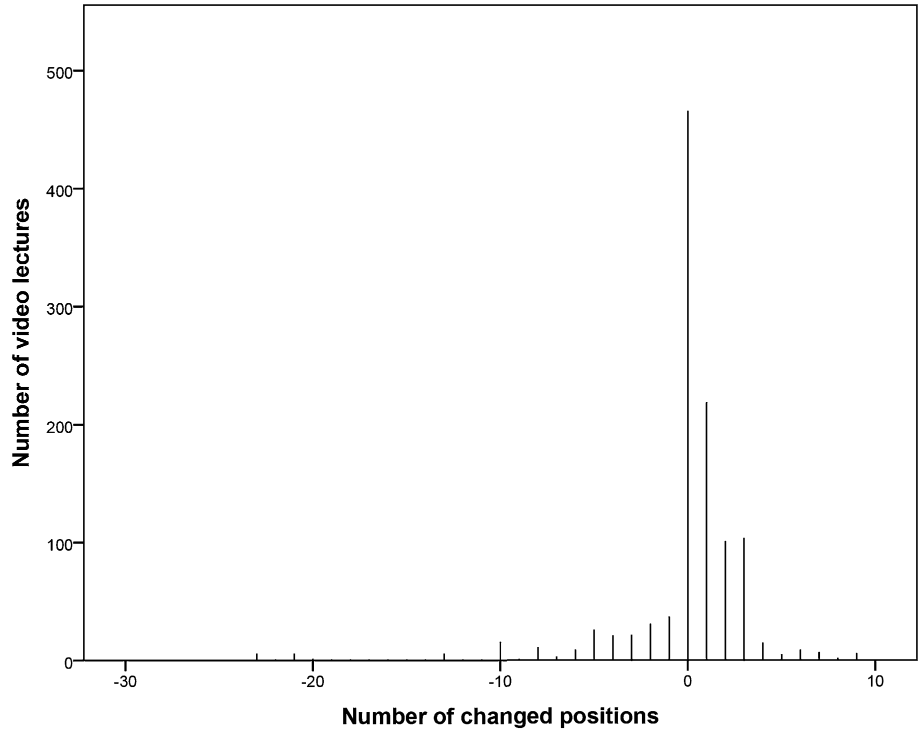


Fig. 4 Distribution of the number of changed positions

Table 2 Video duration

Video lecture	Video1	Video2	Video3	Video4	Video5	Video6
Time duration (min:sec)	6:02	5:26	5:10	8:56	15:12	3:46

video lectures originating from the search query group *database* are randomly provided to the users and they are asked to rank them. The timing of the videos is shown in Table 2. The choice of the number of the videos given to the users is based on the respective number of videos that appear first on the user's screen after a search on YouTube.

5.2 Methodology

A website on an exact YouTube template is given to users, presenting the 6 video lectures being reviewed. Users have enough time at their disposal to watch them. Each user has to rank the video lectures according to how they would like them to appear after a possible search of the word *database* on YouTube. After each user has decided the order of the ranking, their answers are recorded in an online questionnaire. The same questionnaire also includes their answers to questions about each video, i.e., a) whether they have watched the whole or part of each lecture and b) what positive and negative elements they have observed in it, which in their opinion have helped them in their decision for their ranking. Finally, five of the users are interviewed (Yew et al. 2011) to record their attitude toward the videos evaluated.

Assuming that every user's ranking is correct, the ranking of each user is compared to the content similarity video ranking and the social-content similarity video ranking, respectively, using the mean average precision (MAP) measure. MAP provides a single-figure measure of quality across recall levels (Manning and Raghavan 2009). MAP is defined as the mean of average precisions of all the correctly ranked video lectures, assuming that the users ranking is correct. Average precision for each case is defined as the mean of the precision at six values computed after each video lecture is ranked in the right position.

5.3 Experiment

Table 3 presents the ranking of the six videos according to the content similarity and the social-content similarity methods. It can be observed that the two methods have ranked four videos at the same position and differ in the ranking only in the fourth and fifth position of the videos. The MAP between each ranker and the two aforementioned methods are shown in Fig. 5. The percentages show how close each method is to the users' ranking. Thus, the social-content similarity method predicts the user's preference by 37.07% while the content similarity method by 30.41%.

Table 3 Ranking order based on similarity methods

Method	1st position	2nd position	3rd position	4th position	5th position	6th position
Content Similarity	Video1	Video2	Video3	Video4	Video5	Video6
Social-content Similarity	Video1	Video2	Video3	Video5	Video4	Video6

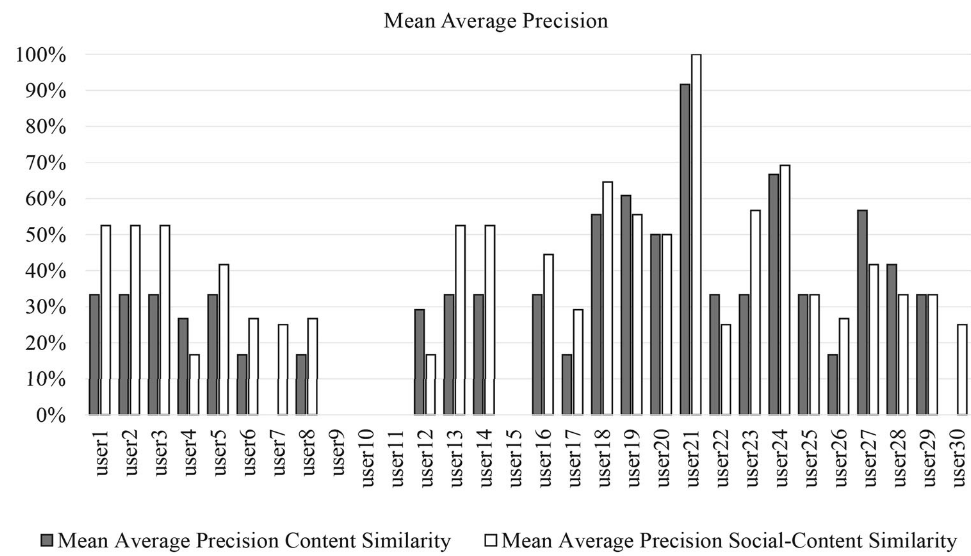


Fig. 5 MAP values between each ranker and two methods

Considering the fact that human behavior is being measured, it can be said that the percentage of social-content similarity method is quite satisfactory (Shortell 2001).

5.4 Results

From the above results, it is noticeable that the social-content similarity method is closer to the ranking order of the 30 users in 17 cases (56.67%), the cosine similarity method in 6 cases (20%), while there are 7 draws (23.33%). There is also a case where the social-content similarity method is identical to the ranking of user 21, as shown in Fig. 5. From the questionnaires that the users have answered it is observed that they choose videos based on their simplicity and comprehensibility, and they reject videos that contain tiring details, according to their opinion. Furthermore, from their personal interviews it can be noticed that the users choose to press the *like* option of a video on YouTube more easily than the *dislike* option. Most of the users who are not satisfied with a video simply stop watching it. Users prefer to comment on a video regarding the information that its content provides, and to respond to other users' comments on the video's subject of interest. The user evaluation experiment is critical for the validation of the accuracy of the proposed theoretical framework. The results highlight the superiority of the social-content similarity ranking method that takes into consideration social attributes, such as the users' opinion.

6 Conclusion

The present paper studies the search for information in open educational video lectures in the content-based social media communities. The research aims at exploiting the views of the users of educational videos. Thus, the popularity of the videos is calculated in a social network and a re-rank of the video lectures is possible by adding the parameter of popularity to the previous ranking of the video lectures based only on their content.

More analytically, opinion mining techniques are applied in order to extract the user's opinion from the comments recorded in video lectures. Moreover, positive and negative parameters of a video are extracted and a comprehensive formula is presented to calculate the popularity of the video lectures. The popularity formula can also be applied to other video categories as well and to other social networks beyond YouTube. The incorporation of popularity in the form of the social weight parameter in the calculation of the score for the ranking of educational video lectures has shown to be particularly strong and has changed the position in more than 64% of the video lectures. Those video lectures change by 84% for no more than 4 positions, higher or lower than the initial ranking, and so the process improves the ranking procedure without making rapid changes in the results except in minimum extreme cases. Finally, a user evaluation experiment is conducted which calculates the MAP between the two methods of content similarity and social-content similarity and of the users who are asked to rank six videos. It is found that the social-content similarity method that takes into account the popularity parameter is by 56.67% closer to the ranking of the users and even once the ranking order is identical to the order suggested by the user. This strengthens the original position for the integration of the social parameters in the ranking of the search results in video lectures. The application of this methodology can be applied to other digital material available in social networks such as blogs, web pages, images.

The search for information is a field that will continue to be explored in the coming years. The growth of digital content and how it is handled is creating new challenges. Social networks can be used to optimize searching, as it is proved in the present research. The methodology followed constitutes a secure framework for future research into social networks.

Finally, the following ideas that can benefit future research are suggested. As far as the freshness of the search results is concerned, it is recommended that at the end of the process described the videos be filtered based on the date of publication from the most recent to the oldest for every twenty videos. Twenty is the number of videos presented on each YouTube page after a search. So the user will have the newest videos at his disposal without significantly reducing their relevance. As far as search is concerned, it is suggested that the history of videos that the user has already watched be also taken into account. Thus, the similarity of the transcripts of the watched videos and the ones that appeared in the search can be calculated. This similarity can be added to the factor of social-content similarity presented so far. In this way, similar videos to what the user has already watched or will probably prefer to watch will increase their ranking position in the process.

References

Ali NH, Ibrahim NS (2012) Porter stemming algorithm for semantic checking. In: Proceedings of 16th international conference on computer and information technology, pp 253–258

Balakrishnan V, Gan CL (2016) Students' learning styles and their effects on the use of social media technology for learning. Telemat Inf 33(3):808–821

Blankenship M (2011) How social media can and should impact higher education. Educ Dig 76(7):39–42

Chatzopoulou G, Sheng C, Faloutsos M (2010) A first step towards understanding popularity in youtube. In: INFOCOM IEEE conference on computer communications workshops, 2010. IEEE, pp 1–6

Chelaru SV, Orellana-Rodriguez C, Altingovde IS (2012) Can social features help learning to rank youtube videos? In: International conference on web information systems engineering. Springer, pp 552–566

Chen H, Zimbra D (2010) Ai and opinion mining. IEEE Intell Syst 25(3):74–80

Deldjoo Y, Elahi M, Cremonesi P, Garzotto F, Piazzolla P, Quadrana M (2016) Content-based video recommendation system based on stylistic visual features. J Data Semant 5:1–15

Fernandez M, d'Aquin M, Motta E (2011) Linking data across universities: an integrated video lectures dataset. In: International semantic web conference. Springer, pp 49–64

Figueiredo F, Almeida JM, Gonçalves MA, Benevenuto F (2014) On the dynamics of social media popularity: a youtube case study. ACM Trans Internet Technol 14(4):24

Gou L, Zhang XL, Chen HH, Kim JH, Giles CL (2010) Social network document ranking. In: Proceedings of the 10th annual joint conference on digital libraries. ACM, pp 313–322

Hofmann M, Klinkenberg R (2013) RapidMiner: data mining use cases and business analytics applications. CRC Press, Boca Raton

Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 168–177

Jungermann F (2009) Information extraction with rapidminer. In: Proceedings of the GSCL symposium sprach technologie und eHumanities. Citeseer, pp 50–61

Khodaei A, Shahabi C (2012) Social-textual search and ranking. CrowdSearch 37(5):3–8

Klobas JE, Mackintosh B, Murphy J (2014) The anatomy of moocs. Massive open online courses: The MOOC Revolution, pp 1–22

Kravvaris D, Kermanindis KL (2015) An analysis of online educational videos in social media based on verbal content. In: 2015 6th international conference on information, intelligence, systems and applications (IISA). IEEE, pp 1–6

Kravvaris D, Kermanindis KL, Chorianopoulos K (2015) Ranking educational videos: the impact of social presence. In: 2015 IEEE 9th international conference on research challenges in information science (RCIS). IEEE, pp 342–350

Krutka DG, Carpenter JP (2016) Why social media must have a place in schools. Kappa Delta Pi Record 52(1):6–10

Liu B (2012) Sentiment analysis and opinion mining. Synth Lect Hum Lang Technol 5(1):1–167

Manca S, Ranieri M (2016) Facebook and the others. potentials and obstacles of social media for teaching in higher education. Comput Educ 95:216–230

Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval, vol 1. Cambridge university press, Cambridge, p 496.

Nie B, Zhang H, Liu Y (2014) Social interaction based video recommendation: recommending youtube videos to facebook users. In: 2014 IEEE conference on computer communications workshops (INFOCOM WKSHPS). IEEE, pp 97–102

Park J, Cardie C (2014) Identifying appropriate support for propositions in online user comments. In: Proceedings of the first workshop on argumentation mining, pp 29–38

Shortell T (2001) An introduction to data analysis & presentation. World Wide Web: http://academic.brooklyn.cuny.edu/soc/courses/712/chap18.html

Siemens G (2013) Massive open online courses: innovation in education. Open educational resources: innovation, research and practice 5

Szabo G, Huberman BA (2010) Predicting the popularity of online content. Commun ACM 53(8):80–88

Telang A, Li C, Chakravarthy S (2012) One size does not fit all: toward user-and query-dependent ranking for web databases. IEEE Trans Knowl Data Eng 24(9):1671–1685

Wang Y, Luo Z, Yu Y (2016) Learning for search results diversification in twitter. In: International conference on web-age information management. Springer, pp 251–264

Willett P (2006) The porter stemming algorithm: then and now. Program 40(3):219–223

Yew J, Shamma DA, Churchill EF (2011) Knowing funny: genre perception and categorization in social video sharing. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 297–306

Yu H, Kim S (2012) Svm tutorial—classification, regression and ranking. In: Handbook of Natural computing. Springer Berlin Heidelberg, pp 479–506

Zachara M, Pałka D (2016) Comparison of text-similarity metrics for the purpose of identifying identical web pages during automated web application testing. In: Information systems architecture and technology: proceedings of 36th international conference on information systems architecture and technology–ISAT 2015–Part II. Springer, pp 25–35

Zhou R, Khemmarat S, Gao L, Wan J, Zhang J (2016) How youtube videos are discovered and its impact on video views. Multimed Tools Appl 75:1–24