Multimedia Tools and Applications (2021) 80:4155–4181
https://doi.org/10.1007/s11042-020-09755-z

A NLP framework based on meaningful latent-topic detection and sentiment analysis via fuzzy lattice reasoning on youtube comments

Hamed Jelodar[1] · **Yongli Wang**[1] · **Mahdi Rabbani**[1] · **Sajjad Bagheri Baba Ahmadi**[1] · **Lynda Boukela**[1] · **Ruxin Zhao**[1] · **Raja Sohail Ahmed Larik**[1]

Received: 23 December 2019 / Revised: 9 July 2020 / Accepted: 28 August 2020 /
Published online: 28 September 2020

Abstract
Social media platforms such as Twitter, Facebook, and YouTube have unique architecture, norms, and culture. These platforms are valuable sources of people's opinions which should be examined for knowledge discovery and user behavior analysis. This paper proposed a novel content analysis to examine user reviews or movie comments on YouTube. In fact, the proposed hybrid framework is based on semantic and sentiment aspects using fuzzy lattice reasoning to meaningful latent-topic detection and utilizing sentiment analysis of user comments of the Oscar-nominated movie trailers on YouTube. Based on the word vector feature, classification algorithms are employed to detect the comments' sentiment level. The results of this study suggest that the hybrid framework could be effective to extract features associated and latent topics with sentiment valence on user comments. In addition, NLP methods can have an impressive role for exploring the relationship between user opinion and Oscar movies comments on YouTube.

Keywords Natural language processing · Topic model · LDA · Social media · YouTube

1 Introduction

Social Media provides a solid platform for finding, sharing and broadcasting videos, images, etc. An important factor in the development of these technologies is the capability to easily create user-generated data. Twitter and YouTube are only a few examples of these

✉ Hamed Jelodar
jelodar@njust.edu.cn

Yongli Wang
yongliwang@njust.edu.cn

[1] School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China

media that are strongly changing the Internet perspective today. These platforms motivate the users to publish and annotate content and more importantly, sharing information with their social media account. However, social media serves as a significant platform for both non-professional content creators and multimedia organizations. For instance, in terms of releasing movies, YouTube role is remarkable in introducing and popularizing new movies. Besides, YouTube users ~~can review~~, comment and share their personal opinions about a movie online [32, 52]. Topic modeling based on LDA is an unsupervised machine learning technique that aims to arrange large corpora of text into candidate topics [52]. These topics explain thematic relations and patterns among text-contents. In this study, also, we designed a semantic model based on Latent Dirichlet Allocation (LDA), which is a probabilistic unsupervised classification method in topic modeling. In fact, for the current research, LDA is utilized to extract semantic topics from YouTube user comments and investigate if they can be mapped to some motivational categories. Therefore, this semantic model attempts to discover the meaningful-topics iteratively based on the co-occurrence of words in documents and demonstrates each document as a mixture of various topics with related weights.

Sentiment models are an interesting technique in social media, since the text content includes various opinions about user comments or movie reviews. Moreover, sentiment models can have a significant role in user-understanding, making topic-recommendation systems and opinion mining based on text content on social media. Also, another important goal of this research, classifying and opinion mining of the user comments on YouTube. There are some related works that try to extract opinions and sentiments on YouTube. For instance, the authors in [32] utilized an analysis of YouTube comments by applying the SentiWordNet thesaurus using a lexical resource including sentiment annotations in order to find dependencies between comments, views, and topic categories. The research of [53] performed an analysis of YouTube comments to find how users participate in text discussions about Minecraft. An aspect-based method has been presented for evaluating Malayalam movie reviews. In addition, in [49], the authors found that regular sentimental analysis is not sufficient, and from another point of view, user comments are very important for analyzing. However, common machine learning techniques such as Naïve Bayes, support vector machines (SVM), and decision trees are typically used as classifiers for identifying sentiments or opinions. Moreover, a group of researchers conducted a sentiment classification technique for detecting the sentence polarity of online reviews and dividing them into pos, neg, or neutral categories [28, 44, 46, 55].

1.1 Motivation and Contribution

In general, sentiment/opinions mining of online users is very helpful in enhancing the quality of the videos and meeting the needs of the people. With sentiment/polarity analysis, substantial comments can be obtained for positive, neutral, and negative sentiments [23, 37, 40]. This can assist in satisfying viewers'experience and improve recommendation systems of movies to the related groups. However, sentiments interpreted by applying a limited resource of language pose which is very complex task via existing sentiment analysis methods. This research improves our perception of social media engagement and moves beyond a usual view of user motivations to have consisted of an evaluation of particular features on YouTube. The main aim of this research is to generate descriptive statistics of YouTube comments and particularly about discussions via YouTube comments based on natural language processing (NLP) methods. Although there have been some quantitative and qualitative studies of YouTube, not enough is known about its uses in general to be able to formulate hypotheses about why discussions might occur. To the best of our knowledge, most of the previous works was only focused on the sentiment/word polarity aspects or some of the work was one-dimensional focused on the semantic aspects. In this research with the difference from past work, we were motivated to investigate both semantic and sentiments aspects in the form of a hybrid framework, which is a NLP method includes text mining, sentiment analysis, and polarity mapping were employed to accomplish the research goals.Moreover, the NLP model combined by a fuzzy lattice reasoning (FLR) model as a novel methodology in order to evaluate the commenting behavior of YouTube users. For this study, we considered an existing datatset of user comments about Oscar-nominated movie trailers on YouTube. In sum, the contributions of this paper can be highlighted as:

- A novel hybrid framework is proposed by combining semantic, sentiment and fuzzy approaches in order to evaluate the commenting behavior of YouTube users.
- A new application area of YouTube comments analysis is presented in our framework to measure public emotions in different topics such as racism, social issues.
- We adopt fuzzy lattice reasoning to handle the prediction sentiment-levels, and the experimental results on real-world data sets indicating the superiority of our model in comparison with several machine learning methods.
- The weakness, limitations of the research framework are discussed along with possible enhancements.
- A novel stop-words with special cover of YouTube movies comments is utilized which can be useful for future research in this area, because the default stop-word lists in the topic modeling is not adequate.

The rest of this research is prepared as follows: In Section 2 we discuss related works on online content and comment analysis on YoutTube. Section 3 describes the hybrid framework proposed in four phases. Section 4 introduces the characteristics of the data set, and extraction of the semantic topics and analyzing of the connection between sentiment in comments community. Then it provides the outputs of large-scale classification experiments. In Section 5 we discuss about the potentials and limitations. Finally, we conclude the remarks and future work in Section 6.

2 Related works

Social media such as Twitter, YouTube are playing an impressive role as information sources for analyzing user behavior and knowledge discovery [3, 13, 16]. The diversity in the ways YouTube is being used for not only watching and entertainment but also for people interaction through commenting and providing information, makes YouTube interesting area of the research. YouTube supplies different useful tools for social interaction, such as the possibility to comment posted videos by online users, as shown in Fig. 1.

- Sentiment analysis on YouTube comments: many studies have addressed the extent to which online social media in contexts of different platforms. Sentiment analysis is a popular method for analysis text-content in social media, which also is defined as an opinion mining. Many researchers used this method for comment analysis on YouTube [4, 6, 15, 42]. For example; in [45], the authors discussed the usage of emotionally loaded language in news as well as its differences among outlets with different political orientations. They investigated the sentiment trajectories of YouTube news channels in major English media outlets. However, they found that proright videos are more famous and negative sentiments are growing sharply, for averaged sentiment

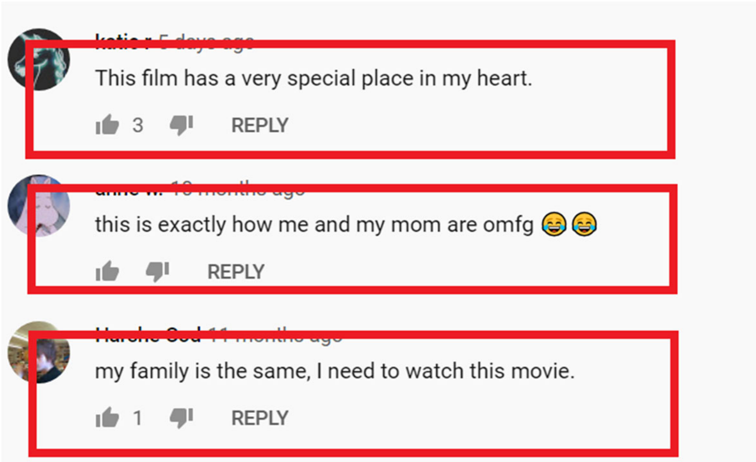

Fig. 1 A sample of Youtbe user comments is about a Oscar movie

ratings of each video. In addition, in [50], the authors developed a deep learning method to detect multi label sentiment and emotion for Bangla sentences from various YouTube videos. Moreover, there are some neural network methods that can be combined with NLP methods for classification, such as [26, 58, 59].

In another work of [36], the authors considered 50 most famous YouTube anti-pro-anorexia and pro-anorexia channels. They collected data from users and commentators, then analyzed 12,161 comments with positive or negative polarity and scores submitted by the online users. In [18], the authors applied two NLP tools to carry-out knowledge discovery from almost seven million words by performing an analysis of reactions to videos about the autonomous vehicle on YouTube. In fact, they addressed the content exploration of these videos by examining the users' attitudes with considering disliking or liking, and commenting patterns. In [9], the authors have examined the influence of various aspects of the video's subject from YouTube user comments in the video retrieval with applying aspect-based opinion analysis. In [24], the authors focused on mood analysis based on content analysis techniques such as spam filtering, in which the best filtering classifiers are identified by applying a labeled data of Youtube comments. Also, other researchers in [27], proposed an application of inverse reinforcement learning from a huge YouTube data set into non-overlapping segments. They found that this group of YouTube users are approximately logically inattentive and also users groups tend to comment on videos that are famous with high view count.

– Semantic extraction on YouTube comments: since the semantic and sentiment methods are not unique to text processing in Natural language processing, therefore we describe some few of papers that used semantic techniques on YouTube user comments. In [38] the authors evaluated the content of videos released in a set of right-wing YouTube channels, and extracted the semantic topics in hate and discrimination issues. Furthermore, in [35], an investigation of the semantic structures of unstructured text contents from eight YouTube channels is presented to analyze the toxic behaviors online of commenters. However, the importance and application of semantic models based on topic models in social media is very attractive in natural language processing. Moreover, topic modelling is a significant technique in NLP, specially when there is no clear classification to associate with text-documents (such as labelling), which manually is too difficult and expensive. However, due to the complexity of the semantic concepts of comment feature on Youtube, topic models can lead to provide meaningful structures, and make a clearer picture of the relationship among behavioural, latent topics, and YouTube user comments. In addition, topic tracking and trend detection is a very valuable capacity of topic models, which can be practical for understanding YouTube user comments. Furthermore, a topic model has special advantages for text classification systems, which allow for soft clustering of comments in Youtube. Based on previous works, also in [10, 11], the authors focused on topic models for rating prediction and extract users' preferences on different aspects from reviews.

– Fuzzy lattice reasoning classifier (FLR): this classification technique [14] produces hyperboxes based on the points included in the training data. A hyperbox matches to a rule which demonstrates that a point determined within this box is a segment of the relevant class. However, training is utilized repeatedly and per point from the training data set impels a novel rule. Moreover, this model calculates the fuzzy degree for each new rule based on contents with available rules. The maximum value of these rates suggests, how the new rules should be mixed. Figure 2 shows a FLR architecture for classification and supervised learning, which consists of three different layers, mean the inputs, middle and decision layer. Another hand, in the testing phase, all the learned rules is computed of the testing pattern. In addition, the class label of the rules or hyperbox with the highest value of inclusion measure is determined to the testing pattern. In this research, one of the important aims is to define an intelligent fuzzy system for classification rule problem in comment features based on sentiment aspects. Inference systems based on fuzzy logic is a technology that produces rules by induction process. The main idea of the fuzzy models in the decision-making process is to improve the

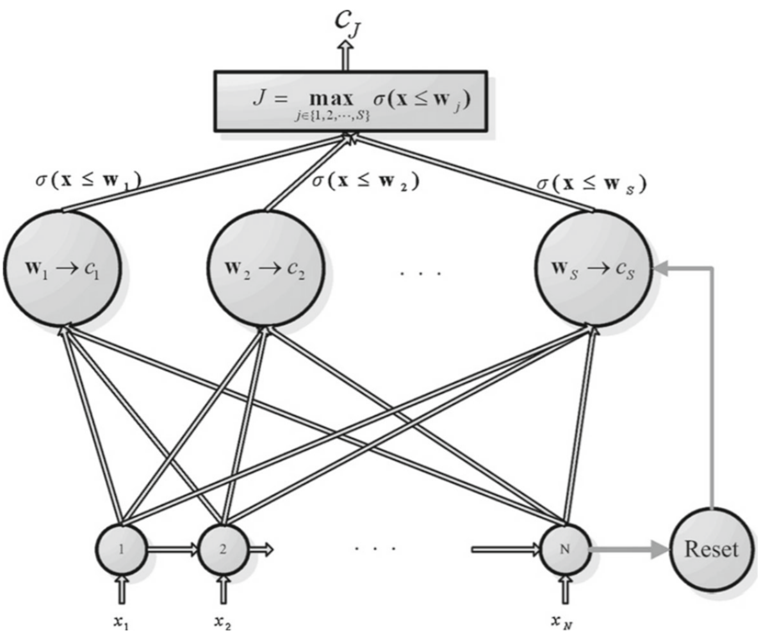

Fig. 2 A simple architecture of FLR model [34]

success rate in the smart standard classifiers. Overall, we consider a FLR model, due to their reliability in rules generation and provides logical results in order to a clever classification.

In general, the problem of detecting affects from YouTube comments has been approached in a number of ways. Regarding our knowledge, most of the previous works only focused on sentiment aspects and also some few works utilized only semantic techniques on YouTube comments. In this study we propose a novel hybrid framework based on natural language processing using LDA topic model, sentence polarity and fuzzy lattice reasoning (FLR), which covers both semantic and sentiment structures on movie comments. Besides, an important aspect of this paper is to detect the latent topics and types of labels that are discussed in order to clarify the opinion of the users from the movies.

3 The NLP-hybrid framework

The analysis of comments constitutes a potentially interesting data source to mine for getting implicit knowledge about videos, user opinion analysis. YouTube commenters can select to be anonymous because even though they have to register an identification to comment, they may use a pseudonym and this looks to be the norm, this may lead to antisocial behavior, such as putting racism comments by a user. In this section, we present the hybrid framework which has been developed and includes all the steps required for semantic extraction and sentiment analysis via fuzzy lattice reasoning for detecting meaningful latent topics and sentiment level analysis, as shown in Fig. 3.

3.1 Preparing data and comments processing: Phases A & B

We first prepared the pure input-data and process the user movie comments. Then remove the noisy information that is wasteful for training a good model by filtering "stop-words" and also we detect unique words via the stemming techniques.

– Stop-words are those simple words that are usually meaningless and they cannot be effective in output, some examples are: 'they', 'or', 'the', 'these', 'and'. Stop-words can be a pre-defined list of terms or they can be related to the text-corpus. Preparing a stop-words list depends on the content, which is applied in the context. In other words, preparing a list for the content of analyzing optional applications in the domain of YouTube movie comments, allows users to apply general and publicly available sources. So far, we have investigated different sources to prepare stop-words lists for different movie topics. Since the domains and texts are different, it is difficult to create a finite list. We provided a novel stop-words with a special cover of YouTube movies comments because the default stop-words lists in NLP is not sufficient. Although, we do not claim to define a complete list but it is considered as a logical list set.
– Stemming is mostly used in the area of natural language processing, where the goal is to improve system efficiency and to decrease the number of distinct words. Stemming is the process of eliminating prefixes and suffixes, leaving the root of the word. For example, the words 'watching', 'watched' and 'watch' would be reduced to the root word 'watch'. This makes sense, because the terms have a same meaning. Furthermore, stemming will decrease the size of the lexicon, and could be effective on computational procedure.

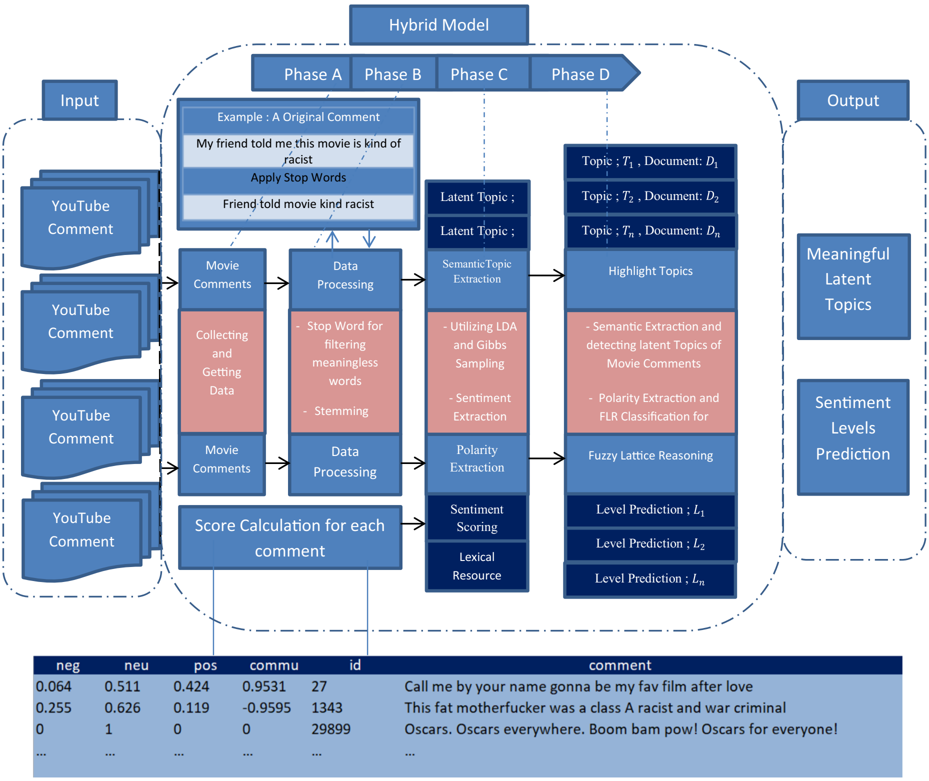

Fig. 3 An overview of the proposed hybrid framework

3.2 Semantic and sentiment analysis: Phase C

3.2.1 Meaningful latent-topic extartcion: Semantic

Comments can be dependent on movie posts of various topics, which those YouTube users discuss. There are different methods to comment analysis natural language processing based on topic modeling which is one of the most frequently applied techniques to semantic extraction and content similarity analysis of user comments of YouTube movies. In this phase of framework, we considered topic modeling via Latent Dirichlet Allocation (LDA) [7, 8] to detect latent topics of YouTube comments in English context. We define a set of documents as YouTube-comments context and words as topics (K), where the discrete topic distributions are drawn from a symmetric Dirichlet distribution. Mathematically, given a vocabulary with N words, $D = w_1, w_2, ..., w_N$, which the procedure for each YouTube-comment document w in a corpus D is as following:

Given the parameters α *and* β the joint distribution over the random variables $(w_m, z_m, \varphi_k, \theta_m)$, which is given by:

$$p\left(w_m, z_m, \theta_m, \varphi_k | \alpha, \beta\right) = p\left(\theta_m | \alpha\right) \\ p(\varphi_k|\beta) \prod_{n-1}^{N_m} p(z_{m,n}|\theta_m)\, p(w_{m.n}|\varphi_{z_{m,n}}), \tag{1}$$

Embedding over θ_m *and* φ_k, summing over $z_{m,n}$, the marginal distribution of a YouTube-comments can be computed:

$$p\left(w_m|\alpha,\beta\right)=\int_{\theta_m}\int_{\varphi_k}p\left(\theta_m|\alpha\right)p(\varphi_k|\beta)\left(\prod_{n=1}^{N_m}\sum_{z_{m,n}}\sum_{z_{m,n}}p(z_{m,n}|\theta_m)p(w_{m,n}|\varphi_{z_{m,n}})\right)d\theta_m d\varphi_k, \tag{2}$$

Finally, taking the produce of the marginal probability of every YouTube-comment in the corpus, the construction process probability of a corpus is determined as follows:

$$P(D|\alpha,\beta)=\prod_{m=1}^{M}\int_{\theta_m}\int_{\varphi_k}p\left(\theta_m|\alpha\right)p(\varphi_k|\beta)\left(\prod_{n=1}^{N_m}\sum_{z_m}p(z_{m,n}|\theta_m)p(w_{m,n}|\varphi_{z_{m,n}})\right)d\theta_m d\varphi_k\ , \tag{3}$$

Gibbs sampling is applied to learn the LDA model, and then each instance is expressed with topic distributions. Therefore, we use a Gibbs sampler to allocate a new label (topic) to the word by sampling:

$$p(z_i=k|z_{i-1},d,w)\propto\frac{n_{wk}+\beta}{\sum_v n_{vk}+\beta_v}\left(n_{dk}+\alpha\right) \tag{4}$$

where $n_{w,k} and n_{dk}$ are the respective counts of topics k with words w or in documents $d\ and\ \alpha\ and\ \beta$ hyper parameters as before.

3.2.2 Comment-polarity calculation: Sentiment

In this phase, we utilize sentiment analysis which is effective to a vast range of problems that are interested in social media data and also other fields such as medical [20, 22, 30], marketing [41, 54] ,and digital humanities [43]. Therefore, we consider this NLP method to find the feelings expressed of online users in YouTube comments. There are a variety of methods and tools to conduct sentiment analysis of comments such as SentiStrength [47, 48], VADER [29]. The Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiment model was considered for sentiment analyses of all comments in this framework. To compute sentiment, VADER utilizes both rule-based sentiment analysis and lexicon analysis to supply a measure of how positive, negative and neutral a bunch of context. VADER maps terms to sentiment by making a dictionary or a lexicon from sentiments. The dictionary assesses the sentiment of sentences and YouTube comments. However, lexical techniques can be utilized to assess either the score of the YouTube comments or the sentiment category.

The score for each comment is based on how they are written and how they are classified, because the user comments in social media are very informal (such as slang, or specific words), and also people tend to apply uppercase letters or different marks in order to give emotions for their words. From each YouTube Comment, VADER produces four sentiment components including positive, neutral, negative and compound. The fourth component is a normalized score of the first three metrics.

Algorithm 1 Hybrid-SemSenti

Input ← a bunch of Youtube movie cooments; $D_i = doc_1, doc_2, doc_3, ..., doc_N$
$PhaseA$ ← YMcomment : A set of comments of different text documents
$PhaseB$ ← Remobve Stop-words from YMcomment.
$PhaseB$ ← Getting new data and utilizing stemming.
$PhaseC$ ← Start to process for semantic analysis:
$PhaseC$ ← LDA_Calculate: Utilizaing and extracting laten topics with learing LDA with Gibbs Sampling
$PhaseC$ ← latent_Scoring: Investigating laten topics and Return the value of the best score of a latent topics as a integer value; Get_valueLatent(distrubation topics, Propration, word-weight)
$PhaseC$ ← Start to process for sentiment scoring:
$PhaseC$ ← sentenceScore ← pos = getPOS(YMcomment);
$PhaseC$ ← sentenceScore ← neg = getNeg(YMcomment);
$PhaseC$ ← sentenceScore ← neu = getNEu(YMcomment);
$PhaseC$ ← YM_SentiLevels= sentenceScore (pos,neg, neu) & SentimentScore-Calculation for each comment
$PhaseD$ ← Labelig latent topics and choosing meaningful topics from the latent_Scoring of Phase C.
$PhaseD$ ← Getting comments sentiment features/scores of YM_SentiLevels from Phase C to Algorithm 2.
Output ← Return the output as a set of latent topics = $topic_1, topic_2, topic_3, ..., topic_N$; as an integer values, and return a set of comment-sentiment levels → $L_1, L_2, L_3, ..., L_N$.

However, the scope of compound score is different in size, from -1 to 1, where -1 indicates least preferred and 1 defines most preferred selected movies. In this research; regarding to Table 1, we consider the score ranges of YouTube comments in order to have a great detecting with more sensitivity, into 23 score-levels. For example, the sentence "Congratulations on winning best picture!" has the following sentiment values: neg: 0.0, neu: 0.145, pos:0.855, compound: 0.9151. This means the comment sentence is positive with a 85% probability and 14% is for neutral. The probability for this user comment to be negative is 0%. The compound value of 0.9151 tells us that this sentence is very positive since the compound value is in the range between (0.90 → + 0.97) and the final score will be $+pos^{18}$.

3.3 Getting meaningful topics & FLR polarity classification: Phase D

After extracting and detecting NLP results of the previous steps, we need to investigate comments' sentiment levels, which utilized a fuzzy lattice reasoning (FLR) [5, 31] in this step, as showed in Algorithm 1. The FLR classifier generates specific fuzzy rules to strengthen the decision on determining the object class for an input query pattern Q. The input data for the classifier was obtained from the outputs of sentiment analysis in Phase C. In fact, this hybrid framework utilizes the advantages of the fuzzy lattice reasoning (FLR) classifier to detect sentiment levels via classification, we annotated each comment with a sentiment label among the twenty-three sentiments (based on Table 1). Therefore, we predict the type of sentiment level of each YouTube comment based on FLR in an intelligent classification.

Mathematically by FLR, Let $L = L_1 \times \times L_N$ be the Cartesian product of N lattices $L_1 \times \times L_N$, namely constituent lattices [54]. A lattice inclusion relation can be defined in L

Table 1 The range of sentiment scores and values leveling

Level Type	Range Number	Polarity Level
Positive Levels	+ 0.98 → + 1	$+pos^{20}$
	+ 0.90 → + 0.97	$+pos^{18}$
	+ 0.80 → + 0.89	$+pos^{16}$
	+ 0.70 → + 0.79	$+pos^{14}$
	+ 0.60 → + 0.69	$+pos^{12}$
	+ 0.50 → + 0.59	$+pos^{10}$
	+ 0.40 → + 0.49	$+pos^{8}$
	+ 0.30 → + 0.39	$+pos^{6}$
	+ 0.20 → + 0.29	$+pos^{4}$
	+ 0.10 → + 0.19	$+pos^{2}$
	+ 0.01 → + 0.09	$+pos^{1}$
Negative Levels	- 0.01 → - 0.09	$-neg^{1}$
	- 0.10 → - 0.19	$-neg^{2}$
	- 0.20 → - 0.29	$-neg^{4}$
	- 0.30 → - 0.39	$-neg^{6}$
	- 0.40 → - 0.49	$-neg^{8}$
	- 0.50 → - 0.59	$-neg^{10}$
	- 0.60 → - 0.69	$-neg^{12}$
	- 0.70 → - 0.79	$-neg^{14}$
	- 0.80 → - 0.89	$-neg^{16}$
	- 0.90 → - 0.98	$-neg^{18}$
	- 0.99 → - 1	$-neg^{20}$
Neutral Level	0	*neutral*

as $(x_1...x_N) \ll (y_1...y_N)$ if and only if $x_1 \ll y_1, ..., x_N \ll y_N$. The meet in $L = L_1 \times \times L_N$ is given by $(x_1, ..., x_N) \wedge (y_1, ..., y_N) = (x_1 \wedge y_1, ..., x_N \wedge y_N)$ moreover the join in L is given by $(x_1, ..., x_N) \vee (y_1, ..., y_N) = (x_1 \vee y_1, ..., x_N \vee y_N)$.The dual of a lattice L is another lattice denoted by L^{∂}, which has, by definition, the same underlying set nevertheless its partial ordering relation ($\ll_{\partial}$) is the converse of L, i.e. a $\ll_{\partial}$ $binL^{\partial}$ if and only if $b \ll ainL$

Based on this method a fuzzy set is denoted by (X, μ) where X is the universe of discourse and μ is a fuzzy membership function $\mu : X \rightarrow [0, 1]$. A fuzzy lattice is presented as a pair $\langle L, \mu \rangle$ where L is a crisp lattice and $\langle L \times L, \mu \rangle$ is a fuzzy set with membership function. $\mu : L \times L \rightarrow [0, 1]$ defined such that $\mu(x, y) = 1$ if and only $x \leq 1$. In other definition, an inclusion measure σin a complete lattice L is a real function $\sigma : L \times L \rightarrow [0, 1]$ such that for $u, w, x, y \epsilon$ L the following conditions are satisfied:

(1) $\sigma(x, O) = 0, x \neq O$
(2) $\sigma(x, x) = 1, \forall x \in L$
(3) $u \leq w \Rightarrow \sigma(x, u) \leq (x, w)$
(4) $x \wedge y \prec x \Rightarrow \sigma(x, y) < 1$

More obviously, it can be interpreted as the fuzzy degree to which x is less than y; therefore notations will be used interchangeably. The FLR classifier evolves rules from the training set by allowing a rule's diagonal size growth up to a maximum threshold size. The algorithm 2 is utilized for the training process on sentiment features of comments.

Algorithm 2 FLR-Sentilevel

Require: The input data as sentiment features of movie comments
$Stage0 \leftarrow$ Let a fuzzy lattice rule engine ξ that consists of R rules. All rules are considered to be "set".
$Stage0 \leftarrow$ The engine could $\leftarrow$ empty
$Stage1 \leftarrow$ Present the next training element in the form to the initially "set" rules of the engine
if no more rules in ξ are "set" **then**
$Stage2 \leftarrow$ append a new rule $u \rightarrow c$, and go to preivious step.
end if
$Stage3 \leftarrow$ Compute the truth table of all the "set" rules in ξ against the antecedent u.
$Stage3 \leftarrow$ The rule $a_j \leftarrow c_j$, that produces the highest value in the truth table, is considered as a candidate winner
if $c = c_j$ and the size of is less **then**
$Stage4 \leftarrow$ A predefined threshold hen replace the candidate winner rule in the engine to Step-1.
else
$Stage4 \leftarrow$ the candidate winner rule from the "set" ones and go to Step-2
end if

The FLR classier can also set out learning without a previous knowledge; however, a previous knowledge can be provided to the FLR classifier in the form of a primary set of rules. During training step, this classification model produces hyper-boxes based on the points included in the training step. A hyper-box concord to a rule which demonstrates that a point placed within this box is a part of the respective class. Training is applied iteratively and per point from the training set includes a new rule. For per new rule, the model computes a fuzzy degree of inclusion with the available rules. The best value of these degrees indicates how new/existing rules should be mixed. For categorizing an unclear data point, the system computes the inclusion degrees for the rule enforced by the unclear point and allocates this point to the class of the rule with a great value.

4 Experimental results and setting

We considered an existing dataset [1] to study the YouTube comments posted for the Oscar-nominated movie trailers, which include 'Call Me by Your Name', 'Darkest Hour, Dunkirk', 'Get Out', 'Lady Bird', 'Phantom Thread', 'The Post', 'The Shape of Water', and 'Three Billboards Outside Ebbin Missouri' movies. The dataset contains 38,855 comments and all the comments posted until the 6th of March 2018, which showed statistical words of available dataset in Table 2.

4.1 Parameters and settings

For learning LDA Topic model and semantic extraction we adopted the Library Mallet [2], a Java-based package for statistical NLP, document classification, clustering, topic modeling,

[1] https://www.kaggle.com/PromptCloudHQ/youtube-reviews-for-oscarnominated-movie-trailers.
[2] http://mallet.cs.umass.edu/topics.php

Table 2 A statistical information of available data

Names of Oscar Movies	Without-Stemming	Stemming
Call Me by Your Name	92,200	89,020
Darkest Hour	111,319	106,781
Dunkirk	1,831,126	1,686,653
Get Out	1,062,211	1,020,914
Lady Bird	139,027	134,357
Phantom Thread	94,363	182,148
The Post	171,564	164,182
The Shape of Water	427,479	356,659
Three Billboards Outside Ebbin Missouri	141,437	136.200

information extraction, and other machine learning applications for text. We configured with 10000 Gibbs sampling iterations and an optimization interval ($\alpha = 0.5$ and $\beta = 0.01$). In addition, word generation is defined by a conditional distribution and described by two parameters, the first is the size of the vocabulary and second is the number of topics. However, there is no standard number for determining the number of topics and we also experimented more with different options of the number of topics in LDA topic modelling, and then the quality of the inferred topics is compared to determine the optimal number of topics, which we determined at 100. Moreover, we extract top 20 words for each topic generated by each model for computing the coherence. Moreover, we used NLTK library[3] in Python to sentiment analysis, remove punctuations, stemming and stop-words processing from the comments. The machine learning and training process via fuzzy lattice reasoning was performed using the Weka Library, which divided the pre-processed data into a training set (70%) and a testing set (30%). All experiments were carried out on a machine running a Windows OS 10 with a Core I5 and 8 GB of memory.

4.2 Results visualization and findings

Regarding the results, we detected meaningful latent topics of all YouTube comments in various aspects and labels. In Fig. 4, we demonstrated the ten top topics discovered along with the 20 terms of each topic. In addition, Table 3 presents the details related to the topics ranking, labels, and propration. The visualization of the relationship among words of the meaningful latent topic is shown in Fig. 5 and the word cloud visualization for various topics based on word-weight is provided in Figs. 6 and 7. In this section, since we have discovered a large number of topics and because of page limitation, we only describe the topics with ranks 1-12, which are as follows:

Topic 87 contains a number of words relating to movie and all comments were submitted to address positive and negative aspects of movie trailers. Based on the sentiment analysis, this topic has positive words, including “good”,“love”,“great”,“amaz” and also some negative words that include“bad” and “fuck”. Additionally, the terms “movi” ,“film” and “watch”are the most highlighted words discovered for this topic, with a words-weight equal to 13.3%, 22% and 34% respectively. Based on the results of Fig. 7 for this topic, it seems that the Youtube users many times mentioned on this topic in “Tree Billiboards

[3] https://www.nltk.org/index.html

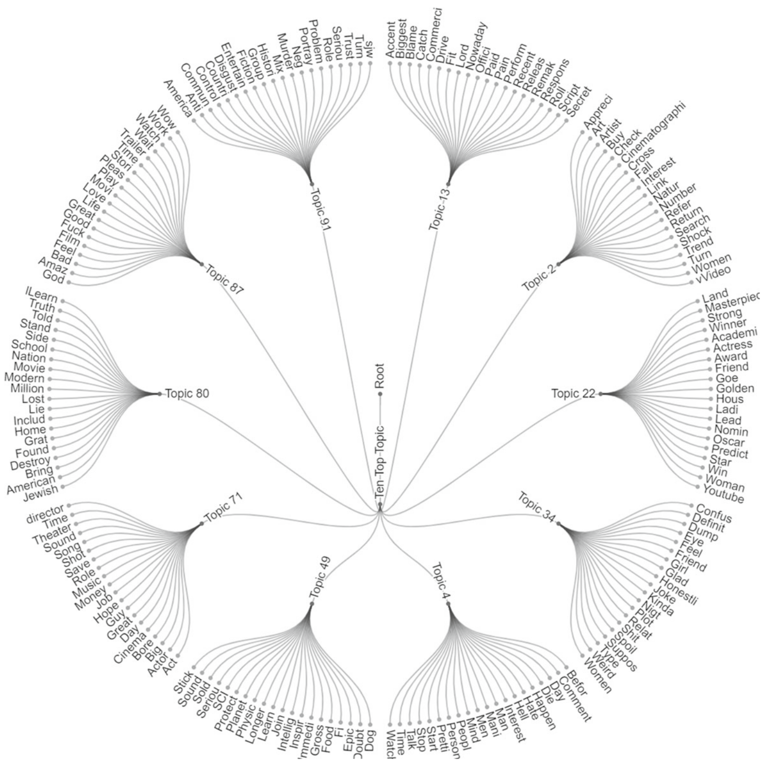

Fig. 4 A circle tree of ten top meaningful latent-topics of all data

outside eddin Missouri", "Lady bird" and "Call Me by Your Name" movies, which the distribution values are 12%, 43% and 18% respectively.

For Topic 4, the topic terms are "people", "watch", "comment", "hate", "befor", "time", "mani", "man", "happen", "stop" and may have words that are more indicative of the categories they are describing about people and person categories. Based on the sentiment analysis, the negative words are more impressive than positive ones, which the negative words such as "hate" and "hell". It seems that the negative sentiments in some comments were expressions of sadness elicited by a performance. Additionally, the terms "people" and "watch", "comment" are the most highlighted words discovered of this topic, which the words-weight are 13.3%, 22%, 34% respectively. Regarding this topic, the YouTube users many times mentioned on this topic in "Get Out", "Call me by your name", "Dunkirk" movies, which the distribution value are 12%, 43%, and 18% respectively.

Topic 71, the top terms of this topic being "actor", "guy", "hope", "director", "music", "act", "bore", "great", "sound", "time", seems that some YouTube users are addressing the song of movies and also many times pointed about directors and actors. However, the movie fans often talk about movies actors and music fans may often talk about the song of movies. In this case, we may find that topics associated with this concept. Regarding this topic, the YouTube users many times discussed in "Dunkirk", "Phantom Thread, "Darkest Hour" movies, which the distribution values are 12%, 43% ,and 18% respectively.

Topic 34, we consider the label 'Relationship' that appears in the engagement content with words like "shit", "girl", "feel", "plot", "suppose", "friend", "weird", "glad",

Table 3 More detail of the latent topics detected of comments, which the rank of each topic determined based on the proparition value of topics

Rank	Topic No.	Label	Propration	Latent Topic Words			
1	87	Movie	4.18848	movi film watch trailer good	love wait great amaz stori	god time play pleas bad	fuck feel wow work life
2	4	people	1.83845	peopl watch com- ment hate befor	time mani man happen stop	interest pretti hell person die	mind talk start day men
3	71	Song	1.38245	actor guy hope director music	act bore great sound time	day cinema theater big save	song shot money role job
4	34	Relationship	0.99427	shit girl feel plot suppos	friend weird glad women dumb	definit confus honestli relat night	kinda spoil joke eye type
5	22	Oscar	0.88301	oscar win award woman nomin	actress lead youtube ladi academi	masterpiec goe winner land star	golden strong predict friend hous
6	80	Nation	0.69042	american home side great movie	lost million truth learn jewish	told school includ destroy found	bring stand lie modern nation
7	2	Cinematographi	0.68129	video interest trend turn art	women fall cinematographi return cross	number shock refer appreci link	artist check natur search buy
8	91	Histor , fiction	0.50149	anti portray countri america histori	sjw mix problem control turn	seriou disgust fiction murder entertain	commun trust group role neg
9	13	Releas	0.46501	releas perform lord drive recent	nowaday offici remak biggest fit	catch blame accent respons script	roll pain com- merci secret paid
10	49	SCI movies	0.36916	inspir intellig doubt stick fi	planet dog sold longer epic	join learn sound immedi gross	sci physic seriou food protect

Table 3 (continued)

Rank	Topic No.	Label	Propration	Latent Topic Words			
11	54	Sexual	0.30886	song	version	relationship	en
				sex	bitch	sexual	children
				bore	emot	masterpiec	franchis
				parent	titl	blue	xd
				deserv	move	la	um
12	95	Racism	0.19818	white	acist	rpeel	good
				movi	horror	shit	scari
				black	race	jordan	evil
				peopl	trailer	racism	racial
				fuck	guy	comedi	show
13	47	Crime stories	0.16096	crimin	iron	domin	deadli
				defend	befor	race	puppet
				refus	netflix	fault	commun
				john	threat	phoni	deliber
				communist	send	great	traitor
14	84	War	0.16022	harri	dunkirk	british	soldier
				nolan	christoph	hardi	peopl
				style	fuck	tom	film
				war	french	histori	die
				movi	fan	german	zimmer
15	79	General	0.11954	youth	shop	authent	heh
				ireland	road	implement	strut
				resolv	mood	fulfil	william
				landri	tempt	urgent	pinch
				crisi	academ	detect	int

"women", "dumb". Regarding this topic, the YouTube users many times mentioned on this topic in "Get Out", "Call Me by Your Name", "The shape of water" movies, which the distribution values are 12%, 43% , and 18% respectively. Some examples of comments of "Get Out" movie, described as follows:

– What is this shit? Fucking ape with a beautiful white woman. JEWS;
– So after this you still want a "romance" with a white woman lmao;
– This movie gonna make a brotha never wanna date white woman again;
– A lot of white girls like Black men;

Topic 22 consists of several words relating to Academic Oscar award and as expected all comments were submitted to discuss about actress and movies. The top terms of this topic are "Oscar", "win", "award", "woman", "nomin", "actress", "lead", "youtube", "ladi", "academi". Additionally, the terms "Oscar" and "win", "award" are the most highlighted words discovered for this topic, which the words-weight are 13.3%, 22%, 34% respectively. Regarding this topic, the Youtube users many times mentioned on this topic in "Three billboards Outside Ebbin", "The Post", "Phantom Thread" movies, which the distribution values are 12%, 43%, and 18% respectively.

Topic 80 also includes words that refer to the nations, which are impressive terms like "American", "jewish", "nation". Topic 2 contains the telling words "video", "interest", "trend", "turn", "art", "women", "fall" and "cinematographi", but it is difficult to describe this topic, it seems the topic discuses about trending movies. Other revealing words in Topic 2 include "art", "cinematographi" and "artist". These terms initially suggest a set of

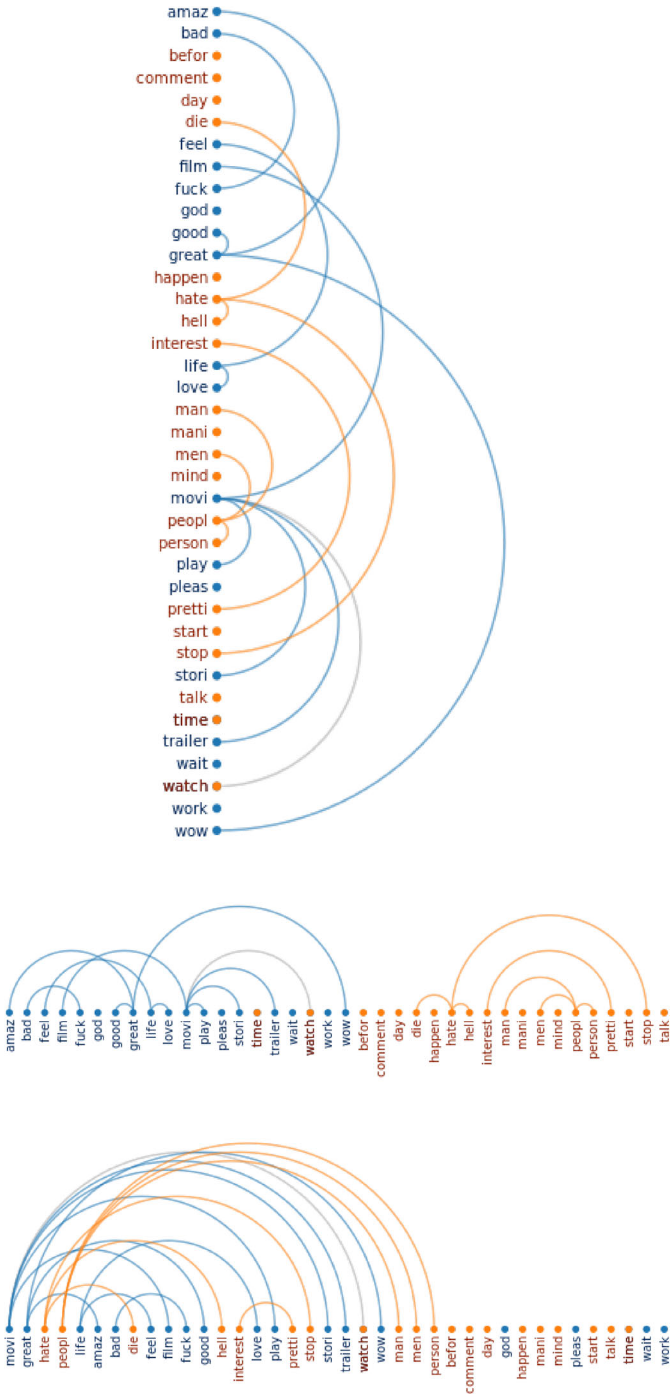

Fig. 5 Relationship among words by considering similarity words in different view of a meaningful latent topic, which this figure describes about the Topic 87. For example; the word "movie" have similar concepts with 'play', 'film', 'trailer', 'stori'

(a) Topic Word 87 (b) Topic Word 4

(c) Topic Word 71 (d) Topic Word 34

(e) Topic Word 22 (f) Topic Word 80

Fig. 6 Topic word clouds based on word-weight of each term; from ranks 1-6 in Table 3

YouTube user comments about visual effects are created by the artist. Topic 91 also includes words that the topic discuses about American history, also sentiment analysis of terms that suggest the negative words are more impressive than positive ones. Regarding this topic, the Youtube users many times mentioned on this topic in “Get Out”, “The Post” and “Darkest hours” movies.

For Topic 13, It is difficult to describe about it, but since the term ‘release is the most highlighted word discovered of this topic, which the words-weight is 21.3%, it seems that the topic refers to the user discussion about ‘release of movies’. Topic 49 contains words like “inspire”, “fi”, “planet”, “epic”, “sci”, “planet” and thus clearly pertains to the inspirational and intelligent aspects of SCI-FI movies. Regarding this topic, the YouTube users many times mentioned on this topic in “Shape of water”, “The Post” and “Dunkirk” movies, which the distribution values are 12%, 43% and 18% respectively. Some examples of comments related to this topic are as follows:

– War time period piece meets sci fi meets romance.....?

(a) Topic Word 2

(b) Topic Word 91

(c) Topic Word 13

(d) Topic Word 49

(e) Topic Word 54

(f) Topic Word 95

Fig. 7 Topic word clouds based on word-weight of each term; from ranks 7-12 in Table 3

– This looks thought provoking, potentially a new sci fi classic. As long as it doesn't turn into a typical slasher movie where the Merman breaks free and decides to kill everybody.
– That must be every left wingers wet dream. First fucking children, then animals and now even sci fi shit. haha;
– First a sci fi film inspired by games then a political drama it is going to be great for speilberg X.

Topic 54 contains words like 'sex','bore','bitch', 'emot', 'relationship', 'sexual' and thus clearly pertains to the sexual issue. The sentiment analysis of terms that suggest the negative words are more impressive than positive ones. Regarding this topic, the YouTube users many times discussed in "The Shape of Water", "The Post" and "Dunkirk" movies, which the distribution values are 12%, 43%, and 18% respectively. Some examples of comments that are relevant to this discussion topic are as follows:

– This is evil another beauty and the best thing having strange relationships with beast and animals. Don't be foolish.
– I got a good Movie script about Actors racing to expose a massive cover up of Sexual Predators in Hollywood.
– A movie about to start the right of breeding with animals , or having intimate relationship!

Topic 95, we can see in this topic that there is a focus on racist-discourse issues. However, it is difficult for a precise definition of racism. The [51] attempted to adopt a simple definition which includes all negative utterances, negative generalizations and insults concerning ethnicity, nationality, religion and culture. Based on the sentiment analysis, the negative words are more impressive than positive ones, which include "fuck", "racist", "horror", "shit". Overall, this topic tended to be more subjective, expressing a sense of community and strong negative sentiment toward persistent structural racism. Some examples of comments related to this topic are as follows:

– Whites are so racist they would rather see a daffy white gal bang a lochness monster rather than a Black man hahahaa.
– Basically a white racist bitch hypnotize niggas to be slaves and work for them and they try to hypnotize the black bf.
– My friend told me this movie is kind of racist.
– Stupid ass racist movie, don't waste your time.

4.3 Sentiment comments & FLR classification

Polarity detection is the most popular sentiment analysis task. In fact, many research works even use the terms "polarity detection" and "sentiment analysis" interchangeably. This is due to the (limited) definition of sentiment analysis as the NLP task that aims to categorize a piece of text as either positive or negative. Regarding the framework of this research, we calculated sentiment scores of comments by utilizing VADER technique as well as we described in Section 3. Moreover, we provided a new score range for various levels that showed in Table 1. We demonstrated the sentiments distribution or polarity levels of all comments in Fig. 8.

In this section, classification algorithms were implemented for learning the results of sentiment results and then, the results of several well-known algorithms are examined and validated, as following:

– Senti-C1 the approach proposed based of this research.
– Senti-C2 (MultilayerPerceptron): A multilayer perceptron is a well-known machine learning model and an artificial neural network (ANN), which is composed of an input layer, output layer and hidden layers. However, these models have a wide range of applications that include performance regression, pattern detection and text classification [39] (Fig. 9).

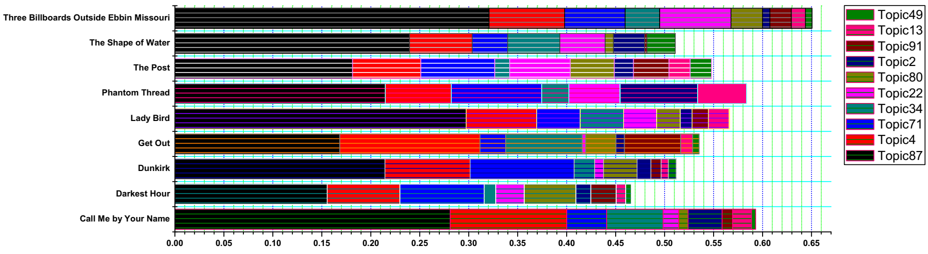

Fig. 8 Distribution topics from different movie pages

- Senti-C3 (NaiveBayes): A naive Bayes is a well-known classifier which is based on Bayesian theory with adoptions of each feature from a specific class is independent of other feature. However, this model can be applied widely in several cases because it has easy implementation and high efficiency. Moreover, Naive Bayes is one of the successful techniques applied in sentiment analysis [33].
- Senti-C4 (RandomTree): The random tree is a decision tree structure based on supervised classification, which is created based on choosing a random subset of features of a dataset. The model can deal with both regression and classification problems [17].
- Senti-C5(IBK): This model is a k-nearest-neighbour classifier that utilizes the uniform distance metric matrices to classify the data. The number of nearest neighbours can be determined automatically by applying leave-one-out cross-validation [2].
- Senti-C6(SVM): Support Vector Machine or SVM is a widely utilized classifier in sentiment classification. However, this model utilizes a kernel function for mapping the

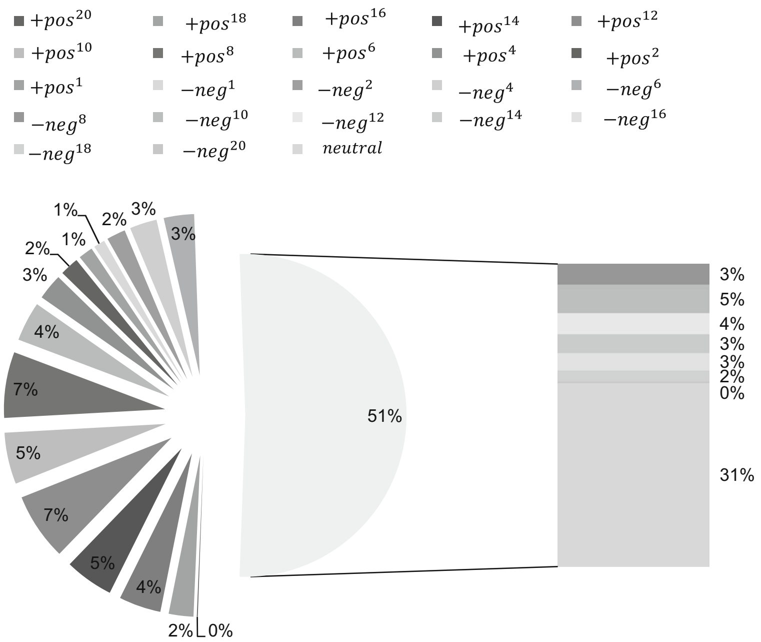

Fig. 9 The detail of the polarity level in various classes of comments

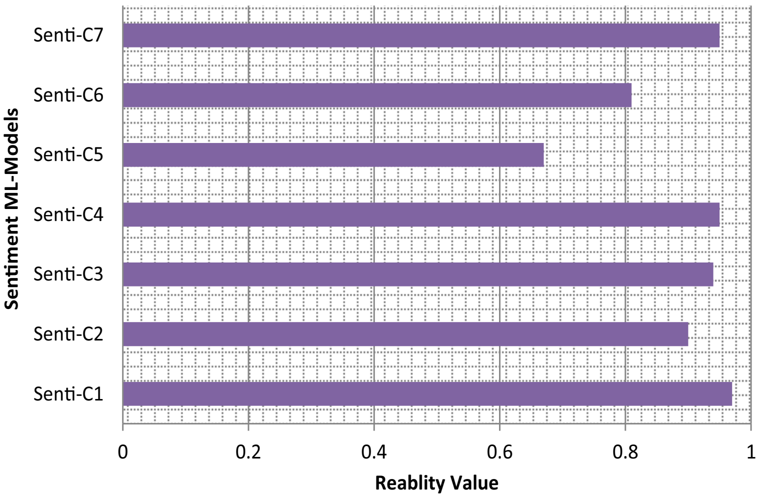

Fig. 10 Statistical comparssion of the reliability the classifiers based on Kappa Statistic

input model to a feature space with high-dimensional and by considering an optimal hyper-plane to divide two-class data [12].
- Senti-C7(WiSARD): This model a multi-class classification method based on weightless neural network (WNNs) in machine learning domaine [19].

The results of utilizing algorithms were examined with different metrics that includes; accuracy and error rates for each topic class (i) generalized, equations 5 and 6 are used:

$$Topic_{Overall\ Accuracy} = \frac{N_{tp} + N_{tn}}{N_{tp} + N_{tn} + N_{fp} + N_{fn}} \tag{5}$$

$$Topic_{Overall\ Error} = 1 - \frac{N_{tp} + N_{tn}}{N_{tp} + N_{tn} + N_{fp} + N_{fn}} \tag{6}$$

where N_{tp}, N_{tn}, N_{fp}, and N_{fn} are the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), respectively. Moreover, we inestigared the models based on other metrics such as reliability or Kappa Statistics(k), Root Mean Square Error (RMSE), and Root Relative Squared Error (RRSE), as following :

$$k = \frac{p_o - p_e}{1 - p_e} \tag{7}$$

$$RMSE = \sqrt{(\frac{1}{n})\sum_{i=1}^{n}(y - \bar{y})^2} \tag{8}$$

$$RRSE(y, \widetilde{y}) = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \widetilde{y}_i)}{\sum_{i=1}^{n}(y_i - \overline{y})}} \tag{9}$$

Where, p_o is the percentage agreement between the classifier models and ground truth, and p_e represents the chance agreement. And also, y is the actual value; $\bar{y}$ is the predicted value; and n is the number of data samples.

However, the FLR techniques can set out learning without a previous knowledge; moreover, a previous knowledge can be provided to the FLR classifier in the form of a primary set

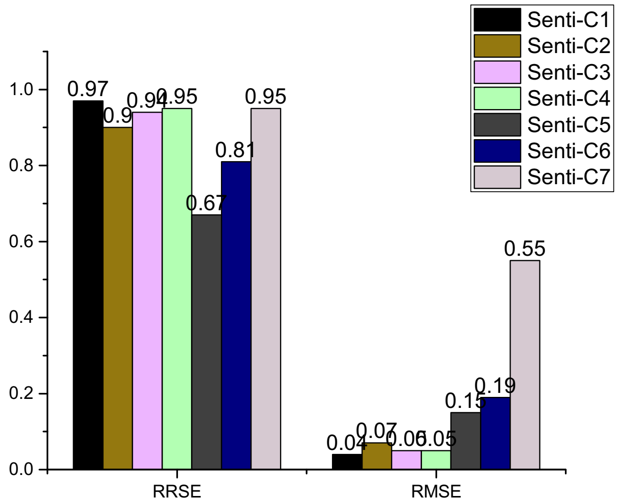

Fig. 11 Statistical comparssion of the RMSE and RRSE for all models

of rules. During training step, this classification model produces hyper-boxes based on the points included in the training step. A hyper-box concord to a rule which demonstrates that a point placed within this box is a part of the respective class. Training is applied iteratively and per point from the training set includes a new rule. For per new rule, the model computes a fuzzy degree of inclusion with the available rules. The best value of these degrees indicates how new/existing rules should be mixed. For categorizing an unclear data point, the system computes the inclusion degrees for the rule enforced by the unclear point and allocates this point to the class of the rule with a great value. Algorithm 1 describes a pseudocode of the hybrid framework for detecting meaningful latent topics and sentiment level analysis, which the input data consists of the number of YouTube movie comments as pure data.

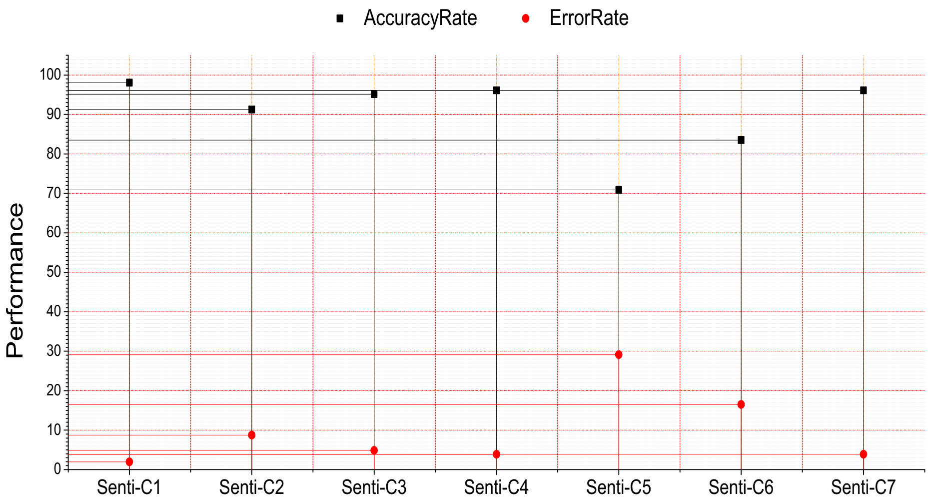

Fig. 12 Performanc of sentiment-levels classification based on accuracy and error rates of classified features

Based on Phases A and B, by using the text-cleaning method to reduce the dimension of the data, the complexity of the phase is approximately O(N). However, for extracting latent topics the memory scales separately by the total number of words in the corpus “N”, and the number of topics, “K” and its time complexity is O(NK) for Phase C. And also by calculating the score for sentiment of each comment, the complexity of the phase is approximately O(N). Based on Phase D, calculated an intelligent classification to detect sentiment levels with thought the Algorithm 2, which is quadratic $O(N_2)$ in the number N of data. Totally, the time complex for this algorithm are O(max(O(n) + O (N) + O(NK) + $O(N_2)$), which will be $O(N_2)$. Figs. 10 and 11 show statistical comparison based on reliablity, RMSE and RRSE metrics with the same parameters. These statistics show the significant results of the proposed framework from other well-know machine learning models. Moreover, the statistical results based on the accuracy and error rates of the sentiment classification presented in Figure 12. It is clear that our framework based on FLR has predicted the sentiment levels or target labels with more accurately than the other traditional models, which the accuracy rate is 98.25%. It should also be noted that some researchers have used deep learning methods to sentiment classification [1, 21, 25], but due to our higher accuracy achieved by hybrid machine learning approaches for our currently used dataset, we postponed to apply deep learning approaches in our future works for different datasets with more complexities, such as emotion recognition [56, 57, 60, 61].

5 Discussion and future works

YouTube is an online media website to the interests of millions with a variety of videos to watch for hours on end. YouTube, as a user-generated content platform, allows users to upload and watch videos, then they can react through a commenting infrastructure. Video content on YouTube has been analyzed in various studies; to the best of our knowledge, this is the first study to comments analysis by considering semantic and sentiment aspects from Oscar-related videos on YouTube. However, analyzing social media such as YouTube could provide meaningful information to understand online users/people about their opinion with comments, which would be difficult to achieve through other traditional ways such as manual methods.

Overall, in Section 4; we provided statistical evidence of the dependency of comment in sentivalues and also we detected meaningful latent topics based on semantic techniques. In this section, we extend the analysis to check whether we can find a dependency of semantics and sentiment aspects for different categories. In this case, we considered an existing data set, which includes 38,855 comments from nine Oscar movies trailer.

About semantic aspects of the model, we extracted the meaningful latent-topics of terms, which are valuable sources about Oscar trailer movies, as shown in Table 3, Figs. 4 and 5. However, a variety of different visualizations were used to interpret the generated LDA results. As we discussed, LDA is a probabilistic model that, when applied to documents, hypothesizes that each document from a collection has been generated as a mixture of unobserved (latent) topics, where a topic is defined as a categorical distribution over words. Regarding the top-ranked topics for the Oscar comments, it is possible to recognize many words probably related to movie, people and song topics, which are first, second and third highest ranked topics respectively. An example for song topic, it seems that some YouTube users are addressing the song of movies and also many times pointed about directors and actors.

Regarding the sentiment aspects of the framework, we examined the strength of positive and negative sentiment expressed in response to user comments on YouTube, the effect of several variables on comments and reply sentiment, and the projected effects that sentiment-based moderation would have had on posted content. The classification is performed twenty-three basic levels of sentiment results, as shown in Fig. 9. We have applied a fuzzy logic resource algorithm and also various machine learning algorithms, the results show that fuzzy logic resource algorithm yielded the best result with the accuracy of 98.25% in sentiment classification on Oscar comments.

This work is limited to English-language content because it was used as a selection criterion. Thus, the results might not apply to videos and comments written in other languages. In addition, this study was limited to the content analysis of videos retrieved on September 10, 2018, from YouTube. Therefore, the external validity of the data is limited and may not be generalized to overall Oscar-related videos available on the Internet. The gap between the period in which the research was being completed and the time frame of our study may have somewhat affected the timeliness of our results. The comments used for this study was from YouTube and was anonymous, which may affect the validity and value of the results. For example, it is virtually impossible to be certain about the identity of those responsible for creating comments because individual users may have (illegally) commented videos using fake email identification. Moreover, the age of users cannot be confirmed because minors can fake their age to gain access to restricted content. Overall, the study suggests that our hybrid framework with both semantic and sentiment techniques based on FLR, generated some valuable information from the YouTube comments. However, these kinds of statics, findings contribute can be useful for a better understanding of the behavior of general YouTube users, or obtain user opinions to help movie directors. Regarding future work, we plan to consider other scientific domains on YouTube, via deep learning techniques that can be used in future for sentiment level classification, and it will be a novel way to retrieve meaningful latent topics of these comments.

6 Conclusion

In this research, we proposed a hybrid framework that covers semantic and sentiment aspects by utilizing the advantage of Fuzzy lattice reasoning for detecting polarity, meaningful latent topics of the movie user comments on YouTube. In fact, we performed extensive experiments of YouTube comments about Oscar movies and demonstrated the strong dependencies between different kinds of sentiments expressed in comments and also detected the topic orientation of the discussed video contents. Regardless of the neutral sentiments, we noticed that most comments and replies were positive, although some topics were more likely than others to elicit negative sentiment. Overall, this study can be considered an alternative approach to assess the user opinions about YouTube comments in Oscar videos by covering both semantic and sentiment features. In other perspectives, YouTube comment analysis can provide further insights on different types of users and on social relationships between users. For instance, this approach is applicable in identifying groups of users with similar interests and in making recommendation systems for recommending contacts or groups to users of the social media systems. However, this study only considered the comenents in English language, and it requires to cover other languages too; besides, the used comments in this research are from the Oscar movies' trailers and there is a need to consider the user comments from the full movies as well.

Acknowledgments This article has been awarded by the National Natural Science Foundation of China (61941113), the Fundamental Research Fund for the Central Universities (30918015103, 30918012204), Nanjing Science and Technology Development Plan Project (201805036), China Academy of Engineering Consulting Research Project(2019-ZD-1-02-02), National Social Science Foundation (18BTQ073), State Grid Technology Project (5211XT190033).

References

1. Abdi A, Shamsuddin SM, Hasan S, Piran J (2019) Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion. Information Processing & Management 56(4):1245–1259
2. Aha D, Kibler D (1991) Instance-based learning algorithms. Machine Learning. 6:37–66
3. Ahmad U, Zahid A, Shoaib M, AlAmri A (2017) HarVis: An integrated social media content analysis framework for YouTube platform. Inf Syst 69:25–39
4. Amarasekara I, Grant WJ (2019) Exploring the YouTube science communication gender gap: A sentiment analysis. Public Underst Sci 28(1):68–84
5. Athanasiadis IN (2007) The fuzzy lattice reasoning (FLR) classifier for mining environmental data. In: Computational intelligence based on lattice theory (pp. 175–193). Springer, Berlin, Heidelberg
6. Bhuiyan H, Ara J, Bardhan R, Islam MR (2017) Retrieving youtube video by sentiment analysis on user comment. In: 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) (pp. 474–478). IEEE
7. Blei DM, Lafferty JD (2009) Topic models. In: Text Mining (pp. 101–124). Chapman and Hall/CRC
8. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. Journal of machine Learning research 3:993–1022
9. Chauhan GS, Meena YK (2019) YouTube Video Ranking by Aspect-Based Sentiment Analysis on User Feedback. In: Soft Computing and Signal Processing (pp. 63–71). Springer, Singapore
10. Cheng Z, Chang X, Zhu L, Kanjirathinkal RC, Kankanhalli M (2019) MMALFM: Explainable recommendation by leveraging reviews and images. ACM Transactions on Information Systems (TOIS) 37(2):1–28
11. Cheng Z, Ding Y, He X, Zhu L, Song X, Kankanhalli MS (2018) A3NCF: an adaptive aspect attention model for rating prediction. In: IJCAI, pp 3748–3754
12. Chidambarathanu K, Shunmuganathan KL (2019) Predicting user preferences on changing trends and innovations using SVM based sentiment analysis. Clust Comput, 1–5
13. Cordero P, Enciso M, Mora A, Ojeda-Aciego M, Rossi C (2015) Knowledge discovery in social networks by using a logic-based treatment of implications. Knowl-Based Syst 87:16–25
14. Cripps A, Nguyen N (2007) Fuzzy lattice reasoning (FLR) classification using similarity measures. In: Computational Intelligence Based on Lattice Theory (pp. 263–284). Springer, Berlin, Heidelberg
15. Cunha AAL, Costa MC, Pacheco MAC (2019) Sentiment Analysis of YouTube Video Comments Using Deep Neural Networks. In: International Conference on Artificial Intelligence and Soft Computing (pp. 561–570). Springer, Cham
16. Curiskis SA, Drake B, Osborn TR, Kennedy PJ (2019) An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. Information Processing & Management
17. Cutler A, Zhao G (2001) Pert-perfect random tree ensembles. Computing Science and Statistics 33:490–497
18. Das S, Dutta A, Lindheimer T, Jalayer M, Elgart Z (2019) YouTube as a Source of Information in Understanding Autonomous Vehicle consumers: Natural Language Processing Study. Transp Res Rec, 0361198119842110
19. De Gregorio M, Giordano M (2018) An experimental evaluation of weightless neural networks for multi-class classification. Appl Soft Comput 72:338–354
20. Denecke K, Deng Y (2015) Sentiment analysis in medical settings: New opportunities and challenges. Artificial intelligence in medicine 64(1):17–27
21. Dogan E, Kaya B (2019) Deep Learning Based Sentiment Analysis and Text Summarization in Social Networks. In: 2019 International Artificial Intelligence and Data Processing Symposium (IDAP) (pp. 1–6). IEEE
22. Edara DC, Vanukuri LP, Sistla V, Kolli VKK (2019) Sentiment analysis and text categorization of cancer medical records with LSTM. Journal of Ambient Intelligence and Humanized Computing, pp 1–17
23. Ernst J, Schmitt JB, Rieger D, Beier AK, Vorderer P, Bente G, Roth HJ (2017) Hate beneath the counter speech? A qualitative content analysis of user comments on YouTube related to counter speech videos. Journal for Deradicalization 10:1–49

24. Ezpeleta E, Iturbe M, Garitano I, de Mendizabal IV, Zurutuza U (2018) A Mood Analysis on Youtube Comments and a Method for Improved Social Spam Detection. In: International Conference on Hybrid Artificial Intelligence Systems (pp. 514–525). Springer, Cham
25. Gao ZY, Chen CP (2019) AI Deep Learning with Multiple Labels for Sentiment Classification of Tweets. In: 2019 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 1–5). IEEE
26. Geng Y, Liang RZ, Li W, Wang J, Liang G, Xu C, Wang JY (2016) Learning convolutional neural network to maximize pos@ top performance measure, ESANN 2017 - Proceedings, pp. 589–594
27. Hoiles W, Krishnamurthy V, Pattanayak K (2019) Rationally Inattentive Inverse Reinforcement Learning Explains YouTube Commenting Behavior. arXiv:1910.11703
28. Hsu WY, Hsu HH, Tseng VS (2019) Discovering negative comments by sentiment analysis on web forum. World Wide Web 22(3):1297–1311
29. Hutto CJ, Gilbert E (2014) Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Eighth international AAAI conference on weblogs and social media
30. Jiménez-Zafra SM, Martín-Valdivia MT, Molina-González MD, Ureña-lópez LA (2019) How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain. Artificial intelligence in medicine 93:50–57
31. Kaburlasos VG, Athanasiadis IN, Mitkas PA (2007) Fuzzy lattice reasoning (FLR) classifier and its application for ambient ozone estimation. International journal of approximate reasoning 45(1):152–188
32. Khan ML (2017) Social media engagement: What motivates user participation and consumption on YouTube? Comput Hum Behav 66:236–247
33. Laksono RA, Sungkono KR, Sarno R, Wahyuni CS (2019) "Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes". In: 2019 12th International Conference on Information & Communication Technology and System (ICTS), pp. 49–54. IEEE
34. Li B, Liu P-Y, Hu R-X, Mi S-S, Fu J-P (2012) "Fuzzy lattice classifier and its application to bearing fault diagnosis". Appl Soft Comput 12(6):1708–1719
35. Obadimu A, Mead E, Nihal Hussain M, Agarwal N (2019) "Identifying Toxicity Within YouTube Video Comment". In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, pp. 214–223. Springer, Cham
36. Oksanen A, Garcia D, Sirola A, Näsi M, Kaakinen M, Keipi T, Räsänen P (2015) Pro-anorexia and anti-pro-anorexia videos on youtube: Sentiment analysis of user responses. Journal of medical Internet research 17(11):e256
37. Orimaye SO, Alhashmi SM, Eu-gene S (2012) Sentiment analysis amidst ambiguities in YouTube comments on Yoruba language (nollywood) movies. In: Proceedings of the 21st International Conference on World Wide Web (pp. 583–584). ACM
38. Ottoni R, Cunha E, Magno G, Bernardina P, Meira W Jr, Almeida V (2018) Analyzing right-wing youtube channels: Hate, violence and discrimination. In: Proceedings of the 10th ACM Conference on Web Science (pp. 323–332). ACM
39. Pal SK, Mitra S (1992) Multilayer perceptron fuzzy sets, and classification. IEEE Trans. Neural Networks 3(5):683–697. https://doi.org/10.1109/72.159058
40. Poché E, Jha N, Williams G, Staten J, Vesper M, Mahmoud A (2017) Analyzing user comments on YouTube coding tutorial videos. In: Proceedings of the 25th International Conference on Program Comprehension (pp. 196–206). IEEE Press
41. Rambocas M, Pacheco BG (2018) Online sentiment analysis in marketing research: a review. Journal of Research in Interactive Marketing 12(2):146–163. https://doi.org/10.1108/JRIM-05-2017-0030
42. Rangaswamy S, Ghosh S, Jha S, Ramalingam S (2016) Metadata extraction and classification of YouTube videos using sentiment analysis. In: 2016 IEEE International Carnahan Conference on Security Technology (ICCST) (pp. 1–2). IEEE
43. Schmidt T, Burghardt M, Dennerlein K, Wolff C (2019) Sentiment annotation in lessing's plays: Towards a language resource for sentiment analysis on german literary texts. Language, Data & Knowledge, 2019
44. Sharma A, Dey S (2012) A comparative study of feature selection and machine learning techniques for sentiment analysis. In: Proceedings of the 2012 ACM research in applied computation symposium (pp. 1–7). ACM
45. Soldner F, Ho JCT, Makhortykh M, van der Vegt IW, Mozes M, Kleinberg B (2019) Uphill from here: Sentiment patterns in videos from left-and right-wing YouTube news channels. In: Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science, pp 84–93
46. Tarımer İ, Çoban A, Kocaman AE (2019) Sentiment Analysis on IMDB Movie Comments and Twitter Data by Machine Learning and Vector Space Techniques. arXiv:1903.11983
47. Thelwall M, Buckley K, Paltoglou G (2012) Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology 63(1):163–173

48. Thelwall M, Buckley K, Paltoglou G, Cai C, Kappas A (2014) SentiStrength. http://sentistrength.wlv.ac.uk
49. Thulasi PK, Usha K (2016) “Aspect polarity recognition of movie and product reviews in Malayalam”. In: 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), pp. 1–5. IEEE
50. Tripto NI, Ali ME (2018) Detecting Multilabel Sentiment and Emotions from Bangla YouTube Comments. In: 2018 International Conference on Bangla Speech and Language Processing (ICBSLP) (pp. 1–6). IEEE
51. Tulkens S, Hilte L, Lodewyckx E, Verhoeven B, Daelemans W (2016) The automated detection of racist discourse in dutch social media. Computational Linguistics in the Netherlands Journal 6:3–20
52. Veletsianos G, Kimmons R, Larsen R, Dousay TA, Lowenthal PR (2018) Public comment sentiment on educational videos: Understanding the effects of presenter gender, video format, threading, and moderation on YouTube TED talk comments. PloS one 13(6):e0197331
53. Walker J, Slater S, Kafai Y (2019) “A Scaled Analysis of How Minecraft Gamers Leverage YouTube Comment Boxes to Participate and Collaborate.”
54. Wu SJ, Chiang RD, Chang HC (2018) Applying sentiment analysis in social web for smart decision support marketing. Journal of Ambient Intelligence and Humanized Computing, pp 1–10
55. Xia H, Yang Y, Pan X, Zhang Z, An W (2019) Sentiment analysis for online reviews using conditional random fields and support vector machines. Electron Commer Res, pp 1–18
56. Yang J, She D, Lai YK, Rosin PL, Yang MH (2018) Weakly supervised coupled networks for visual sentiment analysis. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7584–7592
57. Yang J, She D, Sun M (2017) Joint image emotion classification and distribution learning via deep convolutional neural network. In: IJCAI, pp 3266–3272
58. Zhang G, Liang G, Li W, Fang J, Wang J, Geng Y, Wang JY (2017) Learning convolutional ranking-score function by query preference regularization. In: International conference on intelligent data engineering and automated learning (pp. 1–8). Springer, Cham
59. Zhang G, Liang G, Su F, Qu F, Wang JY (2018) Cross-domain attribute representation based on convolutional neural network. In: International Conference on Intelligent Computing (pp. 134–142). Springer, Cham
60. Zhao S, Jia Z, Chen H, Li L, Ding G, Keutzer K (2019) Pdanet: Polarity-consistent deep attention network for fine-grained visual emotion regression. In: Proceedings of the 27th ACM International Conference on Multimedia, pp 192–201
61. Zhao S, Ma Y, Gu Y, Yang J, Xing T, Xu P, Keutzer K (2020) An End-to-End visual-audio attention network for emotion recognition in user-generated videos. In: Proceedings of the AAAI Conference on Artificial Intelligence, 34(01), pp 303–311