Application of Sentiment Lexicons on Movies Transcripts to Detect Violence in Videos

Badriya Murdhi Alenzi[1], Muhammad Badruddin Khan[2]
Information Systems Department
College of Computer and Information Sciences
Al Imam Mohammad Ibn Saud Islamic University(IMSIU) , Riyadh, KSA

***Abstract*—In the modern era of technological development, the emergence of Web 2.0 applications, related to social media, the dissemination of opinions, feelings, and participation in discussions on various issues have become very easy, which have led to a boom in text mining and natural language processing research. YouTube is one of the most popular social sites for video sharing. This may contain different types of unwanted content such as violence, which is the cause of many social problems, especially among children like aggression and bullying at home, in school and in public places. The research work reports performance of two different sentiment lexicons when they were applied on video transcripts to detect violence in YouTube videos. The automation of process to detect violence in videos can be helpful for censor boards that can use the technology to restrict violent video for a certain age group or can fully block entire video regardless of age. The models were built using the existing sentiment lexicons. The dataset consists of 100 English video transcripts collected from the web and was annotated manually as violent and non-violent. Various experiments were performed on the dataset using English SentiWordNet (ESWN) and Vader Package with different text preprocessing settings. The Vader package outperformed the ESWN by providing 75% accuracy. ESWN results for all POS tagging with 66% accuracy were better than its result for adjectives POS tagging with 58% accuracy.**

Keywords—Sentiment lexicons; sentiment analysis; video transcript; part-of-speech tagging; English SentiWordNet; Vader Package; violence detection

I. Introduction

In recent years, text mining has gained increasing attention as huge amounts of text data (unstructured data) are created by using the web and social networks. The increasing amount of text data has created a need for methods and algorithms which can be used to learn interesting patterns from the data in a scalable and dynamic way. Automated sentiment analysis (the computational study of people‘s opinions and emotions about individuals, issues, events, topics and their attributes) can be used to analyze the text data and find interesting patterns and relationships about different topics. One of the primary concerns in relation to web users is the harmful and inappropriate content on the web [1].

YouTube is a popular video sharing site, where users are allowed to upload, view, share, rate, and comment on videos, and subscribe to other YouTube users. It offers a wide variety of videos that have different contents including TV shows, video clips, documentary films, movie trailers, and educational videos, etc. In August 2017, YouTube was ranked as the second-most popular site in the world, there were 400 hours of content uploaded to YouTube site every minute and one billion hours of content were watched every day [2]. Videos on YouTube carry different content, which may contain many unwanted things such as violence. Violence is the cause of many problems, especially among children like aggression and bullying at home, in school and in public places.

This research work can be used to make video sharing sites like YouTube more suitable and safe from inappropriate content. This will protect children and make parents more comfortable by having control of what their children are watching. There are some studies focused on detecting violence, but there is only one study focused on detecting violence in videos using movies comments [3]. Furthermore, there are limited studies analyzing video transcripts for different purposes but there are no studies focusing on detecting violence in videos using video transcripts (video transcription is the process of translating a video's audio into text) [4]. Sentiment analysis is basically meant to understand opinion and emotions for a particular issue.

In this work, we hypothesize that sentiment analysis on video transcripts can help us in detecting violence in videos. It means that if the sentiment detected for particular movie using its movie transcript is positive, it means that there exist almost no violence in that particular movie. On the other hand, if the sentiment detected is negative, then violence is present in that movie. In the experiments, two sentiment lexicons were applied on video transcripts to automatically classify videos into violent or non-violent classes. The effects of text preprocessing techniques on video transcripts were also examined based on the classifications' accuracy.

The main contribution of this work is the novel application of sentiment lexicons that were developed for sentiment analysis in the field of violence detection.

The structure of the paper is as follows: Section (2) introduces some of the previous studies that are related to the topic of the presented work. Section (3) presents the methodology used to develop this paper study. Section (4) presents an overview of the main results. Section (5) indicates our conclusions and recommendations for future work.

II. Related Work

With the rapid growth of social media on the web, individuals and organizations need to analyze public opinions in these media for best decision making [5]. Sentiment analysis aims to determine if the expression of the text about a certain domain is positive, negative, or neutral emotions [6]. Sentiment analysis has two approaches corpus-based and lexicon-based. The corpus-based approach is a supervised approach using machine learning classifiers which are applied to a manually annotated or labeled dataset. On the other hand, the lexicon-based approach is the unsupervised approach, which states the polarity, semantic orientation, of a word or sentence based on a dictionary [7]. Many studies have been done on sentiment analysis of different types of social media platforms such as YouTube. YouTube is the most popular video sharing site where a huge number of videos are posted every day. M. Wöllmer et al. [8] focused on automatically analyzing the movie reviews of online videos to determine a speaker's sentiment. L. P. Morency, R. Mihalcea,, and P. Doshi [9] classified the polarity of the opinions in online videos by using multimodal sentiment analysis, and explored the mutual use of multiple modalities. S. Poria et al. [10] suggested a new methodology for multimodal sentiment analysis, which explaining a model that uses audio, visual and textual modalities to gathering sentiments from Web videos. M. Thelwall, P. Sud, and F. Vis [11] discovered the reaction of audience to important issues or particular videos by analyzing large samples of YouTube videos' text comments.

One of the primary concerns for web users is the harmful and inappropriate content on the web. There are many researchers who studied different forms of violence. Y. Elovici et al [12] studied a terrorist detection system that was used to monitor a specific group of users' traffic, give an alarm if the access information was not within the group interests, and analyzed the content of information that the suspected terrorists had been accessing. P. Calado et al. [13] discussed text-based similarity metrics combined with link-based similarity measures and used them to classify web documents for anti-terrorism applications. W. Warner and J. Hirschberg [14] developed an approach for detecting hate words on the web and developed a mechanism for detecting methods used to avoid common - dirty words filters. S. Liu and T. Forss [15] used the existing methods of topic extraction, topic modeling and sentiment analysis to develop a content classification model used to detect violence, intolerance, and hateful web content. D. Won, Z. C. Steinert-Threlkeld, and J. Joo [16] developed a visual model used to recognize the activities of protesters by detecting visual attributes, and evaluate the level of violence in an image. The study collected geotagged tweets (tweets assigned to an electronic tag that assigns a geographical location) and their images from 2013 to 2017, and then a multi-task network was used to classify the existence of protesters in an image and predict the visual attributes of the image that observed violence and showed emotions. D. A. AlWedaah [3] attempted to detect violence in cartoon videos by using text mining techniques to the video‘s comments. The study built classifiers by collecting comments for 1,177 YouTube cartoon videos. The classifiers were applied by using RapidMiner and natural language toolkit (NLTK) in Python. The study used classification algorithms such as decision trees (DTs). Support Vector Machine (SVM) and Naïve Bayes were used and it used text preprocessing techniques in order to increase the classifiers performance. NLTK and Naïve Bayes classifier gave the best accuracy of 91.71% and an error rate of 8.29% in predicting video violence.

There are few studies using video transcripts in text mining and sentiment analysis. N. Sureja et al. [17] used the movies‘ subtitles and movie genre such as thriller, comedy, action, drama, and horror to build a sentimental analysis model using lexicons, which are context specific to each considered movie‘s genre. A. Denis et al. [20] presented a preliminary approach to visualize the effects carried by movies by affective analysis of movies' scripts. A. Blackstock and M. Spitz [18] classified movie scripts into genres based on the features of natural-language processing that were extracted from the scripts. The study innovated two evaluation metrics, Partial Credit (PC) Score and F1 score, to analyze the performance of a Maximum Entropy Markov Model and Naive Bayes Classifiers. U. Sinha, and R. K. Panda [19] used movies‘ subtitles to detect emotional scenes; the major emotions included happiness, sad, love, surprise, emotionless, disgust, fear, and anger, by applying Natural Language Processing (NLP) techniques on video subtitle dialogues.

There are some studies that were focused on detecting violence, but there is only one study, to the best of our knowledge, that was focused on detecting violence in videos using movies comments. Furthermore, there are limited studies analyzed video transcripts for different purposes. There are no studies that were focused on detecting violence in videos using video transcripts. This paper focused on using the sentiment lexicons in detecting violence in YouTube movies by analyzing Anime video transcripts which is new input type to be used for the purpose of the study. For the purpose of the research work, the corpus was created containing 100 English Anime video transcripts that were manually annotated as violent and non-violent by three persons including the researcher.

III. Methodology

In the following section, we present the used methodology of the sentiment analysis which includes three phases (data collection, pre-processing, and classification and evaluation). Also we describe the techniques that were used in each phase.

A. *Data Collection*

This is the first step of the methodology that consists of five steps as follows: 1) Anime cartoons were selected based on the requirements of the study from the Web (Anime, a style of Japanese film and television animation, typically aimed at adults as well as children). Three different Anime cartoon series were selected. 2) YouTube was used to watch the Anime episodes, each with the duration of 20 minutes, to divide each episode into individual scenes. 3) The corresponding Anime transcripts of the chosen scene were collected by searching on the World Wide Web. 4) Each scene was annotated manually as violent and non-violent based on its content. 5) The transcripts were saved in Excel file with the annotation (label). 100 scenes were gathered from 68 video transcripts.

1) Data cleaning: After collecting the scenes, they were cleaned by deleting the names from the scenes, which indicate the person who speaks in the scene as the names of the scene's characters are not important in this analysis.

Example:

Raw dialogue:

TAKADA: Misa, I‘m sorry to have to invite you to dinner so late at night.

I‘m afraid it had to wait until after the 9 o‘clock news was finished.

MISA: No problem! I don‘t mind it at all. I‘m a night owl anyway.

Cleaned dialogue:

Misa, I‘m sorry to have to invite you to dinner so late at night.

I‘m afraid it had to wait until after the 9 o‘clock news was finished.

No problem! I don‘t mind it at all. I‘m a night owl anyway.

2) Scenes annotation: In this step, the scenes were annotated in the dataset manually by three persons including the researcher into two classes: violent or non-violent scene after viewing the scenes not based on the text of the scene transcript. Then the scene was saved as a raw text in excel sheet. For the corpus, there were 50 scenes annotated as violent and 50 scenes annotated as non-violent.

3) Data collection issues: The first issue was searching for Anime transcripts on the web as it was a time-consuming process. After the search was complete, the Anime movie needed to be compared with its transcript text to ensure that they were identical. Then, based on the visual view of the movie, the transcript was divided into different scenes. The second issue was that the datasets needed a cleaning process as most of the collected scenes had the character‘s name at the start of each sentence, which were not needed in the analysis. The third issue was that the scene annotation was a complex process. For example, the scene describes a person sitting in a place who tells the other person a violent story. By the visual view, this scene should be annotated as a non-violent scene but by the words the scene should be annotated as a violent scene. Thus, there are some rules that were used to annotate a scene as violent or non-violent. Some of these rules were as follows:

a) The scenes that include a fight with some weapons and the shedding of blood will be annotated as violent.

b) The scenes that include any bully actions will be annotated as violent.

c) The scenes that include characters with awful forms like a monster will be annotated as violent.

B. Preprocessing Stages

The preprocessing phase is very important in the analysis. It reduces the noise in the scenes to improve the performance of the classification process. This section explains the preprocessing stages of the dataset by using Python programming language. The steps of dataset preprocessing were used as follows:

1) Punctuation removal: The string module in python was used to remove the punctuation from the scene transcript before the tokenization step as each punctuation symbol will be considered as a token and will be source of noise and extra overhead for learning algorithm. They were removed because as an individual token, they do not express any feeling.

2) Tokenization: NLTK library was used to tokenize the dataset which contains a package for tokenization. The scenes should be tokenized to list of words, numbers, and symbols before working on them.

3) Stop words removal: Stop words are not useful in the study because they do not have any sentiment. Therefore, they were removed to improve the performance of the classifier. The English stop words corpus which is built-in in the NLTK library was used; it contains 153 words.

4) Word stemming: Porter stemmer (or Porter stemming algorithm) was used to stem the words in the English dataset. It is used to stem the English words by removing the common morphological and inflexional endings. It is used to get the root of each word in the scene.

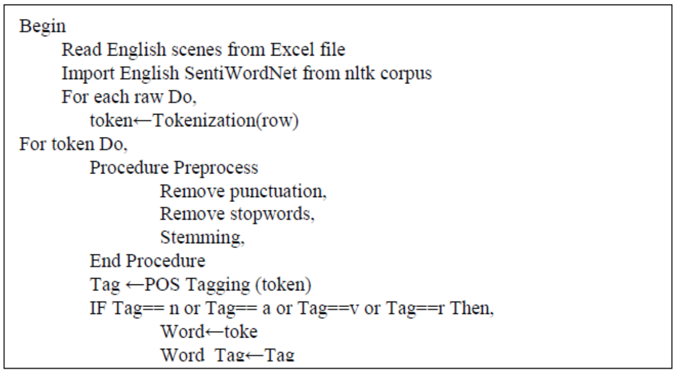
After performed the preprocessing stages on the scene transcripts, the preprocessed scenes were saved in an MS Excel file.

C. Classification

The aim of this work is to use sentiment lexicons to classify movie transcripts into violent / non-violent class. This section describes the classification process of the research work.

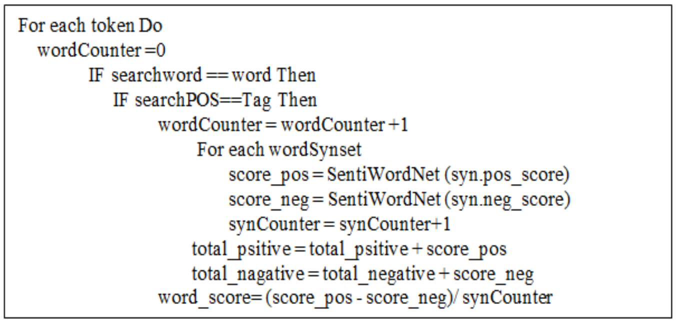
1) English SentiWordNet classification: There are different ways to calculate the sentiment score by using English SentiWordNet (SWN). SentiWordNet lexicon allocates different sentiment weights to different words. The classification process is performed in three stages as shown in the pseudo code of Fig. 1: First, Python preprocessing steps were applied, and then the preprocessed scenes were saved in an MS Excel file. Second, a function was performed, the function read each row in the MS Excel file, which contained a scene and applied the POS tagging for each word, and then saved each word in the scene with its corresponding POS tag in an array. Third, IF function was implemented, where each word that was not tagged as a noun 'n', verb 'v', adjective 'a', or adverb 'r', was excluded from the classification process.

After the words with unwanted tags were removed from each scene, a function was implemented. The steps of the function are as follows: First, the English SentiWordNet was imported from NLTK corpus. Then, the English scenes were read from the Excel file. For each word (or token) in each scene, the word was searched in ESWN with its corresponding synsets. Each word in SWN contains a number of synsets and each synset contains three scores: positive, negative, and neutral. The score of each word was calculated from the average of the word sysnsets scores (calculating the total positive and the total negative score of all the word synsets, subtracted the total negative score from the total positive score, and divided them by the number of the synsets of the word) by using Equation 3.1 and pseudo code of Fig. 2.



```
Begin
    Read English scenes from Excel file
    Import English SentiWordNet from nltk corpus
    For each raw Do,
        token←Tokenization(row)
For token Do,
        Procedure Preprocess
                Remove punctuation,
                Remove stopwords,
                Stemming,
        End Procedure
        Tag ←POS Tagging (token)
        IF Tag== n or Tag== a or Tag==v or Tag==r Then,
                Word←toke
                Word_Tag←Tag
```

Fig 1. Pseudo Code of Preprocessing Steps.



```
For each token Do
   wordCounter =0
        IF searchword == word Then
           IF searchPOS==Tag Then
                 wordCounter = wordCounter +1
                      For each wordSynset
                           score_pos = SentiWordNet (syn.pos_score)
                           score_neg = SentiWordNet (syn.neg_score)
                           synCounter = synCounter+1
                     total_psitive = total_psitive + score_pos
                     total_nagative = total_negative + score_neg
                 word_score= (score_pos - score_neg)/ synCounter
```

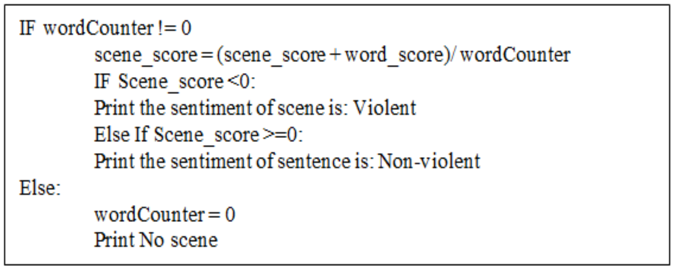
Fig 2. Pseudo Code of Word Classification (ESWN).

$$Score\ (W) = \frac{1}{n}\sum_{i=1}^{n} Score_{pos}(S_i) - Score_{neg}(S_i)$$

Equation 1: Word Score Equation

W is the word, S is the score of synset i for word W, and n is the number of synsets of word W.

After the scores of each word are calculated, the score of each scene is calculated by calculating the average words scores for each scene by using Equation 3.2 and pseudo code of Fig. 3.



```
IF wordCounter != 0
        scene_score = (scene_score + word_score)/ wordCounter
        IF Scene_score <0:
        Print the sentiment of scene is: Violent
        Else If Scene_score >=0:
        Print the sentiment of sentence is: Non-violent
Else:
        wordCounter = 0
        Print No scene
```

Fig 3. Pseudo Code of Scene Classification (ESWN).

$$Total\ Score\ (scene) = \frac{1}{n}\sum_{i=1}^{n} Score\ (W_i)$$

Equation 2: Scene Score Equation

W_i is the score of word i of the scene, and n is the number of words in the scene (scene transcript).

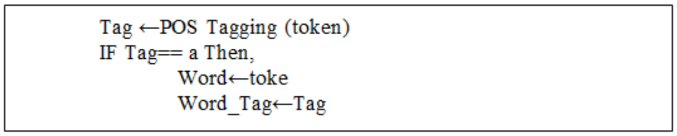
The sentiment score of scene transcript was mapped with violent or non-violent class based on threshold of zero as given in following formula.

Sentiment =
Violent if Total Score (scene) < 0
Or
Non-violent if Total Score (scene) > =0

It can be seen from the above formula, that if the total sentiment score of the scene is greater than or equal to zero then the scene is classified as 'non-violent'. On the contrary, if total score of the scene is lesser than zero then the scene is classified as 'violent'.

Disambiguation Method: A variation to calculate score

To improve the classification accuracy, disambiguation method was used by calculating the sentiment score of the words by taking only the synsets of the word that corresponds with the part of speech tag "adjective". Then the word and scene scores were calculated by using Equations 3.1 and 3.2. In order to understand the disambiguation method, the pseudo code is given in Fig. 4.



```
Tag ←POS Tagging (token)
IF Tag== a Then,
        Word←toke
        Word_Tag←Tag
```

Fig 4. Pseudo code of the Disambiguation Method.

2) Vader classification: The Valence Aware Dictionary and sEntiment Reasoned (VADER) is a python package used with English text only. First, the Sentiment Intensity Analyzer object was loaded from the VADER package, then, the polarity scores method was used to get the sentiment scores of the scenes [21]. By using the Vader package, there was no need to use the preprocessing steps like punctuation removal or tokenization as Vader does not just do simple matching between the words in the text and its lexicon. It also considers certain things about the context of the words and the way the words are written. Moreover, to increase the intensity of the words sentiment, the scenes were analyzed with capitalization and exclamation marks [21]. The Vader package was used to get the positive, negative, neutral, and compound scores for each English scene. The sentiment score of each scene was calculated by using Equation 3.3, by subtracting the negative score from the positive score of each scene.

$$Sentiment\ Score\ (scene) \\ = Score_{pos}(Scene\,) - Score_{neg}(\mathrm{Scene})$$

Equation III: Vader Sentiment Score Equation.

Example of Vader scoring:

The food is good.

Vader scored this sentence as: {'neg': 0.0, 'neu': 0.508, 'pos': 0.492, 'compound': 0.4404}

Capitalization increases the intensity of both positive and negative words.

The food is GOOD.

Vader scored this sentence as: {'neg': 0.0, 'neu': 0.452, 'pos': 0.548, 'compound': 0.5622}

Exclamation marks increase the intensity of sentiment scores.

The food is GOOD!

Vader scored this sentence as: {'neg': 0.0, 'neu': 0.433, 'pos': 0.567, 'compound': 0.6027}

The words which are present before a sentiment word increase or decrease the intensity of both positive and negative words.

The food is really GOOD!

Vader scored this sentence as: {'neg': 0.0, 'neu': 0.487, 'pos': 0.513, 'compound': 0.6391}

If the sentence contains ‘but’, the sentiments before and after the ‘but’ are considered; however, the sentiment after is weighted more heavily than that before.

The food is really GOOD! But the service is dreadful.

Vader scored this sentence as: {'neg': 0.192, 'neu': 0.529, 'pos': 0.279, 'compound': 0.3222}.

Where ‘neg’ mean negative score, ‘neu’ mean neutral score, ‘pos’ mean positive score and ‘compound’ score, is the sum of all of the lexicon ratings, which have been standardized to range between -1 and 1.

D. *Evaluation Measures*

Using the steps mentioned in previous sections, sentiment for each scene script was found using two sentiment lexicons namely ESWN and VADER. In order to check their performances, we used classic performance measures of confusion matrices, precision, recall, F-measure and accuracy methods. A **confusion matrix** is a table that describes the performance of a classification model on a set of test data where the true values are known [22], as in Fig. 5.

- True positives (TP): The cases in which the scenes were predicted as violent and they are not violent.
- True negatives (TN): The cases in which the scenes were predicted as non-violent and they are non-violent.
- False positives (FP): The cases in which the scenes were predicted as violent and actually they are non-violent. (Known as "Type I error.")
- False negatives (FN): The cases in which the scenes were predicted as non-violent and they are actually violent. (Known as “Type II error.") [22].

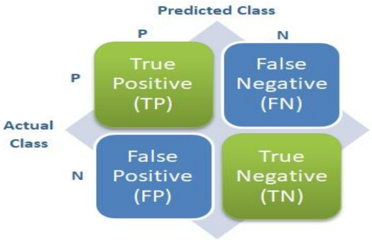


Fig 5. Confusion Matrix.

- **Precision** (positive predictive value): is the percentage of things that were identified positive are really positive [22].

Precision = TP/ (TP+FP)

Equation 4: Precision Equation [22]

- **Recall** (sensitivity) is the percentage of relevant instances that have been retrieved correctly from the total number of relevant instances [22].

Recall = TP / (TP+FN)

Equation 5 Recall Equation [22]

- **The accuracy** of the model which is the **overall success rate** [22].

Accuracy = (TP + TN) / (TP+TN+FP+FN)

Equation 6: Accuracy Equation [22]

- **F-Measure:** it is a harmonic average of obtained precision and recall value [23]. It gives a good indication of the overall performance of a model and it can be calculated by using the following formula:

F-measure = 2 × ((precision × recall) / (precision + recall))

Equation 7: F-measure Equation [22]

IV. Results: Analysis and Discussion

The analysis will be presented after presentation of experiments and their results after performing following steps: 1) Analyze the results after using different preprocessing mechanisms, such as tokenization, stop words removal, and stemming, to get video transcripts with reduced noise and unstructuredness. 2) Follow the methodology phases to build the lexical-based classifier ESWN and measure the classification results. 3) Use Vader, Python package, to classify the English video transcripts.

A. *Experiment 1: Lexical-based Analysis using ESWN Approach*

1) Objective: The objective of this experiment is to get the sentiment scores of the English video transcripts by using the English SentiWordNet lexicon and to examine the performance and accuracy of the sentiment results. In addition, this experiment aims to examine the effect of using POS tagging and different preprocessing stages on the ESWN results performance.

2) Method: The dataset used in this experiment consists of 100 video transcripts, 50 violent scenes and 50 non-violent scenes, which were annotated manually. The experiment is divided into many stages as follows:

- Different preprocessing steps were applied on the video transcripts such as tokenization, punctuation and stop words removal, and stemming.
- Each word in each transcript in the dataset was tagged to suitable POS tagging based on the lexicon.
- The sentiment score of each word in each transcript was calculated by using Equation 3.1 and the sentiment score of the scene was calculated by using Equation 3.2.
- In ESWN, structure the part of speech (POS) tagging is considered an important attribute. Therefore, the scores of the scenes were calculated by using the sentiment scores of only the words which have adjective POS tagging.
- The scenes were annotated as violent and non-violent based on their sentiment scores. If the score was less than zero; the scene was annotated as violent. If the score was greater than or equal to zero; the scene was annotated as non-violent.

For evaluating this experiment, the classification results were compared with the actual labeled dataset to get the performance metrics: confusion matrices, precision, recall, F-measure and accuracy.

3) Results: The following Tables and Graph indicate the results of applying ESWN lexicon on dataset after different preprocessing stages.

TABLE I. ESWN RESULTS AFTER TOKENIZATION ON THE DATASET

Tokenization		Confusion Matrix		Result			
		violent	non-Violent	Precision	Recall	F-Measure	Accuracy
All synsets + all POS tagging	violent	38	12	63.00%	76.00%	69.00%	66.00%
	non-violent	22	28				
POS tagging = adjectives	violent	29	21	55.00%	58.00%	56.00%	55.00%
	non-Violent	24	26				

TABLE II. ESWN RESULTS AFTER TOKENIZATION AND PUNCTUATION & STOP WORDS REMOVAL ON THE DATASET

Tokenization + Punctuation & Stop words removal		Confusion Matrix		Result			
		violent	non-Violent	Precision	Recall	F-Measure	Accuracy
All synsets + all POS tagging	violent	25	25	64.00%	50.00%	56.00%	61.00%
	non-Violent	14	36				
POS tagging = adjectives	violent	25	25	58.00%	50.00%	54.00%	57.00%
	non-Violent	18	32				

TABLE III. ESWN RESULTS AFTER TOKENIZATION, PUNCTUATION & STOP WORDS REMOVAL, AND STEMMING ON THE ENGLISH DATASET

Tokenization + Punctuation & Stop words removal + Stemming		Confusion Matrix		Result			
		violent	non-Violent	Precision	Recall	F-Measure	Accuracy
All synsets + all POS tagging	violent	32	18	63.00%	64.00%	63.00%	63.00%
	non-Violent	19	31				
POS tagging = adjectives	violent	28	22	58.00%	56.00%	57.00%	58.00%
	non-Violent	20	30				

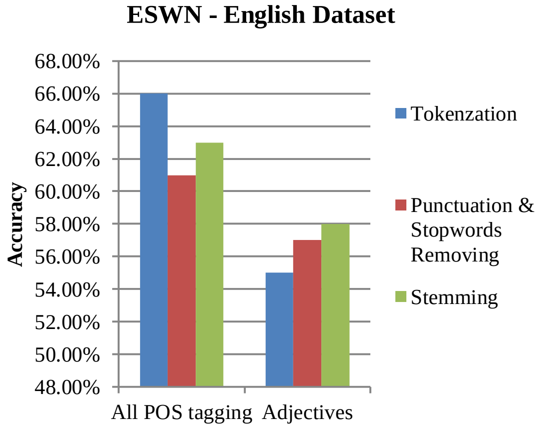


Fig 6. Comparison of Accuracies of ESWN lexicon that were generated for and applied on English dataset.

4) Discussion: The results of the ESWN lexicon performance with and without using the preprocessing operators are shown in a detailed manner in the previous section using Tables 1, 2 and 3, and Fig. 6. The results indicate that the results of ESWN sentiment scores for all POS tagging, nouns, adjectives, verbs, and adverbs, were better than the results of only adjectives POS tagging. However, the results of ESWN sentiment scores for all POS tagging were better before preprocessing while the results of adjectives POS tagging were better after preprocessing.

Punctuation and stop words removal had a negative impact on the ESWN sentiment score of all POS tagging, decreased the accuracy by 5%. However, it had a positive impact on adjectives POS tagging, increased the accuracy by 2%. Moreover, stemming had a negative impact on the ESWN sentiment score of all POS tagging, decreased the accuracy by 3%, while it had a positive impact on adjectives POS tagging, increased the accuracy by 3%.

B. *Experiment 2: Vader Classification*

1) Objective: The objective of this experiment is to get the sentiment scores of the English video transcripts by using Vader, Python sentiment package, and to examine the effect of using different preprocessing stages on the performance of Vader results.

2) Method: The dataset used in this experiment consists of 100 video transcripts, 50 violent scenes and 50 non-violent scenes, which were annotated manually. The experiment was divided into many stages that were applied cumulatively by using Python. The stages are as follows:

- Sentiment Intensity Analyzer object was loaded from the VADER package.
- The polarity scores method was used to get the sentiment scores of the video transcripts, that is, the positive, negative, neutral, and compound scores for each English scene.
- The sentiment scores of each scene were calculated by using Equation 3.3.
- The compound score, the sum of all of the lexicon ratings which have been standardized to range between -1 and 1, for each scene were used as a second sentiment score.
- The scores of the scenes, the one retrieved from Equation 3.3 and the compound scores were annotated as violent and non-violent based on threshold of 0 that is if the score was less than zero; the scene was annotated as violent. If the score was greater than or equal to zero; the scene was annotated as non-violent.

Vader did not need to apply any preprocessing steps on the dataset as Vader did not just do simple matching between the words in the text and its lexicon. Vader package worked on the whole scene and considered certain things about the way the words are written as well as their context. However, the preprocessing mechanisms were used on the video transcripts to discover their effects on the results.

For evaluating this experiment, the classification results were compared with actual annotated dataset and the performance metrics, i.e., confusion matrices, precision, recall, F-measure and accuracy, were retrieved.

3) Results: The following Tables and Graph indicate the results of applying Vader package on dataset after different preprocessing stages.

TABLE IV. THE RESULTS OF USING VADER PACKAGE WITHOUT PREPROCESSING ON THE ENGLISH DATASET

Without preprocessing		Confusion Matrix		Result			
		violent	non-Violent	Precision	Recall	F-Measure	Accuracy
Positive score – Negative Score	violent	35	15	70.00%	70.00%	70.00%	70.00%
	non-violent	15	35				
Compound Score	violent	38	12	71.00%	69.09%	76.00%	72.38%
	non-violent	17	33				

TABLE V. THE RESULTS OF USING VADER PACKAGE AFTER PUNCTUATION & STOP WORDS REMOVAL ON THE ENGLISH DATASET

Punctuation & Stop words removal		Confusion Matrix		Result			
		violent	non-Violent	Precision	Recall	F-Measure	Accuracy
Positive score – Negative Score	violent	36	14	72.00%	72.00%	72.00%	72.00%
	non-violent	14	36				
Compound Score	violent	41	9	71.93%	82.00%	76.64%	75.00%
	non-violent	16	34				

TABLE VI. THE RESULTS OF USING VADER PACKAGE AFTER PUNCTUATION & STOP WORDS REMOVAL AND STEMMING ON THE ENGLISH DATASET

Punctuation & Stop words removal + Stemming		Confusion Matrix		Result			
		violent	non-Violent	Precision	Recall	F-Measure	Accuracy
Positive score – Negative Score	violent	38	12	65.52%	76.00%	70.37%	68.00%
	non-violent	20	30				
Compound Score	violent	40	10	63.49%	80.00%	70.80%	67.00%
	non-violent	23	27				

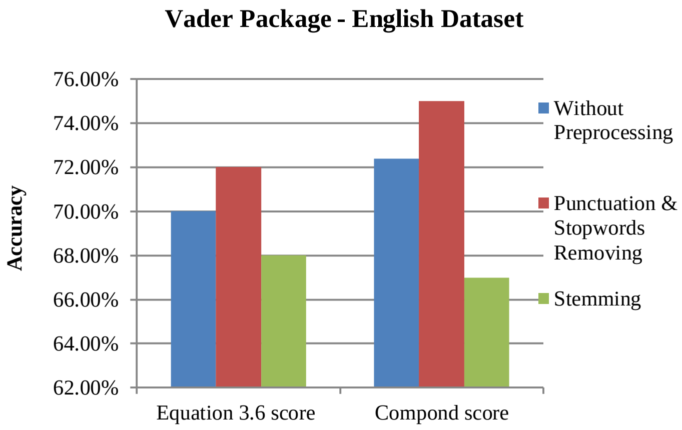


Fig 7. Comparison of accuracies of Vader package that were generated for and applied on dataset.

4) Discussion: The results of the Vader package performance with and without using the preprocessing steps are shown in a detailed manner in the previous section using Tables 4, 5, and 6 and Fig. 7. The results indicate that Vader compound scores of the scenes were better than the results of Equation 3.3 scores.

Punctuation and stop words removal had a positive impact on both the Vader sentiment score of all Equation 3.3 scores, increased the accuracy by 2%, and the compound score, increased the accuracy by 2.62%, while stemming had a negative impact on both Equation 3.3 scores, decreased the accuracy by 2%, and the compound score, decreased the accuracy by 5.38%.

C. Comparison between ESWN and Vader Results

Fig. 8 indicates that violence detection using Vader package was better than ESWN in all settings in terms of accuracy. Similarly Vader outperformed ESWN with respect to other performance metrics namely precision, recall and F-measure.

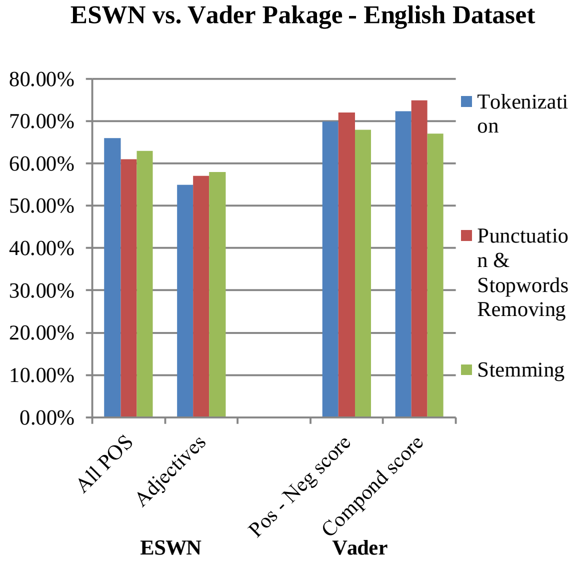


Fig 8. Comparison between the accuracies of ESWN and Vader applied on the dataset.

V. CONCLUSION

There is much research that has studied and discussed existence of harmful and inappropriate content, such as violence, on the web. Also, there are limited studies that focused on detecting violence in YouTube videos. However, those few studies were focused on analyzing the comments of YouTube videos. The studies that have analyzed the transcript of the video are very limited. The main objective of the research is to detect violence in a video at the scene level by mapping the video scene to the video transcript. To achieve this objective, the following approach was followed:

- Manually mapping each movie scene to its video transcript.
- Manually annotating the video transcripts.
- Proposing mechanisms for preprocessing the English video transcripts and using them to reduce the noise of the text. The preprocessing mechanisms used were punctuation and stop words removal and stemming with porter stemming, and POS tagging.
- Proposing a methodology for the lexical-based approach which included using ESWN and Vader, and for extracting sentiment words and calculating the sentiment scores.
- Comparing the classification performance results of the experiments based on two sentiment lexicons.

For Lexical-based classifiers, it is clear that the Vader package outperformed the ESWN by achieving 75% accuracy using settings in which compound scores were used as deciding factor for sentiment assignment on the dataset that was preprocessed with removal of stop words and punctuations. ESWN results for all POS tagging with 66% accuracy were better than its result for adjectives POS tagging with 58% accuracy. This was contrary to our expectations based on the view that adjectives can be main deciding agents to detect violence in a scene transcript.

This study work is the beginning of several new studies on the same topic. There are various aspects required for further studies and analysis. The recommendations for future studies are as follows:

- The dataset size should be increased and other specialized lexicons should be developed to discover violence with better values of different performance metrics.
- Finally, the models which were built in this work should be developed into a system that can be used in different domains such as the YouTube site.

ACKNOWLEDGMENT

This work was supported by King Abdulaziz City for Science and Technology (KACST) Project number AT-200-34.

REFERENCES

[1] J. Han, J. Pei, and M. Kamber, Data mining: concepts and techniques. Elsevier, 2011.

[2] Alexa, "youtube.com Traffic Statistics," Alexa, March 18, 2018.

[3] D. A. AlWedaah, "Detecting violence in YouTube videos using text mining techniques," 2015.

[4] E. GRIFFIN, "3 Reasons Why You Need Video Transcription," February 3, 2015

[5] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in Mining text data: Springer, 2012, pp. 415-463.

[6] T. H. A. Soliman, M. A. M. A. R. Hedar, and M. Doss, "MINING SOCIAL NETWORKS'ARABIC SLANG COMMENTS," in Proceedings of IADIS European Conference on Data Mining, 2013, vol. 22, p. 24.

[7] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, and M. Al-Ayyoub, "Arabic sentiment analysis: Lexicon-based and corpus-based," in Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on, 2013, pp. 1-6: IEEE.

[8] M.. Wöllmer, F. Weninger, T. Knaup, B. Schuller, C. Sun, K. Sagae, & L. P. Morency, (2013). Youtube movie reviews: Sentiment analysis in an audio-visual context. IEEE Intelligent Systems, 28(3), 46-53.

[9] L. P. Morency, R. Mihalcea,, & P. Doshi, (2011, November). Towards multimodal sentiment analysis: Harvesting opinions from the web. In Proceedings of the 13th international conference on multimodal interfaces (pp. 169-176). ACM.

[10] S. Poria, E. Cambria, N. Howard, G. B. Huang, & A. Hussain, (2016). Fusing audio, visual and textual clues for sentiment analysis from multimodal content. Neurocomputing, 174, 50-59.

[11] M. Thelwall, P. Sud, & F. Vis, (2012). Commenting on YouTube videos: From Guatemalan rock to el big bang. Journal of the American Society for Information Science and Technology, 63(3), 616-629.

[12] Y. Elovici, B. Shapira, M. Last, O. Zaafrany, M. Friedman, M. Schneider, & A. Kandel, (2005, May). Content-based detection of terrorists browsing the web using an advanced terror detection system (ATDS). In International Conference on Intelligence and Security Informatics (pp. 244-255). Springer, Berlin, Heidelberg. Springer.

[13] P. Calado, M. Cristo, M. A. Gonçalves, E. S. de Moura, B. Ribeiro-Neto, and N. Ziviani, "Link-based similarity measures for the classification of Web documents," Journal of the Association for Information Science and Technology, vol. 57, no. 2, pp. 208-221, 2006.

[14] W. Warner and J. Hirschberg, "Detecting hate speech on the world wide web," in Proceedings of the Second Workshop on Language in Social Media, 2012, pp. 19-26: Association for Computational Linguistics.

[15] S. Liu and T. Forss, "New classification models for detecting Hate and Violence web content," in Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), 2015 7th International Joint Conference on, 2015, vol. 1, pp. 487-495: IEEE.

[16] D. Won, Z. C. Steinert-Threlkeld, and J. Joo, "Protest Activity Detection and Perceived Violence Estimation from Social Media Images," in Proceedings of the 2017 ACM on Multimedia Conference, 2017, pp. 786-794: ACM.

[17] N. Sureja, "A Review on Movie Script Classification using Sentimental Analysis Approach," 2016.

[18] A. Blackstock and M. Spitz, "Classifying Movie Scripts by Genre with a MEMM Using NLP-Based Features," ed: Citeseer, 2008.

[19] U. Sinha, and R. K. Panda, "Detecting Emotional Scene of Videos from Subtitles," 17th April 2015.

[20] A. Denis, S. Cruz-Lara, N. Bellalem, and L. Bellalem, "Visualization of affect in movie scripts," in Empatex, 1st International Workshop on Empathic Television Experiences at TVX 2014, 2014.

[21] C. H. E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf, 2014.

[22] D. EMC, "data Sience and Big Data Analytics Student Guide," Nov 2013. Y. Sasaki, "The truth of the F-measure," Teach Tutor mater, vol. 1, no. 5, 2007.