Encapsulation Video Classification and Retrieval Based on Arabic Text

Reem A. K. Aljorani[*] and Boshra F. Zopon Al_bayaty

Computer Science Department – Mustansiriyah University

reemi91@yahoo.com

Abstract

Since Arabic video classification is not a popular field and there isn't a lot of researches in this area especially in the educational field. A system was proposed to solve this problem and to make the educational Arabic videos more available to the students. A survey was fulfilled to study several papers in order to design and implement a system that classifies videos operative in the Arabic language by extracting its audio features using azure cognitive services which produce text transcripts. Several preprocessing operations are then applied to process the text transcript. A stochastic gradient descent SGD algorithm was used to classify transcripts and give a suitable label for each video. In addition, a search technique was applied to enable students to retrieve the videos they need. The results showed that SGD algorithm recorded the highest classification accuracy with 89.3 % when compared to other learning models. In the section below, a survey was introduced consisting of the most relevant and recent papers to this work.

Encapsulation Video Classification and Retrieval Based on Arabic Text

Reem A. K. Aljorani and Boshra F. Zopon Al_bayaty

دراسة شاملة عن تصنيف واسترجاع مقاطع الفيديو بالاعتماد على النص باللغة العربية

ريم عبدالكريم جاسم و بشرى فاضل البياتي

قسم علوم الحاسبات – الجامعة المستنصرية

الخلاصة

بما أن تصنيف مقاطع الفيديو العربية ليس مجالًا شائعًا ولا يوجد الكثير من الأبحاث في هذا المجال خاصة في المجال التعليمي. تم اقتراح نظام لحل هذه المشكلة وجعل الفيديوهات التعليمية باللغة العربية متاحة أكثر للطلاب. تم انجاز استبيان لدراسة عدة بحوث وذلك من اجل تصميم وتنفيذ نظام لتصنيف مقاطع الفديو المنطوقة باللغة العربية من خلال استخلاص خصائصها الصوتية باستخدام خدمات Azure المعرفية والتي تنتج ملفات نصية. ثم يتم تطبيق العديد من عمليات المعالجة المسبقة لمعالجة هذه النصوص. كما تم استخدام خوارزمية انحدار الميل العشوائي لتصنيف النصوص الى الفئة المناسبة لكل فيديو. بالإضافة إلى ذلك ، تم تطبيق تقنية بحث تمكن الطلاب من استرجاع مقاطع الفيديو التي يحتاجون إليها. أظهرت النتائج أن خوارزمية SGD سجلت أعلى دقة تصنيف بنسبة 89.3٪ مقارنة بنماذج التعلم الأخرى. أدناه ، تم تقديم دراسة عن أحدث البحوث ذات الصلة بهذا العمل.

الكلمات مفتاحية: تعدين الفيديو، استرجاع الفيديو، تصنيف النص، التعلم الآلي.

Introduction

Data mining is the process of learning new patterns in a large set of data that is extracted from already existed knowledge. Due to the rapid development in information technology and digital multimedia systems, digital data has increased enormously and made available to users rapidly. Videos nowadays are used widely in the social media because of its content richness as it contains several kinds of data such as text, image, audio, and meta-data. Video mining is one of the main problem areas of the data-mining field. The objective of video mining is to discover and learn new patterns from a large data set of videos. Many applications used videos such as education platforms on the web that uses videos to deliver knowledge to all students in the world in the form of lesson tutorials [1].

Encapsulation Video Classification and Retrieval Based on Arabic Text

Reem A. K. Aljorani and Boshra F. Zopon Al_bayaty

Since digital video content consists of all other digital data formats (Images in the form of frames, Audio sounds, and Text) which characterize the features of the video. Thus, videos retrieval depends essentially on these data formats. A video retrieval process is a process of providing the user with a specific video from the database, which is related to the query given as the user needs. Its applications are wide and important in the multimedia field such as video search engines, and video-on-demand broadcasting [2].

In the meanwhile, the research main question is how to design and implement an Arabic video retrieval system that classify educational videos automatically and serves the Iraqi students in which they can retrieve different videos from the created database upon the query they request.

Searching for the requested video is not an easy task which takes a lot of time and resources. Also, most of the searching techniques on the web depend on detecting the hashtags, the keywords in the title of the video or any other information provided manually by the uploader of the video. To retrieve and obtain the requested videos, an automatic and accurate video classification process must be performed first which is the process of attaching the right labels automatically to a video to facilitate the retrieval process also enhance storage solutions depending on the content of the video [3]. As technologies evolve, the world is heading to use this technology in every corner of life. Several educational channels and online course websites were spread significantly in the past few years to make electronic learning more available to all students in the world. Students mainly learn the knowledge by watching these channels and online courses videos. Therefore, it is considered an important learning resource for students [4].

The rest sections in this paper are arranged as follows, Section 2. will go through some of the most related literatures to the presented work, while, section 3. illustrate the methodology of the proposed system, section 4. Shows the results and provides a discussion upon these results also, section 5. analysis the challenges of creating such a system. Finally, the conclusion is presented in section 6.

Encapsulation Video Classification and Retrieval Based on Arabic Text

Reem A. K. Aljorani and Boshra F. Zopon Al_bayaty

Review of Literatures

There is a lot of researchers who dive through the fields of video mining and information retrieval due to their significant importance in many aspects. Despite that, not so many of them focused on video digital contents (Audio and Text) to extract video features especially when the videos are in the Arabic language. This field is considered a rich field and there is a lot to develop. This research work will focus on the most important related works which had a close relation to the research contribution most of them dealing with the classification and retrieval of Arabic digital content.

1. Kastrati, Z. Imran, Ali S. Kurti and Arianit [5], proposed a system based on text transcripts which were extracted from Coursera, which is a Massive open online courses MOOC platform that contains lecture videos in different disciplines, using Web video text tracks (WebVTT) files. A convolutional neural network (CNN) model is proposed which take advantage of text features in video lectures and classify these textual features using two methods. First, embeddings learned from the MOOC dataset and Second, by using transfer learnings. The model used vector support machine VSM to transform the text into a numerical format to be processed by machine learning techniques then the noise was removed in a pre-processing step by eliminating unwanted words and punctuation. A text segmentation was performed to break the text into tokens which is called tokenization. Weights were assigned to each token or word in order to evaluate the distinctive power of these words by using word embedding technique which represents text document into vectors then a transfer learning technique was used for the learning phase by using pre-trained models such as Word2Vec and GloVe models. The methodology used in this paper aims to extract the audio component from the electronic lecture videos and use the generated transcripts to produce keywords which is used for an effective text-based search technique to retrieve video lectures by training different machine learning algorithms to classify them.

2. L. H. Medida and K. Ramani [6], proposed that the e-lecture videos were gathered from various internet sources of MP4 format then converted to audio by using Fast Forward Moving Pictures Expert Group (FFmpeg) of WAV format. A segmentation process was applied to audio tracks to translate speech into text transcripts using the Google Speech Recognition (GSR) library. Python's gensim library was applied to generate summary and keywords extracted from text transcripts. These keywords were used to train the ML model by implementing the Support Vector Machine, Naive Bayes, and Logistic Regression algorithms. A search technique was performed based on the user query which finds the query given in different text documents and retrieve the lecture video with the matched query. Figure 1 below shows the workflow for designing a machine learning text classification model.
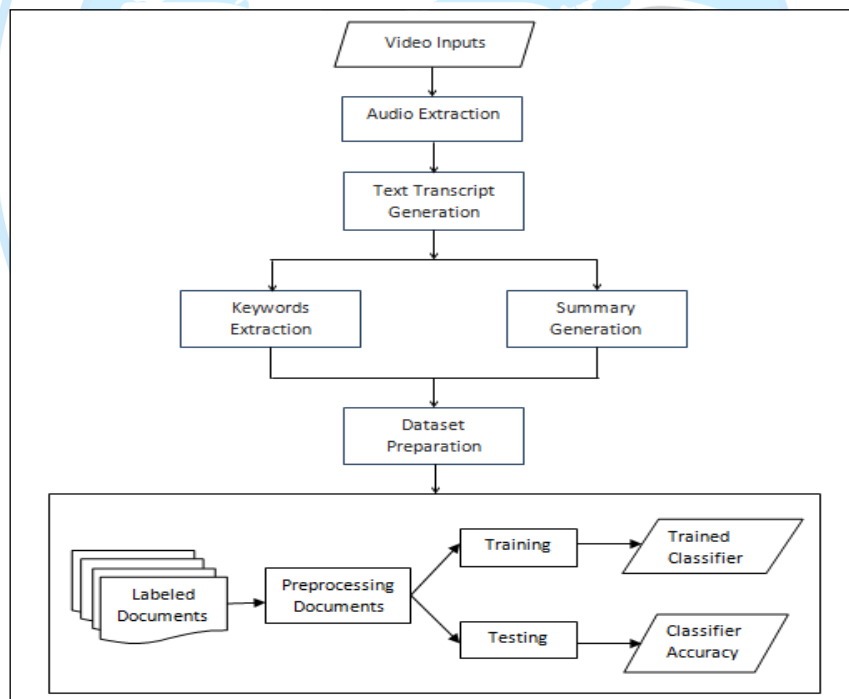
Figure 1: The workflow of the proposed system [6].

3. H. Chatbri et al. [7], suggested an automatic classification for Massive open online database of educational videos. The approach was based on convolutional neural networks (CNN) see Figure 2. The algorithm first extract text transcripts from the video by using speech recognition

techniques then forming images with the extracted text from each video by using a statistical co-occurrence transformation method which was designed and proposed in this paper. At last, the produced images were used as an input to the CNN to give the appropriate labels for each video based on 13 different categories. The Khan Academy dataset was used for training the network. It contains 2,545 videos and each of them was given a label with at most two categories from the 13 categories that existed. The technique proposed in this paper showed a high performance in compare to other techniques similar to this work.
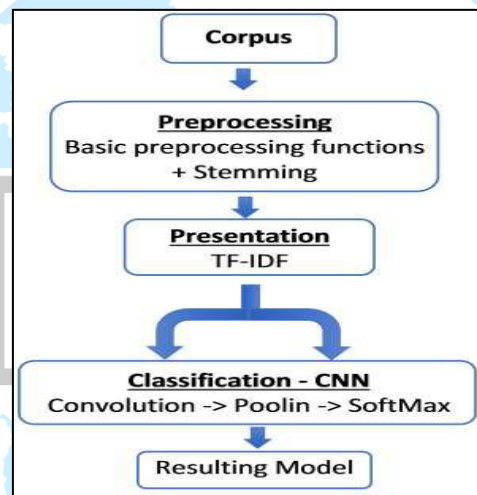

Figure 2: System Architecture [8].

4. S. Boukil et al. [8], a novel method for Arabic text classification was proposed by S. Boukil et al. by using a stemmer algorithm to extract features and choose the most relevant ones to this work then reduce the dimensionality of the feature vectors. Each feature was assigned to be weight by using Tf-Idf algorithm. Convolutional Neural Networks algorithm was implemented in this work to classify Arabic text. A dataset of Arabic newspapers articles with text format was used for training the proposed method. These articles were classified into five categories sport, politic, culture, and economy and diversity. With the text classification problem, the convolution neural network model proved a high performance on large datasets in compare of SVM and other algorithms.

5. K. Sundus, F. Al-Haj and B. Hammo [9], present a deep learning model based on feed-forwarding for Arabic text classification. The model consists of several layers where the first layer is used to calculate the most frequent words which form the vectors generated by term frequency-inverse document frequency (TF-IDF). The second layer uses the first layer output as an input. Adam optimizer was used as an optimization method to reduce the error rate which was computed by a loss function. Two datasets, Khaleej-2004 with 4 categories and a borrowed dataset with 9 categories were used for training and evaluating the efficiency of the proposed method. The method was compared to the logistic regression classification algorithm and the results showed a great improvement in classification accuracy and time building model.

6. D. Kosmajac and V. Keselj [10], used a machine learning algorithm to identify Arabic native language based on support vector machine (SVM) model. A Character and Word N-grams was used to predict the occurrence of character or word in a text document. TF-IDF then was applied to weight the text features. The SVM model was trained using the Stochastic Gradient Descent algorithm. The dataset used by this paper was provided by the INLI Workshop organizers. It consists of six categories collected from social media comments referring to six Indian native languages spoken in different regions in India. The algorithm resulted in distinguishing native language among non-native comments by achieving 89.60% F1 measure.

7. S. Diab [11], proposed a machine learning approach based on the Stochastic Gradient Descent SGD algorithm to classify and enhance text acquired from the Global Terrorism Database. The hyper parameters of the SGD algorithm were fine-tuned by using a Grid search approach in order to give better performance by checking on different methods for feature engineering processes. After applying preprocessing steps to the GTD database which was used to track terrorist attacks incidents, the data was divided into two datasets which is the training dataset with 70% for training purposes and 30% for testing. The Stochastic Gradient Descent SGD was applied to give an enhanced performance over the Logistic Regression, Support Vector Machine, and Perceptron classifiers by optimizing the accuracy and the execution time of the algorithm performance.

8. A. Dandashi, J. Jihad Al Ja'am and S. Foufou [12], suggested a video classification system based on multiple multimedia extracted from the video. A combination of audio based approaches and text based approaches were implemented in this paper to classify Arabic news. The system first, extracts the audio feature from the video then converts the speech into text using Arabic Named Entity Recognition (NER) tools. The system used an Arabic videos dataset extracted from the Aljazeera Arabic news channel which was collected by the author in order to extract audio and text then classify text news. Another dataset was used to classify and test English news videos called TRECVid database. The intended news extracted from videos were the videos concerned with "shooting" and "explosion" categories. An IBM 2011 GALE Arabic speech transcription tool was used to transcribe audio to text. Then the MADAMIRA entity recognition tool was used with a multimodal fusion method aimed to achieve optimal text classification.

9. M. Madhfar et al. [13], evaluated a comparison among six different ML classification models implemented on a large Arabic dataset. The models used in this comparison was Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Logistic Regression, Nave Bayes (NB) Random Forest, and Decision Tree (DT). A series of preprocessing steps were first applied to the Arabic text corpus such as stop word removal and normalization to ignore not so necessary words and punctuations. The corpus was represented as a Bag of Words (BoW) which is a method to represent corpus documents as a matrix. Corpus used in this article contains 111,728 newspapers articles of Arabic documents which consist of five different categories: economy, sport, news, culture, and diversity. Evaluating these models on the mentioned corpus showed that the Logistic Regression model gives the highest accuracy, followed by SGD, SVM, NB, Random Forest, and DT. It also demonstrates that corpus size impact greatly on the performance of the models.

10. A. Elnagar, R. Al-Debsiand and O. Einea [14], introduced a CNN Model consisting of three layers to enhance the performance and computation efficiency by constructing the computed similarities between words. Words were represented by a dense vector and as the learning

process performed, words were take a place in the vector space based on other words in the text document. A pre-trained word embedding model was used in the embedding layer to initialize this layer. To complete the work pre-trained models based on a huge corpus that is made of news articles of SANAD has been created, Arabic Wikipedia, and a variety of Arabic text books. Random weights were assigned to this layer. The output layer, which is the dense layer, would take the number of categories available as its output dimension. Results showed solid performance of all models on SANAD corpus with a minimum accuracy of 91.18%, achieved by convolutional GRU, and top performance of 96.94%, achieved by attention-GRU.

11. Hamroun, Mohamed & Lajmi, Sonia & NICOLAS, Henri & Amous, Ikram [15], proposed a study that deals with the stemming techniques and how these techniques affect the classification process, several classification algorithms were used in this study to give a clear comparison on the impact of stemming techniques, these classification algorithms are support vector machine (SVM), Naïve Bayesian (NB), and K-nearest neighbors (KNN). The feature selection method was used which is the chi-square method to discover the most important and related features. A CNN Arabic corpus was used in this study, which contains a huge number of Arabic websites collected to help the performance of the classification system. The evaluation of the system was done by using the K-fold cross- validation method and Micro-F1. As a result of the study, the ARLStem outperforms the ISRI and Tashaphyne stemmers by enhancing the classification process. In addition, The SVM classifier showed a better performance over the KNN and NB classifiers, by achieving 94.64% Micro-F1 value when using the ARLStem stemmer. Figures below shows the most important phases of this study which is the indexing phase and the retrieval phase. Table 1 display methodologies of each research as well as the accuracy obtained from these methodologies.

Encapsulation Video Classification and Retrieval Based on Arabic Text

Reem A. K. Aljorani and Boshra F. Zopon Al_bayaty

Table 1: Survey Information

REFERENCE NO.	AUTHOR	YEAR OF PUBLICATION	METHODOLOGY	DATASET	ACCURACY %
1	Kastrati, Z. Imran, Ali S. Kurti and Arianit	2019	CNN with transfer learning	Coursera Platform (128 video)	73.08
2	L. H. Medida and K. Ramani	2019	SVM, Naive Bayes, Logistic Regression	lecture video files of MP4 format	12. 5 1 0
3	H. Chatbri et al.	2017	CNN	Khan Academy dataset (2,545 videos)	97.8
4	S. Boukil et al.	2019	CNN	Arabic newspapers articles (111,728 documents)	92.94
5	K. Sundus, F. Al-Haj and B. Hammo	2019	feed-forward deep learning (DL) neural network	Khaleej-2004 (5690 document) Dataset-II(1445 document)	97.6
6	D. Kosmajac and V. Keselj	2017	SVM-SGD	INLI Workshop organizers Dataset	89.6
7	S. Diab	2018	SGD with fine-tuning hyper-parameters	GTD Dataset (170,350 different attack incidents)	88.5
8	A. Dandashi, J. Jihad Al Ja'am and S. Foufou	2016	multimodal fusion component	Arabic news channel (Aljazeera) and social media (raw videos).	—
9	M. Madhfar et al.	2019	SGD, SVM, LR, NB and DT	Hespress , Akhbarona and Assabah newspapers (111,728 Arabic documents)	96
10	A. Elnagar, R. Al-Debsiand O. Einea	2020	CNN, RNN	single-label (SANAD) multi-label (NADIA) Arabic text Datasets	96.94
11	Hamroun, Mohamed & Lajmi, Sonia & NICOLAS, Henri & Amous, Ikram	2019	SVM	CNN Arabic corpus	94.64

The Proposed Methodology

A new system was developed to make Arabic educational videos more available to the students. This work is the first in the Arabic educational field. The system designed and implemented a

model that serves the Iraqi students and the ministry of education by providing a search engine for students to reach out a specific topic they need to understand and provide a convenient environment to search for their most interesting subjects. A dataset of educational videos uploaded by the Iraqi educational television on their channel on the YouTube platform was used.

The audio feature was extracted from the videos in a WAV format by using MoviePy python library, then transcripts were generated using azure cognitive service delivered by Microsoft that converts Arabic audio format into a text document. Texts then were prepared in a preprocessing step to be used in the model by eliminating unwanted words and punctuation. Weights were given to each word using TF-IDF technique then a stochastic gradient decent SGD algorithm using SVM learning model was implemented first, to train the system using the training dataset. Second, to classify each of the videos and figure out to which category the video belongs to.

The training dataset was created manually based on educational documents in the form of text files gathered from Iraqi educational books studied by sixth grade secondary students as well as explanatory bindings which contains a lot of rich information that the students use in their study to be employed for training the algorithm. In addition, educational videos were collected for the prediction stage.

Evaluating the accuracy of the test dataset over the trained system showed a total of 89.3% F1 Score. Also, the SGD technique gave a better performance in compare to other machine learning techniques such as Logistic regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). Figure 3 shows the block diagram of the proposed system steps.

Encapsulation Video Classification and Retrieval Based on Arabic Text

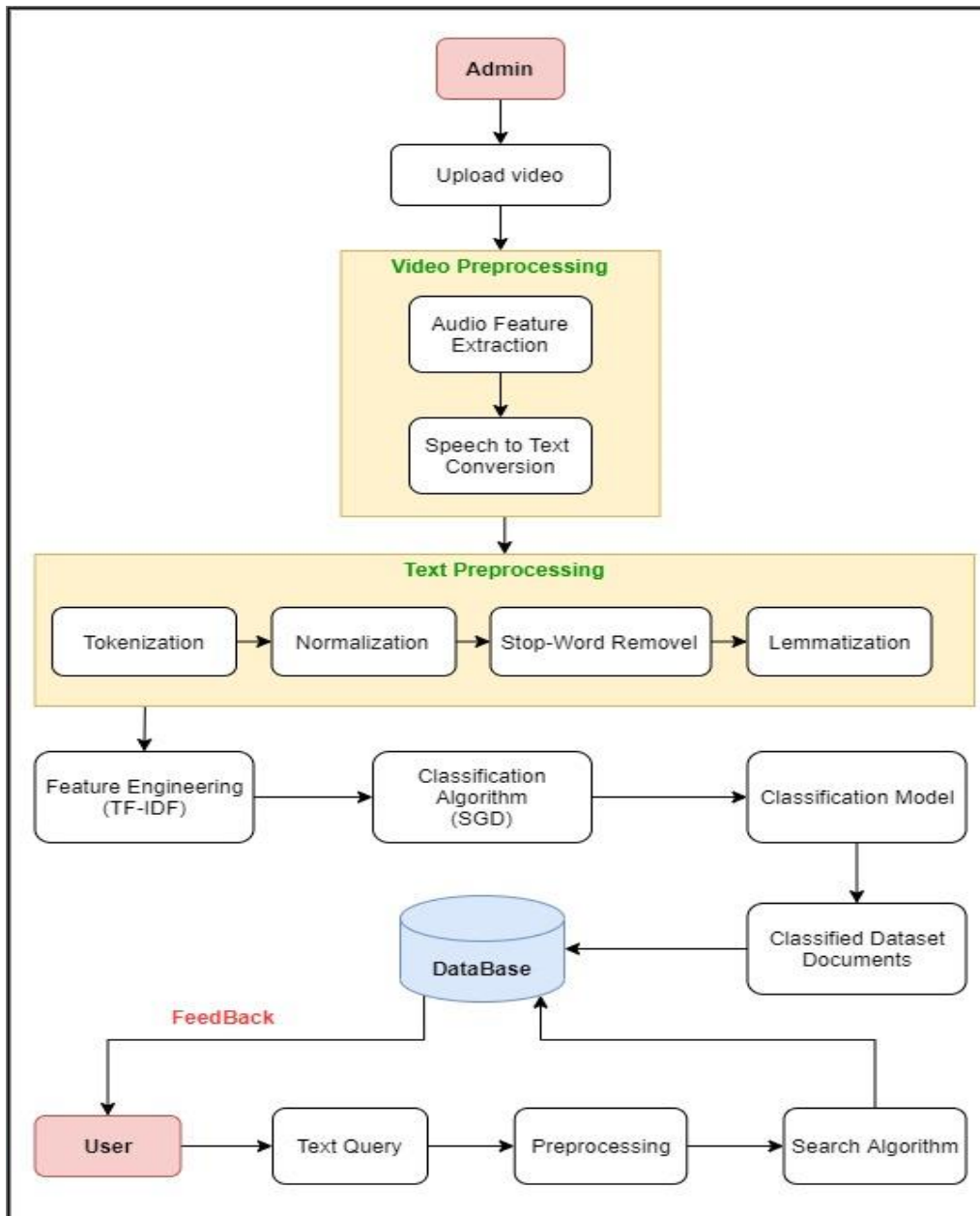**Reem A. K. Aljorani and Boshra F. Zopon Al_bayaty**

Figure 3: Block diagram of the proposed system

Encapsulation Video Classification and Retrieval Based on Arabic Text

Reem A. K. Aljorani and Boshra F. Zopon Al_bayaty

Results and Discussions

In this section, the results of the experiments were presented. A model was trained and evaluated using the same corpus which was created and preprocessed manually as shown in table 2.

Table 2: The size of the Arabic educational corpus used for training and testing the model

NO.	Categories	No. of Documents	No. of Words
1	Biology	10	483,841
2	Chemistry	11	538,837
3	Physics	8	469,105
4	Arabic Grammar & Literature	12	403,093
5	Religion	9	390,942
6	Math	9	390,904

The dataset is composed of six different categories each of which refers to a specific subject for sixth grade secondary students. A total of 70 videos operative in Arabic language was collected to be used in the study. The audio feature was extracted from the videos in a WAV format by using MoviePy python library, then transcripts were generated using azure cognitive service delivered by Microsoft that converts Arabic audio format into a text document.

The educational videos were preprocessed and then predicted by the system to be available for the students. The dataset was divided into 70% for training and 30% for testing. In this experiment, the Scores of precision, recall, and F1 metrics were calculated and compared to different machine learning models such as Support Vector Machine (SVM), Logistic Regression LR, and k-nearest neighbor algorithm KNN. SGD model scores the highest F1 Score with 89.3% as shown in table 3.

Table 3: Comparison between four machine learning algorithms

Model	CA	F1	Precision	Recall
KNN	0.621	0.618	0.726	0.621
SVM	0.257	0.154	0.110	0.257
SGD	0.893	0.893	0.901	0.893
LR	0.803	0.806	0.848	0.803

Classification accuracy gave the highest score as well for the stochastic gradient descent algorithm with 0.893 which makes SGD algorithm the most suitable for the presented work.

Analysis of the Proposed System

The video classification task is a time consuming job [3]. Finding related text data about a certain subject in a massive open online source is a challenging problem [16]. There are many classification algorithms that were used to classify Arabic text documents. Nevertheless, most of these classification algorithms had three main problems:

1. The first problem is feature representation where vectors of the feature space are represented with high dimensionality.
2. The accuracy of the classification algorithm is approximately low measured by the rate of precision.
3. The classification algorithms complexity increases with text classification as the data dimension increases.

Several challenges were found in this work while designing and implementing the proposed system, these challenges are:

1. Creating a training dataset: because of the lack of Arabic educational classification researches, there wasn't any available training dataset that matches the system requirements. Therefore; a training dataset was created manually.
2. Converting speech to text: the process of converting Arabic audio into Arabic text is a time consuming task that depends on the internet speed which makes it a challenge to convert a big data into text.
3. Cost expensive: also speech to text conversion is an expensive task that requires a monthly paying in order to be activated and work properly.

4. Implementing four machine learning approaches on the dataset to obtain better results.

Conclusions

In this paper, a survey on several researches concerning videos classification and retrieval based on different machine learning techniques was proposed with a focus on researches that classify videos and text in the Arabic language. Other papers focus on Education and electronic learning and how to classify and deal with this kind of information in a video to make it easily available to all students. A contribution was introduced to classify Arabic educational videos based on SGD technique then retrieve these videos based on the students query. Limitations and challenges were discussed which led to a conclusion that the distinctive feature of educational videos operative in the Arabic language is the Knowledge spoken by the teacher. Speech gives rich information about educational videos that's why the audio feature was extracted from the video then converted to Arabic text. SGD technique was applied on the extracted text in which accurate results were obtained. A future work for research is to enhance the system platform to be more interactive between the students and the teachers by making the system available online. Also, make the system available for all the students from all grades by classifying videos to classes and subjects.

Acknowledgment

References

1. V. Vijayakumar, R. Nedunchezhian, A study on video data mining, International journal of multimedia information retrieval, 1(3), 153–172, (2012).

2. S. Mohamed, A. Ghalwash, A. Youssif, Journal of Computer and Communications, 6(8), 28–44(2018).

3. Z. A. A. Ibrahim, S. Haidar, I. Sbeity, Journal of Electrical Engineering and Computer Science, 3(2), (2019).

4. F. Zhang, D. Liu, C. Liu,Sustainability, 12(7), 3066(2020).

5. Z. Kastrati, A. S. Imran, A. Kurti. Transfer Learning to Timed Text Based Video Classification Using CNN. In: Proceedings of the 9th International Conference on Web Intelligence, Mining and Semantics (WIMS2019). Association for Computing Machinery, New York, NY, USA, Article 23, 1–9, (2019).

6. L. H. Medida, K. Ramani, Int. J. Eng. Adv. Technol., IJEAT, 8(6), 4820–4827, (2019).

7. H. Chatbri, K. McGuinnes, Automatic MOOC Video Classification using Transcript Features and Convolutional Neural Networks. In: Proceedings of the ACM Workshop on Multimedia-based Educational and Knowledge Technologies for Personalized and Social Online Training (MultiEdTech '17). Association for Computing Machinery, New York, NY, USA, 21–26. (2017). fdxs

8. S. Boukil, M. Biniz, F. El Adnani, L. Cherrat, A.E. El Moutaouakkil, International Journal of Grid and Distributed Computing, 11(9),103-114, (2018).

9. K. Sundus, F. Al-Haj, B. Hammo, A Deep Learning Approach for Arabic Text Classification, In: 2nd International Conference on new Trends in Computing Sciences (ICTCS), Amman, Jordan, 1-7, (2019).

10. D. Kosmajac, V. Keselj, In FIRE (Working Notes), CEUR Workshop Proc., 2036, 118–122(2017).

11. S. Diab, International Journal of Computer Science and Information Security, 16(12), 1–6, (2019).

12. A. Dandashi, J. Jihad Al Ja'am, S. Foufou, International Journal of Computing & Information Sciences, 12(2), 205–212(2016).

13. M. A. H. Madhfar, M. A. H. Al-Hagery, Arabic text classification: A comparative approach using a big dataset, In: 2019 International Conference on Computer and Information Sciences (ICCIS) 1–5, (2019).

14. A. Elnagar, R. Al-Debsi, O. Einea, Information Processing & Management, 57(1), 102121, (2020).

15. M. Hamroun, Arabic Text-Based Video Indexing and Retrieval System Enhanced by Semantic Content and Relevance Feedback, In: 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), Abu Dhabi, United Arab Emirates, (2019).