CASE STUDY REPORT
CSCI-B555 MACHINE LEARNING

Rahul Raghatate
rraghata@iu.edu

Data Science - Spring 2017
School of Informatics and Computing,
Bloomington, IN 47408, U.S.A.

ABSTRACT

The Machine Learning Algorithms have been gone timely customization and modification as per the requirements or situations and have been extensively used to solve complex problems in myriad of domains. Convolutional Neural Network, Hidden Markov Models, Support Vector Machines, K-means are some of these algorithms which have been extensively used in domains of Market Analysis, Big Data, Drug Discovery, and Business Analysis. This report conducts study of these four algorithms already applied for stock market forecast, library management, data security, and big data analytics for multimedia transmission and storage.The subsections summarize the problem and how it was approached along with convincing experiments and results for each of the above mentioned applications.

***Index Terms*—** HMM, K-Means, SVM, CNN, Big Data analytics

1. APPLICATION OF HIDDEN MARKOV MODEL

1.1. Stock Market Forecasting Using Hidden Markov Model: A New Approach [1]

1.1.1. INTRODUCTION

Hidden Markov Models (HMM), finite state machine, have been widely applied in all sectors of industry. Because of its suitability for modelling dynamic systems, HMMs are extensively used for pattern recognition and classification problems. But, using HMM for predicting future events has been proven challenging. One of the challenges in finance industry is forecasting stock price for interrelated markets. In this paper, authors made use of the well-established Hidden Markov Model (HMM) technique to forecast stock price for some of the airlines.

Stock trading involves lot of buying and selling of stocks at lower or higher prices on daily basis. Prediction of perfect timings for this business is one of the challenge in finance industry. Any fluctuation in this market influences our personal and corporate financial lives, and the economic health of a country. Therefore, there is always some risk to investment in the Stock market due to its unpredictable behavior. A high accuracy stock price prediction is of high interests.

1.1.2. BASIC APPROACH

As reported in paper, most of the forecasting research has employed the statistical time series analysis techniques like auto-regression moving average (ARMA) [2] as well as the multiple regression models. However, most of them have their own constraints [1].

HMM that is trained on the past dataset of the chosen airlines is used to search for the variable of interest, behavioral data pattern from the past dataset. Forecasts are prepared basically on interpolating the neighboring values of these datasets. Thus, this applied HMM in stock predictions, offers a new paradigm for stock market forecasting in all industrial sectors.

Therefore,

1. First locate pattern(s) from the past datasets that match with todays stock price behavior.
2. Then, interpolate these two datasets with appropriate neighboring price elements.
3. Forecast tomorrows stock price of the variable of interest [1].

1.1.3. METHODOLOGY

While implementing the HMM, the choice of the model, choice of the number of states and observation symbol (continuous or discrete or multi-mixture) become a tedious task.

1. Model Used: left-right HMM with 4 states.
2. Dataset :For training the model, past one and a half years (approximately) daily data were used and recent

last three months data were used to test the efficiency of the model.

3. Input: 4 features for a stock that is the opening price, closing price, highest price, and the lowest price.

4. Output: The next days closing price

5. Observations here being continuous rather than discrete, authors choose 3 mixtures for each state for the model density. Prior probability pi_i = a random number was chosen and normalized so that $\sum_{i=1}^{N} pi_i = 1$.

6. The observation probability density function was chosen as a three-dimensional Gaussian distribution as the dataset being continuous, the probability of emitting symbols from a state cannot be calculated.
$b_j(O) = \sum c_{jm}[O, \mu_{jm}, U_{jm}] \;\; 1 \leq j \leq N$
where
$O =$ vector of observations being modelled.
$c_j m =$ mixture coeff. for the m-th mixture in state j ,
where $\sum_{m=1}^{M} c_{jm} = 1$
$\mu_{jm} =$ mean vector for the m-th mixture component in state j
$U_{jm} =$ Covariance matrix for the m-th mixture component in state j
N = Gaussian density.

7. Use the training dataset for estimating the parameter set (A, B, π) of the HMM.
$A = \{a_{ij}\}$ transition matrix,
$B = \{b_j(O_t)\}$ observation emission matrix.
Using the trained HMM, likelihood value for current days dataset is calculated. For example, say the likelihood value for the day is x, then from the past dataset using the HMM, locate those instances that would produce the same x or nearest to the x likelihood value i.e locate the past day(s) where the stock behavior is similar to that of the current day.

8. Now, assuming that the next days stock price should follow about the same past data pattern, from the located past day(s), calculate the difference of that days closing price and next to that days closing price. Thus, the next days stock closing price forecast is established by adding the above difference to the current days closing price [1].

1.1.4. RESULTS

Based on training of 4 HMMS and 4 different ANN(Artificial Neural Network) for four different airlines, prediction for next few day's closing price of stocks were more accurate for HMM's as compared to ANN as shown in 3.

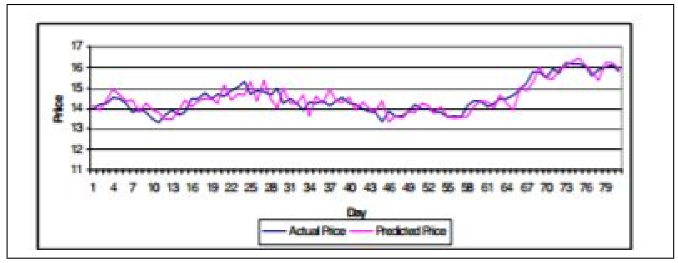

Fig. 1. The current days stock price behavior matched with past days price data [1].

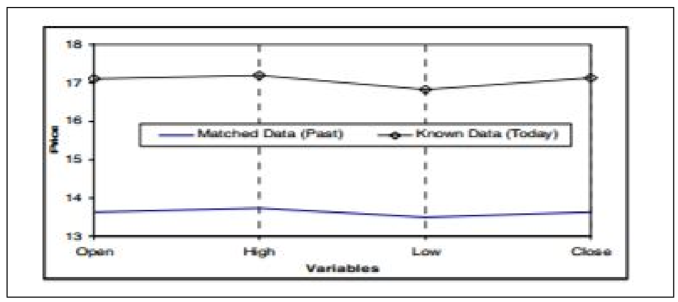

Fig. 2. Actual Vs. Predicted stock price [1].

	ANN (MAPE)	Proposed method (MAPE)
British Airlines	2.283	2.629
Delta Airlines	9.147	6.850
Southwest Airlines	1.673	2.011
Ryanair Holdings Ltd.	1.492	1.928

Fig. 3. Prediction accuracy of ANN and the proposed method [1].

1.1.5. CONCLUSION

Though ANN is well researched and established method that has been successfully used to predict time series behaviour from past datasets. As seen in 3, the mean absolute percentage errors (MAPE) values of the two methods are quite similar. Thus the paper gives a new approach, to predict unknown value in a time series(stock market).

2. APPLICATION OF K-MEANS

2.1. K-means Clustering Algorithm Application in University Libraries [3]

2.1.1. INTRODUCTION

This paper is based on application of K-means Clustering Algorithm to find most frequently used books and literature so as to satisfy demand of the student readers. Most libraries are lacking in data integration and analysis and are more short of generalization and revelations of implied correlation in the statistical data. This results in the capture of readers information at a simpler level and disallowing automatic acquisition of the aid decision making and knowledge.

Data mining of library information has been carried out in finding the implied correlation to guide the rational distribution of library resources and increase the resource utilization and thus improving the service level of the university library.

2.1.2. DATASET DESCRIPTION

Data from ILAS in 2011 at a university library in Chongqing City, including three tables:

- Circulation borrowing: number of readers record, book barcode, borrowing period and borrowing duration.
- Booklist collection: booklist number, book title and CLC number
- Readers record: Reader ID, name, gender, specialty and education, starting time and ending time of library card and number of readers record.

2.1.3. DATA PREPROCESSING

Normalization of data was performed to facilitate the data operation. Normalized book cluster was created for following set of attributes:

- ClassCode (CLC number)
- borrNum (borrowing times)
- renewNum(renew times) and
- averageTime (average borrowing duration).

The ClassCode shall be considered as the cluster object and the borrNum, renewNum and average Time shall be taken as the attributes of the cluster object. Similarly , Tuser(readers borrowing), Treader(readres record), Tbook(booklist database) , Trebor(renew times), averageTime (average borrowing duration) were created.

2.1.4. METHODOLOGY

1. Randomly select k objects from n data objects as the initial clustering center;
2. Distribute the rest objects respectively into their most similar clusters (represented by the corresponding clustering centers) according to their similarity (distance) with these clustering centers;
3. Calculate the clustering centers (mean value of all objects in the cluster) of all gained new clusters;
4. Take the mean square error as the standard measure function and find the error.
5. Repeat above process until the convergence of standard measure function appears [3].

2.1.5. CHARACTERISTICS OF K-CLUSTERS

- All clusters themselves are as compact as possible.
- The separation of the clusters are as extreme as possible.

2.1.6. *Types of clustering*

1. Clustering of time behavior
   Selection of input week, hour and book code as input attributes and then select group dividing number 9, provides categorical clustering of books. This clustering technique can help in finding which category books are borrowed most and at what times and which days of week.
2. Clustering of readers departments
   This clustering algorithm allows the statistical analysis of different department readers interest in borrowing books, thus the characteristics and trends may vary with different departments. The results of this clustering have much to do with the number of students in different departments and specialty nature.
3. Clustering of readers on grade basis
   Besides the clustering of readers departments, clustering to readers borrowing frequency of all grades was also performed, allowing to find which grade students are under highest borrowing frequency, providing insights about study intensity at different college years [3].

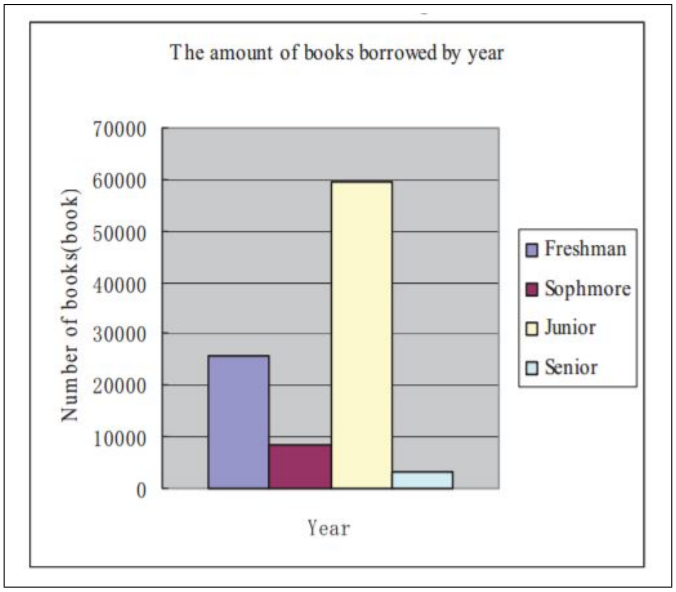

Fig. 4. The number of readers by Year [3].

4. Comprehensive clustering of characteristics
   K-means algorithm application to input attributes of department, gender and grade and select the group-dividing number 3 can give the results about the more frequently borrowed books are from which reading materials?

2.1.7. RESULTS

1. The results of time clusters can be considered as a reference to librarys personalized recommending service. At the time points when the most frequently borrowing appears, library services can adjust the personnel allocation to serve better for readers[3].

2. The analysis of reader clusters studies category of most frequently used books which was literature. One of the reason might be because Literature books always exert a subtle influence on personal qualities of the college students, which thus brings a broader readership. Exam tests books are also heavily used ones, as most students have to make massive exercises for the test preparation and are frequently borrowed[3].

2.1.8. CONCLUSION

Recommendations based on cluster analysis of the readers behavior tendency, predict the readers loss can be made to the library to adjust service strategy, rationalize the staffing, and provide personalized service. The conducted k-means clustering can provide guidance to library to truly realizing the service concept of Readers First and providing them with the most needed resources[3].

3. APPLICATION OF SUPPORT VECTOR MACHINE(SVM)

3.1. An SVM-based Query Monitoring Method for Inference Control [4]

3.1.1. INTRODUCTION

Many widely-used data sharing applications rely on allowing untrusted clients to query against secret data. Inference control for secure data is a key issue [4]. Access control protocols are commonly used to protect secure data. However, series of seemingly trusted queries can be used by malicious user to infer much information of the secure data.

For example, if a pharmacist needs to perform a drug interaction check over a patient, he needs to know only certain part of the patients history (certain allergies). Unlike the patients doctor, the pharmacist should not know the detail sensitive health information. Similar situation are prevalent in personal recommendation systems such as on-line shopping websites [4].

Secure computation is considered as a powerful tool to handle applications with privacy concerns. Secure computation guarantees that each party only knows the result and its own inputs after the computation. The owner of secret data opens the secret data for querying to provide general information while trying to ensure not too much information is leaked from these queries. The information leakage is defined by the number of remaining possible private values that are consistent with all previous query results [4].

The paper propose a method based on Support Vector Machines (SVM) to quickly determine whether a set of queries is leaking too much information. Through experiments over several sample functions, the performance and flexibility of this method has been demonstrated.

3.1.2. METHODOLOGY

1. Classes for incoming queries: leaking too much information. or its okay.

2. The classifier will determine which class the incoming queries belong to.

3. Here, following properties of SVM are utilized to the advantage.

   - Training process which is complicated can be separated from the classifying process which is much simpler.

   - Since the classifier can be trained before actual queries come, classifier can be trained extensively with well-prepared samples and these training samples can be pre-computed.

3.1.3. MODEL EVALUATION

1. The authors tested the performance of SVM-based classification with preliminary experiments when the secret was a string and the query function was Hamming or Levenshtein distances.

2. LibSVM was used as SVM tool.

3. The performance of the SVM was determined by its kernel function, which maps the original data to a higher dimension. One kernel function they used was the classic RBF kernel, defined as:

$$K(x, y) = Ce^{-(\gamma(||x-y||^2))}$$

   where x, y are two data points and C, γ are parameters.

4. When the secret is a string and the query function is Hamming distance represented as H(x, y), a data point is a certain set of queries. In the case where the number of queries is 1, we define the distance function between two sets of queries x and y as:

$$||xy|| = min(H(x, y), N - H(x, y))$$

   where N is the length of the query sequence.

3.1.4. CHARACTERISTICS

1. This SVM classification can be done by a vector multiplication and one comparison.

2. This classifier can be applied to many secret data formats and query functions efficiently.

3. The validity of this classifier lies for similar queries over identical secret leaks with very similar amount of information.

4. Querying with a certain binary sequence X leaks the same amount of information as the querying with the opposite of X. So, two query sequences are considered similar if they are almost identical or almost opposite in the distance function above.

5. When there are multiple queries, define the distance function as a match of queries that has the minimal sum [4].

3.1.5. RESULTS

1. The classification performance is depicted with ROC curves. In an ROC curve, the x-axis and y-axis represent the False Positive Rate (FPR) and True Positive Rate (TPR), respectively.

2. Two experiments with following training samples:
   100 positive and 100 negative samples from 1000 random query sets
   100 positive and 100 negative samples from 500 random query sets [4]

3. Introduced a new parameter r.
   The SVM was trained by extreme samples, i.e. queries that will leak almost all or no information. The parameter r represents the threshold to determine whether the incoming query set is too close to the average information leakage [4].

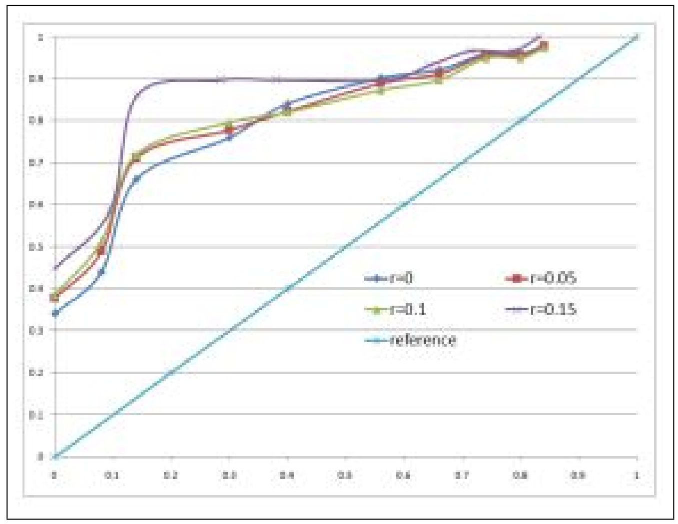

Fig. 5. ROC curve for 10 bit genome sequence/2 queries over Levenshtein distance[4].

4. If the information leakage is moderate or an incoming query set locates within the interval defined by the percentage r around the median, the decision made by the SVM will not be considered and this affects the classification performance to some extent [4]. So, it can detect very dangerous queries and very safe queries.

5. For the dataset, generally the SVM with 200-400 samples can detect more than 90% dangerous queries with less than 30% false positive rate (¡20 % in most cases).

3.1.6. CONCLUSION

Inference control is necessary for multiple query process over single secret data. Since the secret owner can easily prepare more samples to ensure their data safety in a practical application, performance improvement is feasible. Thus, Support Vector Machines (SVM) can be applied to quickly determine whether a set of queries is leaking too much information.

4. APPLICATION OF CONVOLUTION NEURAL NETWORK(CNN)

4.1. Real-time Big Data Analytics for Multimedia Transmission and Storage [5]

4.1.1. INTRODUCTION

The unprecedented exponential growth of data in recent years is no doubt result of increase in usage of digital devices and generation of tremendous data over multiple Internet Services. Therefore, data processing and network overload have become area of concern for businesses and research area of Data Management. Multimedia transmission and storage are a problem of network overload [6], and solving network overload problem usually has two strategies, one is dispersing blocked traffic in advance by optimizing route selection, and the other is recognizing abnormal traffic and abandoning it before transmission. This digital data majorly consists of images and videos from general population which has vast usage of digital devices. There has been intensive research for these problems, extensive study has been published majorly on image analysis using deep learning, but this approach for video analysis is still lagging in solutions [7].

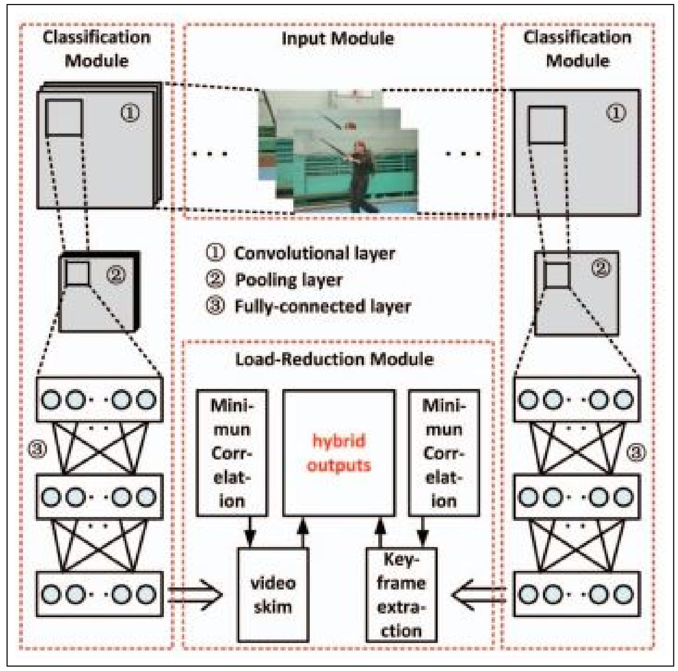

Fig. 6. Hybrid-stream big data analytics model [5].

4.1.2. METHODOLOGY

This paper proposed a hybrid-stream big data analytics model[5] which contains a reliable key frame extraction mechanism and an improved CNN classification algorithm to enhance the classification precision and relieve the load in transmission and storage. The authors proposed an innovative multi-dimensional Convolution Neural Network (CNN) to algorithm to perform big data video analysis by obtaining the importance of each video frame. With the inclusion of reliable decision-making algorithm and key frame extraction mechanism, unimportant frames can be dropped automatically by a series of correlation operations.

The functionality of hybrid-stream model can be summarized as three parts: Input Module, Classification Module and Load-Reduction Module consisting following procedures:

- data preprocessing
- data classification
- data recognition and
- data load reduction.

4.1.3. MODEL EVALUATION

1. Input Model

   To improve the classification precision, raw video resources are divide into two input forms, video frames and video clips. Video clips generally consist of several video frames. The part of input module transforms the raw video resources into another form which is suitable for data classification.

2. Classification Module

   Perceptron is usually used in image classification. The transformed raw pixel data for raw images is fed to transmission network and it correctly classifies these pixel data by dynamically learning its weights and biases. The dynamic learning process is a tough work, as for small changes over the transmission network (such as weighs or biases of any single perceptron), the output of perceptron may flip. The change in the perceptron affects rest part of transmission network. Analysis of the changes in network is difficult. Time dimensionality to connect several neighboring frames is adapted in the approach, thus output with time dimensionality can be written as:

$$f_j^{i,j,k} = \sum_k \sum_{j=1}^{J} \sum_{i=1}^{I} f_i * w_{i,j,k} \quad (1)$$

The above equation is used as 3-dimension convolution kernel instead of 2 (used for images) to compute feature maps pixel by pixel. Consider an greyscale image P Q, leading to P Q input neurons, with the intensities scaled appropriately between 0 and 1. The defined the cost function is:

$$J(W, b) = \frac{1}{2n} ||h_{w,b}(x) - \alpha||^2 \quad (2)$$

The inputs are 3-dimensionality matrixes, while actual neural network conditions determine the output dimension.

Algorithm 1: Classification Module

Initialize w, b randomly
While cost function $<$ threshold **do**
    Compute feature maps pixel by pixel according to the 3-dimension convolution kernel (1), and feed forward
    Calculate cost function (2)
    If cost function $>$ threshold **then**
        Back propagation using method Stochastic Gradient Descent (SGD)
    End if
End while

The cost function J is guaranteed to decrease and proved in paper[5]. Changing J as the form of equation (2), and authors algorithm is associated with vectors, so they denote the gradient vector by ∇J. The change of cost function is analyzed as follows:

$$\Delta J \approx \frac{\partial J}{\partial w}\Delta w + \frac{\partial J}{\partial b}\Delta b \tag{3}$$

$$\nabla J = (\frac{\partial J}{\partial w}, \frac{\partial J}{\partial b})^T \tag{4}$$

$$\Delta J = \nabla J \cdot (\Delta w, \Delta b) \tag{5}$$

$$(\Delta w, \Delta b) = -\eta \nabla J = -\eta(\frac{\partial J}{\partial w}, \frac{\partial J}{\partial b})^T \tag{6}$$

$$\Delta J = -\eta ||\nabla J||^2 = -\eta||(\frac{\partial J}{\partial w}, \frac{\partial J}{\partial b})^T||^2 \leq 0 \tag{7}$$

According to equations (5) and (6), a suitable parameter (leaning rate) can be estimated and can be used to simplify the expression (4), and thus the J will always decrease.

Also the authors has proved usage of factor $\frac{1}{n}$ or $\frac{1}{m}$ in cost function so as to accommodate real time training of data [5].

$$w\prime_k = w_k - \varrho \frac{\partial J}{\partial w_k} \tag{8}$$

$$b\prime_k = b_k - \varrho \frac{\partial J}{\partial b_k} \tag{9}$$

$$\frac{\sum_{j-1}^{m} \nabla J_{xi}}{m} \approx \frac{\sum_x \nabla J_x}{n} = \nabla J \tag{10}$$

$$w\prime_k = w_k - \frac{\varrho}{m} \frac{\partial J}{\partial w_k} \tag{11}$$

$$b\prime_k = b_k - \frac{\varrho}{m} \frac{\partial J}{\partial b_k} \tag{12}$$

In Equation (2), the overall cost function is scaled by a factor $\frac{1}{n}$. According to authors, people often instead of averaging, prefer to sum over the costs of individual training data. However, it is particularly helpful when we cannot know the number of training data in advance. The situation may happen if large number of training data generate in real time. Also, in a similar way, the mini-batch update rule (11) and (12) sometimes omit the $\frac{1}{m}$ term out the front of the sums. Conceptually this makes little difference, since it is equivalent to rescaling the learning rate .

3. Load Reduction Module

   (a) Keyframes and Video Skim
   Classification module labels every video frame and clips information value. These information values are fed as input to load-reduction module. Key frames also called as representative frames (R-frames) are a set of a collection of salient images extracted from the original video. The video skim S is defined as follows:

$$S = F_{k-frames}(Video)$$
$$S = f_{r_1} \bigodot f_{r_2} \bigodot \cdots \bigodot f_{r_\sigma} \tag{13}$$

   where $\bigodot$ is the excerpt assembly and integration operation.

   (b) Key Frame Ratio
   According to authors, this is an optimization problem of finding a suitable set $R = r_1, r_2 \ldots r_\sigma$, which can represent the video using the least frames or clips. Thus,

$$r_1, r_2, \cdots, r_\sigma = \text{argmin}_{r_i}\{D(R, F)|1 \leq r_i \leq n\} \tag{14}$$
$$\sigma = \theta n \tag{15}$$

   where n is the number of frames or clips in the original video sequence, σ is the total number of key frames or clips, D is a dissimilarity measure and F is the output in classification module.

   (c) Minimum Correlation method
   As the inputs, frames or clips, are always sequential elements, the minimal correlation method was selected. Accordingly, it select frames or clips that are dissimilar to each other and can represent the video with the least elements. Therefore,

$$\{r_1, r_2, \cdots, r_\sigma\} = argmin_{r_i}\{\sum_{i=1}^{\sigma-1} corr(f_{r_i}, f_{r_{i+1}})\} \tag{16}$$

   where $corr(f_{r_i}, f_{r_{i+1}})$ is the correlation coefficient of two sequential frames or clips $(f_{r_i}, f_{r_{i+1}})$.

Algorithm 2: Load-reduction Module — Step 1

```
Inputs f_{r_i}, A
Procedure:
  Begin
    While (i < n)
      If (Corr(f_{ri}) < A)
        enter into scene+1
      Else
        still in scene
      End if
    End while
```

Algorithm 3: Load-reduction Module — Step 2

```
Input f_{r_i}
Procedure:
  Begin
    η = α ∑ f/k
    While (scene has not change)
      If (f_{r_i} > η)
        drop the frame
      Else
        store the frame in set S
      End if
    End while
```

The reliable key frame extraction mechanisms implementation process is presented in Algorithm 2 and 3 [5]. Algorithm 2 mainly judges whether the scene changes or not and Algorithm 3 emphasizes the load-reduction. The algorithms first recognize the scene change, because different scenes have different thresholds. Correctly recognizing scene changes can improve the final performance. If continuous frames are all in the same scene, classify these frames into a group and distribute each group a threshold η.

$$\eta = \frac{\sum Importance_{frame}}{number}$$

where $Importance_{frame}$ is processed frames importance which ranges from 0 to 1, and number refer to several correlative frames number.

4.1.4. RESULTS

Performance comparisons with Basic CNN, SVM, Temporal stream ConvNet and Two-stream model are demonstrated in paper. Datasets Used: UCF-101, UCF-101 Expand, HMDB-51 and CCV [8] Input Format: 214×214

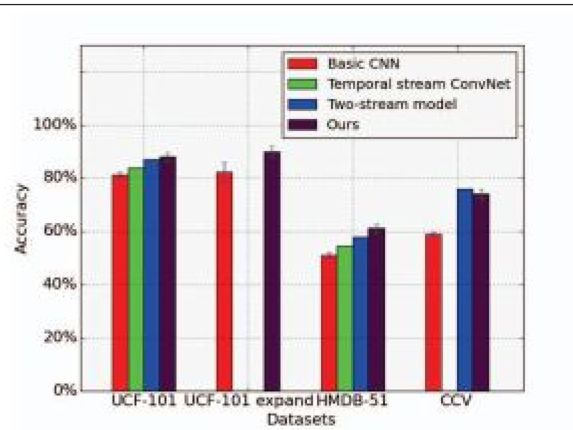

Fig. 7. Comparisons among different models under different data set [5].

The performance of load-reduction module is shown in 9. Authors' used true positive rate (TPR) and false positive rate (FPR) of ROC to compare proposed method with basic CNN model and SVM [5]. As can be seen, for the same data set, proposed method has a better approach to the coordinate (0,1) which is usually called as a perfect classifier. Meanwhile, the approaching speed of our method obviously faster than that of basic CNN and SVM.

4.1.5. CONCLUSION

Simulation results illustrate that the size of the processed video can be significantly reduced. The simulation also shows that proposed model performs steadily and is robust enough which different from conventional deep learning methods to address image analysis problem, performs video analysis. Large-Scale real-time data can be accommodated in the proposed algorithm. Thus, it can lead to a fairly good video stream transmission and storage.

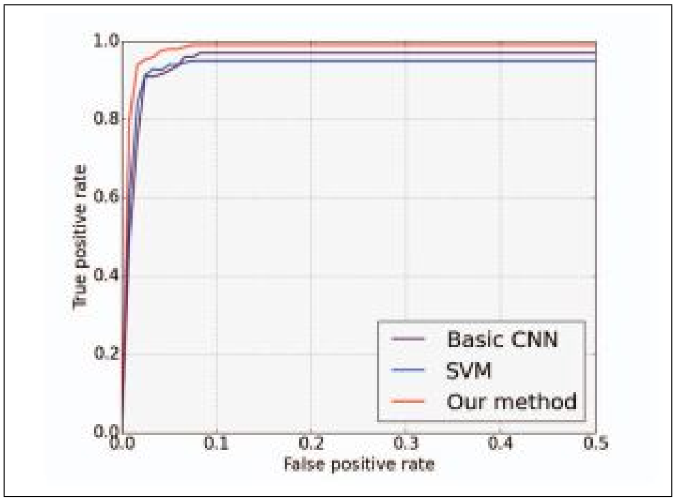

Fig. 8. Comparisons among basic CNN, SVM and our method under UCF-101 [5].

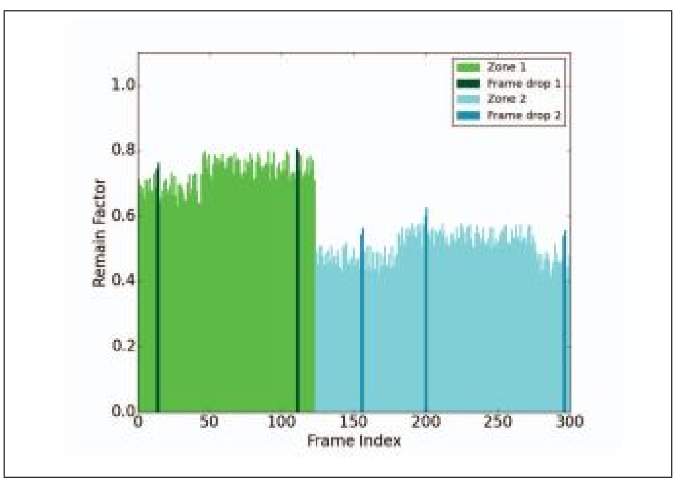

Fig. 9. Two scenes in a 10-second video and its drop frames [5].

5. REFERENCES

[1] Md. Rafiul Hassan and Baikunth Nath, "Stockmarket forecasting using hidden markov model: A new approach," in *Proceedings of the 5th International Conference on Intelligent Systems Design and Applications*, Washington, DC, USA, 2005, ISDA '05, pp. 192–196, IEEE Computer Society.

[2] T. Kimoto, K. Asakawa, M. Yoda, and M. Takeoka, "Stock market prediction system with modular neural networks," in *1990 IJCNN International Joint Conference on Neural Networks*, June 1990, pp. 1–6 vol.1.

[3] R. Wang, Y. Tang, G. Liu, and Y. Li, "K-means clustering algorithm application in university libraries," in *Cognitive Informatics Cognitive Computing (ICCI*CC), 2011 10th IEEE International Conference on*, Aug 2011, pp. 419–422.

[4] Yikan Chen and David Evans, "Poster: An svm-based query monitoring method for inference control," *IEEE Symposium on Security & Privacy*, 2012.

[5] K. Wang, J. Mi, C. Xu, L. Shu, and D. J. Deng, "Real-time big data analytics for multimedia transmission and storage," in *2016 IEEE/CIC International Conference on Communications in China (ICCC)*, July 2016, pp. 1–6.

[6] X. Zhang, R. Yu, Y. Zhang, Y. Gao, M. Im, L. G. Cuthbert, and W. Wang, "Energy-efficient multimedia transmissions through base station cooperation over heterogeneous cellular networks exploiting user behavior," *IEEE Wireless Communications*, vol. 21, no. 4, pp. 54–61, August 2014.

[7] Chao Yan, B. Zhang, and F. Coenen, "Driving posture recognition by convolutional neural networks," in *2015 11th International Conference on Natural Computation (ICNC)*, Aug 2015, pp. 680–685.

[8] Yu-Gang Jiang, Guangnan Ye, Shih-Fu Chang, Daniel Ellis, and Alexander C. Loui, "Consumer video understanding: A benchmark database and an evaluation of human and machine performance," in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval*, New York, NY, USA, 2011, ICMR '11, pp. 29:1–29:8, ACM.