Future Generation Computer Systems 112 (2020) 507–511

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

English text quality analysis based on recurrent neural network and semantic segmentation

Xiaoyu Luo [a], Zhibin Chen [b,*]

[a] *School of Foreign Language, Hunan University of Technology and Business, Changsha 410205, China*
[b] *Admissions and Career Service Office, Hunan Institute of Engineering, Xiangtan 411104, China*

ARTICLE INFO

Article history:
Received 21 January 2020
Received in revised form 11 May 2020
Accepted 17 May 2020
Available online 22 May 2020

Keywords:
Deep learning
Semantic segmentation
Neural network
English text
Quality analysis

ABSTRACT

In recent years, deep learning algorithm based on cyclic neural network and semantic segmentation has performed well in the field of image segmentation. The purpose of this paper is to realize the quality analysis of English text through recurrent neural network and semantic segmentation. This paper proposes an attention based English text quality analysis model based on recurrent neural network. Through the introduction of attention mechanism, the influence of semantics in the text is considered in the analysis of English text quality. The target relies on the quality of English text to determine the text quality of the sentence for a given target object. At present, most English text quality analysis methods are aimed at the traditional semantic analysis tasks. Based on rnn-attention model, a rnn-attention-t model is proposed, which introduces the information of the target object while modeling the text. In addition, considering that the influence of the top and bottom of the target object on the semantic trend is usually different, this paper proposes an rnn-attention-c model, which models the top and bottom of the target object respectively. The experimental data have shown that the quality analysis of English text based on recurrent neural network and semantic segmentation is faster than the traditional method. The experimental results have demonstrated that our method can effectively and quickly confirm the quality of English text, which is about 7% faster than the conventional method.

1. Introduction

The problem of text classification is a classic problem in the field of natural language processing, with the goal of marking the category to which the text belongs. Text classification has a wide range of uses, such as topic labels, emotional classification, and so on. However, good text expression plays a decisive role in performing natural language processing tasks, such as text classification. Traditional text representations are used in word bag models (BOW) or vector spatial models. The word bag model not only loses contextual text information, but also faces a high degree of latitude and sparseness. Vector spatial models are based on the word pack model and are extracted from feature elements (how often documents are used, common information). By calculating the weight of the element (TF–IDF), you can reduce the size and increase the density. In recent years, the distribution based on neural network model sitters more and more popular. The distribution representation based on the neural network model is called the word vector, the word embedded or distributed. The method of representing the word vector in a neural network uses a neural network method to model the context and the relationship between the context and the target word. This method can map text to low-dimensional vector space, which solves the problem of high-latitude and high-latitude thesaurus model. Sparseness avoids the problem of dimensionality while maintaining the correlation between words. Using word vectors as input, and then classifying text using a ring or convolution network, can significantly improve performance.

As Internet is widely used, many information has flooded people's lives. How to find the required information in mass information has become a new topic for people to study [1–3]. In the process of searching information in English, if the target word and the target word with a specific relationship with the subject word can be used as the basis of the search, the problem of information query becomes a problem. the efficiency of identifying semantic relationships will be greatly improved. search [4–6]. For example, someone should look for "search reasons that cause headaches". If there is a tool to automatically search causality, it can reduce search time and improve search performance. in addition to serving as the basis for information search functions, the method of classifying semantic relationships has the potential to create thesaurus, build corpora in various fields, answer questions on social networks, translate text and eliminate information [7,8]. semantic relationship is an important basis for semantic analysis, and semantic relationship technology is a key step in natural

* Corresponding author.
E-mail address: chenzhibin1977@outlook.com (Z. Chen).

https://doi.org/10.1016/j.future.2020.05.027
0167-739X/

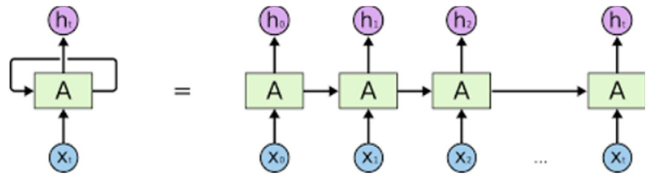

Fig. 1. Basic flow chart of neural network.

language processing (NLP). Therefore, it is necessary to study the classification methods of semantic relations and contribute to the development of related fields.

The innovation of this experiment: this paper proposes a RNN-Attention-T model based on the RNN-Attention model, which introduces the information of the target object when modeling the text. Furthermore, considering that the degree of influence on semantic trends is usually different up and down the target object, this paper proposes a RNN-Attention-C model, which is modeled separately up and down the target object. Experimental data show that the quality analysis of English text based on recurrent neural network and semantic segmentation is faster than the traditional method [9]. Experimental data show that the method can effectively and quickly confirm the quality of English text, about 7% faster than the conventional method.

2. Proposed method

2.1. Recurrent neural network

(1) The concept of a circular neural network

A recurrent neural network (RNN) is a recurrent neural network that recursively receives sequence data as input in the direction of sequence evolution, and all nodes (cyclic units) are linked by chains. The study of recurrent neural networks began in the last century and became one of the deep learning algorithms at the beginning of this century, including bidirectional recurrent neural networks and long-term and short-term term memory. Long-term memory networks (LSTM) are common cyclic neural networks. Recurrent neural networks are structures that repeat over time. And it is widely used in many fields such as natural language processing (NLP) and speech images. The biggest difference between RNN network and other networks is that RNN can achieve some kind of "memory function", which is the best choice for time series analysis. Just as humans can better understand the world with memories of the past. RNN also implements a human brain-like mechanism that preserves some memory of processed information, unlike other types of neural networks that do not retain processed information. Recurrent neural networks have memory, parameter separation and graph element completeness, so they have some advantages in the nonlinear characteristics of training sequences [10]. Recurrent neural networks have applications in the field of natural language processing (NLP), such as speech recognition, language modeling, machine translation, etc. they are also used to predict different time series. a recurrent neural network constructed from a convective neural network (CNN) can be used to handle the form of a computer vision problem containing sequence input. The results are shown in Fig. 1.

(2) Jump connection

The disappearance of jump gradient is a function of time step, so the ability of cyclic neural network to study distance dependence can be improved by jumping. Jump connection is a long-distance connection with several time steps. By introducing jump connection, the state of long-term scale can be better transferred in neural network, and the phenomenon of gradient disappearance can be reduced. This paper studies how to delete the unit loop connection directly when using hop connection, which makes the loop neural network run in the layered structure for a long time.

(3) Leakage block and valve block

There are two disadvantages of the leaking module in the application: firstly, the artificial burden is not the best way to remember the state of the memory system; secondly, the leaking module will not forget the function and will be affected by the information. Overload. In the past, cyclic modules have been fully utilized. After that, it might help to forget it. Based on this, the gate control block is the summary of leakage door, and the types of gate control block include entrance door, exit door and forgetting door. Each gate is a packed neural network, which can be calculated in the algorithm part. Generally, gating block is an effective method to reduce the error dependency of long-distance learning. In a variety of tasks, algorithms using gated blocks (including long-term and short-term storage networks and gated cyclic block networks) have proven to exceed SRN.

(4) Internal calculation of the cyclic network

The main part of cyclic neural network is directed graph, and the elements connected in directed graph decomposition are called RNN elements. The convolution neural network for image classification and detection is concentrated at the image level, so there will be a fully connected layer behind the convolution network to reduce the size of the network and display the required classification information and location information; semantic segmentation is concentrated at the pixel level of the image. We want to input the image and keep the output the same size. Therefore, in FCNN, the fully connected layer behind the common convolution mesh is deleted. In general, the chain connection formed by the cyclic block can be an analog of the hidden layer with direct connection in the neural network, but in another discussion, the "level" of the cyclic neural channel can refer to the time step. Therefore, as a general introduction, the concept of "hidden layer" is avoided here for circular block or whole circular block. The training data provided is input in order.

$$X = \{x_1, x_2 \ldots\ldots x_n\} \tag{1}$$

The expanded length of the circulating neural network is. The sequence to be processed is usually a time series, in which case the evolution direction of the sequence is called "time step". For time steps, the cyclic unit of the Recurrent Neural Network has the following representation:

$$h^t = f(s^{(t-1)}, x^t, \theta) \tag{2}$$

In the formula, H is called the system state of the Recurrent Neural Network. From the perspective of the dynamic system, the system state describes the change of all points in a given space with time. S is an internal state and is related to the system state.

$$S = S(H, X, Y) \tag{3}$$

Because the solution of the current state of the system needs the internal state of the previous time step, the calculation of the loop unit includes recursion. From the perspective of tree structure, all cycle cells of the previous time step are the parent nodes of the cycle cells of the current time step. Where f is the excitation function or the compression neural network with direct connection, the first corresponds to the simple cycle network (SRN), the second corresponds to the gating algorithm and some depth algorithms. The common functions for selecting excitation function include logic function and hyperbolic tangent function. θ is the weight factor in the cyclic unit, which is independent of the time step. That is to say, for a group of training samples, the recurrent neural network uses the common weight coefficient to calculate the output of all time steps. In theory, it is possible for a ring-shaped neural network to contain only cyclic elements, but recurrent neural network usually has an output node, which is defined as a linear function.

$$o^t = vh^t + C \tag{4}$$

where v, c is the weight coefficient. According to the structure of the Recurrent Neural Network, the calculation result of one or more output nodes can obtain the output value after passing the corresponding output function:

$$\hat{y} = g(o) \tag{5}$$

(5) Connectivity

Loop unit–loop unit connection: Also known as "implicit-implicit connection or full connection", the state of the current time step of each cycle unit is determined by the input of the time step and the state of the previous time step:

$$h^t = f(Uh^{t-1} + WX^t + b) \tag{6}$$

u, w is the weight of the loop node. The former is called state–state weight, and the latter is called state-input weight. The loop unit-loop unit connection can be bidirectional, corresponding to a bidirectional Recurrent Neural Network. Context-based connection: Since the closed-loop structure is presented from the perspective of the graph network, this connection is also referred to as a closed-loop connection, in which the system state of the loop unit introduces the true value y^{t-1} of its previous time step. Using a context-based looping neural network as an input to the actual value of the learning sample during training, it is a generating model that approximates the probability distribution of the learning target. Context-based connections come in many forms, and a common class uses the input at that moment, the state at the previous moment, and the true value:

$$h^t = f(ux^{t-1} + wh^{t-1} + Ry^{t-1}) \tag{7}$$

2.2. Output mode

By establishing an output node, the Recurrent Neural Network can have multiple output modes, including sequence-classifier, sequence–sequence, encoder–decoder (asynchronous multiple output), and the like.

(1) Sequence-Classifier

The output mode of the sequence-classifier is suitable for machine learning problems of sequence input and single output, such as text classification. Given learning data and classification labels:

$$X = [X_1, X_2, \ldots X_n], y \in (1, \ldots C) \tag{8}$$

The output node of the loop unit in the sequence-classifier passes directly through the classifier. A common choice is to use the output node of the last time step:

$$\hat{y} = g(o^\tau) \tag{9}$$

Or the mean of all system states in a recursive calculation:

$$\hat{y} = g[o(\bar{h})] \tag{10}$$

A common sequence-classifier uses a fully connected structure.

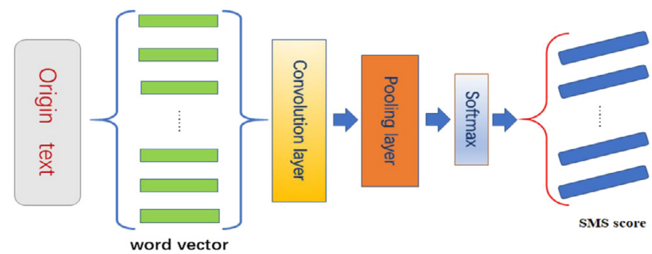

Fig. 2. The CNN-based architecture for text assessment.

2.3. The concept of Bayes

Traditional neural networks have problems such as complex models and over-fitting, which hinder the generalization of neural networks. Bayesian theory solves these problems well. In 1997, Carlin et al. proposed a calculation method based on Bayesian theory. Suppose there is a network model M, the corresponding weights are $w_1, w_2, \ldots, w_n$. The traditional neural network method finds a suitable network model by minimizing the error function to adjust the weight. The disadvantage of this method is that it is easy to produce over-fitting. The prior distribution of weights and the likelihood distribution of data are considered simultaneously in the Bayesian neural network method. By introducing the distribution of super-parameter control weights, the error is minimized and the weight is optimized. According to Bayesian rules and neural network principles, the whole process goes through the following three steps.

Step 1: Under the condition that the network model M observes data D and the initial values of the hyperparameters α and β, the posterior probability of the weight w is calculated, and the weight w_m which satisfies the maximum posterior probability is obtained.

$$\mathrm{p}(\mathrm{w}|\mathrm{D}, \alpha, \beta, \mathrm{M}) = \frac{\mathrm{p}(\mathrm{D}|\mathrm{w}, \alpha, \beta, \mathrm{M})\, \mathrm{p}(\mathrm{w}|\alpha, \beta, \mathrm{M})}{\mathrm{p}(\mathrm{w}|\alpha, \beta, \mathrm{M})} \tag{11}$$

Step 2: Calculate the posterior probability of the hyperparameter and update the hyperparameters α and β based on observed data.

$$\mathrm{p}(\alpha, \beta|\mathrm{D}, \mathrm{M}) = \frac{\mathrm{p}(\mathrm{D}|\alpha, \beta, \mathrm{M})\, \mathrm{p}(\alpha, \beta|\mathrm{M})}{\mathrm{p}(\mathrm{D}|\mathrm{M})} \tag{12}$$

Step 3: By comparing the saliency of each model, the model with the largest a posteriori probability is obtained to determine the optimal network (see Fig. 2).

$$\mathrm{p}(M|D) = \frac{\mathrm{p}(\mathrm{D}|\mathrm{M})\, \mathrm{p}(\mathrm{M})}{\mathrm{p}(\mathrm{D})} \tag{13}$$

In this paper, we utilize Bayesian theory to simplify the feature extraction process. Specifically, for input world vector $\{v_1, v_2, v_3, \ldots, v_n\}$, we acquire features $\{x_1, x_2, x_3, \ldots, x_n\}$. We choose feature x_i to get the quality score f_i. $\mathrm{p}(f_i) = \mathrm{w} * \mathrm{p}(x_i)$, where w is the weight of feature x_i. Through formula (11)∼(13), We can easily choose the features that make the score higher.

2.4. CNN-based text quality assessment

The convolutional neural network is a feedforward neural network whose network structure is composed of an input layer, a convolution layer, a pooling layer, a fully connected layer and an output layer. The convolution layer is a feature extraction layer, and the features of the text are extracted by a filter; the pooling layer is a feature mapping layer, and the features obtained after the convolution layer are sampled to obtain a local optimal value.

In this paper, the text vector is taken as input, and each sentence X is represented as a matrix of n × h, where n represents the length of the words constituting the text sentence, h represents the dimension of the word vector x_i, and the vector of words in the text sentence are trained using the word2vec model. The convolution layer is mainly for learning the local features of text sentences. This layer mainly performs convolution operations on the word vector matrix of the input layer, and operates on each continuous window of size h. The result is expressed as:

$$t_i = f\left(w \cdot x_{i:i+k-1} + b\right) \tag{14}$$

where t_i denotes the corresponding ith eigenvalue after the convolution operation. f() denotes the choice of this layer of convolution kernel function. W represents the weight matrix in the filter, where W? $R^{k\times h}$, k × h represents the size of the selected filter. b denotes the bias term. $x_{i:i+k-1}$ denotes the length from the ith word to the $(i+k-1)$th word in the text sentence. After passing through the convolutional layer, the characteristic matrix t is obtained as:

$$\mathrm{t} = [t_1, t_2, \ldots, t_{n-k+1}]^T \tag{15}$$

where: t∈R^{n-k+1}. In this paper, the maximum pooling method is used for sampling, and the characteristics obtained after the pooling layer are expressed as:

$$t' = \max\left(t_1, t_2, \ldots, t_{n-k+1}\right) \tag{16}$$

In this paper, multiple filters are selected for feature extraction. After the above operation, we can get the text feature.

An RNN using LSTM units can be trained in a supervised fashion, on a set of training sequences, using an optimization algorithm, like gradient descent, combined with backpropagation through time to compute the gradients needed during the optimization process, in order to change each weight of the LSTM network in proportion to the derivative of the error (at the output layer of the LSTM network) with respect to corresponding weight.

2.5. Semantic segmentation

Because the solution of the current state of the system needs the internal state of the previous time step, the calculation of the loop unit includes recursion. From the perspective of tree structure, all cycle cells of the previous time step are the parent nodes of the cycle cells of the current time step. Where f is the excitation function or the compression neural network with direct connection, the first corresponds to the simple cycle network (SRN), the second corresponds to the gating algorithm and some depth algorithms. The common functions for selecting excitation function include logic function and hyperbolic tangent function. θ is the weight factor in the cyclic unit, which is independent of the time step. That is to say, for a group of training samples, the recurrent neural network uses the common weight coefficient to calculate the output of all time steps. In theory, it is possible for a ring-shaped neural network to contain only cyclic elements, but recurrent neural network usually has an output node, which is defined as a linear function.

3. Experimental results and analysis

3.1. Experimental settings

Word representation through vector space has been widely used in semantic field. In this article, words will no longer be represented by just one vector. In tensor space. The term consists of a first-order tensor (vector) and a second-order tensor (parameter matrix), which represent an unsupervised 50 dimensional vector model. Initializes all word vectors to X and. This model predicts the probability of each word appearing in context similar to other vector space models. The word vector represented by this method can display syntax and semantic information. At the same time, in the experiment, each word was also associated with matrix X. The matrix is initialized to Gaussian noise. If the dimension of each vector is n. So the matrix for each word is. Since the vectors and matrices representing words are learned through semantic tags, they will be continuously modified to synthesize vector sequences with predictable distribution. Therefore, matrix initialization is random. It is usually represented by a unit matrix.

3.2. Experimental process

(1) In this paper, a single variable method was used to divide the experimental data into experimental groups and control groups, and the experimental data was collected. The experimental components were divided into three groups. The first group uses the unsupervised structured model training model to use the Recurrent Neural Network, and the second group uses the simplified classification process to apply the semantic segmentation idea for classification research. The third group combines the idea of circular neural network and semantic segmentation method to analyze English text and English pictures, and the control group adopts conventional surgical methods. Record several sets of experimental data and analyze the experimental data.

4. Discussion

4.1. Comparison of data analysis

(1) Compared with traditional neural networks, Recurrent Neural Networks can store information, which makes the circular network successful in tasks such as speech recognition, language modeling and machine translation. Circular networks also have disadvantages: long-term dependence problems, gradient fading or gradient explosion problems. The first group was reformed using an uncontrolled structured learning model and compared to the control group. The results of the improved method are shown in Table 1.

(2) In recent years, with the application of deep learning in the field of natural language processing, a text classification model based on deep learning has been developed. Some scholars have found that the classification method based on convolutional network has achieved good performance in text classification tasks. The experiment used a single variable method, and the second group abandoned the manual tagging by a simple classification process. The experimental results were obtained by collecting experimental data, and the experimental results are shown in Table 2.

4.2. Comparison of methods

(1) In the experiment, we used the Adam optimization method to train our model, and the learning rate was set to 0.001. For the fully connected layer, the L_2 regularization method is used to prevent overfitting and the coefficient is set to 0.0001. The factor for Dropout is set to 0.5. The input batch is 64, the loop layer hidden unit is 128, and for the convolutional layer, the convolution kernel with convolution kernel sizes of 3, 4, and 5 is used, and each size has a convolution kernel of 100, and we performed three. The collection and comparison data collection of group data is recorded as shown in Fig. 3.

(2) In our experiments, we just did a simple parameter setting. In future work, we can use random parameter search or Bayesian optimization to get the best parameters to improve the performance of the model. Similarly, we can use our word vectors

Table 1
Effect under the improved method.

	Speed	Computational efficiency
Unsupervised Structured Approach	1.2 Per second	300 Sessions Per Second
Using the Usual Method	2.1 Per second	150 Sessions Per Second

Table 2
Comparison of experimental results.

	Speed	Internet speed
Easy Classification Process, Discarding Manual Labeling	250 Sessions Per Second	30 Words Per Second
Normal Labeling	Section 2.1 Per Second	25 Sessions Per Second

	Number					
Score range	SVM	RF	NN	NB	LR	BDCNN
0~20	63	63	112	39	109	33
20~40	214	125	178	153	229	85
40~60	448	211	338	625	536	124
60~80	2746	3206	2675	2689	3018	3038
80~100	2103	1969	2271	2068	1682	2294

Fig. 3. Comparison of the collection of three sets of data.

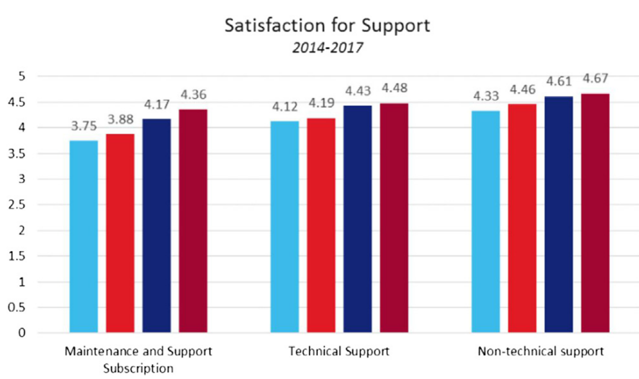

Fig. 4. Satisfaction survey results.

trained by other models of vector words (such as Glo V_e) to measure our model. We also investigated the results of random inspection models and conducted a random survey of 100 industry experts to examine these industry experts. The satisfaction degree of the experimental model is divided into three groups of 10 people each, and the results of the survey are shown in Fig. 4.

5. Conclusions

Layout network representation based on neural network model is becoming more and more popular [11]. The distribution representation of the neural network model is called word vector, and the word embedding or distribution representation. The method of representing word vectors in a neural network uses a neural network method to utilize the relationship between the modeling model and the target word. This method can map text into a low-dimensional vector space, solve the problems of high-latitude and high-latitude vocabulary models, avoid dimensional disasters, and maintain the correlation between words. Using word vectors as input and then sorting the text using a circular or convolutional network can significantly improve performance. This method has high efficiency and good anti-noise performance. Then, each sub-region will be represented by feature vectors such as texture, density and brightness. Train the word vector of each word in the short message data set, and then convert the word vector group of each short message into a two-dimensional matrix. Then, the feature matrix is used as the input of the convolutional neural network, the convolution kernels of different scales are used to extract the multi-scale short message features, and the fusion strategy is used to obtain the local optimal features. Finally, the local optimal characteristics are used to obtain the classification results.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Hunan Province Education Scientific Research Project, China (No. XJT(2019), No. 291(724)), Social and Science Fund of Hunan Province, China (No. 16WLH07).

References

[1] J. Zhang, H. Li, X. Liu, et al., On efficient and robust anonymization for privacy protection on massive streaming categorical information, IEEE Trans. Dependable Secure Comput. 14 (5) (2017) 507–520.
[2] S.P. Damodaran, Mutual information of massive MIMO systems on block Rayleigh-faded channels, Cluster Comput. (2) (2018) 1–8.
[3] J. Martinez-Gil, B. Freudenthaler, T. Natschläger, Automatic recommendation of prognosis measures for mechanical components based on massive text mining, Int. J. Web Inf. Syst. 14 (4) (2018) 480–494.
[4] J.T. Tsiang, B.K. Woo, Comparison of online dementia information in chinese and in english languages, Cureus 9 (10) (2017) e1808.
[5] G. Szommer, Parallel expansions: The role of information during the formative years of the english east India company (1600–1623), Inf. Cult. 53 (3/4) (2018) 303–336.
[6] H. Ge, Research on the chinese foreign english teaching quality assessment with intuitionistic fuzzy information, J. Comput. Theoret. Nanosci. 15 (1) (2018) 278–281.
[7] T. Riede, K. Zuberbühler, The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization, J. Acoust. Soc. Am. 114 (2) (2017) 1132–1142.
[8] S. Saeb, E.G. Lattie, K.P. Kording, et al., Mobile phone detection of semantic location and its relationship to depression and anxiety, Jmir Mhealth Uhealth 5 (8) (2017) e112.
[9] J. Wang, Z. Xu, New study on neural networks: the essential order of approximation, Neural Netw. 23 (5) (2010) 618–624.
[10] X. Liu, Y. Li, Q. Wang, Multi-view hierarchical bidirectional recurrent neural network for depth video sequence based action recognition, Int. J. Pattern Recognit. Artif. Intell. 32 (10) (2018) 1850033.
[11] S. Zhou, L. Chen, V. Sugumaran, Hidden two-stream collaborative learning network for action recognition, CMC-Comput. Mater. Continua 63 (3) (2020) 1545–1561.