Automatic Scoring Model of Subjective Questions Based Text Similarity Fusion Model

Bo Xie and Long Chen[✉]

Zhejiang GongShang University, Hangzhou, China
boxie@mail.zjgsu.edu.cn, 987465580@qq.com

Abstract. AI In this era, scene based translation and intelligent word segmentation are not new technologies. However, there is still no good solution for long and complex Chinese semantic analysis. The subjective question scoring still relies on the teacher's manual marking. However, there are a large number of examinations, and the manual marking work is huge. At present, the labor cost is getting higher and higher, the traditional manual marking method can't meet the demand The demand for automatic marking is increasingly strong in modern society. At present, the automatic marking technology of objective questions has been very mature and widely used. However, by reasons of the complexity and the difficulty of natural language processing technology in Chinese text, there are still many shortcomings in subjective questions marking, such as not considering the impact of semantics, word order and other issues on scoring accuracy. The automatic scoring technology of subjective questions is a complex technology, involving pattern recognition, machine learning, natural language processing and other technologies. Good results have been seen in the calculation method-based deep learning and machine learning. The rapid development of NLP technology has brought a new breakthrough for subjective question scoring. We integrate two deep learning models based on the Siamese Network through bagging to ensure the accuracy of the results, the text similarity matching model based on the birth networks and the score point recognition model based on the named entity recognition method respectively. Combining with the framework of deep learning, we use the simulated manual scoring method to extract and match the score point sequence of students' answers with standard answers. The score recognition model effectively improves the efficiency of model calculation and long text keyword matching. The loss value of the final training score recognition model is about 0.9, and the accuracy is 80.54%. The accuracy of the training text similarity matching model is 86.99%, and the fusion model is single. The scoring time is less than 0.8s, and the accuracy is 83.43%.

Keywords: Subjective question automatic scoring · Text similarity · Siamese network · Named entity recognition · Natural language processing · Machine learning

1 Introduction

The scale of China's online education market is increasing year by year. As a test method for learning effect and knowledge mastery, due to the large number and scale of various

training examinations, the demand of education and training institutions for automatic marking is increasingly strong, so that manual marking can't meet the demand. At present, there is no formed Chinese marking system applied to the market. Because of the complexity of Chinese text and the differences in semantic level, the development of Chinese subjective intelligent marking system is frequently hindered. By reasons of the complexity and the difficulty of natural language processing technology in Chinese text, most of the automatic marking systems stop at the objective question marking and simple English composition marking. Due to the growth of data and the improvement of computing power, deep learning has made a great breakthrough. The deep learning methods based on neural network have been applied into NLP field. At the same time, information extraction, part of speech tagging, named entity recognition and other research directions have been improved, which greatly improves the accuracy of automatic marking.

With the development of computer and network technology, a lot of subjective marking systems about English have sprouted abroad, such as PEG, IEA, Criterion and so on. However, the domestic research on subjective question marking has only been carried out gradually in the past 20 years. At present, no formed Chinese marking system has been applied to the market. Due to the complexity of Chinese text and the differences in semantic level, the development of Chinese subjective question intelligent marking system is frequently hindered.

Three main technical methods about the automatic marking system are introduced at present: the method based on templates and rules, based on the traditional machine learning method, based on the deep learning method.

(1) Rule based and template-based method: this method relies on artificial features and templates, and the trained model does not have generalization. For example, auto mark system [1] makes multiple scoring templates of correct or wrong answers for each question in advance, matches the candidates' answers with the templates one by one, judges the correctness and gives scores, which is in line with people's way of thinking. Bachman et al. Proposed that [2] generate regular expressions automatically according to the reference answers, and each regular expression matches a score. When the students' answers are consistent with the generated expressions, they get a score. This method is suitable for students with low diversity of answers and low difficulty of questions. Jinshui Wang et al. [3]. introduced professional terms in the field of power system analysis into the dictionary to improve the ability of word segmentation of professional terms. At the same time, they introduced ontology and synonym forest in the field of power system analysis to improve the word similarity calculation ability between common words and professional terms. However, the disadvantage is that it costs huge human resources to build the scoring data set, which makes it impossible to comprehensively evaluate Objective to evaluate the effectiveness and universality of the automatic scoring method. Fang Huang proposed [4] to design a new text translation information automatic scoring system based on XML structure. By setting weights, the valuable information in the answers is extracted, the closeness between candidates' answers and standard answers is analyzed, and the corresponding scores are given.

(2) Based on the traditional machine learning method. In traditional machine learning, we usually need to define features manually, and use regression, classification or

a combination of them to get a score. For example, Sultan et al. [5]. constructed a random forest classifier using text similarity, term weight and other features. Kumar et al. [6]. defined a variety of features including key concept weight, sentence length and word overlap features, and scored them by decision tree, and achieved good results on ASAP dataset. Jie Cao et al. [7]. proposed that after preprocessing the student answer text and the reference answer text, the similarity of the topic probability distribution between the student answer and the reference answer can be calculated through LDA model training, so as to realize the evaluation.

(3) With the rapid increase of big data storage capacity and computing power, deep learning has been successfully applied into the field of image recognition and natural language processing. Shuai Zhang [8] Based on the Siamese Network subjective question automatic scoring technology, at the same time input student answers and reference answers for similarity calculation, so as to estimate the score of student answers, improve the similarity calculation method based on sentence surface features, and improve the accuracy. Yifan Wang et al. [9]. used the extended named entity recognition method to extract some keywords from the candidate answers of subjective questions, and used the improved synonym forest word similarity calculation method to calculate the similarity between the candidate keywords and the target keywords in the standard answers of subjective questions. The method solves the problem of low matching efficiency in similarity calculation of long text words and preferentially extracts keywords for similarity calculation, which effectively improves the performance of similarity calculation of key words in shortening the calculation time compared with the traditional word similarity methods.

Subjective question scoring faces many challenges. How to calculate the similarity between standard answers and students' answers is an important problem in subjective question scoring model. Traditional models only consider the surface features of sentences by using words, words and other indicators to calculate text similarity, so the accuracy is not high. There are some researches on the automatic score of composition by analyzing text coherence in China. Due to the limitation of short text in the answer text of subjective question, accuracy is not effectively improved by simply increasing the coherence of the text. In addition, the method of word similarity calculation based on synonym forest has achieved good results in Chinese text, while applying into long text may lead to the decline of the method performance and accuracy.

In order to solve the mentioned problems, this paper proposes a fusion method based on Siamese Network and named entity recognition. On the basis of general lexical features, Siamese Network model is added to judge the similarity between students' answers and reference answers, so as to score students' answers. Compared with other neural network models, Siamese Network is special in that it inputs two subnets at the same time Network, and these two subnetworks share weight. The characteristics of Siamese Network make it have a good effect in measuring similarity. But the disadvantage is that as a kind of neural network, Siamese Network can only get the scoring results, and can't make a reasonable explanation for the scoring results. The extended named entity recognition method is used to extract some keywords from the candidate answers of the subjective questions, and the improved synonym forest word similarity calculation

method is used to calculate the similarity between the candidate keywords and the target keywords in the standard answers of the subjective questions, which improves the performance of the original algorithm and effectively shortens the calculation time.

2 Model Presentation

Neural network can accurately measure the similarity between standard answers and students' answers. To simulate the process of manual scoring and make a reasonable explanation for the results of the model, this paper proposes a text similarity matching model (TSMM) based on Siamese Network, Text similarity matching model and scoring point identification model (SPRM) based on named entity recognition are used to fuse the models. The model is able to score according to the scoring points of user answers and the interpretation in the answers. We adopt a two-pronged strategy: on the one hand, we use deep learning method to extract the scoring points of user answers and highly simulate "manual marking" to realize the judgment of scoring points hit; on the other hand, we use Siamese Network model to compare the standard answers with students' answers. The final subjective score results are obtained through the fusion of dual-strategy model, and the overall route diagram is shown in Fig. 1.
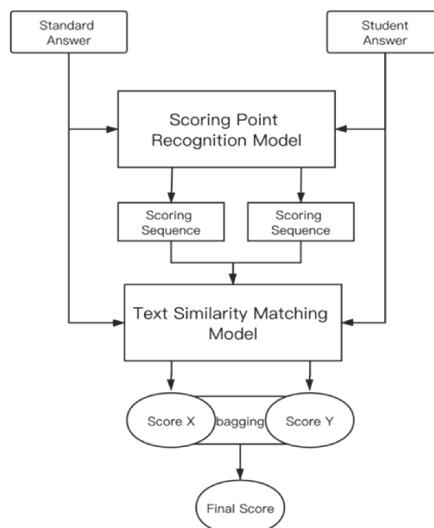
Fig. 1. Overall technology roadmap.

3 Related Technologies

Text similarity calculation is the core of the intelligent evaluation system of subjective questions. The method of text similarity calculation is related to the accuracy and practicability of the whole intelligent evaluation system of subjective questions. The following

is the text similarity calculation technology involved in the development of the subjective question automatic evaluation model, including long-term and short-term memory (LSTM), conditional random field (CRF), pre-training model, Siamese Network and other text similarity models.

3.1 Long Short-Term Memory (LSTM)

The normal RNN has no solution to the long-term memory function. For example, trying to predict the last word of "I majored in logistics when I was in University… I will be engaged in logistics after graduation." Recent information shows that the next word may be the name of an industry. However, if we want to narrow the selection range, we need to include the context of "logistics major" and infer the following words from the previous information. Similarly, in terms of score point prediction, whether the user's answer or the standard answer is a long text, the interval between the relevant information and the predicted position It's quite possible. However, RNNs are incapable of solving this problem. As one of the most popular RNNs, long-short term memory network (LSTM) successfully solved the defects of the original recurrent neural network which has been applied into many fields such as speech recognition, picture description, and natural language processing. LSTM is quite suitable for processing and predicting important events with relatively long interval and delay in time series [10].

3.2 Conditional Random Field (CRF)

In order to make our scoring point recognition model perform better, the marking information of adjacent data can be considered when marking data. This is difficult for ordinary classifiers to do, and also a good place for CRF. CRF is the conditional random field, which represents the Markov random field of another group of output random variables y given a group of input random variables X. the attribute of CRF is to assume that the output random variables establish the Markov random field [11].

The CRF is refered as the speculation of the Maximum Entropy Markov model in the labeling problem. The CRF layer can be used to predict the final result of the sequence labeling task, some constraints are added to guarantee that the predicted label is reasonable. During the training process, these constraints can be adapted consequently through CRF layer [12].

- The first word in the sentence is constantly begun with the name "O-" or "B-", rather than "I-".
- Label stands for name entity (person name, organization name, time, etc.). The label "B-L1 I-L2 I-L3 I-…", L1, L2, L3 are supposed to be entity of the same type.
- A tag sequence that starts with "I-label" is usually unreasonable. A logical sequence would start with "B-label".

These constraints will greatly reduce the probability of unreasonable sequence occurrence in label sequence prediction.

3.3   Pretraining

The pretraining model is a deep learning architecture, which has been prepared to perform explicit assignments on a lot of data. This kind of training is relatively hard to implement, and always requires a great deal of resources. Therefore, the large number of parameters it gets make the model implementation results closer to the actual results. The pretraining model learns a context-dependent representation of each member of an input sentence using almost unlimited text, and it implicitly learns general syntactic semantic knowledge. It can migrate knowledge learned from the open domain to downstream tasks to improve low-resource tasks, and is also very helpful for low-resource language processing [13].

The pretraining model has achieved good results in most of NLP tasks, and the BERT model is a language representation model released by Devlin et al. [14] (Google) in October 2018. the BERT swept the optimal results of 11 tasks in the NLP field, which can be considered as the most important breakthrough in NLP field recently. Because of its flexible training mode and outstanding effect, the BERT model has been deeply studied and applied in many tasks of NLP. This paper applies few BERT modules for pretraining tasks.

3.4   Siamese Network

Siamese Network is a kind of neural network architecture which contains two or more identical subnetworks, which sets the same configuration, same parameters and weights [15]. Parameter updating is carried out in two subnets. The structure of Siamese Network is shown in Fig. 2.

Siamese Networks are popular in tasks involving finding similarities or relationships between two comparable things [15]. Examples of how similar the input or output of two signatures are from the same person verify whether they are. Usually, in such a task,
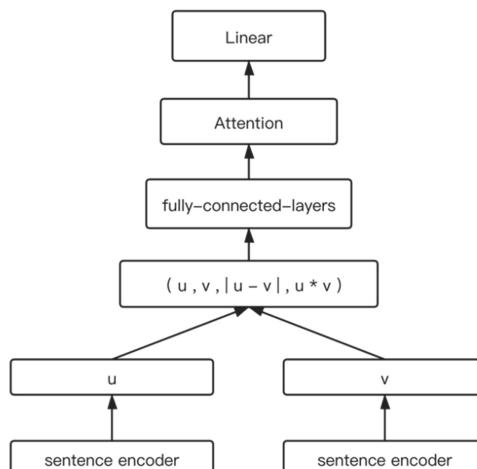

Fig. 2. Schematic diagram of siamese network.

two identical subnetworks are used to process two inputs, and another module will take their output and produce the final output.

The advantages are as follows: 1. Subnet sharing weight means that training needs less parameters, which means that it needs less data and is not easy to over fit. 2. Each subnet essentially produces a representation of its input. It makes sense to use a similar model for the same type of input (for example, matching two images or two paragraphs). Representation vectors with a similar semantics, making them simpler to compare.

4   Model Composition and Fusion

For the sake of scoring user's answers reasonably, this paper proposes an automatic evaluation model of subjective questions, which is composed of text similarity matching model (TSMM) and score point recognition model (SPRM). The TSMM calculates the semantic similarity between the standard answer with the user's answer. The SPRM is used to extract the scores of the answers, which is regard as "manual marking" simulation. Finally, the final subjective score is obtained by the model fusion.

4.1   The Automatic Evaluation Model of Subjective Questions

Input the standard answer text and student answer text into the score recognition model after training respectively, then we can extract the score point sequence of two strings of text, and further match the score points of the two strings of text through the text similarity matching model after training, so as to calculate the score of each score point and accumulate it to get the final score X; at the same time, the standard answer and student answer text are compared Students' answer text is directly input into the text similarity matching model to get the overall similarity, that is, the score Y.

Ensemble learning is a paradigm of machine learning. Training multiple models to solve the same problem and combining them to get better results [16]. One of the most important assumption is that when the weak models are combined correctly, we can get more accurate and more robust models.

Considering that both TSMM and SPRM are homogeneous weak learners, bagging can be used to learn these weak learners independently and in parallel. This method does not operate the model itself, but acts on the sample set. We use the random selection training data, then construct the classifier, and finally combine them. Different from the interdependence and serial operation among classifiers in boosting method, there is no strong dependency between base learners in bagging method, and parallel operation is generated at the same time [16].

We use bagging based method to get the final model fusion result through TSMM and SPRM model, that is, bagging the two scores obtained from the score recognition model and the text similarity matching model to get the final score.

4.2   Scoring Point Recognition Model (SPRM)

Named entity recognition is to identify entities with specific meaning in text. From the perspective of knowledge map, it is to obtain entities and entity attributes from

unstructured text [17]. Therefore, we consider using named entity recognition method to extract score points. Bi-LSTM refers to bidirectional LSTM; CRF refers to conditional random field. In SPRM, Bi-LSTM is mainly used to give the probability distribution of the corresponding label of the current word according to the context of a word, which can be regarded as a coding layer. The CRF layer can add some restrictions on the final prediction labels to ensure that the results are valid. These limitations can be learned from the CRF layer's automatic training data set during the training process. The text sequence is processed by Bi-LSTM model, the output result is transferred to CRF layer, and finally the prediction result is output [18].

The part of preprocessing prediction data, that is, sequence labeling has been completed in data preprocessing.

Take a sentence as a unit, record a sentence with n words as:

$$x = (x_1, x_2, ..., x_n)$$

x_i represents the ID of the ith word of a sentence in the dictionary, thus obtaining the one-hot vector of each word (dimension is the dictionary size).

Look-up layer is the first layer of the model, each word x_i in a sentence is mapped from a one-hot vector to a low dimensional character embedding using a pretrained or randomly initialized embedding matrix $x_i \in R_d$, d is the dimension of embedding. Set dropout to ease over fitting before entering the next layer [19].

Bidirectional LSTM layer is the second layer of the model that automatically extracts sentence features. The char embedding sequence $(x_1, x_2, ..., x_n)$ of each word of a sentence is used as the input of each time step of bidirectional LSTM, and then the hidden state sequence $\left(\overrightarrow{h}_1, \overrightarrow{h}_2, ..., \overrightarrow{h}_n \right)$ of forward LSTM output and the hidden state sequence of reverse LSTM $\left(\overleftarrow{h}_1, \overleftarrow{h}_2, ..., \overleftarrow{h}_n \right)$ output in each position are spliced according to the position $h_t = \left| \overrightarrow{h}_t; \overleftarrow{h}_t \right| \in R^m$ to obtain a complete hidden state sequence

$$(h_1, h_2, ..., h_n) \in R^{n \times m}$$

After dropout is set, a linear layer is connected, and the hidden state vector is mapped from m dimension to k dimension. K is the number of tags in the annotation set, so the automatically extracted sentence features are obtained and recorded as matrix $P = (p_1, p_2, ..., p_n) \in R^{n \times m}$. Each dimension p_{ij} of $p_i \in R^k$ can be regarded as the scoring value of the j-th tag. If softmax is used for P, it is equivalent to k-class classification for each position independently. However, it is impossible to make use of the information that has been labeled when labeling each position, so a CRF layer will be connected to label next [19].

CRF layer is the third layer of the model, which is used for sequence annotation at sentence level. The parameter of CRF layer is a matrix A of $(k + 2) \times (k + 2)$, and A_{ij} represents the transfer score from the i-th tag to the j-th tag. When labeling a location, it can use the previously labeled data. The reason for adding 2 is to add a start state to the beginning of the sentence and an end state to the end of the sentence. If we remember a tag sequence $y = (y_1, y_2, ..., y_n)$ whose length is equal to the length of the sentence,

the score of the model for the tag of Sentence x equal to y is as follows [19]:

$$score(x, y) = \sum_{i=1}^{n} P_{i,y_i} + \sum_{i=1}^{n+1} A_{y-1,y_i}$$

The score of the whole sequence is equal to the sum of the scores of each position, and the score of each position is obtained by combining pi of LSTM output and transfer matrix A of CRF. Then, the normalized probability can be obtained by Softmax:

$$P(y|x) = \frac{\exp(score(x, y))}{\sum_y \exp(score(x, y\acute{}))}$$

By maximizing the log likelihood function in the model training, the log likelihood of a training sample (x, y_x) is given by the following formula:

$$\log P(y^x|x) = score(x, y^x) - \log\left(\sum_y \exp\left(score(x, y\acute{})\right)\right)$$

In the process of prediction (decoding), The Viterbi algorithm of dynamic programming is used to solve the optimal path:

$$y^* = \underset{y}{argmax}\,score(x, y\acute{})$$

The structure is shown in Fig. 3 SPRM structure diagram [20–22]:
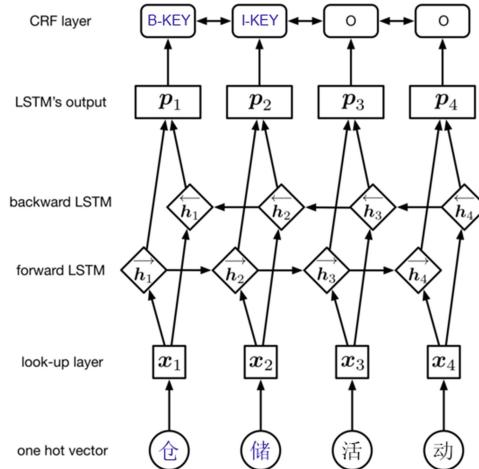
Fig. 3. Scoring point recognition model structure

4.3 Text Similarity Matching Model (TSMM)

The main idea of TSMM is: mapping the input to the target space through a function, and comparing the similarity in the target space using distance. During the training stage,
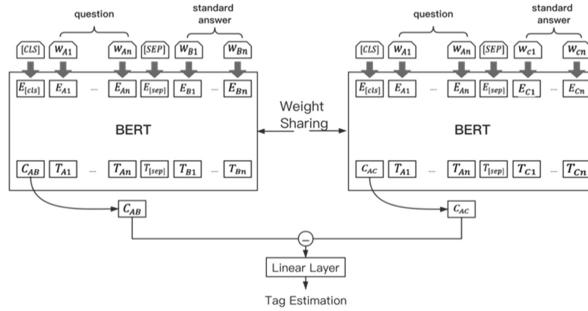
Fig. 4. Text similarity matching model structure.

we minimize the loss function values of a pair of samples from the same category and maximize the loss function values of a pile of samples from different categories. Its feature is that it receives two pieces of text as input instead of one piece of text as input.

It can be summarized as the following three points:

- Input is no longer a single sample, but a pair of samples, no longer give a single sample exact label, and given a pair of sample similarity labels.
- Designed as like as two networks, the network shared weight W, and the distance measurement of output, L1, L2, etc., were carried out in two.
- According to whether the input sample pairs come from the same category or not, a loss function is designed in the form of cross entropy loss.

In the Siamese Network, the loss function is comparative loss, which can effectively deal with the relationship of paired data in the t Siamese Network. The expression of contrastive loss is as follows [23]:

$$L = \frac{1}{2N} \sum_{n=1}^{N} yd^2 + (1 - y)\max(margin - d, 0)^2$$

The specific purpose of Siamese Network is to measure the similarity of two input texts [24]. In the process of training and testing, the encoder part of the model shares weight, which is also the embodiment of the word "Siamese". The choice of encoder is very wide, traditional CNN, RNN and attention, transformer can be used.

After getting the features u and V, we can directly use the distance formula, such as cosine distance, L1 distance, Euclidean distance, to get the similarity between the two texts. However, a more general approach is to build feature vectors based on u and V to model the matching relationship between them, and then use additional models (MLP, etc.) to learn the general text relational function mapping.

5   Experiment and Results

5.1   Experimental Data

The data of this paper comes from the official logistics industry corpus and professional questions provided by China outsourcing service competition in 2020. The data features

are as follows: short answer questions in the field of logistics vocational education are basically noun explanation and concept explanation questions, and the sentence structure is relatively simple; the composition of a piece of data includes serial number, question description, answer, keyword and keyword description, and the data is divided into three parts 600.

For the above 600 pieces of data, we expanded the data according to the score points, and got 5924 pieces of augmented data as the data set for the training of TSMM model. The characteristics of this training set are: it belongs to the field of Logistics Vocational Education, and the data composition includes question number, question, standard answer and user answers with 0 to 10 points.

5.2   Analysis of SPRM Experimental Results

First, we preprocess the existing 600 pieces of data, mainly including sequence annotation, word segmentation, and data cleaning and formatting. For the preprocessed 600 pieces of data, 70% is used as training set, and the remaining 30% is used as test set and verification set.

Table 1. Scoring point recognition model training results

	Accuracy	Precision	Recall
SPRM	80.54%	57.12%	58.75%

Experimental results: the model loss in the training set is reduced from 53.138512 to 0.93004, and the accuracy rate is 80.54%. For SPRM, the processing in each layer is relatively simple compared to the existing work, and there is room for improvement in the future. For instance, the initialization method of word vector embedding we used in the experiment is simple random initialization. Besides, due to the small size of corpus, we can consider the pretraining value on a larger corpus. SPRM may over fit in this case because of the large number of iterations, so it is necessary to draw a verification set for early stopping.

5.3   Analysis of TSMM Experimental Results

For the expanded 5924 data, 70% is used for training set, and the remaining 30% is used for test set and verification set. The loss value of the model is reduced from 174.2736382 to 21.5801761, and the accuracy rate reaches 86.99%. It can be seen that the calculation effect of using twin network to input standard answers and student answers at the same time is higher than that only based on the surface features of sentences.

5.4   Experimental Analysis of the Automatic Evaluation Model for Subjective Questions

After the recognition of the score point sequence by SPRM model, through the word similarity matching calculation based on Synonymy Thesaurus and CNKI, the subjective

score can be obtained, which can be used as the comparison between TSMM model and fusion model. This experiment uses real short answer questions of logistics final examination, a total of 10 questions as experimental data. After scoring by SPRM, TSMM and model fusion, the calculated evaluation indexes are as follows lower.

Table 2. The performance of the grading approaches.

	MSE	RMSE	MAE
SPRM	1.96	1.40	1.16
TSMM	0.80	0.89	0.60
Fusion model	0.32	0.57	0.57

Table 2 compares the calculation results of SPRM, TSMM and fusion model under different indexes. Results show that the fusion model has the advantages of MSE, RMSE, MAE is the minimum, which shows that the fusion model has more advantages than the single model of SPRM and TSMM, and the score sequence of SPRM is interpretable to the fusion model.

References

1. Rudner, L., Gagne, P.: An overview of three approaches to scoring written essays by computer. Practical Assessment **151**(3), 501 (2001)
2. Bachman, L.F., Carr, N., Kamei, G., et al.: A reliable approach to automatic assessment of short answer free responses. In: Proceedings of the19th International Conference on Computational Linguistics - Volume 2. DBLP (2002)
3. Wang, J., Guo, W., Tang, Z.: Automatic scoring method for subjective questions based on domain ontology and dependency parsing. J. Guizhou University (Natural Science) **37**(06), 79–84+124 (2020)
4. Huang, F.: Design of XML structure based automatic scoring system for text translation information. Modern Electron. Tech. **42**(23), 177–181 (2019)
5. Sultan, M.A., Salazar, C., Sumner, T.: Fast and easy short answer grading with highaccuracy. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1070–1075 (2016)
6. Kumar, Y., Aggarwal, S., Mahata, D., et al.: Get IT scored using auto SAS — an automated system for scoring short answers. In: International Conference on Artificial Intelligence, 2019, vol. 33(01), pp. 9662–9669 (2019)
7. Jie, C.A.O., Mengyao, L.I., Dawei, C.H.E.N.: Automatic scoring algorithm of subjective questions based on LDA topic model. Comput. Programm. Skills Maintenance **04**, 119–121 (2020)
8. Zhang, S.: Automatic scoring technology of subjective questions based on twin neural network. Modern Comput. **2020**(05), 23–25 (2020)
9. Yifan, W.A.N.G., Guoping, L.I.: Automated scoring method for subjective questions based on semantic similarity and named entity recognition. Electron. Measur. Technol. **42**(02), 84–87 (2019)

10. Xie, X., Wu, D., Liu, S., et al.: IoT data analytics using deep learning. arXiv preprint arXiv:1708.03854 (2017)

11. Yang, E., Ravikumar, P., Allen, G.I., et al.: A general framework for mixed graphical models. arXiv preprint arXiv:1411.0288 (2014)

12. Panchendrarajan, R., Amaresan, A.: Bidirectional LSTM-CRF for named entity recognition. In: Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation (2018)

13. Tamborrino, A., Pellicano, N., Pannier, B., et al.: Pre-training is (almost) all you need: An application to commonsense reasoning. arXiv preprint arXiv:2004.14074 (2020)

14. Yuanzhi, W.A.N.G., Ziying, C.A.O.: Chinese named entity recognition based on bert-BLSTM-CRF model. J. Anqing Normal Univ. (Natural Sci. Edition) **27**(01), 59–65 (2021)

15. Manocha, P., Badlani, R., Kumar, A., et al.: Content-based representations of audio using siamese neural networks. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) pp. 3136–3140. IEEE (2018)

16. Ganaie, M.A., Hu, M.: Ensemble deep learning: A review. arXiv preprint arXiv:2104.02395 (2021)

17. Adnan, K., Akbar, R.: Limitations of information extraction methods and techniques for heterogeneous unstructured big data. Int. J. Eng. Bus. Manage. **11**, 1847979019890771 (2019)

18. Zhang, M., Geng, G., Chen, J.: Semi-supervised bidirectional long short-term memory and conditional random fields model for named-entity recognition using embeddings from language models representations. Entropy **22**(2), 252 (2020)

19. Ji, B., Liu, R., Li, S., et al.: A hybrid approach for named entity recognition in Chinese electronic medical record[J]. BMC Med. Inform. Decis. Mak. **19**(2), 149–158 (2019)

20. Ma, X.Z., Eduard, H.: End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. Ann Meet Assoc Comput Linguist (ACL) (2016)

21. Dong, C., Zhang, J., Zong, C., et al.: Character-based LSTM-CRF with radical-level features for Chinese named entity recognition. In: International conference on computer processing of oriental languages, vol. 10102, pp. 221–230. Springer, Cham (2017). Doi: https://doi.org/10.1007/978-3-319-50496-4_20

22. Chen, T., Xu, R.F., He, Y.L., et al.: Improving sentiment analysis via sentence type classification usint BiLSTM-CRF and CNN. In: Experts Systems with Applications, pp. 260–270 (2016)

23. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR' 2006), vol. 2, pp. 1735–1742. IEEE (2006)

24. Neculoiu, P., Versteegh, M., Rotaru, M.: Learning text similarity with siamese recurrent networks. In: Proceedings of the 1st Workshop on Representation Learning for NLP, pp. 148–157 (2016)

25. Aderhold, J., et al.: 2001 J. Cryst. Growth 222 701

26. Dorman, L.I.: Variations of Galactic Cosmic Rays (Moscow: Moscow State University Press), p. 103 (1975)

27. Caplar, R., Kulisic, P.: Proc. Int. Conf. on Nuclear Physics (Munich), vol. 1 (Amsterdam: North-Holland/American Elsevier) p. 517 (1973)

28. Szytula, A., Leciejewicz, J.: 1989 Handbook on the Physics and Chemistry of Rare Earths, vol. 12, ed K A Gschneidner Jr and L Erwin (Amsterdam: Elsevier), p. 133 (1989)
29. Kuhn, T.: Density matrix theory of coherent ultrafast dynamics Theory of Transport Properties of Semiconductor Nanostructures (Electronic Materials vol 4) ed E Schöll (London: Chapman and Hall) chapter 6, pp. 173–214 (1998)