2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

AGRICULTURAL TEXT CLASSIFICATION METHOD BASED ON DYNAMIC FUSION OF MULTIPLE FEATURES

Dr.P.Karunakar Reddy[1], Javisetti Divya[2], Ketha Charishma[3],Gugulothu Sunitha[4]

[1]Professor, School of CSE, Malla Reddy Engineering College For Women(Autonomous Institution), Maisammaguda, Dhulapally,Secunderabad,Telangana-500100

[234]UG Student, Department of CSE,Malla Reddy Engineering College for Women, (Autonomous Institution), Maisammaguda,Dhulapally,Secunderabad,Telangana-500100

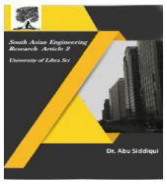
ABSTRACT

The traditional methods of classifying texts consider values from agricultural texts as characters that could lead to a reduction in the semantic meaning while being lost. Most of such numerical features, often necessary in agriculturally related applications, carry specific information that the traditional techniques are unable to fully capture. In this regard, it is proposed that a novel text classification method based on the concept of multivariate feature dynamic fusion be undertaken. The method utilizes multiple windows of Convolutional Neural Networks (CNNs) to extract local semantic information from agricultural texts at different levels. CNNs are very effective in capturing many aspects of the text, from fine-grained details to broader context. However, the crucial numerical features that contain significant semantic value are also extracted manually through an artificial method to create a numerical feature vector. This vector would therefore encapsulate the numerical details of the agricultural text, potentially essential for correct classification. Attention mechanism: This would be proposed to dynamically fuse multiple semantic features derived from both textual and numerical data. The attention mechanism enables the model to focus more on relevant features, thereby having richer and deeper semantic expression for the agricultural text. By integrating the model with textual and numerical information, it can classify agricultural texts better than using only one kind of information. This enhances greatly the ability to mine deep latent semantic features from text and to represent these features so that the accuracy and performance in a text classification task will be even better. This approach incorporates the dynamic fusion of features in attention mechanisms that allow maximum utilization of textual and numerical information.

Keywords: Agricultural Text Classification,Multivariate Feature Dynamic Fusion,Numerical Feature Extraction,Attention Mechanism**.**

I. INTRODUCTION

In recent years, with the rapid development of agricultural information technology, many agricultural technology websites, resource databases and information platforms have emerged. Text as the most common form of data carrying, contains a huge amount of information. How to obtain implicit knowledge quickly and accurately from batch text resources is one of the main tasks of text



classification in the field of agricultural information.

Agricultural text classification is a branch of text classification. It can obtain potentially useful information from massive complex and noisy agricultural text information, and guide practical activities such as agricultural production and research [1], [2]. The agricultural texts that record the phenotypic agronomic traits of wheat are classified according to the cold resistance of wheat, which belongs to the classification of agricultural texts. Traditional text classification methods mainly used machine learning methods such as K-nearest neighbor (KNN) , Naive Bayesian (NB) and Support Vector Machine (SVM) . The traditional method are used for agricultural text classification and effectively improves the performance of text classification, but it requires manual feature extraction to achieve text classification. With the problems of redundancy, sparsity and diversity of agricultural texts, the accuracy and comprehensiveness of text feature expression cannot be guaranteed if manually construct feature engineering. Deep learning can automatically extract key features, does not need complex feature engineering, and has strong adaptability and mobility. It has been widely applicated in image processing and natural language processing. Particularly, deep learning is widely used in text classification and the effect is better. The neural network is also used for the agricultural text classification task. Liang et al. calculated sentence similarity by using word2vec and LSTM for 3007 common rice problems to achieve accurate matching between user problems and common rice problems. Xu et al. constructed a Seq2Seq question-and-answer model based on word2vec and attention mechanism optimization, which significantly improved the accuracy of question-and-answer for rice diseases and insect pests. Zhao et al.used BI-GRU neural network to achieve efficient classification of tomato pest questions. Jin et al. used the method of Bi-GRU combined with MUl-CNN to classify the short text of the problems raised by farmers according to 12 categories of problems such as pests and weeds, market sales and animal diseases. Feng et al.

II. RELATED WORK

S. S. Feng, X. T. Yu, Z. Y. Cheng, Z. D. Xue, and J. Ning (2021).

Exploring applications involving Deep Convolutional Neural Network are the authors in the work titled "Rice knowledge text classification based on Deep Convolution neural network." To carry this type of research effec tively, leveraging local semantics extraction within text at multiple levels is an activity th at can increthe classifier's accuracy when usi ng deep learning techniques, in this context, deep CNNs in txts related to rice.

Y. L. Shi, Y. P. Cui, and Z. G. Du. (2022):

The research, "Classification Method of Agricultural News Text Based on BERT and Deep Active Learning," proposes a hybrid approach that combines the power of BERT with deep active learning for agricultural news text classification. BERT is known for its strong contextual feature extraction, enhancing semantic understanding, while active learning reduces the need for large labeled datasets by

iteratively selecting the most informative samples for training.

M. Zhao, C. C. Dong, Q. X. Dong, and Y. Chen. 2018.

In "Question Classification of Tomato Pests and Diseases Question Answering System Based on BIGRU," the authors focused on question classification for a tomato pest and disease Q&A system. The study used the BiGRU model to capture sequential semantic features from agricultural questions. This approach ensures improved classification performance by understanding context and sequence in the questions, which is crucial for question-answering systems.

N. Jin et al., C. J. Zhao, H. R. Wu, Y. S. Miao, S. Li, and B. Z. Yang 2020:

The work titled "Classification Technology of Agricultural Questions Based on BiGRU_MulCNN" combines BiGRU and Multi-CNN (MulCNN) models for agricultural question classification. The BiGRU component captures sequential dependencies and global contextual features, while the Multi-CNN extracts local features at multiple levels. This combination improves the model's ability to classify agricultural questions effectively, demonstrating the advantages of integrating recurrent and convolutional networks.

III IMPLEMENTATION

The traditional text classification methods tend to treat values in agricultural texts as characters, which causes a huge loss of semantic meaning, especially for numerical features. Numerical values in the agricultural context, such as yield measurements, weather conditions, or phenotypic characteristics, carry critical information that traditional techniques fail to fully capture. To address this, a novel text classification method based on multivariate feature dynamic fusion is proposed. The implementation starts with using the Convolutional Neural Networks with multiple windows to extract the local semantic features from agricultural text at different levels. CNNs are effective in capturing fine-grained word details and broader contextual information. However, crucial numerical features, which typically carry important semantic value, are separately extracted using manual artificial methods to construct a numerical feature vector. This vector supplements the textual data with preserving the semantic significance of numerical values so that they are not overlooked during classification. To integrate textual and numerical features effectively, an attention mechanism is introduced that enables the model to dynamically fuse these features. The attention mechanism focuses on the most relevant aspects of both textual and numerical inputs, enriching the deep semantic representation of the agricultural text. Finally, these fused features are passed through fully connected layers for classification. Thus, this method improves the ability of the model to mine deep latent semantic features by incorporating textual and numerical data, ultimately enhancing accuracy and performance in agricultural text classification tasks. Dynamic feature fusion ensures that all information is optimally utilized, thereby making this approach a giant leap in agricultural text processing.

IV ALGORITHM

1.Decision Tree Classifier

Decision tree classifiers capture decision-making knowledge from training data and are applied in different applications. It works on the principle of recursive partitioning of the dataset according to specific tests until all objects in a subset belong to a single class. A test partitions the data into subsets at each step. The structure takes the form of a tree where each outcome of the test becomes a branch and each final subset a leaf labeled with its class.

2.GRADIENT BOOSTING

Gradient boosting is a machine learning technique for regression and classification tasks that builds models as ensembles of weak learners, typically decision trees. Unlike random forests, it optimizes an arbitrary differentiable loss function and constructs the model stage-wise, where each stage minimizes the residual error of the previous model. Gradient-boosted trees often outperform random forests due to their refined optimization approach.

3. K-Nearest Neighbors (KNN)

KNN is a simple, powerful, non-parametric, and lazy classification algorithm that depends on similarity measures. It does not "learn" from training data until a test example is presented. When new data needs classification, KNN identifies its k-nearest neighbors in the training set based on a chosen distance metric, then assigns the majority class label from those neighbors.

4. Logistic Regression Classifier

Logistic regression is used for modeling the relationship between a categorical dependent variable and independent variables. This is very popular in case of binary classification, like Yes/No or 0/1 outcomes. It can also be used for multinomial logistic regression to solve multi-class problems. Unlike discriminant analysis, logistic regression does not require normality of independent variables. It provides regression equations, goodness-of-fit measures, ROC curves, and model validation, which are very useful for categorical-response variables.

5.Naïve Bayes

Naive Bayes is a supervised learning algorithm that relies on the condition of features being conditionally independent given a class. Though it is quite simple, it is strong, efficient, and effective in big datasets. It is simple to implement and has fast learning time. Performance-wise, it is competitive with other linear classifiers, such as logistic regression and linear SVM, but it's less applied practically because it lacks interpretability.

6.Random Forest

Random forests are ensemble models that build multiple decision trees at the time of training. The final output for classification tasks is a majority vote by all trees, whereas for regression tasks, the average prediction is returned. Through the combination of "bagging" and random feature selection, overfitting is reduced and generalization improved. Because of their accuracy and ease of use, they are extensively used as "black-box" models in business applications.

7.SUPPORT VECTOR MACHINE (SVM)

SVM is a discriminant machine learning technique that finds an optimal hyperplane to separate data classes in a multidimensional feature space. It is robust to overfitting and uniquely solves convex optimization problems, always returning the same model parameters

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

for a given dataset. SVM uses kernel functions to transform data into a higher-dimensional space, ensuring effective separation of classes. Unlike perceptrons or genetic algorithms, SVM consistently finds the best solution without initialization bias.

V.RESULTS

Fig 1: User Remote

Fig 2:Accuracy Bar Chart

Fig 3:Line Chart

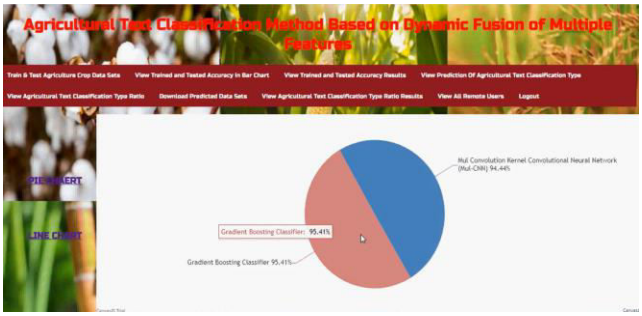


Fig 4:Pie Chart

VI.CONCLUSION

In this paper, since the phenotypic numerical features contained in agricultural texts have practical significance, combined with the experimental tasks of this study, we proposed a classification method of agricultural texts based on dynamic fusion of multiple features. Firstly, three convolution kernels with different sizes are used to obtain more abundant local semantic features at different levels of wheat cold resistance text, and Bi-LSTM with attention-fusion mechanism is used to obtain important semantic expressions between contexts in the global scope. Then independent numerical features are constructed by using the text of wheat traits with numerical expression. Attention mechanism is introduced again in the process of multi-feature dynamic fusion to dynamically adjust the weights of different semantic expressions in the process of feature fusion and capture the key information beneficial to text classification in different levels of features. The performance of agricultural text classification shows that the proposed multi-feature dynamic fusion method fully excavates the potential semantic features in the text data, which is effective for the agricultural text classification task with phenotypic values. The next work will expand the data set and sample equalization to

improve the performance of wheat cold resistance text classification and recognition.

REFERENCES

[1] S. S. Feng, X. T. Yu, Z. Y. Cheng, Z. D. Xue,and J. Ning, ‘‘Rice knowledgetext classification based on deep convolutionneural network,’’ Trans. Chin.Soc. Agricult. Machinery, vol. 52, no. 3, pp. 257–264, 2021.

[2] Y. L. Shi, Y. P. Cui, and Z. G. Du, classification method of agriculturalnews text based on BERT and deep active learning,’’ J. Library Inf. Sci.Agricult., vol. 34, no. 3, pp. 19–29, 2022.

[3] M. Zhao, C. C. Dong, Q. X. Dong, and Y. Chen, ‘‘Question classification oftomato pests and diseases question answering system based on BIGRU,’Trans. Chin. Soc. Agricult. Machinery, vol. 49, no. 5, pp. 271–276,May2018.

[4] N. Jin, C. J. Zhao, H. R. Wu, Y. S. Miao, S. Li,and B. Z. Yang, ‘‘Classificationtechnology of agricultural questions based on BiGRU_MulCNN,’’Trans. Chin. Soc. Agricult. Machinery, vol. 51, no. 5, pp. 199–206,May 2020.

[5] J. D. Liang, B. Cui, H. Jiang, Y. Shen, and Y. Xie , ‘‘Sentence similaritycomputing based on word2vec and LSTM and its application in rice FAQquestion-answering system,’’ J. Nanjing Agricult. Univ., vol. 41, np.5,pp. 946–953, May 2018.

[6] T. Y. Xu, D. Zhao, Y. Zhou, S.Feng,and H. Wang, ‘‘Research on methodof intelligent Q & A for rice pests and diseases based on word2vecand attention-Seq2Seq,’’ J. Shenyang Agricult. Univ., vol. 50, no. 3,pp. 378–384, Mar. 2019.

[7] Z. Liu, H. Huang, C. Lu, and S. Lyu, ‘‘Multichannel CNN with attentionfor text classification,’’ 2020, arXiv:2006.16174.

[8] X. Li and H. Ning, ‘‘Chinese text classification based on hybrid model ofCNN and LSTM,’’ in Proc. 3rd Int. Conf. Data Sci. Inf. Technol., Jul. 2020,p. 20203.

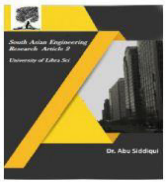
[9] Y. Kim, ‘‘Convolutional neural networks for sentence classification,’’ inProc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2014,pp. 1746–1751.

[10] N. Kalchbrenner, E. Grefenstette, and P. Blunsom , ‘‘A convolutional neuralnetwork for modelling sentences,’’ 2014, arXiv:1404.2188.

[11] R. Wang, Z. Li, J. Cao, T. Chen, and L.Wang, ‘‘Convolutional recurrentneural networks for text classification,’’ in Proc. Int. Joint Conf. NeuralNetw. (IJCNN), Jul. 2019, pp. 1–6.

[12] X. Xue-Feng and Z. Guo-Dong, ‘‘A survey on deep learning for naturallanguage processing,’’ Acta Autom. Sinica, vol. 42, no. 10, pp. 1445–1465,2016.

[13] J. Deng, L. Cheng, and Z.Wang,‘‘Attention-based BiLSTM fused CNNwith gating mechanism



model for Chinese long text classification,'' Comput.Speech Lang., vol. 68, Jul. 2021, Art. no. 101182.

[[14] J. Xie, Y. Hou, Y. Wang, Q. Wang, B. Li, S. Lv , and Y. I. Vorotnitsky,''Chinese text classification based on attention mechanism andfeature-enhanced fusion neural network,'' Computing, vol. 102, no. 3,pp. 683–700, Mar. 2020.

[15] T. Nguyen, Q.-T. Vien, and H. Sellahewa, ''An efficient pest classificationin smart agriculture using transfer learning,'' EAI Endorsed Trans. Ind.Netw. Intell. Syst., vol. 8, no. 26, Apr. 2021, Art. no. 168227.

[16] L. Hashemi-Beni and A. Gebrehiwot, ''Deep Learning for remote sensingimage classification for agriculture application,'' Int. Arch. Photogramm.,Remote Sens. Spatial Inf. Sci., vol. 2020, pp. 22–26, Jun. 2020.

[17] Y. Zhao, L. X. Hui, Z. Y. Cheng, and Z. Yue, ''A study on agricultural textclassification method based on naive Bayesian,'' Water Saving Irrigation,vol. 2, pp. 98–102, Feb.,2018.

[18] F.Wei, D. Qingling, and X. Xiaoyan, ''Classification technique of Chineseagricultural text information based on SVM,'' Trans. Chin. Soc. Agricult.Machinery, vol. 46, pp. 174–179, Dec. 2015.

[19] Y. C. Zhou, T. Y. Xu, and H. B. Deng, ''Agricultural text classificationmethod based on NB and CHI values,'' Jiangsu Agricult. Sci., vol. 46,no. 17, pp. 219–223, Sep. 2018.

[20] H. Y. Li, ''Text recognition and classificationof English teaching contentbased on SVM,'' J. Intell. Fuzzy Syst., Appl. Eng. Technol., vol. 39, no. 5,pp. 1–11, Aug. 2020.

[21] L. Ji, X. Cheng, and L. Kang, ''A SVM-based text classification system forknowledge organization method of crop cultivation,'' in Proc. Int. Conf.Comput. Comput. Technol. Agricult. Berlin, Germany: Springer, 2011,pp. 318–324.