

Ensemble Learning Based on GBDT and CNN for Adoptability Prediction

Yunfan Ye¹, Fang Liu^{1,*}, Shan Zhao², Wanting Hu³ and Zhiyao Liang⁴

Abstract: By efficiently and accurately predicting the adoptability of pets, shelters and rescuers can be positively guided on improving attraction of pet profiles, reducing animal suffering and euthanization. Previous prediction methods usually only used a single type of content for training. However, many pets contain not only textual content, but also images. To make full use of textual and visual information, this paper proposed a novel method to process pets that contain multimodal information. We employed several CNN (Convolutional Neural Network) based models and other methods to extract features from images and texts to obtain the initial multimodal representation, then reduce the dimensions and fuse them. Finally, we trained the fused features with two GBDT (Gradient Boosting Decision Tree) based models and a Neural Network (NN) and compare the performance of them and their ensemble. The evaluation result demonstrates that the proposed ensemble learning can improve the accuracy of prediction.

Keywords: Adoptability of pets, multimodal representation, CNN, GBDT, ensemble learning.

1 Introduction

Millions of stray animals suffer on the streets or are euthanized in shelters every day around the world. If homes can be found for them, many precious lives can be saved, and more happy families created. Therefore, it becomes a challenging and meaningful issue that how to make better use of various types of data, so that the adoptability can be predicted to help corresponding organizations make strategies to improve pet adoption rates and reduce animal suffering.

Started from the question “Can machines think?” first proposed by Turing, researchers have begun studying machine learning for a long time. During this period, many powerful algorithms have been proposed and improved, especially the convolutional neural networks (CNN) [Krizhevsky, Sutskever and Hinton (2012)], which achieved a state-of-the-art

¹ School of Design, Hunan University, Changsha, 410082, China.

² College of Computer, National University of Defense Technology, Changsha, 410073, China.

³ Canvard College, Beijing Technology and Business University, Beijing, 100037, China.

⁴ School of Science and Technology, Macau University, 999078, Macau.

* Corresponding Author: Fang Liu. Email: liufang06@gmail.com.

Received: 20 May 2020; Accepted: 08 June 2020.

performance. Based on advanced machine learning algorithms, many real-life problems can be elegantly solved. For example, an ensemble learning method for wireless multimedia device identification [Zhang, Li, Wang et al. (2018)], a medical expert system to assist doctors in diagnosis [Fang, Cai, Sun et al. (2018); Chen, Yu, Liu et al. (2019)], recommendation system constructed by amounts of user behavior and user data [Davidson, Liebold, Liu et al. (2010); Smith and Linden (2017)], face detection based on multitasking cascaded convolutional networks [Zhang, Zhang, Li et al. (2016)], face recognition [Parkhi, Vedaldi and Zisserman (2015)], Chinese question classification [Liu, Yang, Lv et al. (2019)], etc.

In general, two major factors are driving these applications to succeed, one is definitely data. When machine learning methods are performed to solve real problems, it will be much easier if the data is task-related, organized and abundant. For example, the ImageNet dataset [Deng, Dong, Socher et al. (2009)], which contains 1.2 million images of 1000 semantic classes, is most commonly used for general visual tasks. Apart from that, there are amounts of other famous datasets of different scenes, such as the Places-205 dataset [Zhou, Lapedriza, Xiao et al. (2014)] which contains various indoor and outdoor scenes, the Landmarks [Babenko, Slesarev, Chigorin et al. (2014)], the Product [Bell and Bala (2015)], etc.

The other one is an effective and suitable method that can capture complex connections between features and the target variable. Some of the features may not be distinguishing enough or even have an adverse impact when training. Thus, feature selection (also known as subset selection) is significant because it serves as a fundamental technique to reducing effects from noise or irrelevant variables and still provides good prediction results, offering us guidance about the efficiency and effectiveness of variables for a given classification model. Traditional feature selection methods may calculate the correlation between each feature and the target variable. [Liu, Cai, Xu et al. (2015); Gonzalez-Vidal, Jimenez and Gomez-Skarmeta (2019); Cilia, De Stefano, Fontanella et al. (2019); Tran, Xue and Zhang (2019); Yu, Cai, Wang et al. (2019)] cover a wide range of aspects about features, including feature construction, feature ranking, feature embedding, multivariate feature selection, etc. However, traditional methods mainly rely on artificially designed extractors for specific tasks; thus, its generalization ability and robustness are not satisfactory.

With the development of deep learning, we can now use various neural networks to automatically extract features, especially in the field of computer vision. In this paper, we are faced with a comprehensive task that contains general metadata, texts and images. It is apparent that a pet with a good appearance or a description of “healthy and active, skinny and tall, as the mother fully vaccinated, dewormed”, is more likely to be adopted. Therefore, unlike previous methods that only pay attention to a single kind of information, we fused textual and visual features and employ SVD (Singular Value Decomposition) for dimension reduction. Moreover, we added some new aggregation features through all numerical data and proposed an ensemble model consists of two methods based on GBDT [Friedman (2001)] and a specialized deep neural network. Comparing with each model, the proposed ELGC (Ensemble Learning based on GBDT and CNN) can improve the accuracy in the scene of adoptability prediction.

2 Related work

2.1 EDA and feature engineering

In general, when we are faced with a practical problem with large amounts of mixed clean data and noise, it is better to conduct EDA (Exploratory Data Analysis) to have a comprehensive understanding. As we mentioned above, feature engineering serves as a critical role and will directly affect the final performance. The method of feature engineering may also vary from the task. For example, novel feature engineering techniques for clinical text classification are presented in Garla et al. [Garla and Brandt (2012)], for credit card fraud detection in Bahnsen et al. [Bahnsen, Aouada, Stojanovic et al. (2016)] and for robust visual object tracking [Zhang, Jin, Sun et al. (2018)].

2.2 Pre-trained model

Fortunately, with the development of deep learning, a new direction has been aroused by CNN that obtained a state-of-the-art recognition accuracy. That is, we can directly extract the result of one convolutional layer as features. Although most machine learning algorithms are designed to solve a unique task, we can still transfer knowledge from pre-trained models to improve the learning of new problems. There is a survey on transfer learning here [Weiss, Khoshgoftaar and Wang (2016)]. Now using pre-trained models to reduce consumption and improve efficiency has become a prevalent method in most tasks. In the field of NLP (Natural Language Processing), pre-trained models are also used for solving large-vocabulary speech recognition problems in Dahl et al. [Dahl, Yu, Deng et al. (2011)] and for training high-quality word vector representations in Mikolov et al. [Mikolov, Grave, Bojanowski et al. (2017)]. In the field of CV (Computer Vision), it has also been proved that pre-trained CNN model is a boon for the image classification and recognition tasks, such as medical image analysis problems [Carneiro, Nascimento and Bradley (2015)] and food recognition [Singla, Yuan and Ebrahimi (2016)].

2.3 GBDT based algorithms

As for models, GBDT [Friedman (2001)] is widely used due to its efficiency, accuracy, and interpretability. Based on GBDT, LightGBM [Ke, Meng, Finley et al. (2017)] and XGBoost [Chen and Guestrin (2016)] have been proposed and soon become popular among researchers. For example, A novel cryptocurrency price trend forecasting model is built for guiding investors in constructing an appropriate cryptocurrency portfolio and mitigate risks through LightGBM in Sun [Sun, Liu and Sima (2018)], XGBoost is also used in various fields like DDoS Attack Detection and Analysis [Chen, Jiang, Cheng et al. (2018)], Rock Facies Classification [Zhang and Zhan (2017)], etc. Moreover, they are frequently conducted simultaneously to solve complete tasks and see which performs better, for example, in [Ma, Sha, Wang et al. (2018)], prediction of P2P network loan default was studied on both ways. In [Wang, Zhang and Zhao (2017)], LightGBM and XGBoost have both been performed and compared as miRNA Classification methods in Breast Cancer Patients, and the result demonstrates that in this specific situation, LightGBM was found better performing in several aspects.

3 Methods

In this paper, our proposed ELGC contains multimodal representation, which is constructed from images, natural languages, metadata and some created aggregation features. We employed SVD to reduce dimensions of those features, then LightGBM, XGBoost and NN to train them. In the end, we took the prediction results of the three models as features, performing ridge regression to predict the target variable and used Bayesian optimization and cross-validation to tune the parameters. Fig. 1 illustrates the whole structure of ensemble learning.

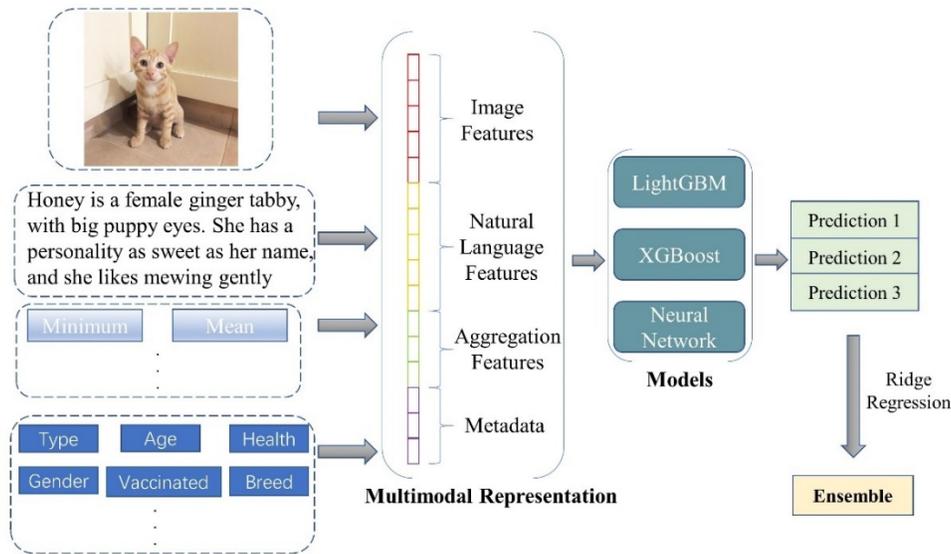


Figure 1: Structure of ensemble learning

3.1 Multimodal representation

For image data, we performed two pre-trained models to help us extract image features. One is pre-trained on ImageNet [Deng, Dong, Socher et al. (2009)] named Dense Convolutional Network (DenseNet) [Huang, Liu, Van Der Maaten et al. (2017)], which connects each layer to every other layer in a feed-forward fashion and can provide us effective general features. The other one comes from the best model of a Kaggle competition ‘cats vs. dogs’, thus features extracted can be of high quality since pets only include cats and dogs also in our task. Meanwhile, to ensure the quality and number of features, we employed SVD to reduce noise and improve predictive accuracy. SVD is an algorithm that decomposes an $m \times n$ matrix called A , into three component matrices, which can be represented as:

$$A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T \quad (1)$$

where U is an $m \times m$ matrix called left singular vector, Σ is an $m \times n$ diagonal matrix and V is an $n \times n$ matrix called right singular vector. Every element on the main diagonal of Σ is called a singular value, which is similar to the eigenvalue in eigenvalue decomposition. Since the singular value in singular value matrix is also arranged from large to small and decreases

particularly fast. That means, we can also use the largest k singular values and the corresponding left and right singular vectors to approximate the matrix, which is calculated as:

$$A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T \approx U_{m \times k} \Sigma_{k \times k} V_{k \times n}^T \quad (2)$$

where k can be much less than n. Therefore, we can adopt SVD to reduce features extracted from DenseNet121 and the best ‘cats vs. dogs’ model. In the end, a total of 64 new features are added according to image data. This process is shown in Fig. 2.

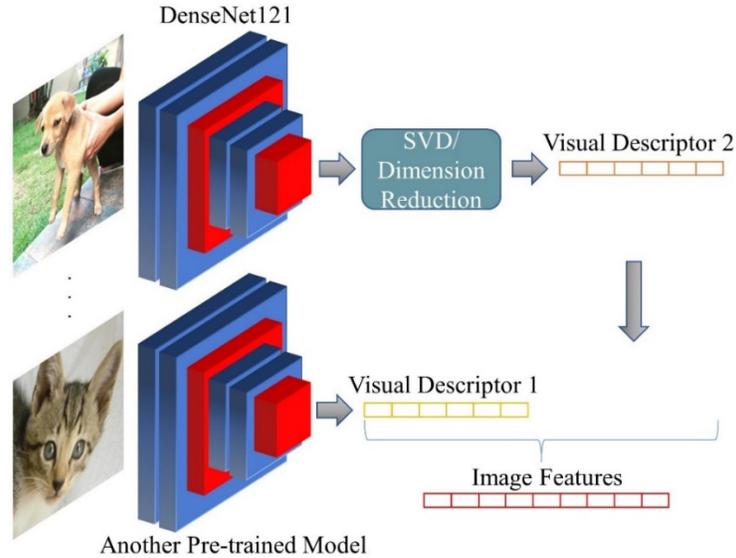


Figure 2: Image features

Similarly, for natural language data, we firstly combined all the descriptions of pets from different files and adopted TF-IDF to get important terms as features, then SVD to reduce dimensions. Term Frequency (TF) refers to how often a given word appears in a document. The importance of a term t_i in a specific document d_j can be given as follows:

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (3)$$

where $n_{i,j}$ is the frequency of a unique term t_i appeared in the document d_j , and the denominator is the sum of frequencies of all unique words in document d_j . However, TF always tend to give high value to commonly appeared but not important words such as “the”, “an” and “but”, without giving enough weight to the more meaningful and unique terms that can indeed represent the document. Therefore, Inverse Document Frequency (IDF) is incorporated which diminishes the weight of terms that occur very frequently in the document set and increases the weight of terms that occur rarely. The process is given by:

$$IDF_i = \lg \frac{|D|}{|\{j:t_i \in d_j\}|} \quad (4)$$

where $|D|$ means the total number of documents, the denominator means the number of documents that contains t_i . At last, the importance of terms can be calculated by the formula $TFIDF_{i,j} = TF_{i,j} \times IDF_i$. Moreover, we adopt a pre-trained word embedding model named wiki-news-300d-1M-subword [Mikolov, Grave, Bojanowski et al. (2017)],

which contains one million vector words with subwords information obtained from the 2017 Wikipedia, UMBC web-based corpus and statmt.org News dataset. At last, after padding them to sequences, the natural language part of feature engineering is accomplished. Fig. 3 illustrates the process.

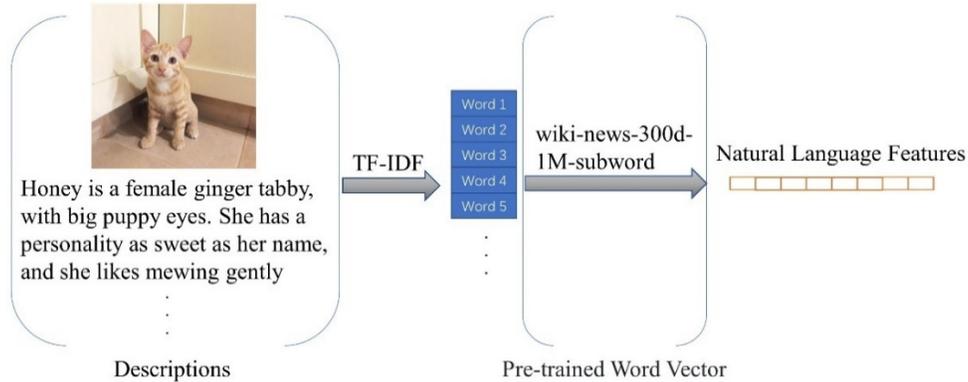


Figure 3: Natural language features

Meanwhile, we can create five new aggregation features that may contain some undiscovered information for each numeric feature such as age, fur length, etc. They are minimum, maximum, mean, standard deviation and variance. Since we cannot determine exactly what features are practical and efficient and what are not for our target variable before enough experiments, thus we can prepare as many features as possible and then discard unimportant features after subsequent analysis. The process of creating new aggregation features is shown in Fig. 4.

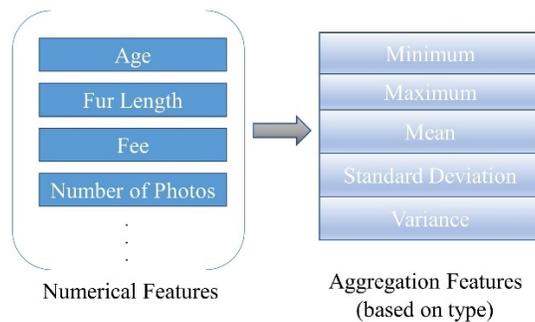


Figure 4: New aggregation features

At last, we simply employed LightGBM to get a baseline, which is a strong classifier composed of several weak classifiers. Since it is also an algorithm based on the decision tree, we can get the importance of each feature and discard those whose importance is zero or very low for keeping retained features of high relevance.

3.2 Model construction and ensemble learning

We have constructed three independent learners, which are LightGBM, XGBoost and NN, together with an ensemble model. The former two methods are improved from GBDT; they

are all based on the decision tree to form a strong classifier by many weak classifiers. In the iteration of GBDT, suppose that the strong classifier learned in the former iteration is $f_{t-1}(x)$, the loss function is $L(y, f_{t-1}(x))$, then our goal is to find out a weak classifier of classification and regression tree to make the loss function of this iteration minimum, which is $L(y, f_t(x) = L(y, f_{t-1}(x) + h_t(x)))$. XGBoost has added regularized learning objectives and used a greedy algorithm to get a sub-optimal tree structure, whereas LightGBM has made some improvements specifically for the case of high feature dimension and large data size, making efficiency and scalability even more satisfactory.

Besides, to add some diversity except decision tree for ensemble learning, we also used part of features to train a neural network including convolutional layers, pooling layers, activation functions, dropout layers, etc. The convolutional layer is the core building block of a CNN, where the parameters consist of several learnable filters. Each filter is convolved across the same volume from the input feature map, computing the dot product and outputting a new one. The pooling layer is simply a block of nonlinear down-sampling. There are several functions such as average pooling and max pooling. For the most commonly used max pooling, the input image will be divided into a set of non-overlapping rectangles, and for each rectangle, replace the whole sub-region by the maximum in it. The activation functions are introduced to add some nonlinear factors for better speed and performance of fitting with relatively fewer layers and nodes. There are lots of activation functions such as $Sigmoid(x) = \frac{1}{1+e^{-x}}$ and $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, in which number e is a mathematical constant approximately equal to 2.71828 and is the base of the natural logarithm. Combining various layers and feature extraction methods, we proposed a neural network that fused numerical features, textual features, etc. The structure is illustrated in Fig. 5.

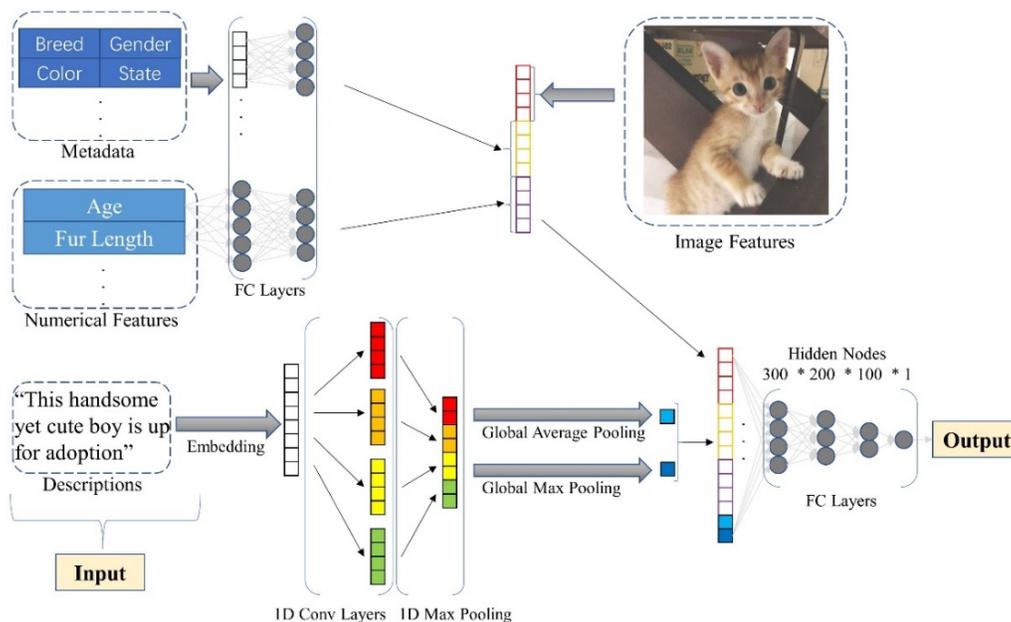


Figure 5: Structure of the neural network

Moreover, we took the prediction results of the above three models as new inputs, performing the ridge regression to predict the target variable as we mentioned before. It is aiming to fit a set of data $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ by $f(x) = w^T x + b$ and make the loss function minimum, which can be represented as:

$$J = \frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2 = \frac{1}{n} \sum_{i=1}^n (w^T x_i + b - y_i)^2 \quad (5)$$

where L2 regularization is added by ridge regression to optimize the problem of overfitting, the loss function is calculated by:

$$J = \frac{1}{n} \sum_{i=1}^n (f(x) - y_i)^2 + \lambda \|w\|_2^2 \quad (6)$$

4 Experiments

4.1 Dataset and evaluation criteria

Our dataset comes from the competition of PetFinder.my Adoption Prediction in Kaggle, which contains about 60,000 images and metadata for training, about 15,000 images and metadata for testing.

About the evaluation criteria, the cost of error must frequently be considered in some cost-sensitive situations like medical and military applications. Therefore, a new methodology named Quadratic Weighted Kappa is proposed for assessing the classifier's accuracy based on Cohen's Kappa statistic in Ben-David [Ben-David (2008)], which is also adopted in our paper for better and strict evaluation.

4.2 Exploratory data analysis

To have an overall preliminary understanding of data, we should first perform some EDA. The target variable in our task is adoption speed which has five levels. From 0 to 4, it means that the time taken for pets to be adopted is getting longer. Zero means pet was adopted on the same day as it was listed and four means no adoption even after 100 days of being listed. Tab. 1 shows the distribution of adoption speed and the numbers of cats and dogs in train and test data are presented in Tab. 2.

Table 1: Distribution of adoption speed

Adoption speed	0	1	2	3	4
Count	409	3090	4038	3259	4197
Percentage	2.73%	20.61%	26.93%	21.74%	27.99%

In addition, we have many categorical features such as Pet Type, Breed, Maturity Size and some numerical features like Fur Length, Age, adoption Fee, PhotoAmt (The number of photos appeared in the advertisement), etc. Besides, there are some NLP and CV data like descriptions and photos of pets whose features should be extracted through various methods. Different features have different predictive powers for the target variable, we can even explore several new features and discard those with low importance.

Table 2: Number of cats and dogs in train and test data

Type	Dog (train)	Cat (train)	Dog (test)	Cat (test)
Count	8132	6861	2100	1872
Percentage	54.24%	45.76%	52.87%	47.13%

4.3 Result and analysis

Before the final experiments, LightGBM is employed to discard features whose importance is very low, thus the retained features can be more effective and relevant. Fig. 6 demonstrates that breed1, age and many image features serve as a more critical role, and surprisingly, our created aggregation features such as the mean of “PhotoAmt”, also achieve miraculous effects for our final predictions.

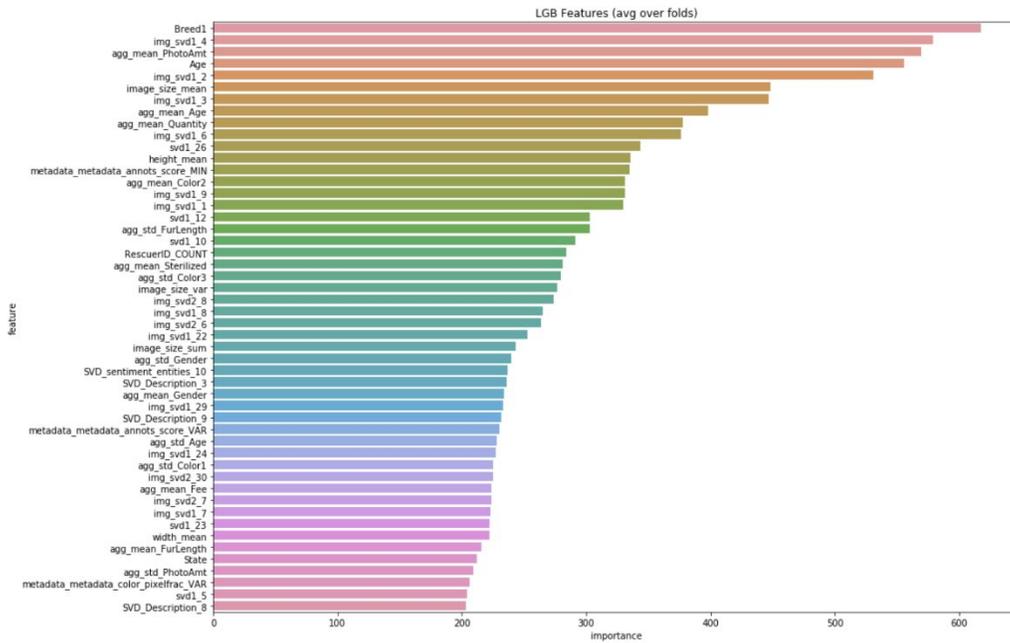


Figure 6: Importance of features

As shown in Tab. 3, when LightGBM, XGBoost, and NN are employed alone, LightGBM performs slightly better than XGBoost, but both of them are acceptable and better than NN. We can see that though the score of NN is relatively low, the ensemble model learning through ridge regression performs the best.

Table 3: Quadratic weighted kappa scores

Criteria	XGBoost	LightGBM	NN	ELGC
Quadratic weighted kappa scores	0.4381	0.4389	0.3536	0.4490

5 Conclusion

With the global trend to consider pets as part of the family, studies about pet adoption have gradually become a research hotspot. In this paper, we firstly performed some EDA and preprocess on our raw data. For images, we adopted two pre-trained models to help extract visual descriptors. For natural languages, we employed TF-IDF to choose keywords from descriptions of pets and used word vectors to get embedding textual features. SVD is performed to reduce dimensions and improve the quality of descriptors. Moreover, based on existing numerical data, some aggregation descriptors are created which are proved effective and critical in this task.

To generate diverse learners for ensemble learning, we took various training strategies and compared the performances of three independent models together with their ensemble. The experimental results show that the ELGC has good classification performance. We can conclude that when a pre-trained model was trained to solve a problem similar to ours, transfer learning is a worthwhile method for feature extraction. Besides, complementarity is more important than pure accuracy, even if the performance of each model is weak like NN here, ensemble learning can combine their diversities and is more likely to improve the accuracy.

Funding Statement: This work is supported by The National Key Research and Development Program of China (2018YFB1800202, 2016YFB1000302, SQ2019ZD090149, 2018YFB0204301).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- Babenko, A.; Slesarev, A.; Chigorin, A.; Lempitsky, V.** (2014): Neural codes for image retrieval. *European Conference on Computer Vision*, pp. 584-599.
- Bahnsen, A. C.; Aouada, D.; Stojanovic, A.; Ottersten, B.** (2016): Feature engineering strategies for credit card fraud detection. *Expert Systems with Applications*, vol. 51, pp. 134-142.
- Bell, S.; Bala, K.** (2015): Learning visual similarity for product design with convolutional neural networks. *ACM Transactions on Graphics*, vol. 34, no. 4, pp. 1-10.
- Ben-David, A.** (2008): Comparison of classification accuracy using cohen's weighted kappa. *Expert Systems with Applications*, vol. 34, no. 2, pp. 825-832.
- Carneiro, G.; Nascimento, J.; Bradley, A. P.** (2015): Unregistered multiview mammogram analysis with pre-trained deep learning models. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 652-660.
- Chen, H.; Yu, J.; Liu, F.; Cai, Z.; Xia, J.** (2019): Archipelago: a medical distributed storage system for interconnected health. *IEEE Internet Computing*, vol. 24, no. 2, pp. 28-38.
- Chen, T.; Guestrin, C.** (2016): Xgboost: a scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785-794.

Chen, Z.; Jiang, F.; Cheng, Y.; Gu, X.; Liu, W. et al. (2018): XGBoost classifier for DDoS attack detection and analysis in SDN-based cloud. *IEEE International Conference on Big Data and Smart Computing*, pp. 251-256.

Cilia, N. D.; De Stefano, C.; Fontanella, F.; di Freca, A. S. (2019): A ranking-based feature selection approach for handwritten character recognition. *Pattern Recognition Letters*, vol. 121, pp. 77-86.

Dahl, G. E.; Yu, D.; Deng, L.; Acero, A. (2011): Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 30-42.

Davidson, J.; Liebald, B.; Liu, J.; Nandy, P.; Van Vleet, T. et al. (2010): The YouTube video recommendation system. *Proceedings of the Fourth ACM Conference on Recommender Systems*, pp. 293-296.

Deng, J.; Dong, W.; Socher, R.; Li, L. J.; Li, K. et al. (2009): Imagenet: a large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248-255.

Fang, S.; Cai, Z.; Sun, W.; Liu, A.; Liu, F. et al. (2018): Feature selection method based on class discriminative degree for intelligent medical diagnosis. *Computers, Materials & Continua*, vol. 55, no. 3, pp. 419-433.

Friedman, J. H. (2001): Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pp. 1189-1232.

Garla, V. N.; Brandt, C. (2012): Ontology-guided feature engineering for clinical text classification. *Journal of Biomedical Informatics*, vol. 45, no. 5, pp. 992-998.

Gonzalez-Vidal, A.; Jimenez, F.; Gomez-Skarmeta, A. F. (2019): A methodology for energy multivariate time series forecasting in smart buildings based on feature selection. *Energy and Buildings*, vol. 196, pp. 71-82.

Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K. Q. (2017): Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700-4708.

Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W. et al. (2017): Lightgbm: a highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, pp. 3146-3154.

Krizhevsky, A.; Sutskever, I.; Hinton, G. E. (2012): Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, pp. 1097-1105.

Liu, J.; Yang, Y.; Lv, S.; Wang, J.; Chen, H. (2019): Attention-based BiGRU-CNN for Chinese question classification. *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-12.

Liu, S.; Cai, Z.; Xu, H.; Xu, M. (2015): Towards security-aware virtual network embedding. *Computer Networks*, vol. 91, pp. 151-163.

Ma, X.; Sha, J.; Wang, D.; Yu, Y.; Yang, Q. et al. (2018): Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms

according to different high dimensional data cleaning. *Electronic Commerce Research and Applications*, vol. 31, pp. 24-39.

Mikolov, T.; Grave, E.; Bojanowski, P.; Puhersch, C.; Joulin, A. (2017): Advances in pre-training distributed word representations. arXiv Preprint arXiv:1712.09405.

Parkhi, O. M.; Vedaldi, A.; Zisserman, A. (2015): Deep face recognition. *British Machine Vision Conference*, vol. 1, no. 3, pp. 6.

Singla, A.; Yuan, L.; Ebrahimi, T. (2016): Food/non-food image classification and food categorization using pre-trained googlenet model. *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, pp. 3-11.

Smith, B.; Linden, G. (2017): Two decades of recommender systems at Amazon. com. *IEEE Internet Computing*, vol. 21, no. 3, pp. 12-18.

Sun, X.; Liu, M.; Sima, Z. (2018): A novel cryptocurrency price trend forecasting model based on LightGBM. *Finance Research Letters*, vol. 32, 101084.

Tran, B.; Xue, B.; Zhang, M. (2019): Genetic programming for multiple-feature construction on high-dimensional classification. *Pattern Recognition*, vol. 93, pp. 404-417.

Wang, D.; Zhang, Y.; Zhao, Y. (2017): LightGBM: an effective miRNA classification method in breast cancer patients. *Proceedings of the International Conference on Computational Biology and Bioinformatics*, pp. 7-11.

Weiss, K.; Khoshgoftaar, T. M.; Wang, D. (2016): A survey of transfer learning. *Journal of Big data*, vol. 3, no. 1, pp. 9.

Yu, G.; Cai, Z.; Wang, S.; Chen, H.; Liu, F. et al. (2019): Unsupervised online anomaly detection with parameter adaptation for KPI abrupt changes. *IEEE Transactions on Network and Service Management*.

Zhang, J.; Jin, X.; Sun, J.; Wang, J.; Sangaiah, A. K. (2018): Spatial and semantic convolutional features for robust visual object tracking. *Multimedia Tools and Applications*, pp. 1-21.

Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. (2016): Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499-1503.

Zhang, L.; Zhan, C. (2017): Machine learning in rock facies classification: an application of XGBoost. *International Geophysical Conference, Qingdao, China*, pp. 1371-1374.

Zhang, Z.; Li, Y.; Wang, C.; Wang, M.; Tu, Y. et al. (2018): An ensemble learning method for wireless multimedia device identification. *Security and Communication Networks*, vol. 2018.

Zhou, B.; Lapedriza, A.; Xiao, J.; Torralba, A.; Oliva, A. (2014): Learning deep features for scene recognition using places database. *Advances in Neural Information Processing Systems*, pp. 487-495.