Comparison of the Performance of Pegasos and Traditional Models in the Task of Sentiment Classification of Product Reviews

Di Zhu

College of Information Engineering, China University of Geosciences, Beijing, China

Abstract. Sentiment classification of a large number of commodity reviews is very important for customer selection and market trend prediction. The key to achieving high accuracy of sentiment classification is to select appropriate models for training. However, most of the existing research literature only uses traditional models or does not have a scientific comparison between models, and only provides training steps for a model that can be used for text sentiment analysis. Therefore, in this study, three models (two traditional models and one new model Pegasos) were selected for comparison. The same datasets and data preprocessing methods were selected for the three models to compare the final emotion classification accuracy, and finally, the experimental result with the highest accuracy of the Pegasos model was obtained. The study also analyzed the performance of different models in the experimental results and found the reasons for the poor performance of traditional models in handling sentiment classification tasks and the reasons for the good performance of Pegasos models. In the process of model hyperparameter tuning, the optimal number of iterations is 25 by comparing the classification accuracy under different iterations.

1 Introduction

With the rapid development of Internet technology, online shopping has become very popular. In recent years, artificial intelligence technology has also made online shopping more intelligent. Through deep learning technology to classify the sentiment of commodity reviews, high-quality goods can be better selected for customers. Therefore, accurately identifying and classifying positive and negative product reviews has become a key problem, and the key to solving this problem is to choose a suitable training model.

At present, there has been a lot of relevant research on text sentiment classification technology, and there are many different implementation methods, but the performance of different methods is different. To put into the application, there is not too much comparison. In the recent research on Amazon sentiment classification, only the implementation steps are introduced, but the superiority of its model is not reflected [1]. In most literature, only one or two common models are used for training, and the accuracy after training is not compared

Corresponding author: 1004211109email.cugb.edu.cn

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

too much, or three novel models are selected, but they are not compared with traditional models, and the advantages of the new models are not shown.

In Kim's research, the CNN model was trained by using movie review datasets, and the accuracy of sentiment classification for new reviews was 86%, but no comparison was made [2]. In Eman's research, the Bayes model and SVM model were trained by using product review datasets, in the end, the review was used to judge whether the product was recommended. The accuracy of both models was 87% [3]. In Munikar's research, three novel models, Beyas, SnowNLP, and Bert are trained by using the tourist evaluation datasets of tourist attractions. It is found that the Bert model has the highest accuracy of emotion analysis, which is about 90% [4]. To sum up, in most literature the comparison was of little significance and there is no recommended model. Even though some studies made comparisons, these studies did not show the superiority of the new model.

This research can be applied in many areas: improving purchasing efficiency, conducting market research and trend analysis, and helping researchers choose suitable models.

In this paper, it trained the Perceptron, Average Perceptron, and Pegasos models respectively, and compared the classification accuracy of the three models. Finally, it showed the advantages of the Pegasos model in the field of text sentiment classification and revealed the shortcomings of traditional models by comparing each model.

2 Method

2.1. Datasets

In this project, two product review datasets on the datasets website Kaggle are selected, Amazon Fine Food Reviews (which contain nearly 500,000 reviews on food on Amazon, including rating, review time, user ID, product ID, and other information). Amazon Product Reviews (reviews of multiple Amazon product categories, such as electronics, books, household goods, etc.) The data will be labeled as +1 and -1 after the model processing, with +1 representing positive comments and -1 representing negative comments.

Table 1 is two sample reviews from the dataset, each describing a customer's experience with sugar-free candy:

Table 1. Contrasting customer reviews for a candy product

	comment	label
Negative comment	Nasty No flavor. The candy is just red, with no flavor. Just plan and chewy. I would never buy them again.	-1
Positive comment	YUMMY! You would never guess that they're sugar-free and it's so great that you can eat them pretty much guilt-free! I was so impressed that i've ordered some for myself to take to the office. These are just EXCELLENT!	+1

2.2. Data pre-processing

First, the toy datasets are loaded as features and labels, then the features are converted into a two-dimensional matrix, and then transposed so that each row of the matrix corresponds to a sample for easy reading by the model. According to the text data of the training set, the dictionary of the bag of words model is constructed. Each dictionary represents a record, including two keys of text and emotion, corresponding to the text content and emotion label respectively. It can implement word-to-index mapping in text, representing all words that appear in the training set. Removing the stop words and regenerating the bag of words model help improve the efficiency and classification accuracy of the following steps. The next step is to traverse each sample in the training set, extract the text and corresponding emotion labels from those samples, and then extract the resulting text and labels into tuples containing all the text and tuples containing all the labels, respectively. Validation sets and test sets are used in the same way to separate text and labels for input to subsequent classifiers.

2.3 Training models

In this research, three models are chosen, including two traditional models, the Perceptron model, the Average Perceptron model, and a new model based on the SVM model called Pegasos.

2.3.1 Perceptron

Perceptron is a simple linear classifier that is mostly used for classification tasks because the model's prediction output is in the form of +1 and -1, which is binary classification, so it adopts this traditional model. The perceptron iterates over the training data and adjusts the weights to try to find the best hyperplane to classify the data points correctly. Specific model building process: the weight θ and bias θ0 are initialized as 0. For each training sample, the predicted value is calculated.

$$y_{pred} = \mathrm{sign}(\theta * x + \theta_0) \tag{1}$$

If the predicted value is inconsistent with the real label, a prediction error occurs, and the weight and bias are updated.

$$\theta = \theta + y * x , \theta_0 = \theta_0 + y \tag{2}$$

After iterating T (T is a hyperparameter, representing the number of iterations) times, return the final θ and θ0. However, this frequent updating of parameters will lead to overfitting of the model, resulting in a decline in accuracy.

2.3.2 Average perceptron

Average Perceptron is an improved version of Perceptron. The average Perceptron takes the average value of all θ and θ0 generated during training as the weight and bias of the final model. Compared with the ordinary perceptron model, the average perceptron can reduce the volatility of the model and has more stable performance. The specific implementation steps are similar to the perceptron model, except that after updating the weights and biases each time, they are added up, and the average value is calculated as the parameter of the final model.

2.3.3 Pegasos

Pegasos is an efficient gradient descent algorithm for linear support vector machines (LSVMS). It can find the hyperplane of classification by using the minimum regularization loss function, and it has a good performance in processing large-scale datasets. Its training time is linear to the number of samples, which means that Pegasos can process large amounts of comment data quickly compared to perceptron models. Specific steps of model construction: initialize the weights and biases. For each iteration, randomly extract a sample (x,y), calculate the subgradient, and update the weights according to the learning rate.

$$\theta = \theta * (1 - \eta * \lambda) + \eta * y * x \, , \; \theta_0 = \theta_0 + \eta * y \tag{3}$$

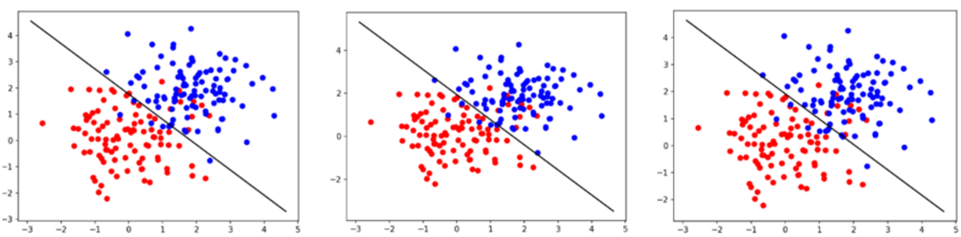
After iterating T times, return the final θ and θ0. With the unique parameter updating mechanism of the Pegasos algorithm, it can balance the training speed and accuracy of the model, so it can handle the binary classification task well.

To make the comparison between the three models more scientific, hyperparameter tuning is carried out on all three models.

3 Result

3.1. Lattice diagram of different model's classification

Fig. 1 is the lattice diagrams of the model results after training through the same data preprocessing method. In the figure, the horizontal and vertical coordinates represent the degree to which the model classifies positive and negative emotions. The smaller the value, the greater the negative emotions of the comments; The larger the value, the greater the positive emotions of the comments.



(a) (b) (c)

Fig. 1 Classified result of different model, (a) Classified Toy Data (Perceptron) (b) Classified Toy Data(Average Perceptron) (c) Classified Toy Data(Pegasos) (Photo/Picture credit: Original).

It can be seen that the three models perform well in the task of toy commodity emotion classification, and the Pegasos model is the best. Different from perceptron and average perceptron models, Pegasos is an SVM-based optimization algorithm that uses regularization technology to effectively control model complexity and prevent overfitting. For data sets with noisy and high-dimensional features such as sentiment classification of product reviews, regularization can help the model better generalize to previously unseen data.

3.2 Weights and biases of each model

Table 2 shows three different models' θ and θ0, because the input data includes both x and y, two values of θ represent x's and y's weight, and θ0 represents bias.

Table 2. θ and θ0 of different models

	θ	θ_0
Perceptron	3.917, 4.164	-8.0
Avg Perceptron	3.478, 3.611	-6.373
Pegasos	0.735, 0.630	-1.220

Table 3 shows three different models' training accuracy and validation accuracy

Table 3. Training accuracy and validation accuracy of different models

	Training accuracy	Validation accuracy
perceptron	0.9157	0.7160
average perceptron	0.9543	0.7900
Pegasos	0.9728	0.7980

From Table 2 and Table 3, the training results of various models can be visually seen and the following conclusions can be drawn:

3.2.1 Perceptron

The large θ value of Perceptron in the table indicates that the model exerts a large weight on the input features, which means that the model is more dependent on the quality of the input data, and the input features have a significant impact on the final classification results. The bias term is negative, indicating that the decision boundary of the model is more inclined to the negative class.

3.2.2 Average perceptron

The training accuracy of the average perceptron model was significantly improved compared with that of the Perceptron model, which indicated that the average perceptron enhanced the stability of the model by averaging the weights of multiple perceptrons, and the verification accuracy was also improved to some extent, and the generalization ability of new data was better. However, both accuracy rates are still lower than the Pegasos model.

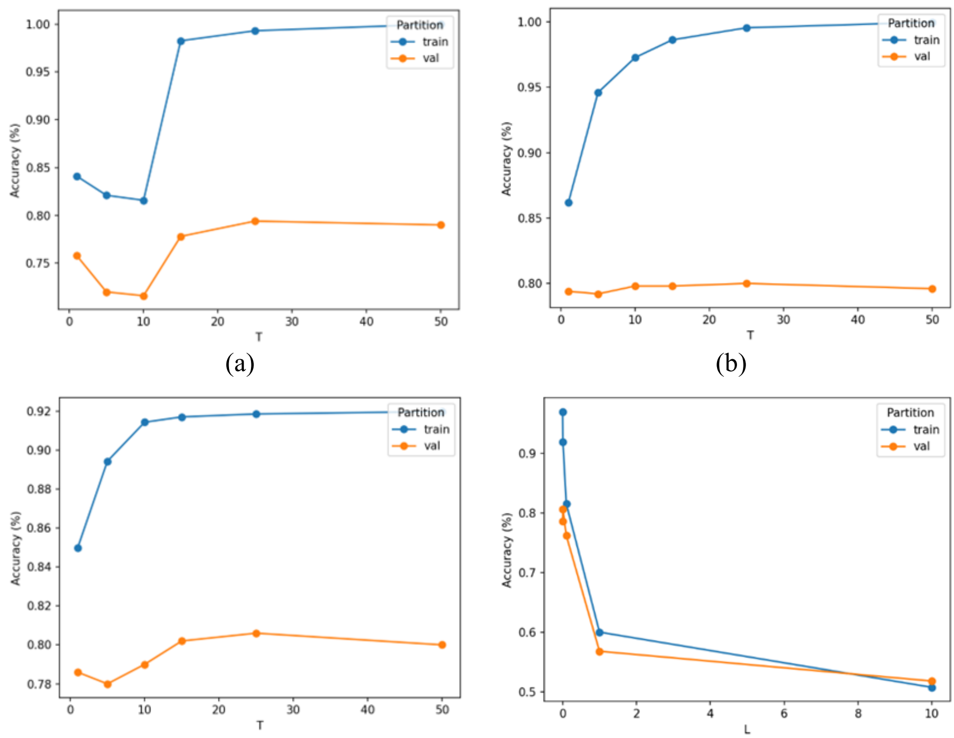
3.2.3 Pegasos

The training accuracy of the Pegasos model is as high as 97%, which is significantly improved compared with the other two models, and the verification accuracy is also the highest among the three models, indicating that the Pegasos model has the strongest generalization ability on new data and better processing in terms of overfitting.

3.3 Classification accuracy vs T & L

Fig. 2 shows the relationship between the classification accuracy of different models and the number of iterations (or regularization parameter.) The horizontal coordinate of the figures represents the number of iteration times (or regularization parameter times), and the vertical coordinate represents the classification accuracy.

In these figures, T represents the number of iterations, L represents the regularization parameter, and L is only valid for Pegasos, which is used to control the complexity of the model and prevent overfitting.



(a) (b)

(c) (d)

Fig. 2 The relationship between classification accuracy and T(or L) of different model, (a) Classification Accuracy vs T (Perceptron) (b) Classification Accuracy vs T (Avg Perceptron), (c) Classification Accuracy vs T (Pegasos) (d) Classification Accuracy vs L (Pegasos) (Photo/Picture credit: Original).

From Fig. 2 (c), when tuning L, Pegasos finds that the optimal value is 0.01, which shows the importance of regularization in this problem. The lower value of L helps the model avoid overfitting better.

Table 4. The verification accuracy of different models with different iterations during the tuning process.

	Perceptron	Avg Perceptron	Pegasos
T=1	0.758	0.794	0.786
T=5	0.72	0.792	0.78
T=10	0.716	0.798	0.79
T=15	0.788	0.798	0.802
T=25	0.794	0.8	0.806
T=50	0.79	0.796	0.8

From Table 4 and Fig. 2, both the Average Perceptron and Pegasos models have higher accuracy on the verification set than Perceptron, and Pegasos has better performance. The optimal number of iterations is 25 for all models, indicating that the training stability and model performance are optimal under this setting. When the number of iterations is 25, the model with the highest verification accuracy is the Pegasos model, which reaches 0.806.

4 Discussion

It can be seen from the experimental results that the three models have obtained good experimental results, among which the Pegasos model has the highest accuracy of comprehensive sentiment classification.

4.1 Perceptron

From Table 1, the Perceptron model has the largest weight, indicating that the classification accuracy of this model is greatly affected by the input text. The efficiency of the Perceptron model may not be high when processing a large number of product review texts, because Perceptron is essentially a linear classifier. A hyperplane is calculated to achieve data classification. However, the task of emotion classification is not completely a linear divisible task, and emotion also has a degree. Perceptron cannot accurately classify non-linearly divisible data. Product review data is a kind of data with a lot of noise points, and ambiguous words will likely cause the wrong marking of the data. Perceptron constantly adjusts the weights through iteration, but the iteration speed of the model is very slow. The high noise and large amount of data not only greatly affects the accuracy, but also results in the low processing speed of the model, which consumes a lot of resources. It can be seen from Table 2 that the validation accuracy of the Perceptron model is the lowest, which is much lower than its training set. According to the iterative mechanism of the model, it may be that the model lacks the bustling ability, overfits the noise of the training set data during the training process, and only focuses on minimizing the classification errors on the data set, which leads to the situation that the training accuracy rate is good and the verification accuracy rate is greatly decreased.

The main problem of the Perceptron model is that the iterative convergence speed is too slow and overfitting often occurs, so the Avg Perceptron model is improved.

4.2 Avg perceptron

From Table 1 and Table 2, the weight, validation accuracy, and training accuracy of the Avg Perceptron model are higher than those of the Perceptron model. It solves the problem of slow convergence to a certain extent. By adding the weights of each iteration and calculating the average value, the weight image of a single noise point is reduced. Faster to find a better classification hyperplane, making the model more stable. Therefore, the model is more stable and has better generalization ability, and there is no overfitting phenomenon like the Perceptron model. It can better maintain classification accuracy when processing new data.

However, the average Perceptron model greatly increases the computational complexity due to the need to update the weights every iteration, especially when processing hundreds of thousands of data such as comments, the processing time is even longer than that of the ordinary Perceptron model, which limits the application of the model in real scenarios. According to Montanari's research, both the Perceptron model and the Avg Perceptron model make it difficult to achieve high accuracy when dealing with such nonlinear and high-noise tasks [5]. However, Roget proposed a method to reduce images of high-noise data. Grover search method was used to make a second improvement on the Perceptron model or Avg Perceptron model, and a quantum machine learning algorithm was used to calculate the error limit generated by the prediction, adjust the prediction results, and reduce the impact of high noise on the image of the training process. In this way, the accuracy of the model can be improved [6].

4.3 Pegasos

According to the experimental results, the Pegasos model achieved the best experimental results. Compared with the Perceptron model, Pegasos no longer updated all samples by traversal, but adopted stochastic gradient descent, and only selected one sample each time to update the model weight, which made the algorithm very efficient in processing large-scale data. In addition, in the process of hyperparameter tuning, not only the optimal number of iterations is obtained through multiple iterations, but also regularization processing is introduced, which makes the model more robust to noise and outliers and prevents the overfitting phenomenon to a certain extent. Since the text data is characterized by high dimensionality and sparsity (the total amount of words in the text is huge, but only a small part of words appear in each sample), the model is more inclined to produce sparse weight vectors, which can accurately identify important words in the text, so that the classification process is not interfered by irrelevant words. However, in the training process, the Pegasos model does not show the theoretical training speed and is not much different from the other two models [7-10]. This may be because only a small part of the datasets used for training are the toy product review datasets, and the datasample base is not large enough. Optimizing the computing mode to parallel computing mode will improve the training speed.

5 Conclusion

In this study, the accuracy of the Pegasos algorithm model and the two traditional models in processing the sentiment classification task of commodity reviews was systematically compared. The regularization processing and gradient descent learning algorithm was introduced into the algorithm model, which increased the generalization ability of the model and performed better in processing the emotion classification task of text. The results show that the Pegasos model has the highest comprehensive classification accuracy, both the speed of the model in processing large-scale data and the ability to reduce the degree of overfitting phenomenon exceed the other two models. Regularization and gradient descent learning algorithms are introduced in the model, which increases the generalization ability of the model and performs well in the task of text emotion classification. However, in product reviews, positive reviews are filled with a lot of false propaganda information. In the future, the Pegasos algorithm model can also adjust model parameters to make it pay more attention to negative evaluation characteristics, improve its ability to recognize negative emotions and make the final classification results more authentic.

References

1. M. Ohammad Abu Kausar, S. Allam Osman Fageeri, A. Rockiasamy Soosaimanickam, Sentiment Classification based on Machine Learning Approaches in Amazon Product Reviews (2023)
2. H. Annah Kim, Y. Oung-Seob Jeong, Sentiment Classification Using Convolutional Neural Networks, Appl. Sci. (2019)
3. E. Man Saeed Alamoudi, N. Orah Saleh Alghamdi, Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings (2021)
4. M. Anish Munikar, S. Ushil Shakya, A. Akash Shrestha, Fine-grained Sentiment Classification using BERT, AITB (2019)
5. A. Ndrea Montanari, Y. Iqiao Zhong, K. Angjie Zhou, Tractability from overparametrization: the example of the negative perceptron (2024)
6. M. Athieu Roget, G. Iuseppe Di Molfetta, H. Achem Kadri, Quantum Perceptron

Revisited: Computational-Statistical Tradeoffs (2021)
7. N. Srebro, S. Hai Shalev-Shwartz, Y. Singer, A.Ndrew Cotter, Pegasos: primal estimated sub-gradient solver for SVM (2007)
8. Xie, P., Huang, S. Sentiment Analysis on Product Reviews: An Empirical Study of Textual Data from Amazon. Journal of Data and Information Science, 5(2), 60-78. (2022)
9. M. Arylou Gabrié, S. Urya Ganguli, C. Arlo Lucibello, R. Iccardo Zecchina, Neural networks: from the perceptron to deep nets, Disordered Systems and Neural Networks, Statistical Mechanics (2023)
10. D. Evi Hawana Lubis, S. Sawaluddin, A. De Candra, Machine Learning Model for Language Classification: Bag-of-words and Multilayer Perceptron, Journal of Informatics and Telecommunication Engineering (2023)