Application, Information System
and Software Development Journal
Appissode Journal
Vol. 1, No. 1, December 2023, Page 15-19
E-ISSN: 3032-5625, P-ISSN: 3032-5269

Emotion Classification Using Support Vector Machine

Rully Ramanda, Muhammad Affandes
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

Corresponding Author: Muhammad Affandes
Informatics Engineering Department
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
Email: affandes@uin-suska.ac.id

Article History

Received Nov 27, 2023
Accepted Dec 9, 2023
Published Dec, 2023

Keywords

Emotions, support vector machine, text mining, data mining, machine learning

Abstract

Emotions have always been significant in how people connect with one another in daily life. This feeling has recently gained importance in human-computer interactions as well. Emotions may be expressed by humans through writing, voice, and facial expressions. Text-to-emotion recognition is a classification task with predefined emotion labels. This study uses the following seven categories of emotions: joy, anger, sadness, fear, disgust, shame and guilt. These categories are drawn from the ISEAR (International Survey on Emotion Antecedent and Reaction) dataset, which consists of 7666 original lines of English sentences with emotion labels. To determine which of the three support vector machine technique kernels—linear, RBF, and polynomial—performed best for text categorization, a comparison of the kernels was also conducted. Rotating models that are created from the outcomes of training on training data employ a variety of metrics. 61.3% of the linear kernel with parameter C = 0.5, 60.3% of the RBF kernel with parameter C = 1 and $\gamma = 2$, and 57.7% of the polynomial kernel with parameter C = 5, $\gamma = 0.8$, and degree = 2 were the accuracy values obtained for each kernel based on the test results. It has been demonstrated that the linear kernel performs better in text categorization than other kernels.

1 INTRODUCTION

Due to their ease of use and technological advancements, computers are now an essential component of daily life. Increased research in the field of human-computer interaction (IMK) results from human dependency on computers. The goal of the scientific discipline of human and computer interaction is to make computers more useful and easier to use for people, therefore fostering better human-computer interaction. Human-computer interaction is used in many fields, from public opinion analysis to online content classification. Sentiment analysis and emotion analysis are two distinct research areas that are crucial to human-computer interaction. People naturally want to interact more socially and meaningfully with computers, just as they do with other people[1].

A unique idea and sensation, a biological and psychological condition, and a collection of inclinations to behave are all referred to as emotions. In essence, emotions are the desire to take action[2]. Numerous scientific fields, including psychology, sociology, philosophy, and even economics, are centered around emotions. Scientists have researched categorizing emotions extensively. Famed psychologist listed the following feelings: fear, contempt, anger, joy, and sorrow. These are regarded as the six fundamental human emotions [3].

Textual information is frequently utilized to communicate ideas or emotional states in addition to explaining events or facts, making it an intriguing topic to study how to recognize emotions from texts. Because of this, sentiment and emotion analysis can greatly benefit from texts. Sentiment analysis focuses on categorizing text as positive or negative, whereas emotion analysis identifies specific feelings that are represented in a text.

The Support Vector Machine (SVM) technique, first presented by Vladimir Vapnik, is a popular choice for text categorization issues[4]. SVM is the third most often used classification technique, after C4.5 and K-Means[5]. The SVM classification approach has demonstrated encouraging results in a variety of applications, from text categorization to handwritten digit identification[6]. In addition, SVM performs very well on data that has several dimensions.

Numerous studies on emotion analysis have been conducted using a variety of subjects and techniques. For example, it applied for classification and five emotional categories (anger, disgust, fear, sorrow, and joy) using the Vector Space Model, which employs ISEAR as a dataset[7]. This study conducted an experiment to evaluate the SVM's degree of accuracy in textual emotion classification.

2 NATURAL LANGUAGE PROCESSING

Natural language processing is an area of study within computer science, artificial intelligence, and linguistics that focuses on how computers and natural human languages—like English or Indonesian—interact. The primary objective of NLP research is to build robots that can comprehend human language, interpret it, and then respond appropriately.

Though there has been NLP study in earlier years, the history of NLP starts in the 1950s. The Turing Test is a test that was developed by Alan Turing, the pioneer of computer technology, in his well-known paper "Computing Machinery and Intelligence," published in 1950. The Turing Test evaluates a machine's (in this example, a computer program's) capacity for intelligent activity. A human judge would converse with both the person and the machine under test in the initial example presentation. Everybody is different from everyone else. It is considered that the machine has passed the test if the jury is unable to distinguish between people and machines.

Speech segmentation, text segmentation, part-of-speech tagging, and word sense disambiguation are a few of the topics covered by NLP research. Even though voice and text may be studied together, speech processing has grown into its own academic discipline.

3 SUPPORT VECTOR MACHINE

A classification technique called support vector machine (SVM) locates the optimal hyperplane with the biggest margin. The border line that divides data into classes is called a hyperplane, and the distance between the hyperplane and the nearest data point inside each class is called a margin. The support vector in each class is the data that is closest to the hyperplane.

An SVM is a binary linear nonprobabilistic classifier since it predicts each input when it belongs to one of the two classes that are already established. This is accomplished by taking an input data set and applying the SVM algorithm. Given a training set of data that has been classified as falling into one of two categories, SVM—being a classifier—is provided with this information. The SVM training algorithm creates a model that forecasts whether fresh data will fit into one or more categories.

The fundamental idea behind Support Vector Machines (SVM) is that they can be used as classifiers for linear problems. However, with the development of the kernel trick method, which looks for hyperplanes by converting the dataset into a higher-dimensional vector space called a feature space and then performing the classification on those features, SVM can now be applied to nonlinear problems. The prediction outcomes will be significantly impacted by the choice of kernel function[8].

SVM looks for the dividing hyperplane between the two classes in this situation. It is also evident from the image that there are several alternate borders (discrimination lines) separating the two classes. By measuring the hyperplane's margin and seeking the biggest margin, the optimal separating hyperplane between the two classes may be found. The line in Figure 2.4b demonstrates that the line with the highest margin (m) and the best location between the two classes is the best hyperplane, and the red and blue points on the dividing line are support vectors. The fundamental step in the SVM learning process is this hyperplane search.

4 SVM ON LINEARLY SEPARABLE DATA

Data that can be divided linearly is called linearly separable data. For instance, $y_i \in \{-1, +1\}$ y1 ∈ {−1, +1} is the class label of data x_i, and $\{x_1, \cdots, x_n\}$ is a dataset. A variety of substitute dividing lines that can divide datasets according to their classes are shown in Figure 2.4. But in addition to being able to divide the data, the optimal dividing line also has the biggest margin. The two classes in Formula (1) are divided by two parallel dividing lines. We obtain: because the first dividing line restricts the first class and the second dividing line restricts the second class.

$$x_i \cdot w + b \geq +1, y_i = +1$$
$$x_i \cdot w + b \leq -1, y_i = -1 \tag{1}$$

w is the plane normal, and b is the plane position relative to the coordinate center. The margin value (distance) is $(1 - b - (-1 - b))/w = 2/\|w\|$. This margin value is maximized while still satisfying equation (1). Multiplying b and w by a constant will produce a margin value multiplied by the same constant. Therefore, constraint (1) is a scaling constraint that can be satisfied by rescaling b and w. Moreover, maximizing $\frac{1}{\|w\|}$ is the same as minimizing $\|w\|^2$ and if both bounding fields in (1) are represented in (2),

$$y_i(x_i \cdot w + b) - 1 \geq 0 \qquad (2)$$

Thus, it is possible to frame the process of finding the optimal separating field with the biggest margin value as a constraint optimization problem, specifically:

$$min\frac{1}{2}\|w\|^2,\ with\, y_i(x_i \cdot w + b) - 1 \geq 0 \qquad (3)$$

As a quadratic programming problem, this method for determining the optimal dividing area always finds the global maximum value of a_i. The decision function value is used to identify the class of test data x once the quadratic programming problem has been solved (a_i value):

$$f(x_d) = \sum_{i=1}^{ns} a_i y_i\, x_i x_d + b \qquad (4)$$

where x_d is the data that has to be categorized, ns is the number of support vectors, and x_i is the support vector.

5 HYPERPARAMETER OPTIMATION

Parameter search, or model selection, is another term for this hyperparameter optimization process. Hyperparameter optimization, as used in machine learning, is a method for choosing and figuring out a collection of hyperparameters as a learning algorithm in order to achieve high generalization. A grid search is the most straightforward method of optimizing hyperparameters. A manually defined subset of a learning algorithm's hyperparameter space is used for grid search. Performance measurements obtained via cross-validation on the training set (Hsu et al. 2010), cited by Yusra (2014), serve as the basis for this grid search method.

Three parameters—C, γ, and d—will be used, depending on the kernel that is employed in this study. By pairing parameters (C, γ, and d) with variable parameter values, the grid search approach with cross-validation is advised to obtain the optimal values of C, γ, and d. These three parameters will be manually calculated, with $0 \leq C \leq 10$ and $0.01 \leq \gamma \leq 10$ being the range of values for each parameter. The goal of training the data set of value pairings (C, γ, and d) is to get the highest value for cross-validation testing accuracy, which will be chosen. Since every data set has the potential to be both training and test data, K-fold cross-validation is a useful method for verifying learning models.

6 EMOTIONS

The term "emotion" is derived from the Latin verb "movere," which means to move. Based on the word's etymology, emotion can be seen as a motivation to take action. An emotion is a specific idea or experience, as well as a biological and psychological condition and a set of behavioral inclinations. An emotion such as joy, anger, sadness, fear, disgust, shame and guilt are examples of emotions. Any and all emotional symptoms, including conflict, tension, worry, irritation, joy, pleasure, and hope, have an impact on a person's bodily alterations.

When a youngster has strong intellectual and cognitive ability but poor emotional development, he will have difficulties in his relationships and in life as a whole. Emotional development is one of the aspects that affects an individual's success in life.

One of the most well-known experts on emotions, Paul Ekman, demonstrates that there are six fundamental human emotions: fear, anger, sorrow, happiness, disgust, and surprise. It is thought that all people, regardless of culture, experience these universal feelings (Ekman P., 1992).

7 RESULTS

The dataset was obtained from the website http://www.affective-sciences.org/researchmaterial/, which contains an emotion dataset in the form of a Microsoft Access database. The data was then transferred into MongoDB after the unicode characters in the sentences in the initial dataset were removed. The following stage will include preprocessing each sentence in the dataset. Python was used for the preparation step, which included cleaning, case folding, tokenization, filtering, and stemming. Following the preprocessing step, 7520 lines are left with the following composition:

Table 1. Dataset composition

Emotion	Count
Joy	1086
Anger	1082
Fear	1083
Disgust	1068
Shame	1056
Sadness	1079
Guilt	1066

TF-IDF sentence weighting with n-gram = 1. A 2-dimensional sparse matrix with 7520 rows and 5374 columns (number of features) is produced as a consequence of the weighting results.

Ten subsets of the data, each with roughly the same amount of data, are created. Using a 10-fold cross validation technique, this subset serves as both training and test data. The Gaussian RBF kernel requires parameters C and γ for the Polynomial kernel plus a degree parameter, whereas the SVM training method using a Linear kernel requires C parameters. The Grid Search method—which involves an iterative training procedure with each pair of parameters and cross validation to discover the pair with the highest level of accuracy—is used to pick the optimal C, γ, and degree values.

The input parameters C = [0.1, 0.5, 0.9, 1, 2, 3, 4] are shown in Table 2. It is determined that parameter C = 0.5 has the best accuracy, 0.60346, and is thus used as a parameter in the Linear kernel training process.

The parameter pair C = 1 and $\gamma = 2$ has the highest accuracy, 0.58551, when the parameters C = [1, 2] and γ = [0.1, 0.2, 0.5, 0.8, 1, 2] are entered. As a result, the parameter pairs C = 1 and $\gamma = 2$ are chosen as parameters in the RBF kernel training process.

By entering the parameters C = [1, 3, 4, 5], γ = [0.1, 0.5, 0.8, 1, 2], and degree = [2, 3, 4], we find that the parameter pair C = 5, $\gamma = 0.8$, and degree = 2 have the highest accuracy, namely 0.56090. As a result, the parameter pairs C = 5 $\gamma = 0.8$ and degree = 2 are taken as parameters in the Polynomial kernel training process.

Table 2. Grid search

C	linear
0,1	0,59801
0,5	0,60346
0,9	0,60173
1,0	0,60066
2,0	0,59295
3,0	0,58976
4,0	0,58830

C	γ	rbf
1	2	0,58551
2	0,1	0,57154
2	0,2	0,58085
2	0,5	0,58245
2	0,8	0,58484
2	1	0,58457
2	2	0,58285

C	γ	degree	poly
5	0,1	2	0,15798
5	0,2	2	0,28032
5	0,5	2	0,54362
5	0,8	2	0,56090
5	0,1	3	0,14441
5	0,2	3	0,14441
5	0,5	3	0,25399

The categorization algorithm, like weighing, makes use of the Scikit-learn module. In the current study, there are two techniques for testing data in the categorization process:

1. Using the K-fold cross validation using a K of 10, with the training data partitioned into ten segments. The classification process will then go through 10 iterations, with 9 segments serving as training data and 1 segment serving as test data respectively.
2. Using a train test split, which divides the whole data into two pieces, training data and test data, If the proportion of test data is 20%, the proportion of training data is 80%.

The next step is to train on the training data to create a model that will later be used for testing on the test data. Linear, RBF, and Polynomial kernels are used in this study.

Classification using Linear kernel

As input to the object classifier, training data is fed into a Linear kernel with C = 0.5. The cross-validation method was repeated ten times, with an average accuracy result of 0.618, or 61.8%. As input to the classifier object, training data is fed into an RBF kernel with values C = 1 and gamma = 2. The cross-validation method was repeated ten times, yielding an average accuracy result of 0.603, or 60.3%. A Polynomial kernel with parameters C = 5, = 0.8, and degree = 2 is used to train data for unigram features. The cross-validation technique has an average accuracy of 0.577, or 57.7%.

Using 7520 ISEAR data points, it was discovered that the model created for each kernel was better at recognizing melancholy and disgust than other emotions. In terms of accuracy, the SVM classifier model developed by cross-validation achieved a maximum accuracy of 61.8% for the Linear kernel function, 60.3% for the RBF kernel function, and 57.7% for the Polynomial kernel function. In this study, linear kernels were shown to be more accurate than RBF and polynomial kernels.

REFERENCES

[1] C. Nass, J. Steuer, and E. R. Tauber, “Computers are social actors,” in *Conference companion on Human factors in computing systems - CHI ’94*, Boston, Massachusetts, United States: ACM Press, 1994, p. 204. doi: 10.1145/259963.260288.

[2] D. Goleman, R. E. Boyatzis, and A. McKee, *Primal leadership: realizing the power of emotional intelligence*. Boston, Mass: Harvard Business School Press, 2002.

[3] P. Ekman, “Are there basic emotions?,” *Psychol. Rev.*, vol. 99, no. 3, pp. 550–553, 1992, doi: 10.1037/0033-295X.99.3.550.

[4] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
[5] M. R. Widyanto and R. Wijayanti, “Membership Function Modification in Fuzzy SVM using Combination of Distance Feature, Correlation, and Depth of Data,” *IJECS*, vol. 13, no. 5, 2013.
[6] P.-N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to data mining*, Second edition. New York: Pearson, 2019.
[7] T. Danisman and A. Alpkocak, “AISB 2008 Symposium on Affective Language in Human and Machine,” in *Affective Language in Human and Machine*, The Society for the Study of Artificial Intelligence and Simulation of Behaviour, pp. 53–59.
[8] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, 1st ed. Cambridge University Press, 2000. doi: 10.1017/CBO9780511801389.

BIOGRAPHY OF AUTHORS

Rully Ramanda

Rully Ramanda is a ex-student at the Universitas Islam Negeri Sultan Syarif Kasim Riau in the Informatics Engineering Department. He is passionate about machine learning research.

Muhammad Affandes

Muhammad Affandes teaches at Universitas Islam Negeri Sultan Syarif Kasim Riau in the Department of Informatics Engineering. His current study focuses on software development, utilizing microservices and web-based technologies.