Received: 13 May 2023 | Revised: 7 July 2023 | Accepted: 31 July 2023

DOI: 10.1002/sam.11639

WILEY

RESEARCH ARTICLE

A neutral zone classifier for three classes with an application to text mining

Dylan C. Friel[1] | **Yunzhe Li**[2] | **Benjamin Ellis**[1] | **Daniel R. Jeske**[1] | **Herbert K. H. Lee**[2] | **Philip H. Kass**[3]

[1]Department of Statistics, University of California, Riverside, California, USA

[2]Department of Statistics, University of California, Santa Cruz, California, USA

[3]Department of Population Health & Reproduction, University of California, Davis, California, USA

Correspondence
Daniel R. Jeske, University of California, Riverside, CA, USA.
Email: daniel.jeske@ucr.edu

Abstract

A classifier may be limited by its conditional misclassification rates more than its overall misclassification rate. In the case that one or more of the conditional misclassification rates are high, a neutral zone may be introduced to decrease and possibly balance the misclassification rates. In this paper, a neutral zone is incorporated into a three-class classifier with its region determined by controlling conditional misclassification rates. The neutral zone classifier is illustrated with a text mining application that classifies written comments associated with student evaluations of teaching.

KEYWORDS

classification, neutral zone, sentiment analysis, text mining, Word2Vec

1 | INTRODUCTION

Classification of observations into groups is an objective for many applications. For example, patients might be classified as diseased or not or loan applications might be classified as high risk or not. While classification problems often involve only two categories, any number of classes may be of interest. A common procedure for classification is to obtain the probabilities that an observation belongs to each of the possible classes, and then assign it to the class with the largest probability. A drawback of hard classification boundaries is the forced classification of ambiguous observations into a specific class. Forcing a definitive classification decision when the probabilities of each class are very close to each other can lead to misclassifications that have undesirable consequences. Consider for example, the consequence of declaring an ambiguous patient as diseased when they are not or, perhaps more catastrophically, declaring a patient as not diseased when they are.

If a particular classifier has a high misclassification rate, a practitioner might try alternative classifiers or try to find new features that improve the separation of the classes. Unfortunately, there will be situations where neither of these approaches works. Introducing a "neutral zone" between the classification boundaries is an alternative where a definitive classification decision for ambiguous cases is delayed. When an observation has class probabilities that lead to ambiguity about which of the classes is likely, it will be assigned a label of "neutral." A practitioner may subsequently engage in follow-up investigations of observations labeled as neutral before making a final classification decision. Labeling the observation as neutral accurately reflects what is known about it and the follow-up investigation provides an opportunity to prevent a misclassification. While follow-up will reduce the misclassification rate, the work involved does add cost to the overall decision. The tradeoff between reduced misclassifications and the cost of follow-up depends on the consequences of making misclassifications, which in healthcare applications for example, are typically severe.

Incorporating neutral zones into classification problems has been explored in the literature. A neutral zone

based upon the cost of misclassifications has been developed for both two classes [1, 2] and three classes [3]. Recognizing the difficulty of ascertaining costs of misclassification, and the comparatively straightforward interpretation of conditional misclassification rates, a neutral zone for two classes that controls the false positive and false negative rates was developed in [4, 5]. The methodological contribution of this paper is the development of a first-of-its-kind neutral zone for a three-class classifier that does not make explicit assumptions of the class-conditional distributions while controlling each of the six conditional misclassification rates and, consequently, reduces the overall misclassification rate. The rest of the paper is outlined as follows. Section 2 presents the motivating application for the development of a three-class neutral zone classifier. In Section 3, we present the formulation of the classifier. Section 4 returns to our motivating application and shows successful implementation of the classifier. Finally, Section 5 offers a summary of the work presented in this paper.

2 | MOTIVATING APPLICATION

The motivating application for this work is an objective to classify written comments associated with student evaluations of teaching as reflecting positive, negative, or ambiguous feelings about a student's overall experience in the class. The data we use are comments written by undergraduate students for teaching evaluations at the University of California, Santa Cruz (UCSC) and the University of California, Riverside (UCR). Student evaluations are an important factor when evaluating the effectiveness of instructors. These evaluations consist of both Likert scale questions as well as open-ended questions where the student may leave comments in their own words.

Issues with ratings and potential biases of the Likert-scale questions have been actively researched [6–10]. The frequency of specific words in written evaluations of medical students has been examined for gender and ethnicity bias [11]. Whereas numerical evaluations from instructor reviews are presented in summary form, written comments are typically presented verbatim in the order in which they were recorded. There may be hundreds of evaluations from a single course. The reviewer of the comments is left on their own to extract the overall message of the comments. In the worst cases, the comments may be glossed over or selectively chosen to support a preconceived narrative. If the comments can be classified as positive, negative, or other they could be sorted to assist the reviewer in getting a more representative understanding of the comments. Reviewers wanting to read all comments could do so more systematically with the sorted ordering.

The labels we use for the comments are defined as follows. A positive comment has the overall interpretation that the instructor is doing a good job and the student would recommend this instructor to other students. A negative comment conveys that the instructor is not doing a good job and the student would not recommend this instructor to other students. Comments that are mixed with both positive and negative remarks or that provide no evaluation of the instructor are labeled as other.

Our data set comprises 104,143 comments from evaluations conducted at UCSC and 34,749 comments from evaluations at UCR. The courses where these comments originated were medium to large enrollment undergraduate classes in both STEM and non-STEM fields and were taught between fall 2018 and summer 2021. To obtain the true label for each comment, a team of three undergraduate students was employed and trained how to identify each type of comment, and the label was determined via majority rules voting. In cases where there was no majority, a graduate student researcher made the final determination of the true label. The comments from UCSC (UCR) resulted in approximately 63% (66%) positive comments, 13% (15%) negative comments, and 24% (19%) other comments.

After obtaining the true labels, C, multiple options for a text classifier were explored including a naïve Bayes classifier [12] and classifiers based on sentiment analysis [13, 14]. We found that a multinomial logistic regression classifier [15] with features extracted via the Word2Vec [16–19] algorithm was the most effective choice for our application. Details for the Word2Vec feature extraction will be explained further in Section 4. Letting p_0, p_1, and p_2 denote the predicted probabilities of the classes negative, positive, and other, respectively, the standard logistic regression classifier would be defined as

$$\widehat{C} = \begin{cases} \text{Negative} & p_0 > p_1 \text{ and } p_0 > p_2 \\ \text{Positive} & p_1 > p_0 \text{ and } p_1 > p_2 \\ \text{Other} & p_2 > p_0 \text{ and } p_2 > p_1 \end{cases}$$

The results from such a classifier applied to the UCSC and UCR data (fit on training data and applied to an independent test set of data) are presented in Table 1 and Table 2, respectively. The overall misclassification rates are about 20%, and it can be seen that there is an imbalance in the conditional misclassification rates. The goal for this application is to incorporate a neutral zone into the classifier that improves the balance of the conditional misclassification rates and lowers the overall misclassification rate.

TABLE 1 Class-conditional classification rates for a standard logistic regression classifier of comments from student evaluations of teaching at UCSC.

True label	Predicted label			Conditional misclassification
	Positive	Negative	Other	Rate
Positive	0.921	0.021	0.058	0.079
Negative	0.232	0.515	0.253	0.485
Other	0.277	0.127	0.596	0.404

Note: Overall misclassification rate: 0.212.

TABLE 2 Class-conditional classification rates for a standard logistic regression classifier of comments from student evaluations of teaching at UCR.

True label	Predicted label			Conditional misclassification
	Positive	Negative	Other	Rate
Positive	0.927	0.026	0.047	0.073
Negative	0.201	0.602	0.197	0.398
Other	0.401	0.218	0.381	0.619

Note: Overall misclassification rate: 0.224.

3 | NEUTRAL ZONE CLASSIFIERS

Our starting point assumes that we have the probabilities an observation belongs to each of three classes. While our application utilizes a multinomial logistic regression model for this purpose, these probabilities can be obtained in a variety of other ways including via a neural network classifier, a classification tree, or a Bayes classifier. Traditionally, an observation would be assigned to the class with the largest probability. A drawback of this approach is the adherence to a hard boundary when the probabilities for each class are close. Incorporation of a neutral zone creates regions for the probabilities such that observations that fall into these regions are classified as neutral due to the lack of convincing evidence for a definitive classification. Observations classified as neutral are left for further investigation through follow-up. We next explore the alternatives for constructing the neutral zone boundaries.

3.1 | Symmetric neutral zones

Yu et al. [3] developed a minimum cost neutral zone classifier for three classes where a neutral zone region between classes is uniformly created by a single constant, L. The experimenter determines L based on the cost of misclassification. We can take this approach, but instead choose L to achieve desired conditional misclassification rates. Letting N denote the label for the neutral zone, the symmetric neutral zone classifier is defined as

$$\widehat{C} = \begin{cases} 0 & p_0 > p_1 + L \text{ and } p_0 > p_2 + L \\ 1 & p_1 > p_0 + L \text{ and } p_1 > p_2 + L \\ 2 & p_2 > p_0 + L \text{ and } p_1 > p_1 + L \\ N & \text{otherwise} \end{cases} \quad (1)$$

If $L \in [0, 1]$ starts at zero and is increased toward one, the conditional misclassification rates go to zero. Therefore, if we find the first L such that $P(\widehat{C} = i | C = j) \leq \alpha_{ij}$ for $i, j = 0, 1, 2$, and $i \neq j$, then there always will be a solution. The optimal L is the smallest L such that each conditional misclassification rate is less than or equal to its target size. Figure 1A sketches the general shape of the symmetric neutral zone classifier using $L = 0.3$ for illustrative purposes. While this symmetric approach allows a uniform upper bound on the conditional misclassification rates, it generally will not substantially improve an imbalance of the conditional misclassification rates.

3.2 | Asymmetric neutral zones

Rather than using a single L to define neutral zone regions, an alternative is to separately choose an L for each pairwise decision boundary. The asymmetric neutral zone classifier is defined as

$$\widehat{C} = \begin{cases} 0 & p_0 > p_1 + L_{01} \text{ and } p_1 > p_2 \text{ or } p_0 > p_2 + L_{02} \text{ and } p_2 > p_1 \\ 1 & p_1 > p_0 + L_{10} \text{ and } p_0 > p_2 \text{ or } p_1 > p_2 + L_{12} \text{ and } p_2 > p_0 \\ 2 & p_2 > p_0 + L_{20} \text{ and } p_0 > p_1 \text{ or } p_2 > p_1 + L_{21} \text{ and } p_1 > p_0 \\ N & \text{otherwise} \end{cases} \quad (2)$$

where $L_{ij} \in [0, 1]$ represents the margin when deciding on class i over class j. Each L_{ij} enters the classification rule when class j is the second-most likely category after class i. Figure 1B sketches the general shape of the asymmetric neutral zone classifier using the six values of L_{ij} shown for illustrative purposes.

It is of interest to note that the geometrical area of the neutral zone as a proportion of the entire classification region is $\sum_i \sum_j \frac{L_{ij}}{12}(2 - L_{ij})/(1/2)$ for $i, j = 0, 1, 2$ and $i \neq j$. This proportion is 0.51 for the symmetric neutral zone in Figure 1A and 0.42 for the asymmetric neutral zone in Figure 1B. While this area may be used to roughly compare the size of alternative neutral zone classifiers, it differs from the proportion of observations that fall within the neutral zone due to the fact that the latter depends on the underlying class-conditional distributions of the features.

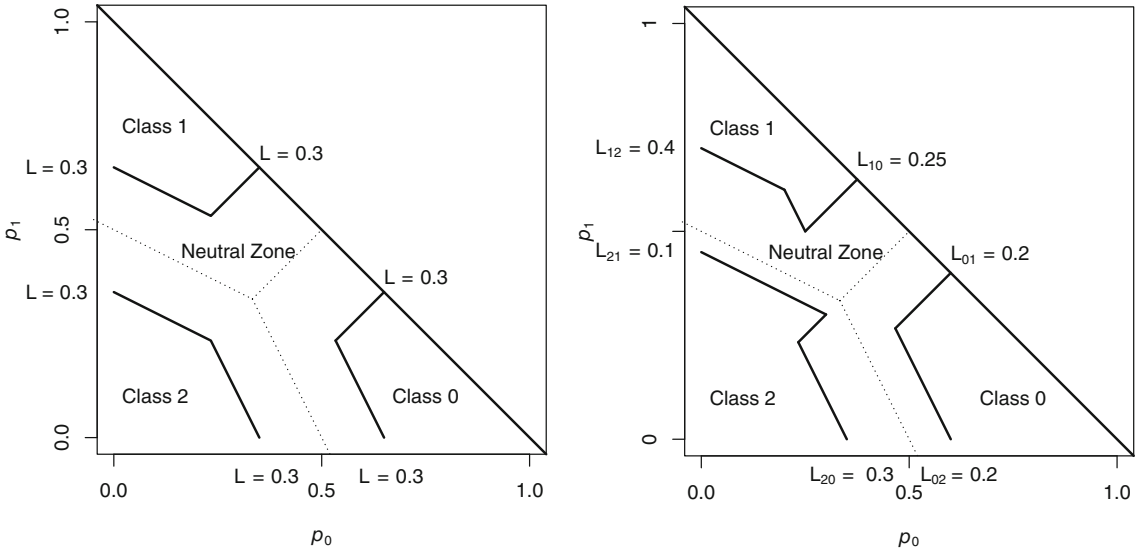

FIGURE 1 Symmetric (A) and Asymmetric (B) neutral zone classifiers. Dotted lines represent the boundaries of the no neutral zone classifier.

3.3 | Controlling conditional misclassification rates

Conditional misclassification rates of the proposed neutral zone classifier can be controlled by selecting L_{ij} such that $P(\widehat{C} = i|C = j) \leq \alpha_{ij}$. For each i, the pair of L_{ij}'s are found jointly since a single L_{ij} affects only two of the six conditional misclassification rates. For example, (L_{01}, L_{02}) are found from the equations $P(\widehat{C} = 0|C = 1) \leq \alpha_{01}$ and $P(\widehat{C} = 0|C = 2) \leq \alpha_{02}$ and similarly for (L_{10}, L_{12}) and (L_{20}, L_{21}). If $\alpha_{ij} = \alpha$, for all i, j, and some constant α, better balance of the conditional misclassification rates will be achieved. The optimal set of L_{ij} are those that give conditional misclassification rates closest to the target rates without exceeding them. Both the symmetric and the asymmetric neutral zones will either give the same predicted class as the traditional classifier or change the predicted class to neutral. Thus, no new misclassifications are introduced by using the neutral zone classifier.

3.4 | Grid search

A straightforward approach to finding the L_{ij} is to use a grid search as follows. First, the p_0, p_1, and p_2 probabilities are obtained for all the observations in the training data set. As explained in the previous section, we find the L_{ij} two at a time. Consider the case of finding L_{01} and L_{02}. For each (L_{01}, L_{02}) on a unit grid, use the predicted classes for the training data to estimate $P(\widehat{C} = 0|C = 1)$ and $P(\widehat{C} = 0|C = 2)$. Then choose the (L_{01}, L_{02}) that gives the conditional misclassification rates closest to, while less than, α_{01} and α_{02}. Perform this same search similarly to find (L_{10}, L_{12}) and (L_{20}, L_{21}) to obtain the set of six optimal L_{ij}.

3.5 | Feature space representation

In some situations, the classifier in Equation (2) can be inverted to display the decision boundaries in the feature space. We illustrate this in a Bayes classification setting with two dimensions. Let π_i represent the prior class probabilities. Suppose the features in each class follow the probability density function $f_i(x)$. Then $p_i = \pi_i f_i(x) / \sum_{j=0}^{2} \pi_j f_j(x)$ are the posterior class probabilities. These probabilities are used in $\widehat{C}$ from Equation (2) to obtain the predicted classes. Letting A_0, A_1, A_2, and A_N denote the regions in the feature space that correspond to the predicted labels 0, 1, 2, and N, respectively, we have

$$\begin{aligned} A_0 =& \{x\colon p_0 > p_1 + L_{01}, p_1 > p_2\} \\ & \cup \{x\colon p_0 > p_2 + L_{02}, p_2 > p_1\} \end{aligned}$$

$$\begin{aligned} A_1 =& \{x\colon p_1 > p_0 + L_{10}, p_0 > p_2\} \\ & \cup \{x\colon p_1 > p_2 + L_{12}, p_2 > p_0\} \end{aligned}$$

$$\begin{aligned} A_2 =& \{x\colon p_2 > p_0 + L_{20}, p_0 > p_1\} \\ & \cup \{x\colon p_2 > p_1 + L_{21}, p_1 > p_0\} \end{aligned}$$

$$A_N = \overline{A_0 \cup A_1 \cup A_2}$$

The six conditional misclassification probabilities are calculated as

$$P(\widehat{C} = i|C = j) = \int_{A_i} f_j(x)\, dx, \quad i, j \in \{0, 1, 2\}, i \neq j$$

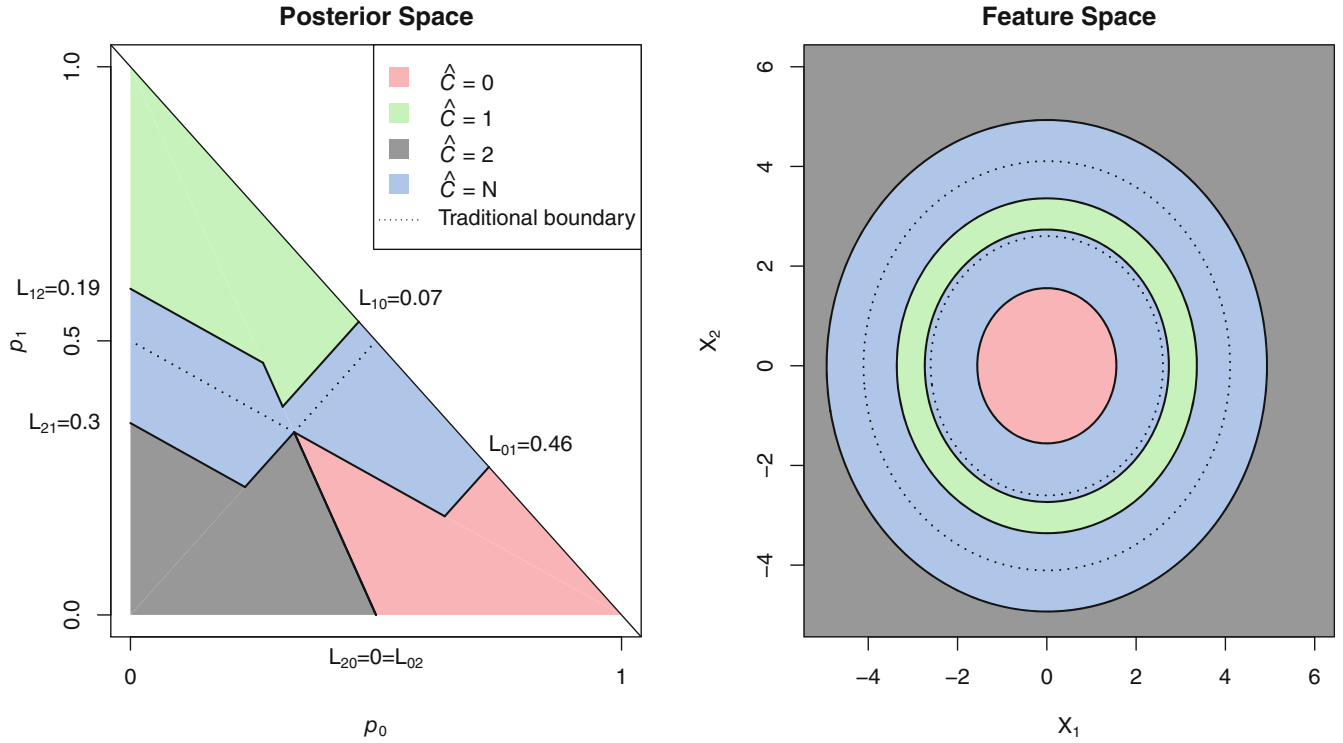

FIGURE 2 Neutral zone in the posterior and feature space for $X \sim N(\vec{\mu}_i, \Sigma_i)$ where $\vec{\mu}_0 = (1,1)$, $\vec{\mu}_1 = (3,3)$, $\vec{\mu}_2 = (5,5)$, $\Sigma_0 = I_2$, $\Sigma_1 = 2 \times I_2$, $\Sigma_2 = 3 \times I_2$, and $\pi_i = 1/3$ for $i = 0, 1, 2$. The L_{ij} are found to give conditional misclassification probabilities less than or equal to 0.1. The dotted lines show the no neutral zone classifier boundaries.

In addition, the conditional neutral zone rates are given by

$$P(\widehat{C} = N | C = j) = \int_{A_N} f_j(x)\, dx, \quad j \in \{0, 1, 2\}$$

Numerical integration techniques can be used in conjunction with the grid search explained in Section 3.4 to estimate the conditional misclassification rates and determine the L_{ij}.

When the $f_i(x)$ are bivariate normal, the regions A_0, A_1, A_2, and A_N can be graphed in the feature space. This is demonstrated in Figure 2 and it can be seen how the neutral zone forms around the areas of ambiguity between the distributions in the feature space. The spherical boundaries in Figure 2 are a consequence of unequal, but diagonal, covariance matrices.

4 | EXAMPLE APPLICATION

4.1 | Word2Vec

We return to our motivating application from Section 2. When working with text data, we first need to transform the text into numeric values. As was mentioned in Section 2, we found feature extraction based upon Word2Vec to be the most effective for our purposes. The main purpose of Word2Vec is to try to predict words that are written together. Word2Vec is a mapping based on a neural network and was originally proposed with a choice of two algorithms: continuous bag-of-words (CBOW) and skip-gram. We focus on the former, where the algorithm attempts to predict a "center" word based on given "context" words. A step in the CBOW algorithm, displayed visually in Figure 3, which is central to our application, is the mapping of words to numerical features.

The CBOW algorithm is trained by moving through each word in each comment, treating them as a center word, w. The context words are determined in a window around w. The window size is inputted by the user. The input layer of the neural network consists of one-hot vectors $b_1, b_2, \ldots, b_c$ representing the context words. The one-hot vectors have length d, where d is the number of words in the entire corpus of comments, and are zero everywhere except for a one at the position of the word in a dictionary formed from the corpus. These input vectors are used to extract rows from a to-be-determined $d \times m$ matrix, W, where m is inputted by the user. An element-wise sum on the extracted rows creates the latent vector u_w. Then matrix multiplication is performed with u_w and another to-be-determined matrix U. The result is a vector, v, which is inputted to a softmax function that uses a normalization

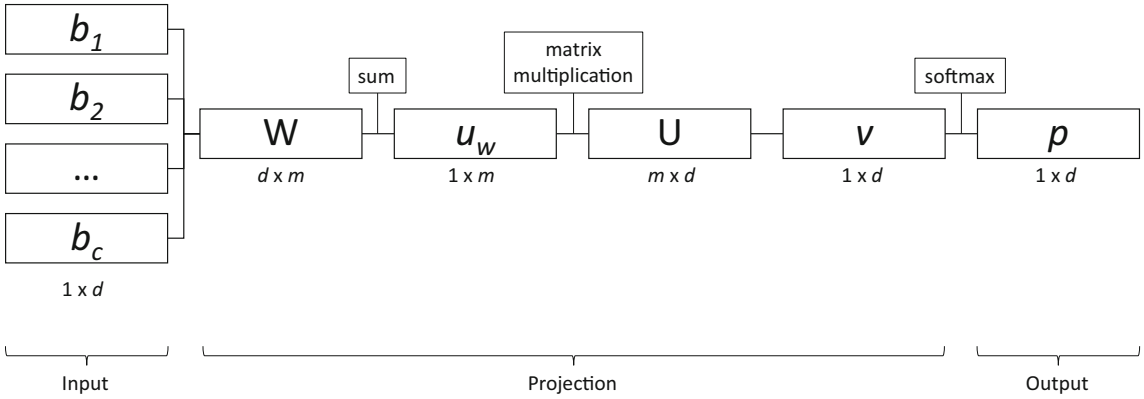

FIGURE 3 Map of the CBOW algorithm used in Word2Vec.

transformation to create a vector, p, of length d that consists of the probability that each word is the center word. This vector of probabilities is used with the one-hot vector of the true center word to compute loss. The cumulative loss is an aggregation of the loss from performing this process with each word as the center word. The fitting process solves for W and U which minimize the aggregated loss.

There are several hyperparameters that may be adjusted in Word2Vec. These include the window size (number of words to consider around the center word), the number of features (length of the latent vector), and an occurrence threshold (number of times a word must occur to be considered one of the d words). We have used the default values of 5, 50, and 5, respectively, which are recommended in the R package word2vec [19].

In our application, we do not need to predict words given context words. Instead, we are interested in the matrix of word embeddings, W, from within the projection layer of Word2Vec. With the fitted W, we have a matrix where each row represents a word in our corpus and the columns represent numerical features. For each word in a comment, we extract the corresponding rows of W to get a matrix of features for the comment. After normalizing the vector of column sums to account for the length of the comment, we obtain a numeric vector of length m that can be used as the features in a multinomial logistic classifier. In the following sections, we use five-fold cross-validation to evaluate the accuracy of the multinomial logistic regression classifier. Each training set is used to fit the Word2Vec model, fit the multinomial logistic regression model, and find the L_{ij}'s.

4.2 | UCSC data

First, we analyze 104,143 comments from instructor evaluations at the University of California, Santa Cruz. Recall that the comments labeled as negative, positive, and other have been defined as Class 0, Class 1, and Class 2, respectively. The results of a largest-probability classifier using multinomial logistic regression were presented earlier in Table 1. In this section, we incorporate the asymmetric neutral zone to mitigate adverse consequences of misclassified comments and both lower and balance the conditional misclassification rates. We choose to implement the asymmetric neutral zone for our application since the imbalance of the conditional misclassification rates shown in Tables 1 and 2 is quite pronounced. We set each $\alpha_{ij} = \alpha = 0.1$ as our target conditional misclassification rates.

Table 3 displays the five-fold cross-validation estimates of the conditional misclassification rates of the asymmetric neutral zone classifier for the UCSC data. In two cases the conditional misclassification rates are much lower than the target. As seen in Table 1, these two conditional misclassification rates were lower than the target before incorporating a neutral zone, which explains why the corresponding L_{ij}'s are zero. The other four conditional misclassification rates are approximately equal, showcasing the ability of the asymmetric neutral zone to achieve better balance in the conditional misclassification rates than without its use. The overall misclassification rate of the classifier is about 10% compared to about 20% that was seen in Section 2 when the neutral zone was not employed. The improved accuracy is the result of approximately 20% of the comments getting classified as neutral because they are too ambiguous to be confidently assigned to any of the classes.

Figure 4 displays the asymmetric neutral zone classifier in the posterior space, fitted with an 80–20 train-test split of the data. The points plotted in this figure are the 20,828 observations from test set. The area of this neutral zone as a proportion of the classification region is 0.35.

TABLE 3 Asymmetric neutral zone classifier applied to classification of comments from student evaluations of teaching at UCSC.

True label	Predicted label				Conditional misclassification rate
	Positive	Negative	Other	Neutral	
Positive	0.763	0.017	0.045	0.174	0.063
Negative	0.100	0.459	0.097	0.344	0.196
Other	0.100	0.100	0.470	0.330	0.200

Note: Overall misclassification rate: 0.114. Overall neutral rate: 0.234.

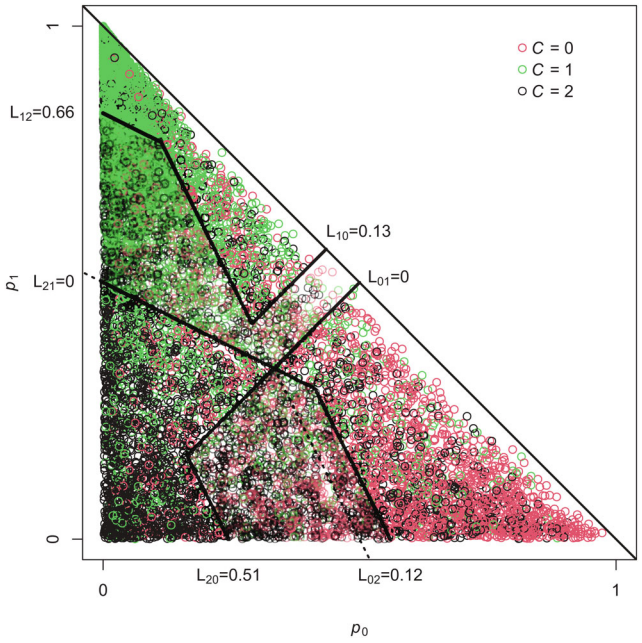

FIGURE 4 Asymmetric neutral zone applied to a test set of comments from UCSC. Neutral zone is indicated by the transparent points. Dotted lines show the no neutral zone boundaries.

4.3 | UCR

Next, we incorporate neutral zones into the multinomial logistic classifier from the 34,739 comments from University of California, Riverside. We apply the same target of 0.1 for the conditional misclassification rates. The results from the implementation of the asymmetric neutral zone are displayed in Table 4. All conditional misclassification rates have been lowered appropriately to be less than or equal to the target value. As with the UCSC data, there are two instances in Table 2 where the conditional misclassification rate is much lower than the target because these two rates were lower than the target before the neutral zone was implemented. The other four conditional misclassification rates are approximately equal. The overall misclassification rate is lowered to about 10% from about 20% in Table 2 with approximately 30% of the observations being classified as neutral. Notice that while the conditional misclassification rates are roughly comparable for the two campuses, the set of L_{ij}'s needed to achieve that are different. The UCR data leads to slightly more comments being labeled as neutral.

Figure 5 displays the asymmetric neutral zone classifier in the posterior space with an 80–20 train-test split of the data, with the test set observations as an overlay. The area of this neutral zone as a proportion of the classification region is 0.30.

While the fitted classifiers from the two campuses are similar, some difference between the two campuses might have been anticipated. Consider, for example, how the prompts for the written comments. At UCR a single, broad question is used asking the student to "comment on how the instructor's teaching helped your learning of the material in this course." On the other hand, UCSC used multiple and more targeted questions to prompt comments from students. As different universities tend to have unique cultures, it is recommended that each university that desires use our approach for labeling student comments written for instructor evaluations fit the neutral zone classifier to their own training data. R code is provided in the supplementary material to create the asymmetric neutral zone classifier from any set of training data.

5 | SUMMARY

In this paper, we have developed two alternative neutral zone classifiers for the three-class setting which recognize and respond to the difficulty of classifying ambiguous observations and by doing so are able to lower the overall misclassification rate and improve the balance of the conditional misclassification rates. This is achieved by labeling ambiguous observations as neutral, giving follow-up investigation the chance to resolve the ambiguity. For most applications we would recommend use of the asymmetric neutral zone. The three-class neutral zone classifiers are the first classifiers to control the six conditional misclassification rates and require no assumptions about the class-conditional distributions. The classifiers may be employed in any three-class scenario where the probabilities for each class are obtained from any of a variety of methods that create them.

TABLE 4 Asymmetric neutral zone classifier applied to classification of comments from student evaluations of teaching at UCR.

True label	Predicted label				Conditional misclassification rate
	Positive	Negative	Other	Neutral	
Positive	0.734	0.015	0.036	0.215	0.051
Negative	0.078	0.396	0.100	0.427	0.178
Other	0.098	0.103	0.281	0.518	0.201

Note: Overall misclassification rate: 0.098. Overall neutral rate: 0.303.

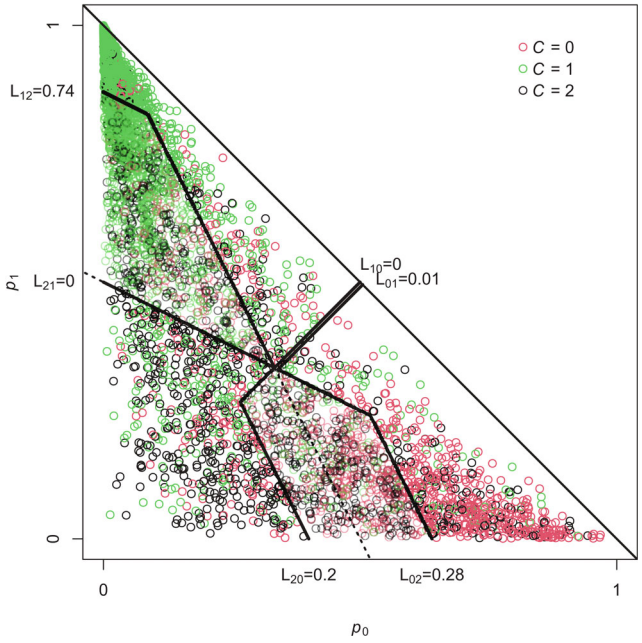

FIGURE 5 Asymmetric neutral zone applied to a test set of comments from UCR. Neutral zone is indicated by the transparent points. Dotted lines show the no neutral zone boundaries.

This work was motivated by student comments written for instructor evaluations. We have shown how Word2Vec and multinomial logistic regression may be combined to analyze text data with three classes. The neutral zone classifier in this setting assists a reviewer in the reading of many comments by providing at a glance the frequency of comments that are classified as positive, negative, or other. The predicted labels also allow the comments to be grouped so that they can be presented to reviewers in sorted order, which aids the selection of a representative sample of the comments for full reading.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Daniel R. Jeske https://orcid.org/0000-0002-0214-7992
Herbert K. H. Lee https://orcid.org/0000-0003-1087-4150

REFERENCES

1. D. R. Jeske, Z. Liu, E. Bent, and J. Borneman, *Classification rules that include neutral zones and their application to microbial community profiling*, Commun Stat Theory Methods 36 (2007), no. 10, 1965–1980. https://doi.org/10.1080/03610920601126514.
2. S. Benecke, D. R. Jeske, P. Ruegger, and J. Borneman, *Bayes neutral zone classifiers with applications to nonparametric unsupervised settings*, J. Agric. Biol. Environ. Stat. 18 (2013), 39–52.
3. H. Yu, D. R. Jeske, P. Ruegger, and J. Borneman, *Neutral zone classifiers using a decision-theoretic approach with application to DNA array analyses*, J. Agric. Biol. Environ. Stat. 15 (2010), 474–490.
4. D. R. Jeske and S. Smith, *Maximizing the usefulness of statistical classifiers for two populations with illustrative applications*, Stat. Methods Med. Res. 27 (2018), no. 8, 2344–2358.
5. D. R. Jeske, Z. Zhang, and S. Smith, *Construction, visualization and application of neutral zone classifiers*, Stat. Methods Med. Res. 29 (2020), no. 5, 1420–1433.
6. A. Boring, K. Ottoboni, and P. Stark, *Student evaluations of teaching (mostly) do not measure teaching effectiveness*, ScienceOpen Res (2016), 1–11. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1.
7. A. Ho, L. Thomsen, and J. Sidanius, *Perceived academic competence and overall job evaluations: Students' evaluations of African American and European American professors*, J. Appl. Soc. Psychol. 39 (2009), 389–406. https://doi.org/10.1111/j.1559-1816.2008.00443.x.
8. F. Mengel, J. Sauermann, and U. Zolitz, *Gender bias in teaching evaluations*, J. Eur. Econ. Assoc. 17 (2019), no. 2, 535–566. https://doi.org/10.1093/jeea/jvx057.
9. J. Miller and M. Chamberlin, *Women are teachers, men are professors: A study of student perceptions*, Teach. Sociol. 28 (2000), no. 4, 283–298. https://doi.org/10.2307/1318580.
10. S. J. Ceci, S. Kahn, and W. M. Williams, *Exploring gender bias in six key domains of academic science: An adversarial collaboration*, Psychol. Sci. Public Interest 24 (2023), no. 1, 15–73.
11. D. Ross, D. Boatright, M. Nunez-Smith, A. Jordan, A. Chekroud, and E. Z. Moore, *Differences in words used to describe racial and gender groups in medical student performance evaluations*, PLoS One 12 (2017), no. 8, e0181659. https://doi.org/10.1371/journal.pone.0181659.
12. T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, Springer Inc., New York, 2001.
13. F. A. Nielsen, *AFINN. Informatics and mathematical modelling*, Technical University of Denmark, Lyngby, Denmark, 2011 http://www2.compute.dtu.dk/pubdb/pubs/6010-full.html.
14. T. W. Rinker. Sentiment: Calculate text polarity sentiment version. 2.9.0 2021 https://github.com/trinker/sentimentr.

15. M. Kutner, C. Nachtsheim, J. Neter, and W. Li, *Applied linear statistical models*, McGraw-Hill Irwin, Chicago, 2005.
16. T. Mikolov, K. Chen, G. S. Corrado, and J. Dean. Efficient estimation of word representations in vector space. 2013 International Conference on Learning Representations.
17. T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, Distributed representations of words and phrases and their compositionality. *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'13)*, Curran Associates Inc., Red Hook, NY, USA, 2013, pp. 3111–3119.
18. Y. Goldberg and O. Levy. *word2vec explained: Deriving Mikolov et al.'s negative-sampling word-embedding method*. *CoRR*, abs/1402.3722. arXiv:1402.3722 2014 https://doi.org/10.48550/arXiv.1402.3722.
19. J. Wijffels, *Word2vec: Distributed representations of words*, R Package Version 0.3.4 (2021), https://CRAN.R-project.org/package=word2vec.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: D. C. Friel, Y. Li, B. Ellis, D. R. Jeske, H. K. H. Lee, and P. H. Kass, *A neutral zone classifier for three classes with an application to text mining*, Stat. Anal. Data Min.: ASA Data Sci. J. **16** (2023), 560–568. https://doi.org/10.1002/sam.11639