International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

Ordinal Based Classification Techniques: A Survey

Anoosha M Rao[1], Vinayak S[2], Swetha S[3]

[1]Student, Dept of Information Science and Engineering, RV College of Engineering, Bangalore-560 059, India
[2] Student, Dept of Information Science and Engineering, RV College of Engineering, Bangalore-560 059, India
[3]Assistant professor, Dept of Information Science and Engineering, RV College of Engineering, Bangalore, India

---

Abstract - *Ordinal classification is a type of supervised learning problem where the target variables exhibit an inherent sense of ordering among them. This differs from a regular classification problem where an input vector is classified into unordered categories. Ordinal classification problems are prevalent in the field of machine learning and pattern recognition. It is used in facial recognition systems, classification of text etc., It plays a major role in the analysis of data in behavioral sciences. Various algorithms have been proposed to solve the ordinal classification problem. This paper aims to present the advances made in this field and review the methods proposed to perform the same. A comprehensive study of the existing techniques has been made and several conclusions regarding the performance of these algorithms have been drawn.*

Key Words: Classification problem, ordinal classification, regression, error-weighted classification, binary decomposition, neural network, discriminant analysis, cumulative link modelling.

1.INTRODUCTION

One of the classic problems of machine learning is categorizing various sets of data into predefined labels. Classification is a type of supervised learning. A general classification problem simply expects the input to be categorized into one of the previously specified classes, each of which is independent of one another. However, ordinal classification involves classifying the input vector into categories possessing an inherent ordering. Consider the example of a student rating his/her knowledge of machine learning. The output variables are Outstanding, Very Good, Good, Mediocre and Poor. These values are categorical but clearly Outstanding > Very Good > Good > Mediocre > Poor. Another property of ordinal categories is that the distance between the adjacent categories are unknown. Also, the distance between categories is not necessarily equal i.e. 'Outstanding' is certainly better than 'Very Good' and 'Good' is better than 'Mediocre'. But, 'Good' might be a lot better than 'Mediocre' while 'Outstanding' might be slightly better than 'Very Good'.

In a general classification problem, if an input vector v is wrongly classified into class i or to class j while it actually belongs to class k (i not = j not = k), then the penalty for improper classification is the same for both cases i and j. However, the penalty for wrong classification varies for every class in case of ordinal classification. In the example considered previously, the penalty for classifying a vector actually belonging to the 'Outstanding' category to 'Very Good' category is lesser than that of classifying it to 'Poor' category. This penalty variation for erroneous classification is one of the important factors to be considered in ordinal classification.

2. ORDINAL CLASSIFICATION METHODS

Ordinal classification can be performed using various approaches. In this section, we explore different methods used to for the same. These methods have been divided into three major groups, namely, traditional approaches, binary decomposition and threshold models. Traditional approaches generally involve simplification of ordinal problems into other standard problems in the field of pattern recognition. This includes regression and error-weighted classification. Binary decomposition deals with splitting the target classes into several binary variables whose values are deduced by other classification models. A method involving many models and a neural networks approach fall into this category. A variety of methods to perform ordinal classification are included under threshold models. A hierarchical classification of the methods mentioned above is represented pictorially in Figure 1. Each of these methods is dealt with in detail in the following sections.

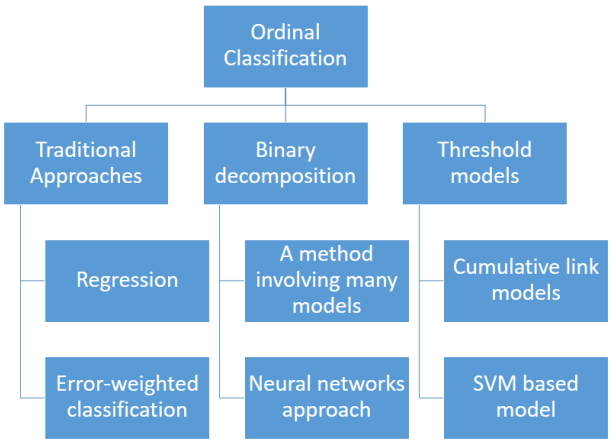

Fig -1: Hierarchy of ordinal classification method

2.1 Traditional approaches

This involves the transformation of an ordinal problem into a standardized problem with a known solution. There are mainly two methods that are traditional in nature. They are:

2.1.1 Regression

In this method, the ordinal data labels are mapped to certain real numbers [1]. The data labels are now real numbers and hence standardized regression-based algorithms and techniques are utilized for further classification. However, this technique is not devoid of shortcomings. Firstly, in ordinal data, the distance between the classes is unknown due to which it is difficult to convert ordinal numbers into a real valued entity. Secondly, the usage of these real values might hamper the regression algorithm's performance. Thirdly, regression algorithms are more concerned about the absolute weight of the label rather than the relative positioning of a particular label with respect to other labels [2]. Hence, applying regression techniques do not give correct results. Monedero [3] has proposed a method in which real numbers are chosen by observing inter-class distances of all pairs of data labels, and not randomly.

2.1.2 Error-Weighted Classification:

In this approach, the quantification of the error due to incorrect classification is carried out in accordance to ordinal scale. Tu [4] utilizes the concept of the relative distance between the predicted and actual classes while Kotsiantis [5] treats the cost variable as an absolute entity. The entire process is modelled using a cost values matrix where each element M[i][j] represents the cost of the ijth element. There are several other methods to choose the nature of the cost variable. Lin [6] assigns an asymmetric Gaussian nature to the cost variable. The major drawback of the error-weighted classification method is that there is no method to determine which nature of the cost variable is most suitable for a particular classification problem.

2.2 Binary decomposition

Another method of performing ordinal regression involves dividing the final ordinal target categories into various binary variables. The values of these variables shall be determined by one or more than one classification models.

2.2.1 A method involving many models

In the process of ordinal classification, it can be assessed whether a certain class is greater or lesser than a particular level. Clearly, this is a dual category classification problem. One can determine whether a given pattern y is greater than a specified level l. Assuming that there exist n ordinal categories and that an input must belong to one of these classes, one can assess the target that the input may belong to by carrying out the aforementioned binary classification on every class. When the outputs of all the n classes are combined, the categorical label to which the input belongs can be effectively determined [7]. In this method, a binary classifier is used on each of the n target labels. The decision regarding the final target is made depending on the probability output values of the binary classifiers.

Although the above-mentioned method performs the required classification, it does not take penalties of wrong classification into account. This point was noted by Waegeman [8] and in his work, weights were introduced such that the increase in distance of the predicted class from the actual target class increased the penalty on the system.

2.2.2 A Neural Networks Approach

In a general model of a neural network, there exists a set of input nodes, one or more hidden node layers followed by a final layer of output notes. For a general class of binary classification problems, the output layer consists of a single node whose value is either 0 or 1. But for a problem consisting of n output classes, a neural network having n nodes in the output layer is used. The normal method of training a neural network can be used if the expected output value is simply categorical. Since the problem at hand is ordinal, it is necessary to consider the relationship among the n output nodes.

Generally, for a given input vector v, a target class is determined based on a simple naïve Bayes classifier and the usage of appropriate activation functions in the hidden layer. The regular method does not take into account the inherent ordering among the output classes. Assuming that there are n output classes, there is no relationship between the conditional probability P(C1/v), P(C2/v) and P(C3/v) for regular categorical data. But it can be established that for ordinal categories, if the input vector belongs to class 3, then the probability of it belonging to class 2 is lesser than that of class 3. Thus, in mathematical notation, it can be said that – Given a set of n ordinal target categories, if an input vector v belongs to a class I, the $P(C1/v) < P(C2/v) < P(C3/v) - - - - - - -< P(Ci/v)$. Also, $P(Cn/v) < P(Cn-1/v) < P(Cn-2/v) < - - - P(Ci/v)$. This was proposed by Joaquim Pinto da Costa and Jaime S. Cardoso [9].

This ordering among the output probabilities ensures that the ordinal nature of the target categories has been taken into consideration. Various evaluation metrics can be used for this model as well. Root mean square error, Spearman's coefficient and mean absolute error can be used to check the accuracy of the proposed model.

2.3 Threshold Modelling approach

In the process of regression, if the response to a non-discrete variable is ordinal in nature, then, such a variable is called as a latent variable. The modelling approaches which work on this assumption, are called threshold- based approaches [10]. Threshold modelling is the most common approach employed to handle ordinal data.

Threshold methods are seen as an enhancement to the traditional modelling approaches. The main distinguishing factor between the two modelling schemes is that, in threshold modelling, the distance between the various classes is not known initially and is determined during further stages. Threshold modelling is also similar to the 0-1 decomposition modelling. The main distinguishing factor is that threshold modelling uses only one mapping scheme for each class, in a given interval.

2.3.1 Cumulative link modelling

The POM (Proportionally odds model) [11], belonging to the CML (Cumulative Link Modelling) approach [12] involves the probabilistic analysis of thresholds, with respect to the ordinal scale. This process has been formalized by William [13] and the formalized model is known as the generalized ordered partial proportional odds models. Peterson [14] applies probabilistic concepts only to a specialized group in the existing variable set. This approach is known as Partial proportion modelling. Tutz [15] developed the generalized semi-parametrically structured ordinal modelling approach. A non-linear POM approach was suggested by Mathieson [16]. This model was enhanced (using evolutionary algorithms) by Dorado [17]. Gestel [18] combined the concepts of logistic regression for ordinal values and non-linear kernels and came up with the Internal rating model.

Rennie [19] introduced the ordistic modelling technique as an enhancement to Cumulative Link Modelling. In this method, parameters such as logistic loss and hinge loss (considered for binary classification) are modelled based on probabilistic measures and are later, formalized. This method is a bi-threshold based modelling approach. In general, N-1 thresholds partition the data line into N parts.

2.3.2 Support vector machine-based modeling:

Support Vector Machine (SVM) is the most widely used classifier for ordinal based classification. This is because of two reasons. Firstly, SVM provides a good general performance and an appreciable value for accuracy. Secondly, the SVM structure can be easily modified to incorporate threshold-based structures.

HerbRich in his papers [20,21] laid down an SVM based approach where features were considered in pairs and a new dataset was created, containing all the values of x d ij = xi − xj and yij = O(yi) − O(yj), where yi , yj ∈ Set of classes. Shashua and Levin [22] wrote SVM based algorithms to maximize the boundaries between adjacent classes and to maximize the arithmetic total of the margins between classes. Both Herbrich's and Shashua's methods have two drawbacks. Firstly, the thresholds are not individually defined due to which the model is inadequate in fully representing an ordinal system. Secondly, there might be a mismatch in the relative ordering of data, as the ordinal inequality has not been taken into account. Keerthi and Chu [23,24] overcame these drawbacks by applying an external constraint on the problem of optimization, by considering only neighbouring class-labels for determining the threshold. The optimal solution constraint was satisfied by allowing each category to contribute error weights to the SVM hyperplane.

Carrizosa and B. Martin-Barragan [25] divided SVM classifier-based errors into two types: Upgrade errors and Downgrade errors. Upgrade errors are the errors caused by the prediction of a higher valued class label while Downgrade errors are the errors caused by the prediction of a lower valued class label. Now, the two major objectives are, to maximize both these margins at the same time and to show that all the Pareto-optimal results can be found out by solving the degree two optimization problem.

The major drawback of SVM based ordinal classification approach is that it is computationally intensive while handling large datasets. Zhao and Wang [26] introduced the concept of block quantized SVM. An incremental version of the same was proposed by Tsang, Kwok and Cheung [27].

2.3.3 Other threshold-based approaches:

Discriminant analysis for ordinal regression [28] involves the increasing of the distance between the classes while decreasing the size of each class. Discriminant analysis uses various statistical concepts such as covariance and standard deviation. Augmented binary classification [29,30] is a technique, where, an ordinal regression problem is transformed into a binary classification problem by following 3 steps. The first step uses a coding matrix in which all possible input patterns are mapped into binary classification equivalent. The second step involves training the above model with the objective of reducing the loss due to incorrect classification. The third step involves the framing of a problem-specific classification rule to carry out the final prediction. Perceptron learning [29] is a threshold modelling technique that utilizes generalized Bayesian concepts.

3. CONCLUSIONS

A generic classification approach cannot be applied while handling ordinal data. Extreme care should be taken to ensure that the relative ordering of ordinal data should be maintained at every step of the classification process. Traditional approaches involve regression and error-value based classification techniques. They are, however primitive

techniques and cannot be applied on real world data seeking efficient outputs. The many-models scheme is computationally intensive as it involves performing binary classification for all the n different categories of the output. The neural networks approach works well when ordinal classification needs to be performed on a very large dataset. Its ordering of the Bayes probabilities ensures that the ordinal nature of the data is not affected. Various threshold modelling schemes have been suggested and the results obtained out-perform the aforementioned schemes' performance. Hence, there is a need to understand the nature of the problem and apply the relevant threshold modelling scheme to ensure efficient ordinal value-based classification.

REFERENCES

[1] V. Torra, J. Domingo-Ferrer, J. M. Mateo-Sanz, and M. Ng, "Regression for ordinal variables without underlying continuous variables," Information Sciences, vol. 176, no. 4, pp. 465–474, 2006.

[2] E. F. Harrington, "Online ranking/collaborative filtering using the perceptron algorithm," in Proceedings of the Twentieth International Conference on Machine Learning (ICML2003), 2003.

[3] J. Sanchez-Monedero, P. A. Guti ´ errez, P. Tino, and C. Herv ´ as- ´ Mart´ınez, "Exploitation of pairwise class distances for ordinal classification." Neural Comput., vol. 25, no. 9, pp. 2450–2485, 2013.

[4] H.-H. Tu and H.-T. Lin, "One-sided support vector regression for multiclass cost-sensitive classification," in Proceedings of the Twenty-Seventh International Conference on Machine learning (ICML2010), 2010, pp. 49–56.

[5] S. B. Kotsiantis and P. E. Pintelas, "A cost sensitive technique for ordinal classification problems," in Methods and applications of artificial intelligence (Proc. of the 3rd Hellenic Conference on Artificial Intelligence, SETN), ser. Lecture Notes in Artificial Intelligence, vol. 3025, 2004, pp. 220–229.

[6] H.-T. Lin and L. Li, "Reduction from cost-sensitive ordinal ranking to weighted binary classification," Neural Comput., vol. 24, no. 5, pp. 1329–1367, 2012.

[7] E. Frank and M. Hall, "A simple approach to ordinal classification," in Proceedings of the 12th European Conference on Machine Learning, ser. EMCL '01. London, UK: Springer-Verlag, 2001, pp. 145–156.

[8] W. Waegeman and L. Boullart, "An ensemble of weighted support vector machines for ordinal regression," International Journal of Computer Systems Science and Engineering, vol. 3, no. 1, pp. 47– 51, 2009.

[9] da Costa J.P., Cardoso J.S. (2005) Classification of Ordinal Data Using Neural Networks. In: Gama J., Camacho R., Brazdil P.B., Jorge A.M., Torgo L. (eds) Machine Learning: ECML 2005. ECML 2005 . Lecture Notes in Computer Science, vol 3720. Springer, Berlin, Heidelberg.

[10] Verwaeren, W. Waegeman, and B. De Baets, "Learning partial ordinal class memberships with kernel-based proportional odds models," Computational Statistics & Data Analysis, vol. 56, no. 4, pp. 928–942, 2012.

[11] P. McCullagh, "Regression models for ordinal data," Journal of the Royal Statistical Society. Series B (Methodological), vol. 42, no. 2, pp. 109–142, 1980.

[12] A. Agresti, Categorical Data Analysis, 2nd ed. John Wiley and Sons, 2002.

[13] R. Williams, "Generalized ordered logit/partial proportional odds models for ordinal dependent variables," Stata Journal, vol. 6, no. 1, pp. 58–82, March 2006.

[14] B. Peterson and J. Harrell, Frank E., "Partial proportional odds models for ordinal response variables," Journal of the Royal Statistical Society, vol. 39, no. 2, pp. 205–217, 1990, series C.

[15] G. Tutz, "Generalized semiparametrically structured ordinal models," Biometrics, vol. 59, no. 2, pp. 263–273, 2003.

[16] M. J. Mathieson, "Ordinal models for neural networks," in Proceedings of the Third International Conference on Neural Networks in the Capital Markets, ser. Neural Networks in Financial Engineering, J. M. A.-P. N. Refenes, Y. Abu-Mostafa and A. Weigend, Eds. World Scientific, 1996, pp. 523–536.

[17] M. Dorado-Moreno, P. A. Gutierrez, and C. Herv´as-Mart ´ ´ınez, "Ordinal classification using hybrid artificial neural networks with projection and kernel basis functions," in 7th International Conference on Hybrid Artificial Intelligence Systems (HAIS2012), 2012, p. 319–330.

[18] T. Van Gestel, B. Baesens, P. Van Dijcke, J. Garcia, J. Suykens, and J. Vanthienen, "A process model to develop an internal rating system: Sovereign credit ratings," Decision Support Systems, vol. 42, no. 2, pp. 1131–1151, 2006.

[19] J. D. Rennie and N. Srebro, "Loss functions for preference levels: Regression with discrete ordered labels," in Proceedings of the IJCAI multidisciplinary workshop on advances in preference handling. Kluwer Norwell, MA, 2005, pp. 180–186.

[20] R. Herbrich, T. Graepel, and K. Obermayer, "Large margin rank boundaries for ordinal regression," in Advances in Large Margin Classifiers, A. Smola, P. Bartlett, B. Scholkopf, and D. Schuur- ¨ mans, Eds. Cambridge, MA: MIT Press, 2000, pp. 115–132.

[21] R. Herbrich, T. Graepel, and K. Obermayer, "Support vector learning for ordinal regression," in Proceedings of the Ninth International Conference on Artificial Neural Networks (ICANN 99), vol. 1, 1999, pp. 97–102.

[22] A. Shashua and A. Levin, "Ranking with large margin principle: two approaches," in Proceedings of the Seventeenth Annual Conference on Neural Information Processing Systems (NIPS2003), ser. Advances in Neural Information Processing Systems, no. 16. MIT Press, 2003, pp. 937–944.

[23] W. Chu and S. S. Keerthi, "Support Vector Ordinal Regression," Neural Comput., vol. 19, no. 3, pp. 792–815, 2007.
[24] W. Chu and S. S. Keerthi, "New approaches to support vector ordinal regression," in In ICML'05: Proceedings of the 22nd international conference on Machine Learning, 2005, pp. 145–152.
[25] E. Carrizosa and B. Martin-Barragan, "Maximizing upgrading and downgrading margins for ordinal regression," Mathematical Methods of Operations Research, vol. 74, no. 3, pp. 381–407, 2011.
[26] B. Zhao, F. Wang, and C. Zhang, "Block-quantized support vector ordinal regression," IEEE Trans. Neural Netw., vol. 20, no. 5, pp. 882–890, 2009.
[27] I. W. Tsang, J. T. Kwok, and P.-M. Cheung, "Core vector machines: Fast SVM training on very large data sets," J. of Machine Learning Research, vol. 6, pp. 363–392, 2005.
[28] B.-Y. Sun, J. Li, D. D. Wu, X.-M. Zhang, and W.-B. Li, "Kernel discriminant learning for ordinal regression," IEEE Trans. Knowl. Data Eng., vol. 22, no. 6, pp. 906–910, 2010.
[29] H.-T. Lin and L. Li, "Reduction from cost-sensitive ordinal ranking to weighted binary classification," Neural Comput., vol. 24, no. 5, pp. 1329–1367, 2012.
[30] L. Li and H.-T. Lin, "Ordinal regression by extended binary classification," in Advances in Neural Information Processing Systems, no. 19, 2007, pp. 865–872.