

Machine learning methods for ordinal classification with additional relative information

Mengzi Tang¹, Raúl Pérez-Fernández^{1,2}, and Bernard De Baets¹

¹ KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure links 653, 9000 Gent, Belgium

² UNIMODE, Department of Statistics and O.R. and Mathematics Didactics, University of Oviedo, C/ Federico García Lorca 18, 3307 Oviedo, Spain.
mengzi.tang@ugent.be, raul.perezfernandez@ugent.be,
bernard.debaets@ugent.be

Ordinal classification problems naturally appear in many research areas, such as medical research [2] and social sciences [3]. Machine learning methods deal with large manually-labeled datasets for solving these problems. Unfortunately, the performance of any machine learning classification method is usually limited by the amount of absolute information (i.e., examples with given class labels) and it is usually time-consuming and costly to collect a large amount of absolute information. Fortunately, gathering a large amount of additional relative information (i.e., preference orders for couples of examples) is much easier. Hence, the main challenge of this work is to combine a small amount of absolute information and a large amount of relative information for ordinal classification.

The importance of considering additional information has already been validated [6]. For example, in the field of soft-label classification [7,8,10], a basic assumption is that there is some uncertainty associated with the class label that should be associated with each of the examples [6]. Experts are invited to assign class labels to examples and provide the corresponding additional information in the form of probability scores reflecting how certain they are on the class label that should be specified. A natural way of solving the soft-label classification problem is to modify classical machine learning methods, e.g., logistic regression and support vector machines. Inspired by this idea, for our problem setting in which there is a small amount of absolute information and a large amount of relative information, we modify classical machine learning methods by combining both types of information.

Different methods have been proposed for solving ordinal classification problems. In addition to some classical methods [4], such as naive methods, ordinal binary decomposition methods and threshold methods, distance metric learning methods [1] are very popular. For example, Nguyen et al. [5] recently considered ordinal information as local constraints and proposed a method that incorporates these constraints into a distance metric learning task. Their experiments demonstrated that exploiting the order between class labels could lead to learning a better distance metric.

Copyright ©2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

In this paper, firstly, we modify some classical machine learning methods by incorporating additional relative information. Subsequently, similarly to the assumption that close examples are assumed to have the same class label, we assume that close couples tend to have the same order relation, an assumption that has been successfully validated in previous work [9]. Inspired by the work of Nguyen et al. [5], we learn a more suitable distance metric than the standard Euclidean distance metric by imposing different distance constraints for absolute and relative information, and incorporate the learned distance metric into the method of k -NN proposed in [9] for ordinal classification. Finally, we modify and test some classical machine learning methods and compare them with the proposed distance metric learning method on some benchmark datasets. The experiments show the benefits of considering additional relative information.

Keywords: Machine learning methods · Ordinal classification · Distance metric learning · Absolute information · Relative information

References

1. Bellet, A., Habrard, A., Sebban, M.: Metric learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* **9**(1), 1–151 (2015)
2. Doyle, O.M., Westman, E., Marquand, A.F., Mecocci, P., Vellas, B., Tsolaki, M., Kloszewska, I., Soininen, H., Lovestone, S., Williams, S.C., Andrew, S.: Predicting progression of alzheimers disease using ordinal regression. *Plos One* **9**(8), e105542 (2014)
3. Fullerton, A.S., Xu, J.: The proportional odds with partial proportionality constraints model for ordinal response variables. *Social Science Research* **41**(1), 182–198 (2012)
4. Gutiérrez, P.A., Pérez-Ortiz, M., Sánchez-Monedero, J., Fernández-Navarro, F., Hervás-Martínez, C.: Ordinal regression methods: survey and experimental study. *IEEE Transactions on Knowledge and Data Engineering* **28**(1), 127–146 (2016)
5. Nguyen, B., Morell, C., De Baets, B.: Distance metric learning for ordinal classification based on triplet constraints. *Knowledge-Based Systems* **142**, 17–28 (2018)
6. Nguyen, Q., Valizadegan, H., Hauskrecht, M.: Learning classification with auxiliary probabilistic information. In: *Proceedings of the 11th International Conference on Data Mining (ICDM)*. pp. 477–486. IEEE, Vancouver, BC (2011)
7. Nguyen, Q., Valizadegan, H., Hauskrecht, M.: Learning classification models with soft-label information. *Journal of the American Medical Informatics Association* **21**(3), 501–508 (2014)
8. Nguyen, Q., Valizadegan, H., Seybert, A., Hauskrecht, M.: Sample-efficient learning with auxiliary class-label information. In: *AMIA Annual Symposium Proceedings*. vol. 2011, p. 1004. American Medical Informatics Association (2011)
9. Tang, M., Pérez-Fernández, R., De Baets, B.: Fusing absolute and relative information for augmenting the method of nearest neighbors for ordinal classification. *Information Fusion*, accepted (2019)
10. Xue, Y., Hauskrecht, M.: Efficient learning of classification models from soft-label information by binning and ranking. In: *Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference*. pp. 164–169. Marco Island, Florida (May 2017)