Classificação I – 6ª Feira, 6 de Abril, Sala 1 (9h00)

Predicting Flight Departure Delay at Porto Airport: a Comparative Study

Hugo Alonso[1]
[1] *Universidade de Aveiro, CIDMA e Universidade Lusófona do Porto, hugo.alonso@ua.pt*

Abstract: Managing an airport is very complex. Decisions are often based on common sense and influence several variables, such as flight delay. In a previous work, we conducted a preliminary study where we were interested in predicting flight departure delay at Porto Airport. The problem was treated as an ordinal classification task and we built classifiers based on the unimodal paradigm. Here, we consider other approaches to ordinal classification and compare the performance of the resulting classifiers.

Keywords: Flight delay prediction, Ordinal classification.

The decisions taken in the management of an airport are often based on common sense and influence several variables, such as flight delay. Reducing this delay presents the advantage of decreasing costs and increasing the quality of the service provided to the passengers.

In (Alonso & Loureiro, 2015), we considered for the first time the challenge of predicting flight departure delay at Porto Airport. The problem is as follows. Given information about a flight that will departure from the airport, such as its arrival delay and ground operation time, we are interested in predicting in which of the following intervals the departure delay will fall:]-∞,0],]0,15],]15,30],]30,60] and]60,+∞[minutes. Since these intervals can be viewed as naturally ordered classes, the prediction problem can be treated as an ordinal classification task. In the aforementioned work, we developed classifiers based on a suitable approach to ordinal classification, called the unimodal model (Pinto da Costa et al., 2008). The main idea behind this machine learning paradigm is that the random variable class associated with a given query should follow a unimodal distribution, so that the order relation between the classes is respected. In this context, the output of a classifier where the *a posteriori* class probabilities are estimated is obliged to be unimodal. One way to impose unimodality is to assume a unimodal discrete distribution, like the binomial, and let the classifier estimate its parameters. The results we obtained with this approach were good and encouraged us to continue our investigation.

Here, we consider other approaches to ordinal classification, namely the one presented in (Frank & Hall, 2001) and the data replication method (Cardoso & Pinto da Costa, 2007), and compare the new prediction results with the ones shown in our previous work. The new approaches are briefly described next.

Suppose that the ordinal class variable takes values C1<...<CK. In (Frank & Hall, 2001), the authors proposed to use K-1 binary classifiers to address the K-class ordinal problem. In order to train those classifiers, K-1 datasets are derived from the original

dataset. The i-th classifier is trained to discriminate C1,...,Ci from Ci+1,...,CK. Given an unseen instance, the *a posteriori* probabilities of the original K classes are estimated by combining the outputs of the K-1 binary classifiers for that instance. In turn, Cardoso and Pinto da Costa (2007) proposed to use a single binary classifier to address the K-class ordinal problem. In order to train that classifier, the original dataset is extended with additional variables, based on a data replication process. Given an unseen instance, the binary classifier is applied to K-1 different replicas of that instance in the extended feature space, producing K-1 outputs. Those outputs are then combined in order to obtain a predicted class, among the original K classes, for the unseen instance.

Acknowledgments: This work was supported by the Portuguese Foundation for Science and Technology (FCT), through CIDMA - Center for Research and Development in Mathematics and Applications, within the project UID/MAT/04106/2013.

References

Alonso, H., & Loureiro, A. (2015). Predicting Flight Departure Delay at Porto Airport: a Preliminary Study. In *Proceedings of the 7th International Joint Conference on Computational Intelligence*, 3 (pp. 93-98).

Cardoso, J. S., & Pinto da Costa, J. F. (2007). Learning to classify ordinal data: the data replication method. *Journal of Machine Learning Research*, 8, 1393-1429.

Frank, E., & Hall, M. (2001). A simple approach to ordinal classification. In *Proceedings of the 12th European Conference on Machine Learning*, 1 (pp. 145-156).

Pinto da Costa, J. F., Alonso, H., & Cardoso, J. S. (2008). The unimodal model for the classification of ordinal data. *Neural Networks*, 21, 78-91.