

Understanding telecom customer churn with machine learning: from prediction to causal inference^{*}

Théo Verhelst¹, Olivier Caelen², Jean-Christophe Dewitte²,
Bertrand Lebichot¹, and Gianluca Bontempi¹

¹ Machine Learning Group, Computer Science Department,
Université Libre de Bruxelles, Brussels, Belgium
{tverhels,gbonte}@ulb.ac.be

² Data science team, Orange Belgium
{olivier.caelen,jean-christophe.dewitte}@orange.be

Telecommunication companies are evolving in a highly competitive market where attracting new customers is much more expensive than retaining existing ones [3]. Retention campaigns can be used to prevent customer churn, but their effectiveness depends on the availability of accurate prediction models. Churn prediction is notoriously a difficult problem because of the large amount of data, non-linearity, imbalance and low separability between the classes of churners and non-churners. In this paper, we discuss a real case of churn prediction based on Orange Belgium customer data.

In the first part of the paper we focus on the design of an accurate prediction model. The large class imbalance between the two classes is handled with the EasyEnsemble algorithm [4] using a random forest classifier. The dataset contains 73 variables and about 7.6 million entries, covering a 5 months time window in 2018. The classification model is trained on the first 4 months of data and evaluated on the last month. We also assess the impact of different data preprocessing techniques including feature selection and engineering. Results show that feature selection can be used to reduce computation time and memory requirements, though engineering variables does not necessarily improve performance.

In the second part of the paper we explore the application of data-driven causal inference, which allows to infer causal relationships between variables purely from observational data. More specifically, we applied 5 different causal inference methods, namely PC [6], Grow-shrink (GS) [5], Incremental Association Markov Blanket (IAMB) [7], Minimum interaction maximum relevance (mIMR) [2] and D2C [1]. PC infers the set of causal graphs faithful to the dataset, GS and IAMB infer the Markov blanket of the churn variable, and mIMR and D2C return the direct causes of churn. Two implementations of the mIMR algorithm are used: one based on histograms to estimate mutual information, and another assuming Gaussian variables, thus allowing a closed-form formula for the computation of the mutual information. The results of these algorithms (summa-

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

rized in figure 1) are varied and are consistent with prior knowledge of the causes of churn. We conclude that the bill shock and the wrong tariff plan positioning are putative causes of churn.

Finally, we present a novel method to evaluate, in terms of the direction and magnitude, the impact of causally relevant variables on churn. We evaluate the average probability of churn predicted by the learning algorithm on the dataset, before and after a shift of the values of the variable of interest. The difference between these two average probabilities is a measure of the effect of a manipulation of the variable on churn. This method is based on the assumption that no latent variables are confounding factors of churn and the variable under inspection. Results show that, on the one hand, some variables such as the tenure and the number of contracts are observed to be monotonically associated with the churn probability. On the other hand, some variables have a non-monotonic causal influence on churn. For example, variables related to the amount paid by the customer and the data usage cause more churn when they are increased, but the opposite is not true.

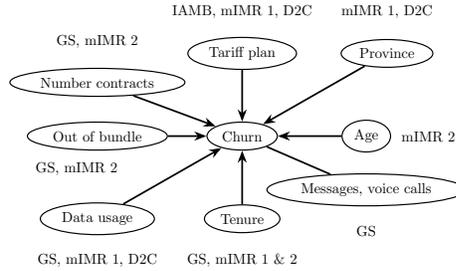


Fig. 1. Summary of results of causal inference. mIMR 1 stands for the histogram-based estimator, and mIMR 2 for the estimator with Gaussian assumption.

References

1. Bontempi, G., Flauder, M.: From dependency to causality: a machine learning approach. *The Journal of Machine Learning Research* **16**(1), 2437–2457 (2015)
2. Bontempi, G., Meyer, P.E.: Causal filter selection in microarray data. In: *Proceedings of the 27th international conference on machine learning (icml-10)*. pp. 95–102 (2010)
3. Hadden, J., Tiwari, A., Roy, R., Ruta, D.: Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research* **34**(10), 2902–2917 (2007)
4. Liu, X.Y., Wu, J., Zhou, Z.H.: Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **39**(2), 539–550 (2009). <https://doi.org/10.1109/tsmcb.2008.2007853>
5. Margaritis, D., Thrun, S.: Bayesian network induction via local neighborhoods. In: *Advances in neural information processing systems*. pp. 505–511 (2000)
6. Spirtes, P., Glymour, C.: An algorithm for fast recovery of sparse causal graphs. *Social science computer review* **9**(1), 62–72 (1991)
7. Tsamardinos, I., Aliferis, C.F., Statnikov, A.R., Statnikov, E.: Algorithms for large scale markov blanket discovery. In: *FLAIRS conference*. vol. 2, pp. 376–380 (2003)