



Subgroup Discovery for Treatment Optimization

Natalia V. Korepanova¹

International Laboratory for Intelligent Systems and Structural Analysis,
at National Research University Higher School of Economics, Moscow, Russia

Abstract

Modern medicine aspire to improve the effectiveness of treatment for some diseases through, so called, personalized medicine. However, totally personalized medicine or personalized treatment of even one disease is a very ambitious goal. Subgroup analysis of patients is a preliminary step to the total personalization. Several completely different views on the principles and usefulness of subgroup analysis for treatment personalization exist. This paper is limited to data-driven subgroup discovery, when collected data analyzed for significant treatment-biomarker interactions in post-hoc manner, and presents a brief overview of key methods for this type of subgroup analysis.

Keywords: personalized medicine, subgroup discovery, subgroup analysis.

1 Introduction

The general idea of treatment personalization is that the treatment should be tailored to a specific patient. As there is no unique definition of treatment personalization, it consolidates a variety of approaches. One of them is identification and analysis of subgroups where one of the compared treatment strategies outperforms others. Two types of subgroup analysis are distinguished in clinical trial literature: confirmatory and exploratory [1]. The main difference is that while confirmatory subgroup analysis focuses on assessment of predefined subgroups, exploratory analysis deals with subgroups defined or derived from the data in post-hoc manner and is intended to retrieve new knowledge. According to [1], exploratory analysis has several subtypes. One of them is subgroup discovery, which allows us to choose the most promising subgroups among the large family of candidate subgroups by employing data mining and machine learning algorithms. This paper presents the brief overview of subgroup discovery techniques.

The rest of the paper is organized as follows. In Section 2, we provide a reader with main definitions and notation. Section 3 presents a taxonomy of methods for data-driven subgroup discovery and overviews methods of each class of this taxonomy, and Section 4 concludes the paper.

2 Preliminaries

Let n be the number of patients in an analyzed trial, and p potential biomarkers are used to define subgroups. Let $\mathbf{x}_i = \{x_{i1}, \dots, x_{ip}\}$ denote observed values of biomarkers for the i -th patient evaluated before the start of the treatment. Let t_i denote the treatment which was

assigned to the i -th patient. Control and experimental treatments are compared, and we assume $t_i = 0$ if the i -th patient received control treatment, and $t_i = 1$, if the i -th patient received experimental treatment. Let y_i denote an outcome evaluated on the i -th patient, and higher y_i corresponds to a better response. By \mathbf{X} , Y and T we denote random variables corresponding to \mathbf{x}_i , y_i and t_i . A subgroup $S(\mathbf{X})$ is defined by a rule which selects a subset of the overall population based on the vector of potential biomarkers \mathbf{X} .

Let $f(\mathbf{x}, t) = E(Y|\mathbf{X} = \mathbf{x}, T = t)$ denote the expected response of a patient as a function of potential biomarkers \mathbf{X} and treatment assignment T at \mathbf{x} and t , respectively. As the higher values of Y are preferable, we assume that the goal of personalization is for any \mathbf{x} to choose t which maximizes $f(\mathbf{x}, t)$. In [2] a mild condition $E(Y|\mathbf{X}, T) = h(\mathbf{X}, z(\mathbf{X})T)$, where $z(\cdot)$ is monotone and $h(\cdot)$ is monotone with respect to its second argument, is proposed. Under this condition the choice of treatment depends only on the second argument $z(\mathbf{X})T$. The particular case is when $f(\mathbf{X}, T)$ is modeled by $g(q(\mathbf{X}) + l(z(\mathbf{X})T))$, where $g(\cdot)$ and $l(\cdot)$ are monotone functions. In this class of models prognostic biomarkers are those that contribute only to $q(\mathbf{X})$, whereas predictive biomarkers also contribute to $z(\mathbf{X})$. Hence, under mentioned condition the choice of treatment depends only on predictive biomarkers and the goal of the methods for treatment personalization is to recover $z(\mathbf{X})$. Then, one possible way to define the subgroup of interest is $z(\mathbf{X}) > \delta \Rightarrow \mathbf{X} \in S$, where δ is some appropriate threshold.

One of the easiest examples of f is $f(\mathbf{X}, T) = f(\mathbf{X}, 0) + z(\mathbf{X})T$. Here $z(\mathbf{X})$ can be rewritten as $f(\mathbf{X}, 1) - f(\mathbf{X}, 0)$.

3 Principled approaches to subgroup discovery

A general taxonomy of principled approaches to subgroup discovery and biomarker evaluation was proposed in [3] and slightly modified in [1]:

- Global outcome modeling methods (modeling $f(\mathbf{X}, T)$).
- Global treatment effect modeling methods (modeling $z(\mathbf{X})$).
- Local modeling methods (direct search for subgroups with higher values of $z(\mathbf{X})$ without modeling it).

While in [1] a detailed overview on the methods of the proposed classes was presented, this paper briefly retells it and supplements it with several groups of methods which were not mentioned, but still fit the taxonomy.

Before we move on to the first class of the taxonomy we should mention basic methods, such as univariate regression approach and tree-based regression approach. In univariate regression approach, a series of regression models with treatment, biomarker and treatment-biomarker interaction components of the expected response are fitted. The resulting biomarkers are those biomarkers for which the interaction component of the corresponding regression model is significant at a pre-specified significance level. These resulting biomarkers are used to define subgroups such that each subgroup is defined by one resulting biomarker, no interaction between the biomarkers are taken into account. Such models do not work well when predictive component $z(\mathbf{X})$ depends on two independent biomarkers, for instance, $z(\mathbf{X}) = X_1 X_2$, where X_1 and X_2 are two independent continuous biomarkers.

In tree-based regression approach, which is usually based on CART methodology [4], both treatment indicator and potential biomarkers are used as predictors, and the outcome variable is used as a target. Basic methods of this type allow us to find more complex interactions and

to define subgroups through several biomarkers. As this approach does not distinguish between patient’s features and a treatment variable, the resulting model may not contain treatment, which means that it is not a model of treatment personalization. Another case of not resulting in a model of treatment personalization is when only prognostic biomarkers are found due to the higher effect of the prognostic biomarkers on the outcome than of the predictive ones (which seems to be likely in practice). In [1] a reader can also find some artificial examples of datasets when basic methods do not perform well. Hence, neither univariate regression approach nor tree-based regression approach to modeling the outcome variable are not good solutions for treatment personalization and should be used only as basic benchmarks for other methods.

Global outcome modeling. The goal of the methods from this class is to fit the model of response with a potentially large number of variables selected from candidate biomarkers and treatment-biomarker interactions. The resulting global model is then used to estimate the differential treatment effect (e.g. $f(\mathbf{X}, 1) - f(\mathbf{X}, 0)$) and select subsets of patients by defining thresholds for the estimated differential treatment effect. Parametric methods in this class of methods are often based on regularized regression models. For instance, in [5] the authors assume that prognostic and predictive components should be regularized separately as the influence of predictive components is usually lower than of prognostic ones. A large range of methods of Bayesian nature exists in this class too (e.g. [8]).

Another group of methods in this class deals with the concept of potential outcomes. Two potential outcomes $\tilde{Y}_i(1)$ and $\tilde{Y}_i(0)$ for the i -th patient are hypothetical outcomes of treatment assignments $T = 1$ and 0 to the i -th patient, respectively. If the i -th patient received treatment $T = t$ then $\tilde{Y}_i(t) = Y_i$ and $\tilde{Y}_i(1 - t)$ is unobserved and is estimated from the data. Potential outcomes can be estimated both parametrically (e.g. in [6] through Cox regression model) and non-parametrically (e.g. in [7] through random forest).

Global treatment effect modeling. In contrast to global outcome modeling, methods of this class do not model prognostic effects, they focus only on predictive biomarkers and model treatment contrast as a function of biomarkers. As for parametric approaches, in [12] it was shown that assuming equal randomization to treatment arms we can consider the problem of fitting $z(\mathbf{X})$ as a classical supervised learning problem.

A bunch of methods of this class is based on recursive partitioning procedures which split the space of biomarkers into non-overlapping subgroups. That procedures greedily maximize the difference between two parts of a split with respect to $z(\mathbf{X})$ at every step. The examples of such methods are interaction trees [9], model-based recursive partitioning [10], qualitative interaction trees [11] and so on. Here we could say that these methods provide us with a piecewise models of $z(\mathbf{X})$.

In this class of methods we also should notice separately methods of optimal treatment regimes modeling. A treatment regime (individual treatment rule) $d(\mathbf{X})$ is defined as the function that maps the vector of biomarkers to one of the treatment strategies ($T = 0$ or 1). Optimal treatment regime is defined as $d_{opt}(\mathbf{X}) = E[\tilde{Y}(1)d(\mathbf{X}) + \tilde{Y}(0)(1 - d(\mathbf{X}))]$, the rule that maximizes an expected reward [13]. The observation that the problem of fitting optimal treatment regime may be transformed into the problem of fitting optimal treatment contrast led to outcome-weighted learning methodology for estimating optimal treatment rules [14].

Another group of methods came from modeling of marketing data and is called uplift modeling. Uplift is defined as $m(\mathbf{X}) = E(Y|\mathbf{X}, T = 1) - E(Y|\mathbf{X}, T = 0) = f(\mathbf{X}, 1) - f(\mathbf{X}, 0)$, which is a variant of $z(\mathbf{X})$. The first uplift modeling methods modeled $f(\mathbf{X}, 1)$ and $f(\mathbf{X}, 0)$ independently [15], therefore they were from the class of methods modeling global outcome. Nevertheless, we consider uplift modeling as a version of global treatment effect modeling because today there

is a series of uplift modeling methods which model directly the uplift function, for instance: regression-based [16], tree-based [17], SVM-based [18], and even ensemble methods [19].

Local modeling. This is the third class of methods for subgroup discovery. The methods which can be referred to this class do not intend to model $z(\mathbf{X})$ or $f(\mathbf{X}, T)$ over the entire space of potential biomarkers, but study subsets of this space and try to identify subsets with desirable properties.

One of the examples of local modeling methods is subgroup identification based on differential effect search [20] which has a recursive partitioning nature. At each step parent subgroup is split into two child subgroups such that the treatment effect in one child subgroup differs from the effect in the other one as much as possible, and the splitting process stops when some predefined condition has been satisfied. The main difference from modeling global treatment effect is that there is no need to predict $z(\mathbf{X})$ on the whole feature space, it is just enough to find regions with high treatment effect.

In [21] another idea how to discover subgroups in a tree-based manner locally is presented. To begin, a measure of patient similarity is defined and is computed for all pairs of patients assigned to different treatment regimes. Some patients are allowed to be incomparable. With the help of Deferred-Acceptance procedure [22] with similarity relation using for preferences pairs of similar patients receiving different treatment are obtained. As Deferred-Acceptance does not aim at finding stable and optimal matching, the obtained pairs may not consist of the most similar patients, and this fact can be thought, on the one hand, as a disadvantage of the method or, on the other hand, as preservation from overfitting, that should be explored experimentally. To decide which treatment is better for each obtained pair of patients, a meaningful rule is defined and each pair is marked with the corresponding label (control treatment is better, experimental treatment is better, or no difference). For instance, for continuous Y , $t_i = 0$ and $t_j = 1$, where i and j are similar patients from one of the obtained pairs, we say that control treatment is better when $y_i \geq y_j$, and experimental treatment is better when $y_i < y_j$. Finally, to obtain interpretable descriptions of subgroups a classification tree is learned with given marks.

The last example of local modeling methods is a pattern-structure based method [23]. Its key difference from other recursive partitioning algorithms is that it does not optimize some measure of goodness of split locally, but try to arrange locally obtained subgroups into a global structure (lattice) and derive the most promising subgroups from it. The general idea is to reduce an exhaustive search of all possible subgroups in a considered class of subgroups through mining only those specifications of subgroups, which contain maximal number of biomarkers describing the same subset of patients. However, the pattern-structure based search of subgroups remains exponential in time complexity.

4 Conclusion

Treatment personalization is tightly bound to statistics, data mining and machine learning. Although these disciplines progress rapidly, their adaptation to medical solutions often faces more problems than to other areas as the cost of errors is human health and life. Moreover, some medical problems may require the invention of new methods of data analysis. In this paper, we show the variety of approaches to data-driven subgroup discovery for treatment personalization which aims at exploratory identification of subgroups of patients with significant prevalence in effect of one treatment strategy over another. Hence, in this paper we find a place for uplift modeling, matching-based and pattern-structure-based methods in the taxonomy of data-driven

exploratory subgroup discovery methods proposed earlier in [1].

Acknowledgments

The article was supported within the framework of project 16-29-1298216 by the Russian Foundation for Basic Research.

References

- [1] Lipkovich, I., Dmitrienko, A., D'Agostino, R.B.: Tutorial in Biostatistics: Data-Driven Subgroup Identification and Analysis in Clinical Trials. *Stat. Med.* (2016)
- [2] Xu, Y., Yu, M., Zhao, Y.Q., Li, Q., Wang, S., Shao, J.: Regularized Outcome Weighted Subgroup Identification for Differential Treatment Effects. *Biometrics*. 71, 645–653 (2015)
- [3] Lipkovich, I., Dmitrienko, A.: Biomarker identification in clinical trials. In: Carini, C., Chang, M. (eds.) *Clinical and Statistical Considerations in Personalized Medicine.*, pp. 211–264. Chapman and Hall/CRC Press: New York (2014).
- [4] Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: *Classification and Regression Trees*. Wadsworth: Belmont, CA (1984)
- [5] Imai, K., Ratkovic, M.: Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*. 7, 443–470 (2013)
- [6] Cai, T., Tian, L., Wong, P., Wei, L.J.: Analysis of randomized comparative clinical trial data for personalized treatment selections. *Biostatistics*. 12, 270–282 (2011)
- [7] Foster, J.C., Taylor, J.M.C., Ruberg, S.J. Subgroup identification from randomized clinical trial data. *Stat. Med.* 30, 2867–2880 (2011)
- [8] Gu, X., Yin, G., Lee, J.J.: Bayesian two-step lasso strategy for biomarker selection in personalized medicine development for time-to-event endpoints. *Contemp. Clin. Trials*. 36, 642–650 (2013)
- [9] Su, X., Zhou, T., Yan, X., Fan, J., Yang, S.: Interaction trees with censored survival data. *The International Journal of Biostatistics*. 4, Article 2 (2008)
- [10] Zeileis, A., Hothorn, T., Hornik, K.: Model-based recursive partitioning. *J. Comput. Graph. Stat.* 17, 492–514 (2008)
- [11] Dusseldorp, E., Van Mechelen, I.: Qualitative interaction trees: a tool to identify qualitative treatment-subgroup interactions. *Stat. Med.* 33, 219–237 (2014)
- [12] Tian, L., Alizaden, A.A., Gentles, A.J., Tibshirani, R.: A simple method for detecting interactions between a treatment and a large number of covariates. *J. Am. Stat. Assoc.* 109, 1517–1532 (2014)
- [13] Qian, M., Murphy, S.A.: Performance guarantees for individualized treatment rules. *The Annals of Statistics*. 39, 1180–1210 (2011)
- [14] Zhao, Y., Zheng, D., Rush, A.J., Kosorok, M.R.: Estimating individualized treatment rules using outcome weighted learning. *J. Am. Stat. Assoc.* 107, 1106–1118 (2012)
- [15] Radcliffe, N.J., Surry, P.D.: Real-world uplift modeling with significance-based uplift trees. *Portrait Technical Report TR-2011-1, stochastic solutions*, Tech. rep. (2011)
- [16] Jaskowski, M., Jaroszewicz, S.: Uplift modeling for clinical trial data. In: *ICML, 2012 workshop on machine learning for clinical data analysis*. Edinburgh, Scotland (2012)
- [17] Rzepakowski, P., Jaroszewicz, S.: Decision trees for uplift modeling. In: *Proceedings of the 10th IEEE International conference on data mining (ICDM)*, pp. 441–450 Sydney, Australia (2010)
- [18] Zaniewicz, L., Jaroszewicz, S.: Lp - Support vector machines for uplift modeling. *Knowl. Inf. Syst.* (2017)
- [19] Soltys, M., Jaroszewicz, S., Rzepakowski, P.: Ensemble methods for uplift modeling. *Data Min. Knowl. Disc.* 29, 1531–1559 (2015)

- [20] Lipkovich, I., Dmitrienko, A., Denne, J., Enas, G.: Subgroup identification based on differential effect search (SIDES): a recursive partitioning method for establishing response to treatment in patient subpopulations. *Stat. Med.* 30, 2601–2621 (2011)
- [21] Korepanova, N., Kuznetsov, S.O., Karachunskiy, A.I.: Matchings and Decision Trees for Determining Optimal Therapy. In: *Analysis of Images, Social Networks and Texts Third International Conference, AIST 2014, Yekaterinburg, Russia, April 10-12, Revised Selected Papers*, pp. 101–110. Springer International Publishing (2014)
- [22] Roth, A.: Deferred acceptance algorithms: history, theory, practice, and open questions. *Int. J. Game Theory.* 36, 537–569 (2008)
- [23] Korepanova, N., Kuznetsov, S.O.: Pattern Structures for Treatment Optimization. In: *CLA 2016: Proceedings of the Thirteenth International Conference on Concept Lattices and Their Applications*, pp. 217–228. CEUR Workshop Proceedings (2016)