Statistics and Its Interface Volume 11 (2018) 421–422

Discussion on "Double sparsity kernel learning with automatic variable selection and data extraction"

Yuan Huang and Shuangge Ma*

Chen et al. (2018) presented a kernel learning method with double sparsity penalties to achieve variable selection and data extraction simultaneously. In this article, we highlight the authors' contributions and provide several remarks that may be worth further discussions and exploration.

Keywords and phrases: Kernel learning, Data extraction, Variable selection, Tuning parameter selection.

We first would like to congratulate Chen, Zhang, Kosorok, and Liu for their exceptional contribution. Although there has been extensive literature on high-dimensional learning, this study complements and significantly advances from the existing ones in multiple important aspects. First, most of the existing studies have focused on learning with linear predictor effects. There are a few that can accommodate nonparametric effects. However, usually a sieve approach, with a set of (sometimes subjectively) specified basis functions, is adopted. The proposed learning, in contrast, is based on the Reproducing Kernel Hilbert Spaces (RKHS) technique, which is popular under the "classic" nonparametric learning paradigm. It is a great pleasure to see that this work, along with a few others, successfully adapt the RKHS technique in high-dimensional learning. The second aspect, which is quite interesting but has been much less studied, is data extraction. It is argued that by extracting "more useful" data points, prediction performance can be improved. To achieve the two different goals simultaneously, the authors proposed combining two penalties in an additive manner. Consistency properties are very nicely established. Although there is no "surprise," as the authors pointed out which we highly agree with, it is "comforting" to observe similar consistency properties with a more complicated learning objective and penalties.

As a feature selection technique, similar to those in the literature, the proposed approach can be the most useful for data with high-dimensional covariates. This is shown in Theorem 2, where a $\log(p)/\sqrt{n}$ rate is established. As discussed above, this study has an equal emphasis on data extraction. Intuitively, it may not be desirable to discriminate observations when the sample size is small. Thus, at least to us, the most appropriate scenario for the proposed approach may be the one with a large sample size and ultrahigh-dimensional covariates. As can be partly seen from the numerical studies in this article, such data may not be popular at this moment. But, with the progression of the big data era, we firmly believe that such data will soon become common.

Several aspects, which have been studied in this article, may be worth additional discussions and exploration. *Nonparametric learning* is an important feature of this study. As suggested in this study and other published, the RKHS, sieve, and other techniques may all provide satisfactory theoretical results. However, our personal experience is that the application aspect may need additional efforts. Specifically, most data analyzed in the literature with high-dimensional covariates have limited sample sizes. As a result, the estimated nonparametric curves are often overly smoothed, and the boundary problem can be more serious. *Data extraction* is a relatively new topic and has been explored in only a few recent publications. Feature extraction is easy to comprehend: it can reduce the noise level as well as the number of parameters to be estimated, both of which can improve estimation. Data extraction, in contrast, may be less lucid. In this study, as well as others published, there is no concern on data mixture or contamination. As such, the goal of data extraction is not to systematically remove undesirable observations. Under the IID assumption, our personal intuition is that each observation may contain useful information for a small region of the sample space. Thus removing an observation may lead to a loss of information on that small region. In addition, removing observations leads to a reduced sample size. Of course, it can be desirable to remove observations when there is a high level of redundancy (some observations are "crowded together"). However, this should have a very low probability if the observations are i.i.d. random. The authors provided very convincing discussions from a prediction perspective. It will be interesting for readers including us to see more results on data extraction, including, for example, diagnostic tools which may suggest the necessity of data extraction, theoretical results firmly establishing the improvement in prediction (and corresponding data/model conditions), as well as more finite-sample numerical studies. Another "old" problem is *tuning parameter selection*.

*Corresponding author.

As shown in this study, as well as a few others, to achieve more complex analysis goals, it is inevitable to have more regularizations (penalties in this case) and thus more tunings. For data with small sample sizes and high-dimensional covariates, the presence of multiple tunings and associated problems, such as high computational cost and lack of stability, have been challenging. With a nonparametric nature of the proposed approach, such problems can potentially be even worse. For numerical convenience, the authors proposed to fix the value of λ_3. As a more generic problem, how to deal with multiple tunings still remains open.

We firmly believe that the proposed strategy of combining the RKHS technique and data extraction will pave the road for multiple follow-up studies. The proposed method and computational algorithm, with the ever increasing data volume, will have many practical applications. Theoretical developments will also shed insights into other relevant studies. We would like to appraise the authors again for their contribution to SII and the broad statistics community.

Received 2 March 2018

REFERENCES

[1] Chen, J., Zhang, C., Kosorok, M. R., and Liu, Y. (2018). Double sparsity kernel learning with automatic variable selection and data extraction. *Statistics and Its Interface*. In press.

Yuan Huang
Department of Biostatistics
University of Iowa
USA
E-mail address: yuan-huang@uiowa.edu

Shuangge Ma
Department of Biostatistics
Yale School of Public Health
USA
E-mail address: shuangge.ma@yale.edu