Density Ratio Estimation: A New Versatile Tool for Machine Learning

Masashi Sugiyama

Department of Computer Science, Tokyo Institute of Technology

A new general framework of statistical data processing based on the ratio of probability densities has been proposed recently and gathers a great deal of attention in the machine learning and data mining communities [1–17]. This density ratio framework includes various statistical data processing tasks such as non-stationarity adaptation [18, 1, 2, 4, 13], outlier detection [19–21, 6], and conditional density estimation [22–24, 15]. Furthermore, mutual information—which plays a central role in information theory [25]—can also be estimated via density ratio estimation. Since mutual information is a measure of statistical independence between random variables [26–28], density ratio estimation can be used also for variable selection [29, 7, 11], dimensionality reduction [30, 16], and independent component analysis [31, 12].

Accurately estimating the density ratio is a key issue in the density ratio framework. A naive approach is to estimate each density separately and then take the ratio of the estimated densities. However, density estimation is known to be a hard problem and thus this two-step approach may not be accurate in practice. A promising approach would be to estimate the density ratio directly without going through density estimation. Several direct density ratio estimation methods have been proposed so far, including kernel mean matching [3], logistic regression [32, 33, 5], the Kullback-Leibler importance estimation procedure [8, 9], least-squares importance fitting [10, 17], and unconstrained least-squares importance fitting [10, 17]. Note that the importance refers to the density ratio, derived from *importance sampling* [34]. Furthermore, a density ratio estimation method incorporating dimensionality reduction has also be developed [35], which is shown to be useful in high-dimensional cases. A review of the density ratio framework is available in [36].

The density ratio approach was shown to be useful in various real-world applications including brain-computer interface [4, 37], robot control [38–41, 15], speaker identification [42, 43], natural language processing [14], bioinformatics [11], visual inspection of precision instruments [44], spam filtering [5], and HIV therapy screening [45].

The author thanks MEXT Grant-in-Aid for Young Scientists (A) 20680007, SCAT, and AOARD for their support.

References

1. Zadrozny, B.: Learning and evaluating classifiers under sample selection bias. In: Proceedings of the Twenty-First International Conference on Machine Learning, New York, NY, ACM Press (2004) 903–910
2. Sugiyama, M., Müller, K.R.: Input-dependent estimation of generalization error under covariate shift. Statistics & Decisions **23**(4) (2005) 249–279

3. Huang, J., Smola, A., Gretton, A., Borgwardt, K.M., Schölkopf, B.: Correcting sample selection bias by unlabeled data. In Schölkopf, B., Platt, J., Hoffman, T., eds.: Advances in Neural Information Processing Systems 19. MIT Press, Cambridge, MA (2007) 601–608
4. Sugiyama, M., Krauledat, M., Müller, K.R.: Covariate shift adaptation by importance weighted cross validation. Journal of Machine Learning Research **8** (May 2007) 985–1005
5. Bickel, S., Brückner, M., Scheffer, T.: Discriminative learning for differing training and test distributions. In: Proceedings of the 24th International Conference on Machine Learning. (2007) 81–88
6. Hido, S., Tsuboi, Y., Kashima, H., Sugiyama, M., Kanamori, T.: Inlier-based outlier detection via direct density ratio estimation. In Giannotti, F., Gunopulos, D., Turini, F., Zaniolo, C., Ramakrishnan, N., Wu, X., eds.: Proceedings of IEEE International Conference on Data Mining (ICDM2008), Pisa, Italy (Dec. 15–19 2008) 223–232
7. Suzuki, T., Sugiyama, M., Sese, J., Kanamori, T.: Approximating mutual information by maximum likelihood density ratio estimation. In Saeys, Y., Liu, H., Inza, I., Wehenkel, L., de Peer, Y.V., eds.: JMLR Workshop and Conference Proceedings. Volume 4 of New Challenges for Feature Selection in Data Mining and Knowledge Discovery. (2008) 5–20
8. Sugiyama, M., Nakajima, S., Kashima, H., von Bünau, P., Kawanabe, M.: Direct importance estimation with model selection and its application to covariate shift adaptation. In Platt, J.C., Koller, D., Singer, Y., Roweis, S., eds.: Advances in Neural Information Processing Systems 20, Cambridge, MA, MIT Press (2008) 1433–1440
9. Sugiyama, M., Suzuki, T., Nakajima, S., Kashima, H., von Bünau, P., Kawanabe, M.: Direct importance estimation for covariate shift adaptation. Annals of the Institute of Statistical Mathematics **60**(4) (2008) 699–746
10. Kanamori, T., Hido, S., Sugiyama, M.: Efficient direct density ratio estimation for nonstationarity adaptation and outlier detection. In Koller, D., Schuurmans, D., Bengio, Y., Botton, L., eds.: Advances in Neural Information Processing Systems 21, Cambridge, MA, MIT Press (2009) 809–816
11. Suzuki, T., Sugiyama, M., Kanamori, T., Sese, J.: Mutual information estimation reveals global associations between stimuli and biological processes. BMC Bioinformatics **10**(1) (2009) S52
12. Suzuki, T., Sugiyama, M.: Estimating squared-loss mutual information for independent component analysis. In Adali, T., Jutten, C., Romano, J.M.T., Barros, A.K., eds.: Independent Component Analysis and Signal Separation. Volume 5441 of Lecture Notes in Computer Science., Berlin, Springer (2009) 130–137
13. Quiñonero-Candela, J., Sugiyama, M., Schwaighofer, A., Lawrence, N., eds.: Dataset Shift in Machine Learning. MIT Press, Cambridge, MA (2009)
14. Tsuboi, Y., Kashima, H., Hido, S., Bickel, S., Sugiyama, M.: Direct density ratio estimation for large-scale covariate shift adaptation. Journal of Information Processing **17** (2009) 138–155
15. Sugiyama, M., Takeuchi, I., Suzuki, T., Kanamori, T., Hachiya, H.: Least-squares conditional density estimation. Technical Report TR09-0004, Department of Computer Science, Tokyo Institute of Technology (Feb. 2009)
16. Suzuki, T., Sugiyama, M.: Sufficient dimension reduction via squared-loss mutual information estimation. Technical Report TR09-0005, Department of Computer Science, Tokyo Institute of Technology (Feb. 2009)
17. Kanamori, T., Hido, S., Sugiyama, M.: A least-squares approach to direct importance estimation. Journal of Machine Learning Research (2009) to appear.
18. Shimodaira, H.: Improving predictive inference under covariate shift by weighting the log-likelihood function. Journal of Statistical Planning and Inference **90**(2) (2000) 227–244

19. Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: LOF: Identifying density-based local outliers. In Chen, W., Naughton, J.F., Bernstein, P.A., eds.: Proceedings of the ACM SIGMOD International Conference on Management of Data. (2000)
20. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Neural Computation **13**(7) (2001) 1443–1471
21. Hodge, V., Austin, J.: A survey of outlier detection methodologies. Artificial Intelligence Review **22**(2) (2004) 85–126
22. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
23. Takeuchi, I., Le, Q.V., Sears, T.D., Smola, A.J.: Nonparametric quantile estimation. Journal of Machine Learning Research **7** (2006) 1231–1264
24. Takeuchi, I., Nomura, K., Kanamori, T.: Nonparametric conditional density estimation using piecewise-linear solution path of kernel quantile regression. Neural Computation **21**(2) (2009) 533–559
25. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, Inc., New York, NY, USA (1991)
26. Kraskov, A., Stögbauer, H., Grassberger, P.: Estimating mutual information. Physical Review E **69**(6) (2004) 066138
27. Hulle, M.M.V.: Edgeworth approximation of multivariate differential entropy. Neural Computation **17**(9) (2005) 1903–1910
28. Suzuki, T., Sugiyama, M., Tanaka, T.: Mutual information approximation via maximum likelihood estimation of density ratio. In: Proceedings of 2009 IEEE International Symposium on Information Theory (ISIT2009), Seoul, Korea (Jun. 28–Jul. 3 2009) 463–467
29. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research **3**(Mar) (2003) 1157–1182
30. Song, L., Smola, A., Gretton, A., Borgwardt, K.M., Bedo, J.: Supervised feature selection via dependence estimation. In: Proceedings of the 24th International Conference on Machine learning, New York, NY, USA, ACM (2007) 823–830
31. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)
32. Qin, J.: Inferences for case-control and semiparametric two-sample density ratio models. Biometrika **85**(3) (1998) 619–639
33. Cheng, K.F., Chu, C.K.: Semiparametric density estimation under a two-sample density ratio model. Bernoulli **10**(4) (2004) 583–604
34. Fishman, G.S.: Monte Carlo: Concepts, Algorithms, and Applications. Springer-Verlag, Berlin (1996)
35. Sugiyama, M., Kawanabe, M., Chui, P.L.: Dimensionality reduction for density ratio estimation in high-dimensional spaces. Neural Networks to appear.
36. Sugiyama, M., Kanamori, T., Suzuki, T., Hido, S., Sese, J., Takeuchi, I., Wang, L.: A density-ratio framework for statistical data processing. IPSJ Transactions on Computer Vision and Applications (2009) to appear.
37. Li, Y., Koike, Y., Sugiyama, M.: A framework of adaptive brain computer interfaces. In: Proceedings of the 2nd International Conference on BioMedical Engineering and Informatics (BMEI09), Tianjin, China (Oct. 17–19 2009) to appear.
38. Hachiya, H., Akiyama, T., Sugiyama, M., Peters, J.: Adaptive importance sampling with automatic model selection in value function approximation. In: Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (AAAI2008), Chicago, Illinois, USA, The AAAI Press (Jul. 13–17 2008) 1351–1356
39. Hachiya, H., Akiyama, T., Sugiyama, M., Peters, J.: Adaptive importance sampling for value function approximation in off-policy reinforcement learning. Neural Networks (2009) to appear.

40. Akiyama, T., Hachiya, H., Sugiyama, M.: Active policy iteration: Efficient exploration through active learning for value function approximation in reinforcement learning. In: Proceedings of the Twenty-first International Joint Conference on Artificial Intelligence (IJCAI2009), Pasadena, California, USA (Jul. 11–17 2009) to appear.
41. Hachiya, H., Peters, J., Sugiyama, M.: Efficient sample reuse in EM-based policy search. In: Machine Learning and Knowledge Discovery in Databases. Lecture Notes in Computer Science, Berlin, Springer (2009) to appear.
42. Yamada, M., Sugiyama, M., Matsui, T.: Covariate shift adaptation for semi-supervised speaker identification. In: Proceedings of 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2009), Taipei, Taiwan (Apr. 19–24 2009) 1661–1664
43. Yamada, M., Sugiyama, M., Matsui, T.: Semi-supervised speaker identification under covariate shift. Signal Processing (2009) to appear.
44. Takimoto, M., Matsugu, M., Sugiyama, M.: Visual inspection of precision instruments by least-squares outlier detection. In: Proceedings of The Fourth International Workshop on Data-Mining and Statistical Science (DMSS2009), Kyoto, Japan (Jul. 7–8 2009) 22–26
45. Bickel, S., Bogojeska, J., Lengauer, T., Scheffer, T.: Multi-task learning for HIV therapy screening. In McCallum, A., Roweis, S., eds.: Proceedings of 25th Annual International Conference on Machine Learning (ICML2008), Helsinki, Finland, Omnipress (Jul. 5–9 2008) 56–63