entropy regularization. Meanwhile, the semi-supervised information for metric learning problems can also be given in terms of a set of pairwise similarity and dissimilarity constraints [1, 2, 5, 28]. A well-known metric learning method with these constrains was proposed by Xing et al. [28]. Following this work, there are several emerging works [1, 2, 5] which study the metric learning problems by exploiting the given relevant constraints. Additionally, there are some works [6, 10–12] that address the multiple instance metric learning problem where the training dataset is provided as a set of labeled bags. They aim to learn a distance metric, which makes bags that share a label closer, and pushes bags that do not share any label apart [6, 11]. However, the above discussed metric learning works fail to deal with the probabilistic class labels.

Learning from such probabilistic information is of great importance [29]. Some works in other fields [9, 15, 16, 18, 29] also consider how to learn models from the probabilistic labels. However, the problem settings in these papers are quite different from ours. For example, the authors in [9] present class ratio models, which take as input an unlabeled set of data and predict the proportions of instances in the set belonging to different classes.

7 CONCLUSIONS

In this paper, we first propose an instance-level metric learning mechanism (InML), based on which the distance metrics can be learned directly from the instance-wise probabilistic labels. Compared with the existing metric learning methods, InML can fully utilize the probabilistic information and learn a more accurate metric. For the cases where the datasets are associated with group-wise probabilistic labels, we design a group-level metric learning mechanism (GrML), which can learn distance metrics directly from the group-wise probabilistic labels with high accuracy. Both theoretical analysis and extensive experiments on real-world datasets are provided to demonstrate the advantages of the proposed metric learning mechanisms.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and helpful suggestions. This work is supported in part by the US National Science Foundation under grants IIS-1218393 and IIS-1514204. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Mahdieh Soleymani Baghshah and Saeed Bagheri Shouraki. 2009. Semi-Supervised Metric Learning Using Pairwise Constraints. In *Proceedings of the International Joint Conference on Artificial Intelligence*. 1217–1222.

[2] Aharon Bar-Hillel, Tomer Hertz, Noam Shental, and Daphna Weinshall. 2005. Learning a mahalanobis metric from equivalence constraints. *Journal of Machine Learning Research* 6, Jun (2005), 937–965.

[3] Qiong Cao, Zheng-Chu Guo, and Yiming Ying. 2016. Generalization bounds for metric and similarity learning. *Machine Learning* 102, 1 (2016), 115–132.

[4] Olivier Chapelle, Vikas Sindhwani, and Sathiya S Keerthi. 2008. Optimization techniques for semi-supervised support vector machines. *Journal of Machine Learning Research* 9, Feb (2008), 203–233.

[5] Jason V Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S Dhillon. 2007. Information-theoretic metric learning. In *Proceedings of the 24th international conference on Machine learning*. ACM, 209–216.

[6] Matthieu Guillaumin, Jakob Verbeek, and Cordelia Schmid. 2010. Multiple instance metric learning from automatically labeled bags of faces. In *Proceedings of the European conference on Computer Vision*. Springer, 634–647.

[7] Mengdi Huai, Chenglin Miao, Qiuling Suo, Yaliang Li, Jing Gao, and Aidong Zhang. 2018. Uncorrelated Patient Similarity Learning. In *Proceedings of the 2018 SIAM International Conference on Data Mining*. SIAM, 270–278.

[8] Yinjie Huang, Cong Li, Michael Georgiopoulos, and Georgios C Anagnostopoulos. 2013. Reduced-rank local distance metric learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 224–239.

[9] Arun Shankar Iyer, J Saketha Nath, and Sunita Sarawagi. 2016. Privacy-preserving class ratio estimation. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 925–934.

[10] Rong Jin, Shijun Wang, and Zhi-Hua Zhou. 2009. Learning a distance metric from multi-instance multi-label data. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*. IEEE, 896–902.

[11] Marc T Law, Yaoliang Yu, Raquel Urtasun, Richard S Zemel, and Eric P Xing. 2017. Efficient multiple instance metric learning using weakly supervised data. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.

[12] Dewei Li and Yingjie Tian. 2016. Multi-view metric learning for multi-instance image classification. *arXiv preprint arXiv:1610.06671* (2016).

[13] Weiwei Liu and Ivor W Tsang. 2015. Large Margin Metric Learning for Multi-Label Prediction.. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 15. 2800–2806.

[14] Gang Niu, Bo Dai, Makoto Yamada, and Masashi Sugiyama. 2014. Information-theoretic semi-supervised metric learning via entropy regularization. *Neural computation* 26, 8 (2014), 1717–1762.

[15] Giorgio Patrini, Richard Nock, Paul Rivera, and Tiberio Caetano. 2014. (Almost) no label no cry. In *Advances in Neural Information Processing Systems*. 190–198.

[16] Peng Peng, Raymond Chi-Wing Wong, and Phillp S Yu. 2014. Learning on probabilistic labels. In *Proceedings of the 2014 SIAM International Conference on Data Mining*. SIAM, 307–315.

[17] Filipe Rodrigues, Francisco Pereira, and Bernardete Ribeiro. 2014. Gaussian process classification and active learning with multiple annotators. In *International Conference on Machine Learning*. 433–441.

[18] Stefan Rueping. 2010. SVM classifier estimation from group probabilities. In *Proceedings of the 27th international conference on machine learning (ICML-10)*. 911–918.

[19] Shai Shalev-Shwartz and Shai Ben-David. 2014. *Understanding machine learning: From theory to algorithms*. Cambridge university press.

[20] Kihyuk Sohn. 2016. Improved deep metric learning with multi-class n-pair loss objective. In *Advances in Neural Information Processing Systems*. 1857–1865.

[21] Jimeng Sun, Fei Wang, Jianying Hu, and Shahram Edabollahi. 2012. Supervised patient similarity measure of heterogeneous patient records. *ACM SIGKDD Explorations Newsletter* 14, 1 (2012), 16–24.

[22] Tao Sun, Dan Sheldon, and Brendan OâĂŹConnor. 2017. A Probabilistic Approach for Learning with Label Proportions Applied to the US Presidential Election. In *Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, 445–454.

[23] Qiuling Suo, Fenglong Ma, Ye Yuan, Mengdi Huai, Weida Zhong, Aidong Zhang, and Jing Gao. 2017. Personalized Disease Prediction Using a CNN-Based Similarity Learning Method. In *Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*.

[24] Tian Tian and Jun Zhu. 2015. Max-margin majority voting for learning from crowds. In *Advances in Neural Information Processing Systems*. 1621–1629.

[25] Dong Wang and Xiaoyang Tan. 2014. Robust Distance Metric Learning in the Presence of Label Noise. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 1321–1327.

[26] Kilian Q Weinberger, John Blitzer, and Lawrence K Saul. 2006. Distance metric learning for large margin nearest neighbor classification. In *Advances in neural information processing systems*. 1473–1480.

[27] Kilian Q Weinberger and Lawrence K Saul. 2009. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research* 10, Feb (2009), 207–244.

[28] Eric P Xing, Michael I Jordan, Stuart J Russell, and Andrew Y Ng. 2003. Distance metric learning with application to clustering with side-information. In *Advances in neural information processing systems*. 521–528.

[29] Felix X Yu, Dong Liu, Sanjiv Kumar, Tony Jebara, and Shih-Fu Chang. 2013. SVM for learning with label proportions. *In Internalization conference on machine learning* (2013).

[30] Pourya Zadeh, Reshad Hosseini, and Suvrit Sra. 2016. Geometric mean metric learning. In *International Conference on Machine Learning*. 2464–2471.

[31] Mengting Zhan, Shilei Cao, Buyue Qian, Shiyu Chang, and Jishang Wei. 2016. Low-rank sparse feature selection for patient similarity learning. In *Proceeding of the 2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE.

[32] Dengyong Zhou, Qiang Liu, John Platt, and Christopher Meek. 2014. Aggregating ordinal labels from crowds by minimax conditional entropy. In *International conference on machine learning*. 262–270.