Label Distribution Learning

Xin Geng[1] and Zhi-Hua Zhou[2]

[1] Key Lab of Computer Network and Information Integration, MOE,
Southeast University, Nanjing 211189, China
[2] National Key Lab for Novel Software Technology,
Nanjing University, Nanjing 210023, China

1 Introduction

Learning with ambiguity has been a hot topic for years in the areas of machine learning and data mining. Multi-label learning (MLL) [8] has been successfully applied to many real-world problems with ambiguity because it can well model the question "*what describes the instance*", i.e., which labels can be used to describe the instance. However, MLL can hardly deal with the more general question "*how to describe the instance*", i.e., the exact role of each label is also involved in the description of the instance. Surprisingly, the real-world data where such label role matters might be more common than most people think because when multiple labels are associated to the same instance, it is unlikely that they happen to be equally important to the instance. For example, it was discovered [9] that one particular kind of protein (instance) could be related to several cancers (labels), and the expression levels of the protein in these related cancer cells are generally different – higher expression level (label roles) indicates closer relationship. Another example is the mixture emotion conveyed by facial expressions. According to Plutchik's wheel of emotions theory [7], there are a small number of basic emotions, and all the other emotions occur as combinations, mixtures or compounds of the basic emotions and can exist in varying degree of intensity or levels of arousal. Each facial expression (instance), on the contrary to the single emotion assumption of most existing expression recognition methods, usually reflects a mixture of basic emotions (labels) with different intensities (label roles). For such data, a natural way to label an instance $\boldsymbol{x}$ is to assign a real number $d_{\boldsymbol{x}}^{y}$ to each possible label y, representing the degree to which y describes $\boldsymbol{x}$. For example, if $\boldsymbol{x}$ represents a protein, y represents a cancer, then $d_{\boldsymbol{x}}^{y}$ should be the expression level of the protein $\boldsymbol{x}$ in the cancer y. Without loss of generality, assume that $d_{\boldsymbol{x}}^{y} \in [0, 1]$. Further suppose that the label set is complete, i.e., using all the labels in the set can always fully describe the instance. Then, $\sum_y d_{\boldsymbol{x}}^{y} = 1$. Such $d_{\boldsymbol{x}}^{y}$ is called the *description degree* of y to $\boldsymbol{x}$. For a particular instance $\boldsymbol{x}$, the description degrees of all the labels (suppose there are c different labels), $\boldsymbol{d} = [d_{\boldsymbol{x}}^{y_1}, \ldots, d_{\boldsymbol{x}}^{y_c}]^{\mathrm{T}}$, constitute a data form similar to probability distribution. So, it is called *label distribution*. The learning process on the instances labeled by label distributions is therefore called *label distribution learning* (LDL).

2 Methods and Applications

Currently, our work on LDL follows two different tracks. The first track is from the classification point of view [4, 6, 2]. In detail, the mapping from the instance $\boldsymbol{x}$ to the label distribution $\boldsymbol{d}$ is modeled by a conditional probability mass function $p(y|\boldsymbol{x})$. Given an instance $\boldsymbol{x}$, its label distribution is generated by calculating $p(y_j|\boldsymbol{x})$ for $j = 1, \ldots, c$. Since the way of calculating $p(y_j|\boldsymbol{x})$ label by label is analogous to the classification process, such LDL methods are categorized as the classification style. The second track is from the regression point of view [1]. In detail, the mapping from the instance $\boldsymbol{x}$ to the label distribution $\boldsymbol{d}$ is modeled by a multivariate regressor $\boldsymbol{d} = f(\boldsymbol{x})$. Different from regular regression problems, the output of $f(\boldsymbol{x})$ is multivariate under the constraint of label distribution, i.e., $d_{\boldsymbol{x}}^{y_j} \in [0, 1]$ and $\sum_j d_{\boldsymbol{x}}^{y_j} = 1$.

By the definition of LDL, both single-label learning (only one label is with the description degree 1, and all others are with 0s) and multi-label learning (all the relevant labels are with equal description degrees, and all the irrelevant labels are with 0s) can be viewed as special cases of LDL. Thus, LDL is a general learning framework with many potential applications. Particularly, the following three scenarios might characterize the advantages of LDL.

1. There is a natural measure of description degree that associates the labels with the instances. For example, genetic analysis [2], and crowd opinion prediction [1].
2. When there are multiple, inconsistent labeling sources for one instance, a good way to incorporate all the sources is to generate a label distribution for the instance. One typical application is multilabel ranking for nature scene images [3].
3. Some labels are highly correlated with other labels. LDL provides a new way to explicitly model such correlation. Typical applications of this kind include age estimation [4, 6], head pose estimation [5], and crowd counting [10].

References

1. Geng, X., Hou, P.: Pre-release prediction of crowd opinion on movies by label distribution learning. In: Proc. Int'l Joint Conf. Artificial Intelligence. p. in press. Buenos Aires, Argentina (2015)
2. Geng, X., Ji, R.: Label distribution learning. In: Proc. 2013 Int'l Conf. Data Mining Workshops. pp. 377–383. Dallas, TA (2013)
3. Geng, X., Luo, L.: Pre-release prediction of crowd opinion on movies by label distribution learning. In: Proc. Int'l Conf. Computer Vision and Pattern Recognition. pp. 3742–3747. Columbus, OH (2014)
4. Geng, X., Smith-Miles, K., Zhou, Z.H.: Facial age estimation by learning from label distributions. In: Proc. 24th AAAI Conf. Artificial Intelligence. pp. 451–456. Atlanta, GA (2010)
5. Geng, X., Xia, Y.: Head pose estimation based on multivariate label distribution. In: Proc. Int'l Conf. Computer Vision and Pattern Recognition. pp. 1837–1842. Columbus, OH (2014)
6. Geng, X., Yin, C., Zhou, Z.H.: Facial age estimation by learning from label distributions. IEEE Trans. Pattern Anal. Machine Intell. 35(10), 2401–2412 (2013)
7. Plutchik, R.: Theories of Emotion, chap. A general psychoevolutionary theory of emotion, pp. 3–33. Academic Press, New York (1980)
8. Tsoumakas, G., Katakis, I.: Multi-label classification: An overview. Int'l Journal of Data Warehousing and Mining 3(3), 1–13 (2007)
9. Wulfkuhle, J.D., Liotta, L.A., Petricoin, E.F.: Proteomic applications for the early detection of cancer. Nature Reviews Cancer 3(4), 267–275 (2003)
10. Zhang, Z., Wang, M., Geng, X.: Crowd counting in public video surveillance by label distribution learning. Neurocomputing p. in press (2015)