Multi-Label Classification using Class Relations Based on Higher-Order MRF Optimization

Ryosuke Furuta Yusuke Fukushima Toshihiko Yamasaki Kiyoharu Aizawa
Department of Information and Communication Engineering, The University of Tokyo
{furuta, fukushima, yamasaki, aizawa}@hal.t.u-tokyo.ac.jp

Abstract

In multi-label classification problems, in which multiple labels are predicted for an input data, the labels assigned to a particular data are often correlated. Although a method based on first-order MRF optimization has recently been proposed in order to treat such correlations, it can only deal with the relationships between two labels. Therefore, we propose a method that is the extension of the previous method to higher-order MRF. Experimental results show that our method can successfully deal with the relationships among arbitrary combination of labels and achieves higher accuracy than the previous method.

1. Introduction

Multi-label classification is a task where multiple binary labels ("yes" or "no") are assigned to input data for each pre-defined class. Most methods proposed for multi-label classification assign each label independently using a binary classifier such as support-vector machine (SVM) trained separately to predict a single label. However, in many cases of multi-label classification, the labels assigned to a particular data are often correlated. To tackle the problem, some CRF/MRF-based methods that deal with the such correlations have been recently proposed. Yamaguchi *et al*. [4] proposed a CRF-based method that leverages the Pearson correlation between attributes of clothing pairs. Yamasaki *et al*. [5] proposed an MRF-based method that considers the relationships among different label types and the relationships among different feature types within a single joint optimization framework, and applied their method to predicting user impressions on a video presentation. However, their methods [4, 5] can treat only pairs of labels because their formulations are both first-order CRF/MRF.

Therefore, in this paper, we extend the method by Yamasaki *et al*. [5] to a higher-order MRF in order to treat arbitrary combinations of labels, in other words, combinations of more than three labels. Experimental results show that our method achieves higher accuracy than their method on their TED dataset. In addition, we present the results on another task: predicting attribute labels on an image (aPascal & aYahoo dataset [1]).

2. Label assignment based on first-order MRF

In this section, we briefly review the label assignment method based on first-order MRF proposed by [5].

Let $l_i \in \{0,1\}$ be the binary label of the i-th class, and $\boldsymbol{l} \in \{0,1\}^n$ be the labeling of all classes $(i = 1, \ldots, n)$. Here, $l_i = 1$ means that the i-th class is positive, and $l_i = 0$ corresponds to negative meaning. We model the labeling problem as a Markov random field (MRF):

$$E(\boldsymbol{l}) = \sum_i \phi_i(l_i) + \beta \sum_{i<j} \psi_{i,j}(l_i, l_j). \tag{1}$$

β balances the unary and pairwise terms. The unary term ϕ_i is the sum of the classification scores by the m classifiers learned by different types of features.

$\psi_{i,j}$ is the pairwise term that represents the relationship between the i-th and the j-th classes, which is defined as:

$$\psi_{ij}(l_i, l_j) = m(1 - \frac{N_{ij}^{l_i l_j}}{N_{ij}}), \tag{2}$$

where N_{ij} is the number of training data that both the i-th and the j-th classes are labeled. $N_{ij}^{l_i l_j}$ is the number of training data in which l_i is assigned to i-th class and l_j is assigned to j-th class. By minimizing the Eq. (1), the optimal labeling that takes account of both the decisions of classifiers and the relationships between labels can be obtained.

3. Extension to Higher-Order MRF

We extend the Eq. (1) to $(k-1)$-th order energy function in order to treat the combinations of k labels.

$$\begin{aligned} E(\boldsymbol{l}) = \sum_i \phi_i(l_i) + \beta_1 \sum_{i_1<i_2} \psi^1_{i_1 i_2}(l_{i_1}, l_{i_2}) + \cdots \\ + \beta_{k-1} \sum_{i_1<i_2<\cdots<i_k} \psi^{k-1}_{i_1 i_2 \cdots i_k}(l_{i_1}, l_{i_2}, \cdots, l_{i_k}). \end{aligned} \tag{3}$$

Table 1. Accuracy comparison on each dataset.

	TED	aPascal & aYahoo
Only SVMs	89.2%	91.9%
First-order MRF [5]	93.3%	**92.2%**
Higher-order MRF (ours)	**93.7%**	**92.2%**

The $(k-1)$-th order term $\psi^{k-1}_{i_1 i_2 \cdots i_k}$ is the extension of the first-order pairwise term in Eq. (2) and defined as

$$\psi^{k-1}_{i_1 i_2 \cdots i_k}(l_{i_1}, l_{i_2}, \cdots, l_{i_k}) = m\left(1 - \frac{N^{l_{i_1} l_{i_2} \cdots l_{i_k}}_{i_1 i_2 \cdots i_k}}{N_{i_1 i_2 \cdots i_k}}\right), \quad (4)$$

where $N_{i_1 i_2 \cdots i_k}$ is the number of training data that all of i_1-th to i_k-th classes are labeled. $N^{l_{i_1} l_{i_2} \cdots l_{i_k}}_{i_1 i_2 \cdots i_k}$ is the number of training data in which labels $l_{i_1}, l_{i_2}, \cdots, l_{i_k}$ are assigned to $i_1, i_2, \cdots, i_k$ classes, respectively.

We can convert the $(k-1)$-th order energy function in Eq. (3) to a first order energy function while keeping the global minimum of the energy unchanged by using ELC reduction proposed by [2]. After that, we can obtain the global minimum of the energy by using QPBO [3].

4. Experimental Results

We conducted the experiments on two different datasets. We used first-order, second-order, and third-order terms in our experiments (i.e., $k = 4$ in Eq. (3)). The parameters of our method were tuned for each dataset.

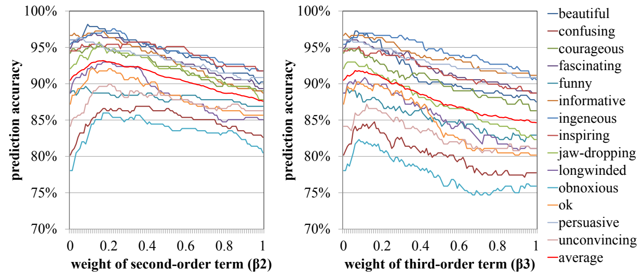
4.1. TED dataset

We applied the proposed method to impression prediction of oral presentations [5]. The setting of this experiment is same as [5]. 1,646 presentation videos and six types of features proposed by [5] were used. 14 types of impression labels were predicted. For each impression voted by viewers, the top/bottom 10% videos are labeled as positive/negative instances. We employed the SVM with a radial basis function (RBF) kernel as the classifier. The accuracy was calculated by the leave-one-out method.

The left column of Table 1 shows the average prediction accuracy of the 14 types of impression labels on TED dataset. We observe that the best performance is achieved 93.7% by our higher-order MRF.

The performance of the impression prediction is shown as a function of β_2 in Fig. 1(a) where β_1 and β_2 are fix to 0. We observe that the prediction accuracy is improved as the β_2 is increased up to a certain point ($\beta_2 = 0.2$) and gradually get degraded because the label relationships become more dominant than the label outputs from the classifiers.

The performance of the impression prediction is shown as a function of β_3 is shown in Fig. 1(b) when β_1 and β_2 is fixed to 0. We observe that the similar transition of the prediction performance to Fig. 1(a), and that both second-order and third-order terms are effective for the prediction.



(a) function of β_2 (b) function of β_3

Figure 1. Prediction accuracy as functions of β_2 and β_3.

4.2. aPascal & aYahoo dataset

aPascal & aYahoo dataset [1] consists of 4,340 images from Pascal VOC 2008 and 2,237 images from Yahoo. There are 12,695 and 2,644 objects in those images respectively, and binary labels for 64 types of attributes and bounding box information are assigned to each object. We cropped each object using the bounding box information and made 15,339 new images in total. We used the 4,096-D deep feature of the pre-trained AlexNet. We employed the linear SVMs, and used 6,340 images of aPascal training set for training the SVMs. We used 6,355 images of aPascal validation set for calculating pairwise and higher-order terms, and used all 2,644 images of aYahoo for test.

The right column of Table 1 shows the average prediction accuracy of 64 types of attribute labels. The accuracy of the baseline method is 91.9%, which predicts each label independently. The first-order MRF [5] improves the accuracy to 92.2%. We observe that our higher-order MRF achieves comparable level of accuracy.

5. Conclusion

We proposed an MRF-based label assignment method that considers the relationship among arbitrary combination of labels. Our method achieved better or comparable results on two different datasets, compared with the previous method that is based on the first-order MRF [5]. For further improvement of accuracy, the feedback from MRF to classifiers is one of the future directions.

References

[1] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *Proc. CVPR*, 2009. 1, 2

[2] H. Ishikawa. Higher-order clique reduction without auxiliary variables. In *Proc. CVPR*, 2014. 2

[3] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary mrfs via extended roof duality. In *Proc. CVPR*, 2007. 2

[4] K. Yamaguchi, T. Okatani, K. Sudo, K. Murasaki, and Y. Taniguchi. Mix and match: Joint model for clothing and attribute recognition. In *Proc. BMVC*, 2015. 1

[5] T. Yamasaki, Y. Fukushima, R. Furuta, L. Sun, K. Aizawa, and D. Bollegala. Prediction of user ratings of oral presentations using label relations. In *Proc. ACMMM-ASM*, 2015. 1, 2

This work was supported by Grants-in-Aid for Scientific Research (No. 26700008 and 16J07267) from JSPS and Microsoft IJARC core10.