More Correlations Better Performance: Fully Associative Networks for Multi-label Image Classification

Yaning Li , Liu Yang*.
College of Intelligence and Computing . Tianjin University. Tianjin 300350
lyn2018216114@tju.edu.cn. yangliuyl@tju.edu.cn

Introduction

Recent researches demonstrate that correlation modeling plays a key role in high-performance multi-label classification methods. However, existing methods do not take full advantage of correlation information, especially correlations in feature and label spaces of each image, which limits the performance of correlation-based multi-label classification methods. With more correlations considered, in this study, a Fully Associative Network (FAN) is proposed for fully exploiting correlation information, which involves both visual feature and label correlations. Specifically, FAN introduces a robust covariance pooling to summarize convolution features as global image representation for capturing feature correlation in the multi-label task. Moreover, it constructs an effective label correlation matrix based on a re-weighted scheme, which is fed into a graph convolution network for capturing label correlation. Then, correlation between covariance representations (i.e., feature correlation) and the outputs of GCN (i.e., label correlation) are modeled for final prediction.

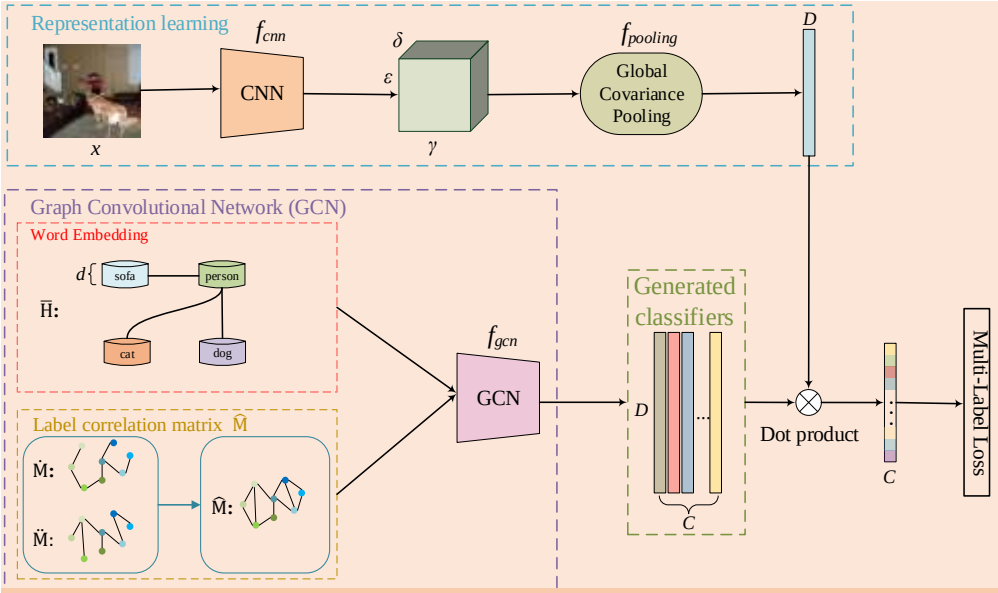

Fig.1 The overall FAN framework for multi-label image recognition

Algorithm

Algorithm 1 Multi-label Image Classification via Fully Associative Networks

Input: Training data $\boldsymbol{X}$ and the corresponding label $\boldsymbol{Y}$, test data $\dot{\boldsymbol{X}}$, word-embedding $\bar{\boldsymbol{H}}$.

Output: Predicted results of test data

Init: Pre-process the input images, pre-processing $\boldsymbol{X}$ to get label correlation matrix $\dot{\boldsymbol{M}}$.

repeat
- Reshape the $\varepsilon \times \gamma \times \delta$ tensor of last convolution layer to a matrix $\boldsymbol{S}$ consisting of $\eta = \gamma\varepsilon$ feature of δ-dimension
- Perform second-order pooling $\boldsymbol{\Omega}=\boldsymbol{S}\bar{\boldsymbol{I}}\boldsymbol{S}^T$ with $\bar{\boldsymbol{I}} = \frac{1}{\eta}(\boldsymbol{I} - \frac{1}{\eta}\boldsymbol{1})$
- Pre-normalize to get $\boldsymbol{A}$ use Eq.(8)
- **for** $k = 1; k \leq \rho; k++$ **do**
  - $\boldsymbol{B}_0 = \boldsymbol{A}, \boldsymbol{Z}_0 = \boldsymbol{I}$
  - **if** $k < \rho$ **then**
    - Get $\boldsymbol{B}_k, \boldsymbol{Z}_k$ with Eq.(6) and Eq.(7)
  - **else**
    - Perform post-compensation $\hat{\boldsymbol{X}}$ with Eq.(9)
- Obtain the learned classifiers $\hat{\boldsymbol{H}}$ from $\bar{\boldsymbol{H}}$ and $\hat{\boldsymbol{M}}$ through GCN
- Apply the learned classifiers to get the representation by $\hat{\boldsymbol{Y}} = \hat{\boldsymbol{H}}\hat{\boldsymbol{X}}$
- Calculate multi-label loss function L with Eq.(3)

until convergence

Get the predicted label by using the trained classifier on the test data

return Predicted results $\hat{\boldsymbol{Y}}$.

Label Correlation Matrix

$\dot{M}$: label pairs matrix of all training data

$\ddot{M}$: All samples in the training set are input into CNN, and the image level matrix is obtained by covariance operation

$$\ddot{M} = Y^T\tilde{X}\tilde{X}^TY$$
$$\tilde{M} = \alpha\dot{M} + (1-\alpha)\ddot{M}$$

Conclusion

Fully Associative Network (FAN) which captures feature and label correlation by introducing a robust covariance pooling and graph convolution network with an effective label correlation matrix