Improved Image Annotation and Labelling through Multi-label Boosting

Authors: Matthew Johnson and Roberto Cipolla
Presenter: Duy Vu

Problems

- Training:
  - Each image is provided with a list of descriptive words.
  - Each image is segmented into many segments. The label of each segment is not defined in the training data.

- Testing:
  - A list of descriptive words for a given image.
  - A label for each segment of the given image.

Applications

- Recognize multiple objects in one image
- Enable users search images using descriptive text

Challenges

- No information about the correspondence between a segment and a word.
- Multi-class and multi-label classification problem.
- Use of multi-label training data which are more available than single labeled images.

MLBoost

- A basic statistical learning model to link blobs and descriptive words. Blobs are color feature vectors extracted from segments.
- Multi-label boosting algorithm (MLBoost).
- How to revise the basic model into a weak learner for MLBoost.
Linking Blobs to Words

• A mixture of models.
• Each model which represents a hidden concept is a joint probability distribution of blobs and words.
• We make up the hidden concepts and assume that blobs and words are only linked through these concepts.

\[ P(w, b \mid c) = P(w \mid c)P(b \mid c) \]

Linking Blobs to Words

• Using the model for labeling:

\[ P(w \mid b) = \sum_c P(w, c \mid b) = \sum_c P(w \mid c)P(c \mid b) \]

• Using the model for annotation:

\[ P(w \mid B) = \sum_c P(w, c \mid B) = \sum_c P(w \mid c)P(c \mid B) \]

\[ B = \bigcup_{b \in \text{image}} b \]

Linking Blobs to Words

• Need to learn \( P(w \mid c), P(c \mid b), \) and \( P(c \mid B) \) from the training data.
• Use iterative EM algorithm to deal with hidden variables (hidden concepts).
• E-step:

\[ P(c \mid b) \propto P(b \mid c)P(c) \]
\[ P(c \mid B) \propto \sum_{b \in B} P(c \mid b) = \sum_{b \in B} P(b \mid c)P(c) \]
\[ P(c \mid w, B) \propto P(w \mid c)P(c \mid B) \]

Linking Blobs to Words

• M-step:

\[ P(c) \propto \sum_c \left[ \sum_b P(c, b) + \sum_w P(c \mid w, d) \right] \]
\[ P(b \mid c) \propto \text{Normal}(\mu_c, \sigma_c) \]
\[ \mu_c = \frac{\sum_b b \times P(c \mid b)}{\sum_b P(c \mid b)} \]
\[ \sigma_c = \frac{\sum_b (b - \mu_c)^T \times (b - \mu_c) \times P(c \mid b)}{\sum_b P(c \mid b)} \]
\[ P(w \mid c) \propto \sum_{b \in B} P(c \mid w, B) \]

Multi-label Boosting

• Boosting algorithm is to learn different weak learners through re-weighting training data.
• Place more weights on training examples that are misclassified by the current weak learner.
• This also means trying to guide the weak learner towards the classes that are hard to classify.

Multi-label Boosting

• After training the current learner, we evaluate it on training set using the pseudoloss measure.
• And then re-weight training examples using pseudoloss.
• Pseudoloss for Multi-class single-label boosting:

\[ ploss_i = \frac{1}{2} (1 - b(x_i, y_i) + \sum_i q(i, y) b(x_i, y)) \]
Multi-label Boosting

- Pseudoloss for Multi-class multi-label boosting:
  \[ p_{loss} = \frac{1}{2} \left( 1 - \frac{\sum_{i \in Y_j} h(x_i, y_i)}{|Y_j|} + \sum_{i \in \neg Y_j} q(i, y)h(x_i, y) \right) \]
- Provide the weak learner with the example weighting function \( D_t \) and the label weighting function \( q_t \).

The Weak Learner

- E-step:
  \[ P(c \mid w, B) \propto \sum_{b \in \hat{B}} [1 - q_t(b, w)] P(w \mid c) P(b \mid c) P(c) \]
- \( q_t(b, w) = 0 \) if \( w \) is a correct label for \( b \)
- M-step:
  \[ \mu_c = \frac{\sum_{b \in \hat{B}} c(b) \times D_t(b)}{\sum_{b \in \hat{B}} P(c \mid b) \times D_t(b)} \]
  \[ \sigma^2_c = \frac{\sum_{b \in \hat{B}} (b - \mu_c)^2 \times P(c \mid b) \times D_t(b)}{\sum_{b \in \hat{B}} P(c \mid b) \times D_t(b)} \]

Experiments

- 1881 images from the Corel database.
- 1667 ≈ 90% for training.
- Annotate each test image with \( N \) words (\( N \) is true number of words assigned to the image).
- Compute the percentage of correct annotation for each image.
- Count the number of images whose percentage of correct annotation from 0%-20%,…, 80%-100%

Annotation Results

Labeling Results

- No quantitative evaluation measure was presented since hand labeling each segment of all images is necessary.

Advantages and Disadvantages

- Advantages:
  - Exploit the availability of the captions of images
- Disadvantages:
  - Time complexity
  - Space complexity
Conclusion

• Both of two papers are on linking images and words.
• The first paper uses a complex hierarchical statistical model (LDA) to capture the joint probability distribution of blobs and words.
• The second paper uses a simpler model, a 1-level mixture of models, but it boosts the performance of this model through the multi-label boosting framework.

Code

• The binary code of MLBoost is available at http://mi.eng.cam.ac.uk/~mj293/mlboost/

Thank You!