Application of related data automatic semantic annotation technology in Internet of things

Lianwang Zhao, Hai Huang

School of Computer Science ,Beijing University of Posts and Telecommunications, Beijing,China

Key words: semi-supervised learning; video correlation; video annotation; semantic relevance; linear neighborhood propagation

Abstract. specific to semi-supervised learning method based on graph ignoring the problem of video correlation in research and application of multimedia, a kind of video annotation algorithm based on related kernel mapping linear neighborhood propagation is put forward. Firstly, the propagation coefficient of the iteration annotation is calculated in the algorithm with the kernel function according to the adjusted distance of semi-supervised learning; secondly, the sample of the low-lever feature space is obtained by using the propagation coefficient; thirdly, the correlation table between the semantic concepts is constructed according to video correlation modeling; finally, the structure of the nearest neighborhood graph is constructed; use the labeled video information to carry out iterative propagation to unlabeled video so as to complete video annotation. Experimental results show that: this algorithm can not only improve the accuracy of video annotation but also can make up for the lack of the number of video data labeled.

Introduction

With the use of storage devices and digital devices and the development of multimedia technology, video data shows the trend of geometric growth. How to organize and retrieve the video data efficiently has become an urgent problem to be solved. At the same time, in the index and retrieval of video, the biggest problem is how to narrow the "semantic gap" between the low-level features of the video and the needs of users.

When organizing and retrieving the video, the common method is to extract "metadata" that can make semantic level representation of video content from video and retrieve the video by using the metadata. Video semantic annotation (often referred to as video concept detection, high level semantic feature extraction and so on) is a basic way to obtain these metadata. Meanwhile, video annotation based on content can complete the mapping from the mapping from the low-lever feature of the video to the semantic concept so that the "semantic gap" is divided into two smaller gaps [1]; namely: (1) the gap between the low-lever feature of the video and the semantic concept; (2) the gap between semantic concepts to user needs.

Introduction of LNP algorithm

Marker propagation algorithm is one of the semi-supervised learning methods [7] based on graph and solves classification problems by using the form of propagating annotation on the graph. In the literature [8], linear neighborhood propagation is put forward, which is a kind of marker propagation algorithm. Generally speaking, all the semi-supervised learning algorithms are directly or indirectly based on the so-called cluster hypothesis: (1) adjacent points may belong to a class; (2) the possibility of points that belong to the same cluster or the structure of the submanifold belonging to the same class is also very big. Linear neighborhood propagation inherits the basic assumptions: each sample can be reconstructed with linear weighting by its neighborhood sample; semantic labeling of the sample can also be reconstructed with linear weighting by its neighborhood sample; the reconstructed weighting coefficient is the same as coefficient of sample reconstruction. The formula is described as follows:

$$x_i = \sum_{x_j \in N(x_i)} \alpha_j x_j \Rightarrow f_i = \sum_{x_j \in N(x_i)} \alpha_j f_j$$

In the formula, f_i is the semantic annotation; $N(x_i)$ is the neighborhood sample of x_i

LNP algorithm approximates the entire graph G through a series of overlapping linear neighbor hood blocks.

Edge weight W in the figure is obtained by the solution of standard quadratic programming problem. Then, combine the edge weight to form the weight matrix of G. LNP algorithm can be divided into two steps: (1) structure of the neighborhood graph; (2) spread the label with data to the remaining data without label.

Assumed that $X = \{x_1, x_2 ... x_l, x_{l+1} ... x_n\}$ are n data objects in the space of Rd; $L = \{1, -1\}$ is the label set; the previous l data points of $x_i \in X (1 \le i \le l)$ are the data labeled; the corresponding label is $L_i \in L$; the latter $n - l$ points are the data not labeled. The corresponding graph of X is $G = (V, E)$; vertex set $V = X$; E is the set of edges; for the relationship between data points x_i and x_j, edge weight is $w_{i,j}$.

The whole graph G is constructed by using the neighborhood information of each data point in LNP algorithm [8]. For convenience, it is assumed that each data point can be linearly constructed by its neighborhood and can achieve the best. Therefore, the objective of optimization is:

$$\varepsilon = \min \sum_i \left\| x_i - \sum_{j: x_j \in N(x_i)}^n w_{ij} x_j \right\|^2$$

Where, x_j is the j neighborhood point of x_i; $N(x_i)$ is the set of k neighborhood points of x_i; w_{ij} is the contribution of x_j to x_i. Add two constraints, $\sum_{j: x_j \in N(x_i)}^n w_{ij} = 1, w_{ij} \ge 0$, it can be seen clearly that the greater the degree of similarity between x_j and x_i is, the greater value of w_{ij} is. Therefore, the similarity between x_j and x_i is measured by w_{ij} It is known that $w_{i,j} \ne w_{j,i}$; matrix $\mathbf{G}^i$ of Gram of data point x_i is defined as: $\mathbf{G}^i_{jk} = (x_i - x_j)^{\mathrm{T}}(x_i - x_k)$; the further derivation is as follows:

$$\varepsilon_i = \left\| x_i - \sum_{j: x_j \in N(x_i)}^n w_{ij} x_j \right\|^2$$

$$= \left\| \sum_{j: x_j \in N(x_i)}^n w_{ij}(x_i - x_j) \right\|^2$$

$$= \sum_{j,k: x_j, x_k \in N(x_i)}^n w_{ij} w_{ik} (x_i - x_j)^{\mathrm{T}} (x_i - x_k)$$

$$= \sum_{j,k: x_j, x_k \in N(x_i)}^n w_{ij} \mathbf{G}^i_{jk} w_{ik}$$

Reconstruction weights for each data point can be obtained by the solution of label quadratic programming problem:

$$\min w_{ij} \sum_{j,k: x_j, x_k \in N(x_i)}^n w_{ij} \mathbf{G}^i_{jk} w_{ik}$$

$$\text{s.t.} \sum_{j:x_j \in N(x_i)}^{n} w_{ij} = 1, w_{ij} \geq 0$$

After finishing solving such n quadratic programming problems, weight matrix $\mathbf{W}$ can be constructed. The initialization of weight matrix is:

$$\mathbf{W}_{ij} = \begin{cases} w_{ij} & x_j \in N(x_i) \\ 0 & x_j \notin N(x_i) \end{cases}$$

Experimental result

How to make a reasonable evaluation for the performance of label is a key issue in the design and implementation of a label system and is also the purpose of our experiment. System performance in this paper is evaluated by Average Precision (AP) and Average Recall (AR) of official performance measurement. The former indicates the degree of accuracy which reflects the ability to reduce the noise of the system; the latter reflects the ability of whether the label is comprehensive and whether there is a missing detection or not. Mean Average Precision (MAP) and Mean Average Recall (MAR) can be obtained by averaging APs and ARs in 10 concepts, which is the final evaluation measure. The calculation method [14] is as follows:

$$MAP = \frac{1}{k} \sum_{i=1}^{k} precision[c_i], \quad MAR = \frac{1}{k} \sum_{i=1}^{k} recall[c_i]$$

Where, k is the total number of semantic, c_i is the i semantic. $precision[c_i]$, $recall[c_i]$

The calculating formula is:

$$precision[c_i] = \frac{N_{correct}[c_i]}{N_{pLabel}[c_i]}$$

$$recall[c_i] = \frac{N_{correct}[c_i]}{N_{label}[c_i]}$$

Where, $N_{correct}[c_i]$ is the number of video shots in test set correctly labeled by c_i; $N_{pLabel}[c_i]$ is the number of video shots in test set labeled by c_i; $N_{label}[c_i]$ is the number of actual video shots in test set related to c_i.

The experiment is carried out on a microcomputer with XP Windows in the paper. Experiment with standard video database set of TRECVID 2007 is designed to test the performance of linear neighborhood propagation based on related kernel mapping. In the experiment, $k-NN$ is used to search for the proximal point. However, the process of searching the neighborhood node consumes a lot of time, which is an essential part of semi-supervised learning algorithm based on graph. When the experiment is made, first take a small part of the samples to carry out the experiment; by setting different values of k, calculate the time consumed in the experiment so as to obtain the optimal value of k. It can be obtained from the conclusion that generally the optimal k is 30.

In order to evaluate the performance of video annotation, we compare the CKLNP algorithm with two kinds of widely used semi-supervised learning methods (SVM algorithm [15] and LNP [8] algorithm).

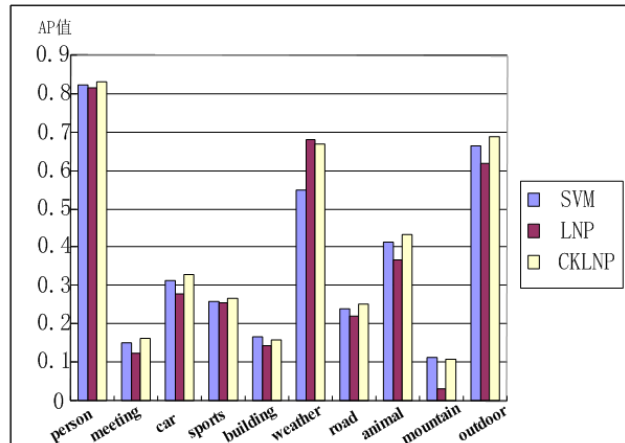With AP as the standard, the results are shown in Fig. 2.

Fig.2 Comparison Results of the Three Algorithms on AP

From Fig.2, with AP as the standard, compared with CKLNP algorithm, LNP algorithm shows a good performance in detecting most of the semantic concepts except for the slight deficiency in the semantic concept of weather, of which the main reason is that it is difficult to detect the semantic concept of weather and the experimental result is relatively random. For the semantic concepts of person, meeting, car, sports, weather, road, animal and outdoor, the AP value obtained by the CKLNP algorithm is higher than that obtained by the SVM algorithm. Through the above data analysis, CKLNP algorithm is better than the other two algorithms on the whole and it can be obtained that CKLNP algorithm has improved precision ratio.

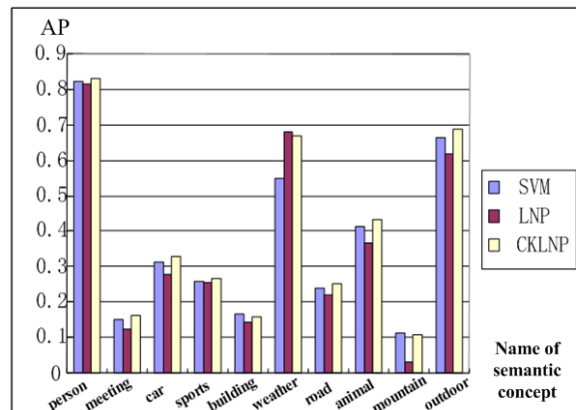With AP as the standard, the results are shown in Fig. 3.

Fig.3 Comparison Results of the Three Algorithms on AP

From Fig.3, AP values obtained by the CKLNP in the detection of semantic concepts of meeting, building, road and mountain is approximately equal to those of the classical SVM algorithm; however, AR values obtained from other semantic concepts are all higher than those obtained by the SVM algorithm [15]. AR value obtained by LNP algorithm [8] is lower than that obtained by the CKLNP algorithm. Therefore, the CKLNP algorithm is better than LNP algorithm and SVM algorithm in the whole range, which also proves that the CKLNP algorithm is beneficial to improve the full annotation ratio of video.

Conclusion

Specific to the characteristic of using a large number of unlabeled samples to improve learning, a new kind of semi-supervised learning method CKLNP based on graph is put forward to carry out automatic video annotation. The successful application of kernel techniques in pattern recognition neighborhood is taken into account and the relevance between semantic concepts is integrated into the existing LNP algorithm in the method. A related table is also constructed in the paper to restrain the correlated noise and ensure the orientation of semantic concepts, which makes the incidence relevance between semantic concepts more accurate. Experimental results show that: CKLNP

method proposed in this paper has good performance in the application of video annotation. Tensor learning method based on Classification at present has attracted the attention of researchers; it is not difficult to take into account the relevance between semantic concepts in tensor learning; so the future research topics will be how to integrate the relevance between semantic concepts into tensor learning.

Reference

[1] Ying Liang, Yingying Yi, Qiufen Sun. (2014). The Impact of Migration on Fertility under China's Underlying Restrictions: A Comparative Study Between Permanent and Temporary Migrants, Social Indicators Research, 116(1): 307-326.

[2] Liu Y, Yang J, Meng Q, et al. Stereoscopic image quality assessment method based on binocular combination saliency model[J]. Signal Processing, 2016, 125: 237-248.

[3] Li Su, Hanwei Wang, Jingwei Miao ,Ying Liang.(2015).Clinicopathological Significance and Potential Drug Target of CDKN(2)A/p(16) in Endometrial Carcinoma, Scientific Reports, 5:13238, 2015. 1-7.

[4] Jiang D, Xu Z, Lv Z. A multicast delivery approach with minimum energy consumption for wireless multi-hop networks[J]. Telecommunication Systems, 2015: 1-12.