CEUR-WS.org/Vol-2125/paper_205.pdf

IPL at imageCLEF 2018: A kNN-based Concept Detection Approach

Leonidas Valavanis and Theodore Kalamboukis

Information Processing Laboratory,
Department of Informatics,
Athens University of Economics and Business,
76 Patission Str, 10434, Athens, Greece
valavanisleonidas@gmail.com,tzk@aueb.gr
http://ipl.cs.aueb.gr/index_eng.html

Abstract. In this paper we present the methods and techniques performed by the IPL Group for the Concept Detection subtask of the ImageCLEF 2018 Caption Task. In order to automatically predict multiple concepts in medical images, a k-NN based concept detection algorithm was used. The visual representation of images was based on the bag-of-visual-words and bag-of-colors models. Our proposed algorithm was ranked 13th among 28 runs and our top run achieved F1 score 0.0509.

Key words: Bag of Visual Words, Bag of Colors, image annotation, knn

1 Introduction

Visual Concept Detection or Automated Image Annotation of medical images has become a very challenging task, due to the increasing number of images in the medical domain. This immense number of medical images leads to a situation where automatic image annotation becomes more and more important. It has gained a lot of attention through various challenges and researchers that try to automate the understanding of the image and its content and provide useful insights that could be beneficial for clinicians. Detecting image concepts can be particularly useful in Content Based Image Retrieval (CBIR) because it allows us to annotate images with semantically meaningful information bridging the gap between low level visual features and high level semantic information and improving the effectiveness of CBIR. The main idea of image annotation techniques is to learn automatically semantic concept from a large number of training samples, and use them to label new images. As the amount of visual information increases, the need for new methods for searching it, also increases.

This year our group participated only in the concept detection task. Details of this task and the data can be found in the overview papers [1], [7] and the web page of the contest [1]. In the next section we present a detailed description of the

[1] http://http://www.imageclef.org/2018/caption

modelling techniques. In section 3, the data representation and preprocessing is presented as well as the results of our submitted runs. Finally, Section 4 concludes our work.

2 A k-NN based concept detection algorithm

kNN is one of the simplest and very effective algorithm in classification problems. Given an unknown item, J, (testing image) we calculate its distance from all images in a training set and assign the item to its nearest category. The decision is based on the number of neighbours we take to assign a score to categories. This is a difficult decision and a weak point of the algorithm.

In the following we give the definition of the annotation problem and a brief description of the algorithm we used in our submitted runs.

Let $\mathcal{X} = \{x_1, x_2, ..., x_N\}$ a set of images and $\mathcal{Y} = \{y_1, y_2, ..., y_l\}$ the set of concepts (labels or annotations). Consider a train set $T = \{(x_1, Y_1), ..., (x_N, Y_N)\}$ where $Y_i \subseteq Y$ is a subset of concepts assigned to image x_i. For each concept y_i we define its semantic group by the set $T_i \subseteq T$, such that $T_i = \{x_k \in T : y_k \in Y_i\}$. The annotation problem is then defined by the probability $p(y_i|J)$. Given an unannotated image, J, the best label will be

$$y^* = \arg\max_i p(y_i|J) \tag{1}$$

Several relations have been used to define the probability $p(y_i|J)$ [2]. In the following it is defined by the score:

$$score(y_k, J) = \sum_{x_j \in T_k} dist(J, x_j) \tag{2}$$

were $dist(.)$ can be any distance (L1, L2 etc) between the visual representations of the images J and x_j. If we sort the images inside T_k, with the highest score on the top, we can take any number of images in the summation 2. Due to the imbalance property of the images between the semantic groups of the concepts usually we consider only a subset of the nearest neighbours in the summation of equation 2. The scores in (2) are converted to probabilities using a soft-max function with a decay parameter, w, which nullifies the distances of most of the images and only few of them, the nearest ones, contribute in the summation. In our experiments the value of the parameters, like the parameter w, were estimated on experiments with the CLEF 2017 data set. The algorithm we have described in matrix form is written by the matrix multiplication:

$$score(y_k, J) = J^T X_v^T Y \tag{3}$$

were X_v is a matrix $N \times m_v$ with N the size of the train set and m_v the number of visual features and Y a $N \times numOfConcepts$ binary and very sparse matrix. The entry $Y\{i\}(j)$ denotes the $j-th$ conceptID of the image i. For computational efficiency, eq. 3 was implemented with a very fast Matlab function using the cosine distance -equivalent to Euclidean distance- for normalized to unity vectors.

As we already mention due to the imbalance property of the annotation problem the algorithm benefits those concepts with high frequency occurrence in the training images. From several experiments with smaller data, plotting the distribution of relevance of the concepts, sorted by their values of DF (Document Frequency) versus the distribution of retrieval we observed that, concepts with low frequency in the train set are downgraded while concepts with high frequency are benefited by the algorithm. Thus the algorithm was modified by normalizing the outcome from eq. 3 by the value $DF(y_k)/avgDF$. This last step improved the performance of the algorithm significantly.

3 Experimental Setup

3.1 Image Visual Representation

One important step in the process of concept detection is the visual representation of images. Images are represented using two models, the Bag-of-visual Words (BoVW) model and a generalized version of the Bag-of-Colors (QBoC) model based on the quad tree decomposition of the image. The BoC model was used for classification of biomedical images in [3] and it was combined successfully with the BoVW-SIFT model in a late fusion manner. In a similar vein, we based our approach to the BoVW and QBoC models for the concept detection of images. In this section, we give a brief insight of these descriptors.

Bag-of-visual Words (BoVW) The BoVW model has shown promising results in the field of classification and image retrieval. The DenseSIFT visual descriptor was used to implement the BoVW model in our runs. This process includes the extraction of the SIFT keypoints [4] from a dense grid of locations at a fixed scale and orientation of the images. The extracted interest points are clustered, by k-means, to form a visual codebook of a predefined size. In our runs the size of the codebook was 4,096. The final representation of an image is created by performing a vector quantization which assigns each extracted key-point of an image to its closest cluster in the codebook.

Generalized Bag-of-Colors Model(QBoC) A generalized version of the BoC model was proposed in [5]. The approach introduces spatial information in the representation of an image by finding homogeneous regions based on some criterion. A common data structure which has been used for this purpose is the quadtree. A quad-tree recursively divides a square region of an image into four equal size quadrants until a homogeneous quadrant was found or a stopping criterion is met. This approach uses the simple BoC model [6] to form a vocabulary or palette of colors, which is then used to extract the color histograms for each image. Similar colors within a sub-region of the image are quantized into the same color, which is the closest color(visual word) in the visual codebook. In our runs we have used two different palettes of size 100 and 200. These palettes result to 1500 and 3000 total color features depending on the number of levels in the quad-tree.

3.2 Preprocessing and normalization

The TF-IDF weights of visual words were calculated for each model separately and the corresponding visual vectors were normalized using the Euclidean norm. The similarities between test and train images were combined for both representations of the images in a late fusion manner using weights w1=0.65 for the DenseSIFT descriptor and w2=0.35 for the QBoC. The values of these parameters were chosen based on our experiments on several other image collections. Finally another parameter of our algorithm is the decay parameter (w) of the softmax function. Several values of w were used in our runs based on experimentation with the CLEF-2017 data set. This year's data contain a set of 223305 training images and a test set of 10000 images with 111156 discrete concepts. The training data are represented with two matrices, one of 223305 × 4096 for the dense SIFT representation and the other of 223305 × 3000 for the QBoC representation. Similarly are represented the images in the test set. These data demand more that 13GB of memory. To overcome our memory limitations we implemented a parallel knn algorithm splitting the matrices into 10 blocks that are accommodated in RAM.

3.3 Submitted Runs and Results

To determine the algorithm's optimal parameters we experimented with the ImageCLEF 2017 caption task dataset. In this year's contest we submitted eight visual runs for the concept detection task. For all runs we used Dense Sift with 4.096 clusters and GBoC with 200 clusters. The parameter w is between 200 and 300. The parameter annot denotes the number of predicted concepts. The results are presented in table 1.

Run_ID	**F1 Score**	**Annot Parameter**
DET_IPL_CLEF2018_w_300_gboc_200	0.0509	70
DET_IPL_CLEF2018_w_300_gboc_200	0.0406	40
DET_IPL_CLEF2018_w_300_gboc_200	0.0351	30
DET_IPL_CLEF2018_w_200_gboc_200	0.0307	30

Table 1. IPL submitted visual runs on Concept Detection Task.

The choice of the parameters w, and number of concepts (annot) is a matter for further investigation. It seems that there is a trade off between the values of these parameters which are set experimentally. A large value of w, may lead the model to over-fitting while a large value of annot reduces the accuracy. Our choice of annot was based on the observation that on average each image in the train set contains 30 concepts.

4 Conclusions

In this paper we presented the automated image annotation experiments performed by the IPL Group for the concept detection task at ImageCLEF 2018 Caption task. A k-NN based concept detection algorithm was used for the automatic Image Annotation of medical images. A normalization step was proposed on the scores in eq. (3) which improved significantly the performance of kNN. The results so far with our new knn approach are encouraging and several new directions have emerged which are currently under investigation.

References

1. de Herrera, A.G.S., Eickhoff, C., Andrearczyk, V., Müller, H.: Overview of the imageclef 2018 caption prediction tasks. In: CLEF working notes, CEUR. (2018)
2. Verma, Y., Jawahar, C.V.: Image annotation using metric learning in semantic neighbourhoods. In: Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III. (2012) 836–849
3. de Herrera, A.G.S., Markonis, D., Müller, H.: Bag–of–colors for biomedical document image classification. In: Medical Content-Based Retrieval for Clinical Decision Support. Springer (2013) 110–121
4. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2) (2004) 91–110
5. Valavanis, L., Stathopoulos, S., Kalamboukis, T.: Fusion of bag-of-words models for image classification in the medical domain. In: Advances in Information Retrieval - 39th European Conference on IR Research, ECIR 2017, Aberdeen, UK, April 8-13, 2017, Proceedings. (2017) 134–145
6. Wengert, C., Douze, M., Jégou, H.: Bag-of-colors for improved image search. In: Proceedings of the 19th International Conference on Multimedia 2011, Scottsdale, AZ, USA, November 28 - December 1, 2011. (2011) 1437–1440
7. Bogdan Ionescu, Henning Müller, Mauricio Villegas, Alba García Seco de Herrera, Carsten Eickhoff, Vincent Andrearczyk, Yashin Dicente Cid, Vitali Liauchuk, Vassili Kovalev, Sadid A. Hasan, Yuan Ling, Oladimeji Farri, Joey Liu, Matthew Lungren, Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Liting Zhou, Mathias Lux, Cathal Gurrin: In: Overview of ImageCLEF 2018: Challenges, Datasets and Evaluation In: Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018), 2018, LNCS Lecture Notes in Computer Science, Springer, September 10-14, 1437–1440, Avignon, France