ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

Pairwise Image Ranking with Deep Comparative Networks

Aymen Cherif[1] and Salim Jouili[1] *

Eura Nova
{aymen.cherif, salim.jouli}@euranova.eu - Belgium

Abstract. We focus our work on instance-level image retrieval. We approach this problem from the point of view of learning to rank. We explore the idea of using the pairwise ranking model instead of providing a similarity measure between a query and a candidate document. We also investigate the ability of this model to learn high level query-document joint features. Our preliminary results show that the end-to-end approach is not only able to learn a better preference function, but also to drive the model to learn better high level features.

1 Inroduction

Information retrieval (IR), content based information retrieval (CBIR) and learning to rank are three highly related research area. However, few studies in the past have tried to address them together. In CBIR systems the query is defined as an example. For instance, in image retrieval, the query is an image and the expected system output is a list of the most similar images within a database. In learning to rank, one don't focus only on the quality of the extracted documents, but take into account the quality of the returned rank.

In both cases, system performances mostly rely on the quatlity of the features used to represent images and the metrics used to compare them. In learning to rank, the features are handcrafted by domain expert in order to capture the query-document dependencies [1]. The goal to make a new set of high level query-document features. We will refer to that as the query-document joint features. This can be easily achieved in text retrieval (i.e. computing term frequencies between queries and documents). However, it is less obvious in Image retrieval applications.

In this work we propose a CBIR system that uses a pairwise ranking model and deep convolutional neural networks. We investigate the ability of our system to learn higher level query-document joint features and compare it with a metric learning approach. We show that our architecture not only outperforms the traditional IR systems, but also drives the system to produce better representations for the query-document features. We first briefly describe the state of the art, then we propose an architecture based on pairwise ranking model. Finally,

*The elaboration of this scientific paper was supported by the Ministry of Economy, Industry, Research, Innovation, IT, Employment and Education of the Region of Wallonia (Belgium), through the funding of the industrial research project Jericho (convention no. 7717).

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

we argue with experimental results that the proposed model can improve the performances of IR systems.

2 Related work

2.1 CBIR state of the art

CBIR is a research area that has been extensively studied in the past decades. A big part of the state of the art addresses the problem by working on low level descriptors to represent images. Some are global descriptors such as texture features, GIST, others are local ones such as SIFT, SURF, etc. These representations are then used as input to metric measures for the retrieval task [2, 3].

Learning this metric that measures the query-document relatedness has become a popular technique in CBIR systems. Metric learning [4] consists of learning a malahanobis distance in order to bring closer similar examples and push further different examples. Metric learning has been applied successfully in information retrieval [5].

Several recent work were driven by the success of convolutional neural networks (CNN) in the task of image classification [6, 7]. Such results have shown that the CNN architectures are capable of learning much better features compared to the one extracted by descriptors. However, using directly a pretrained CNN classifier in CBIR is not useful because the high level features are not optimized for the retrieval task. Instead, several contributions have proposed to use CNNs in a metric learning approach. For example, in [8] a Siamese architecture is trained to predict a similarity score for a pair of documents. In [9] a triplet network is trained by comparing an anchor document, a positive document and a negative document.

2.2 Learning to rank overview

Learning to rank methods fall into three major categories. First, the pointwise [10] approach where a similarity measure attributes a score for a pair of query-document. The score is then used to rank the list of documents. Second, the pairwise [11] approach where the idea is to learn a preference function that compares two documents using their query-document joint features. This preference function is used to rank a list of candidates. And finally the list wise [12] approach where the algorithm gets a list of documents as input and output a ranked version of this list. For more details we invite the reader to refer to [1].

To our knowledge, existing CBIR methods fall into the category of point wise ranking. Learning a similarity measure between images allow to compute a score between the query image and a candidate image. Thus, the final ranking list doesn't take into account the relationship between candidate images. The pairwise approach attempts to solve this issue by training a preference function.

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

The list wise approach pushes further this boundary. But in practice, pairwise methods reported the best results in larning to rank state of the art[11]. This can be explained by two factors: (1) the complexity of the objective function in list wise approaches and therefore the difficulty of training such a model (2) and the difficulty to obtain datasets for list wise ranking task.

As far as we know the best reported results in learning to rank are attributed to the comparative neural networks [11] which is a pair wise approach. In the next section we will describe how this architecture can be leveraged in the context of image retrieval.

3 Proposed method

3.1 Using siamese and triplet network

In the siamese architecture, a neural network $N_W(.)$ parameterized with weights W produces a feature vector for a given image. The representation of the query image and the candidate images are then combined to measure a score of similarity. The combination can be for example the Euclidean norm (i.e. $\|N_W(x_1) - N_W(x_2)\|$). Training such a network can be done by minimizing a loss function that penalises large scores for similar images:

$$y\|N_W(x_1) - N_W(x_2)\|^2 + (1-y)max(0, M - \|N_W(x_1) - N_W(x_2)\|^2)$$

where y is a label indicating if two images are similar or not. M is an arbitrary margin.

Using triplet network is similar to siamese network with the difference that this time three images are provided at training time (an anchor, a positive and a negative images). The network can be trained with the following triplet loss that aims to reduce the distance between the positive image with the ancor and increases the distance of the negative images with the ancor:

$$max(\|N_W(x_a) - N_W(x_p)\|^2 - \|N_W(x_a) - N_W(x_n)\|^2 - M, O)$$

where x_a, x_p and x_n are repectively the ancor, positive and negative images. In siamese and triplet cases, the outcome of the trainning phase is a single neural network $N_w(.)$ that can be used as pointwise learning to rank algorithm.

3.2 Comparative neural network for pair wise ranking

One of the most successful techniques in pairwise ranking is the comparative neural networks introduced in [11]. The key idea is to define constraints on the network weights such that the network behavior will be similar to a pseudo-preference function. Hence we preserve the properties of reflexivity, equivalence, anti-symmetry and Transitivity (see [11] for details). The comparative neural network acts as a binary classifier where the labels are the preference order (i.e. $x \succ y$ or $y \prec x$). The inputs are representation for two candidate documents. Both the candidate image and the query image are represented by the same set of high level features.

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

3.3 Toward end-to-tend pairwise ranking model

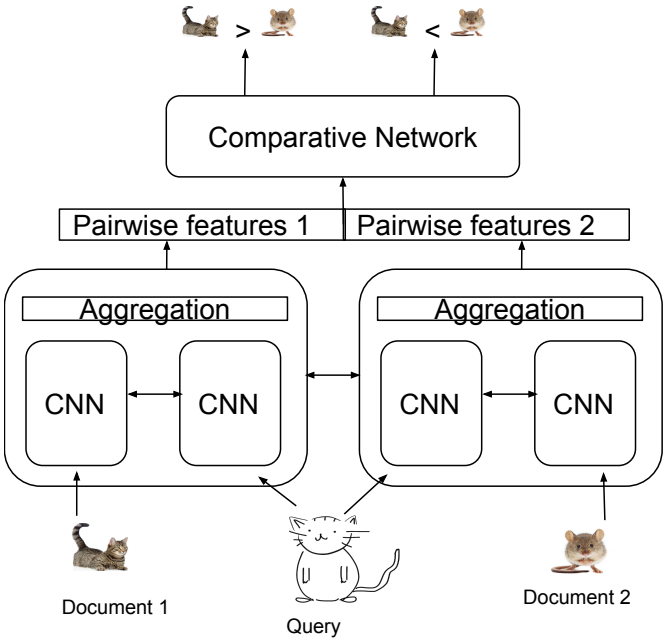

Fig. 1: Architecture of the proposed model

Figure 1 describes the architecture of our model. We define two levels: (1) a lower level uses a combination of siamese networks working on image pixels, (2) and a higher level that uses a comparative neural network as described in [11] to predict the preference between two candidate images.

In an usual siamese (or triplet) network, the output is obtained by euclidean distance. In our case we want to obtain a representation vector of two images. We tested as aggregation two possibilities: (1) computing the mean value from the two vectors and (2) a projection of a dense layer after concatenating the representation of each image. In our preliminary experiments the two options reported similar results. Thus, for simplicity, we use the first option in our reported experiments. We also used shared weights in the network, n ot only inside a single siamese network, but also in the left and right part (figure 1). The reason is that we want to keep the reflexivity property of the comparative network. This architecture acts as two nested siamese networks.

Our main idea is that the backpropagation of comparative network error will drive the low level network to extract the interesting features for the comparison

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

task. To test this idea we used two different settings. In the first setting, we used the usual train-test splitting as in CBIR applications. In the second setting, we split by category, such that the image labels during training are different from the ones during test.

4 Experimental results

For experimentations, we create an image retrieval dataset from a classification dataset. From each label we choose 10% of the images as queries. For each query we associate randomly 20 candidate images (10 positive and 10 negative images) randomly chosen. In addition to that we used two different split strategy. The first is a split by instance where the training and test sets use different images. The second strategy consists of using images of different labels in the train and test sets.

We compared our pairwise model with a pointwise model based on the siamese network using the same MNIST dataset. The two models use the same CNN architecture as a base neural network. The table 1 summarizes the results by measuring the precision@k (up to $k = 10$). We have found that both models perform well at initial position (k = 1 or 2) and fail as we go at the end of the list. However, our model shows more robustness at middle positions on the list. This robustness is still valid when we testwith the label split strategy.

These results suggest that the pairwise model using the error from the comparative network worked better than the pointwise Siamese based model in learning joint feature from the query and the candidate images. However, we often faced the problem of over-fitting when training this pairwise model. The model require more triplets to be generated in order to learn efficiently. Possibly a better strategy of selecting the examples should improve the capacity of the model to generalize.

	Point wise model		Pair wise model	
p@k	normal split	label split	normal split	label split
1	1.	1.	1.	.98
2	.9	.86	1.	.93
3	.8	.72	.9	.92
4	.7	.61	.8	.79
5	.64	.55	.78	.75
6	.63	.55	.75	.76
7	.62	.54	.73	.69
8	.62	.53	.66	.64
9	.61	.53	.63	.55
10	.59	.51	.59	.53

Table 1: Precision @ k on mnist data

ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

5 Conclusion and perspective

In this paper, we addressed the content based image retrieval problem by proposing a new pairwise ranking model by means of a combination of siamese neural network and the comparative neural network. The preliminary experimentations showed promising results that need to be explored further. As an immediate follow up we suggest to investigate this model on a larger dataset having a larger number of categories. Finding better strategies to better construct the pairs and the triplets may help for more efficient training.

References

[1] Tao Qin and TY Liu. Introducing LETOR 4.0 Datasets. *arXiv preprint arXiv:1306.2597*, 2013.

[2] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. SURF: Speeded up robust features. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 3951 LNCS, pages 404–417, 2006.

[3] D.G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, pages 1150–1157 vol.2, 1999.

[4] Aurélien Bellet, Amaury Habrard, and Marc Sebban. A Survey on Metric Learning for Feature Vectors and Structured Data. *BMVC2015*, page 57, 2015.

[5] Brian Mcfee and Gert Lanckriet. Metric Learning to Rank. *Icml*, pages 775–782, 2010.

[6] Fang Zhao, Yongzhen Huang, Liang Wang, and Tieniu Tan. Deep semantic ranking based hashing for multi-label image retrieval. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 07-12-June-2015, pages 1556–1564, 2015.

[7] Xueyi Zhao, Xi Li, and Zhongfei Zhang. Multimedia Retrieval via Deep Learning to Rank. *Signal Processing Letters, IEEE*, 22(9):1487–1491, 2015.

[8] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 539–546, 2005.

[9] Florian Schroff, Dmitry Kalenichenko, and James Philbin. FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 07-12-June-2015, pages 815–823, 2015.

[10] Tie-Yan Y Liu. Learning to Rank for Information Retrieval. *Found. Trends Inf. Retr.*, 3(3):225–331, 2009.

[11] Leonardo Rigutini, Tiziano Papini, Marco Maggini, and Franco Scarselli. SortNet: Learning to rank by a neural preference function, 2011.

[12] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to Rank : From Pairwise Approach to Listwise Approach. *Proceedings of the 24th international conference on Machine learning*, pages 129–136, 2007.