Large Scale Online Learning of Image Similarity Through Ranking

Gal Chechik* Varun Sharma* Uri Shalit† Samy Bengio*

*Google
Mountain View, CA, USA
{gal,vasharma,bengio}@google.com

†Hebrew University
Jerusalem, Israel
uri.shalit@mail.huji.ac.il

Learning a measure of similarity between pairs of objects is a fundamental problem in machine learning. Pairwise similarity plays a crucial role in classification algorithms like nearest neighbors, and is practically important for applications like searching for images that are similar to a given image or finding videos that are relevant to a given video. In these tasks, users look for objects that are both visually similar and semantically related to a given object.

Unfortunately, current approaches for learning semantic similarity are limited to small scale datasets, because their complexity grows quadratically with the sample size, and because they impose costly positivity constraints on the learned similarity functions. To address real-world large-scale AI problem, like learning similarity over all images on the web, we need to develop new algorithms that scale to many samples, many classes, and many features.

The current abstract presents OASIS, an *Online Algorithm for Scalable Image Similarity* learning that learns a bilinear similarity measure over sparse representations. OASIS is an online dual approach using the passive-aggressive family of learning algorithms with a large margin criterion and an efficient hinge loss cost. Our experiments show that OASIS is both fast and accurate at a wide range of scales: for a dataset with thousands of images, it achieves better results than existing state-of-the-art methods, while being an order of magnitude faster. Comparing OASIS with different symmetric variants, provides unexpected insights into the effect of symmetry on the quality of the similarity. For large, web scale, datasets, OASIS can be trained on more than two million images from 150K text queries within two days on a single CPU. Human evaluations showed that 35% of the ten top images ranked by OASIS were semantically relevant to a query image. This suggests that query-independent similarity could be accurately learned even for large-scale datasets that could not be handled before.

1 The Similarity Learning Model and Algorithm

We focus on a similarity learning problem that only assumes a supervised signal about the *relative similarity* of image pairs. Given a set of images, each represented as a vector of features $p_i \in \mathbb{R}^d$, we assume that for every image p_i, we have access to images that are similar to p_i and images that are less similar. Formally, for a small fraction of image pairs we have a relevance measure available $r_{ij} = r(p_i, p_j) \in \mathbb{R}$, which states how strongly p_j is related to p_i. This relevance measure could encode the fact that two images share the same label or match the same query. We do not assume that values of r are precise but only that they correctly capture ordering among pairs. Our goal is to learn a similarity measure $S_\mathbf{W}$ with the form:

$$S_\mathbf{W}(p_i, p_j) \equiv p_i^T \; \mathbf{W} \; p_j \tag{1}$$

with parameters $\mathbf{W} \in \mathbb{R}^{d \times d}$. Importantly, if image vectors p_i are sparse, then $S_\mathbf{W}$ can be computed very efficiently even when d is large. We propose an online algorithm based on the Passive-Aggressive (PA) family of learning algorithms introduced by [CDK$^+$06]. Here we consider an algorithm that uses triplets of images p_i, p_i^+, p_i^- that obey $r(p_i, p_i^+) > r(p_i, p_i^-)$. We define the hinge loss function for all triplets:

$$L_\mathbf{W} = \sum_{(p_i, p_i^+, p_i^-)} l_\mathbf{W}(p_i, p_i^+, p_i^-) \quad \text{with} \quad l_\mathbf{W}(p_i, p_i^+, p_i^-) = \max\left\{0, 1 - S_\mathbf{W}(p_i, p_i^+) + S_\mathbf{W}(p_i, p_i^-)\right\}. \tag{2}$$

To minimize $L_{\mathbf{W}}$, we apply the Passive-Aggressive algorithm iteratively to optimize $\mathbf{W}$. First, $\mathbf{W}$ is initialized to some value $\mathbf{W}^0$. Then, at each training iteration i, we randomly select a triplet (p_i, p_i^+, p_i^-), and solve the following convex problem with soft margin:

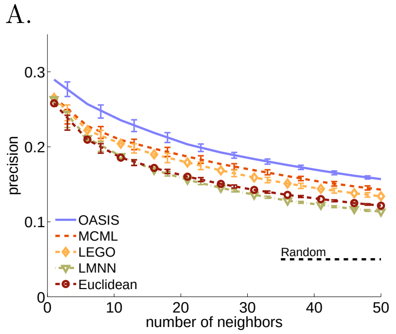
$$\mathbf{W}^i = \underset{\mathbf{W}}{\operatorname{argmin}} \ \frac{1}{2}\|\mathbf{W} - \mathbf{W}^{i-1}\|^2_{Fro} + C\xi \quad \text{s.t. } \ l_{\mathbf{W}}(p_i, p_i^+, p_i^-) \leq \xi \quad \text{and} \quad \xi \geq 0 \tag{3}$$

where $\|\cdot\|_{Fro}$ is the Frobenius norm. At each iteration i, $\mathbf{W}^i$ optimizes a trade-off between remaining close to the previous parameters $\mathbf{W}^{i-1}$ and minimizing the loss on the current triplet $l_{\mathbf{W}}(p_i, p_i^+, p_i^-)$. The *aggressiveness* parameter C controls this trade-off. Eq. 3 can be solved analytically and yields a very efficient parameter update rule. Unlike previous approaches for similarity learning, OASIS does not enforce positivity or even symmetry during learning, since projecting the learned matrix onto the set of symmetric or positive matrices *after training* yielded better generalization (not shown). The intuition is that positivity constraints help to regularize small datasets but harm learning with large data.

2 Experiments

We have first compared OASIS with small-scale methods over the standard *Caltech256* benchmark. Fig. 1 compares the performance of OASIS to other recently proposed similarity learning approaches over 20 of the 256 Caltech classes. All hyper-parameters of all methods were selected using cross-validation. OASIS outperforms the other approaches, achieving higher precision at the full range of first to top-50 ranked image. Furthermore, OASIS was faster by 1-4 orders of magnitude than competing methods (Fig. 1B). For the purpose of a fair comparison with competing approaches, we tested both a Matlab implementation and a C implementation of OASIS for this task. Finally, Fig. 1C compares the runtime of OASIS with a clever fast implementation of LMNN [WS08], that maintains smaller active set of constraints, but still scales quadratically. OASIS scales linearly on a web-scale dataset described below.

A.



B.

Method	Training (minute)
OASIS(M)	45± 8
OASIS(M,C)	0.15± 0.02
MCML(M,C)	7425± 106
LEGO(M)	533± 49
LMNN(M,C)	631± 40

C.

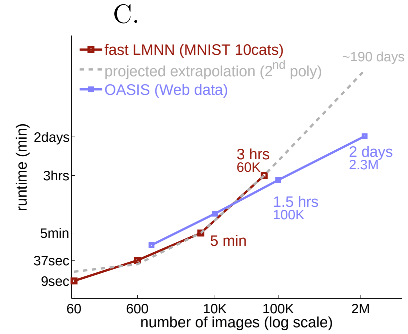


Figure 1: **A.** Comparison of the precision of OASIS, LMNN [WBS06], MCML [GR06], LEGO [JKDG08] and the Euclidean metric in feature space. Each curve shows the precision at top k as a function of k neighbors. Results are averages across 5 train/test partitions (40 training images, 25 test images). **B.** Run time in minutes for methods on panel A. M means Matlab, while M,C means core components implemented in C. **C.** Run time as a function of data set size for OASIS and a fast implementation of LMNN [WS08].

Our second set of experiments is two orders of magnitude larger than the previous experiments. We collected a set of ~150K text queries submitted to the Google Image Search system. For each of these queries, we had access to a set of relevant images, each of which associated with a numerical relevance score. This yielded a total of ~2.7 million images, which we split into a training set of 2.3 million images and a test set of 0.4 million images. Overall, training took ~3000 minutes (2 days) on a single CPU. Fig. 2 shows the top five images as ranked by OASIS on two examples of query-images in the test set.

References

[CDK+06] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer. Online passive-aggressive algorithms. *Journal of Machine Learning Research (JMLR)*, 7:551–585, 2006.

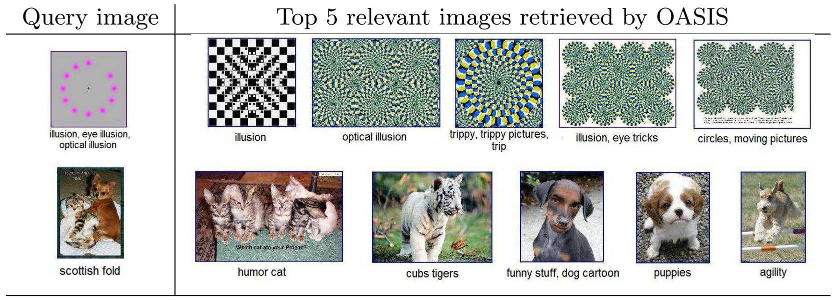


Figure 2: Examples of successful cases from the Web dataset using OASIS.

[GR06] A. Globerson and S. Roweis. Metric Learning by Collapsing Classes. *Advances in Neural Information Processing Systems*, 18:451, 2006.

[JKDG08] P. Jain, B. Kulis, I. Dhillon, and K. Grauman. Online metric learning and fast similarity search. In *Advances in Neural Information Processing Systems*, volume 22, 2008.

[WBS06] K. Weinberger, J. Blitzer, and L. Saul. Distance Metric Learning for Large Margin Nearest Neighbor Classification. *Advances in Neural Information Processing Systems*, 18:1473, 2006.

[WS08] K. Weinberger and L. Saul. Fast Solvers and Efficient Implementations for Distance Metric Learning. In *Proc. of 25th International Conference on Machine Learning (ICML)*, 2008.