Evolutionary Data Purification for Social Media Classification

Stuart James, John Collomosse
Centre for Vision Speech and Signal Processing
University of Surrey, United Kingdom

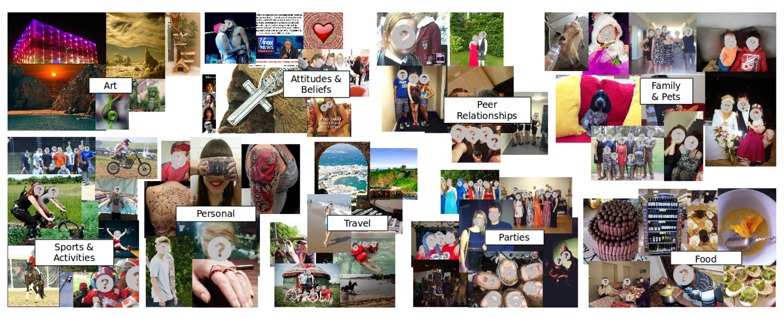
EPSRC
Engineering and Physical Sciences Research Council

Abstract

We present a novel algorithm for the semantic labeling of photographs shared via social media. Such imagery is diverse, exhibiting high intra-class variation that demands large training data volumes to learn representative classifiers. Unfortunately image annotation at scale is noisy resulting in errors in the training corpus that confound classifier accuracy. We show how evolutionary algorithms may be applied to select a 'purified' subset of the training corpus to optimize classifier performance. We demonstrate our approach over a variety of image descriptors (including deeply learned features) and support vector machines.

Dataset

In this work we study the challenging problem of Facebook image annotation. We select this platform due to the abundance of unlabeled social imagery, distinct from well-studied platforms such as FlickR for which images are associated with carefully curated keyword tags, in contrast we use purely visual data.



Twenty college-aged participants were recruited from the same geographic region, and consented to the harvesting of all photographic content within their private Facebook profiles. Building upon an anthropological study of this age group [1], a set of nine high-level semantic concept groups were identified reflecting common themes within posts, namely: ***Art, Attitude & Beliefs, Family & Pets, Food, Friends, Travel, Celebrations; Personal style and self-imagery (e. g. selfie) and Sports***.

Given the high within-class diversity of the dataset, the corpus is boosted using weakly labeled auxiliary content harvested from Google Image Search. An image trawler was implemented to identify additional images based on the above keyword concepts. Since only the top few results for a given keyword are typically relevant, we exploit the WordNet taxonomy [24], applying the 'Is-A' relationship to construct a syn-set of related keywords. We harvest an additional 23k images evenly distributed across the nine concept groups using this method

Visual Representations

We explore four different representations of the items from the dataset including deep and shallow representations, briefly outlined:

Method 1: *CNN Features*
AlexNet model trained on ILSVRC2012 [3] extracted from FC7 layer [4]

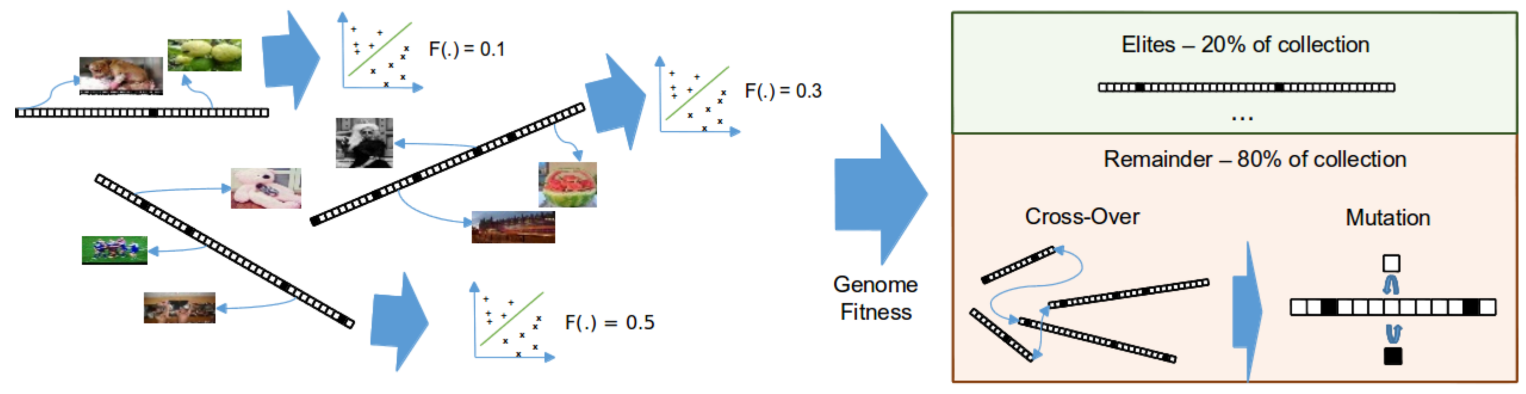
Method 2: Optimised CNN Features
AlexNet model trained on ILSVRC2012 fine-tuned on training data from dataset

Method 3: *SIFT with BoVW*
Dense SIFT regularly spaced (4 pixels) with a BoVW representation K=2000

Method 4: *PHOW-Color with BoVW*
Computed over HSV densely (4 pixels) with a BoVW representation K=2000

Evolutionary Data Purification

Our problem is to select that optimal subset for each of the classifiers. We do so by iteratively turning on/off items of training data, and evaluating the trained model against the validation set. Clearly the space of training data configurations is very high dimensional and this optimality criterion makes the space also very turbulent. Stochastic searches that model evolutionary processes, such as Genetic Algorithms (GAs), are often cited among the best search strategies in such situations; large regions of problem space can be covered quickly, and local minima more likely to be avoided [5].



A population of individuals, each corresponding to a data selection solution, are initially seeded at random. We represent the population as Q x N matrix X(t) coding for a population of Q = 50 individuals at generation t , each of which has genome N bits in length, one per per data item. Individuals are bred at together to produce successive generations i=[2; 250] of the population through processes modeling genetic cross-over, genome mutation, and fitness proportionate reproduction

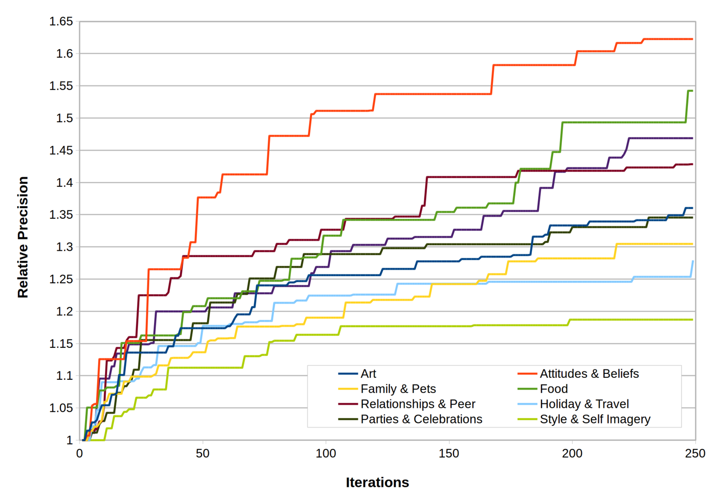
I. Fitness-proportionate Selection
Each member of the population trains Linear SVM (M_i) using only the training data indicated by positive bits in X. Classifiers are evaluated on validation data, where the response is normalized by a sigmoid,

$$\mathcal{S}(v_j;M_i) = \frac{1}{1+\exp(-12\mathcal{N}(M_i(v_j))-0.5)} \quad \text{where} \quad d_j = \rho(v_j) \geqslant 0.5. \quad \text{yields} \quad F(i;t) = 1 - \frac{\sum_{j=1}^{|V|} d_j \oplus l_j}{|\mathcal{V}|}$$

The best 20% fitness (F) get a 'free-pass' to next generation X_{t+1} .

II. Genome Cross-Over
A split point [1,N] is identified randomly within the genome and the two parents are spliced together at that point to a single new genome. This process results in a new combined configuration of training samples to be selected.

III. Genetic Mutation
Offspring from crossover are subjected to mutation to induce diversity in the population. Bits of the offspring's genome and flipped with a 1% chance.



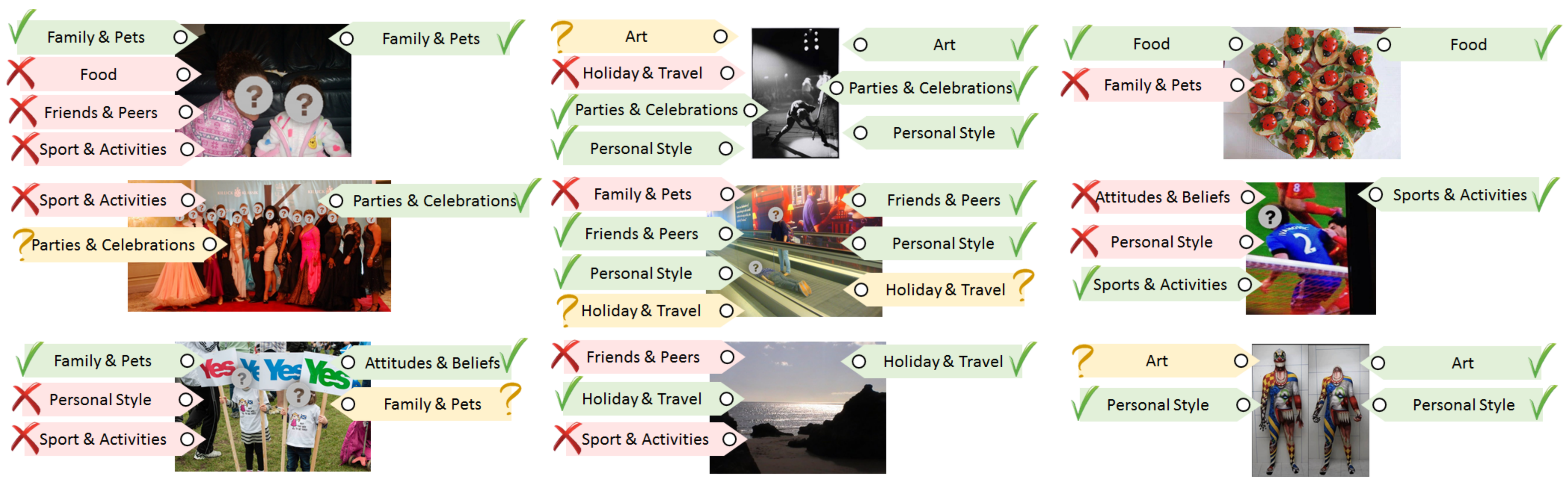
In the graph, iterations of GA training show the relative performance gain of the best performing genome over the validation data. Graph showing one fold for the 9 concepts for the Optimized CNN feature type (Method 2)

Evaluation

We evaluate classifier performance using **5-fold cross-validation** over our social media dataset. The data is split evenly into training, validation and test data yielding 1.7k images in each partition. The auxiliary data from Google was used only to bolster the training set (to 23k images).

The Table shows the mean results over the folds. We highlight the top performing configuration per concept group for the shallow and deep representations. Comparing SVM classifiers using both linear kernels and non-linear (RBF) kernels.

In the best case, GA purification yields a mean precision over all concepts of **32%** representing a **67%** relative improvement over non-purified training data. Despite the intractability of including a non-linear kernel within the GA optimization loop, using the identified training data subset to learn a nonlinear classifier as a final step always equals or betters the performance of the system. This is shown over all configurations except Optimized CNN where the RBF is only able to get a marginal improvement due to higher precision of results pre-purification. The Linear SVM over Optimized CNN features performs best of all, in terms of both precision and computational performance.



Visual examples of label annotation using Method 2 where the initial classifier labels are displayed on the left of the image and the final labels after GA is performed are presented on the right.

Descriptor	**Method 1: CNN features**				**Method 2: Optimized CNN features**				**Method 3: PHOW-Color with BoVW**				**Method 4: SIFT with BoVW**			
Phase	Initial		Post GA		Initial		Post GA		Initial		Post GA		Initial		Post GA	
Classifier Kernel	Linear	RBF	Linear	RBF	Linear	RBF	Linear	RBF	Linear	RBF	Linear	RBF	Linear	RBF	Linear	RBF
Art	13.8 ± 3.68	28.8 ± 1.97	29.9 ± 1.05	36.0 ± 4.60	23.9 ± 2.19	29.7 ± 1.70	**35.6 ± 1.51**	31.2 ± 1.85	15.0 ± 1.70	18.1 ± 0.39	22.6 ± 0.85	**37.4 ± 12.3**	14.2 ± 0.77	18.4 ± 0.54	21.4 ± 1.19	26.6 ± 3.05
Attitudes & Beliefs	4.96 ± 0.88	8.49 ± 0.47	10.6 ± 0.54	**16.9 ± 2.05**	7.46 ± 0.36	13.0 ± 2.50	16.5 ± 1.47	14.5 ± 3.25	4.83 ± 0.52	16.7 ± 2.55	**36.8 ± 6.74**	20.0 ± 3.83	4.97 ± 0.69	14.3 ± 2.89	27.8 ± 10.4	29.4 ± 9.33
Family & Pets	31.2 ± 2.41	29.1 ± 1.00	32.3 ± 1.07	34.2 ± 2.21	29.7 ± 1.47	40.9 ± 1.66	**46.8 ± 3.97**	43.1 ± 2.23	16.1 ± 0.41	19.7 ± 0.29	24.2 ± 0.98	**25.0 ± 1.97**	16.8 ± 0.41	21.3 ± 1.91	24.4 ± 1.57	24.0 ± 4.52
Food	7.64 ± 1.42	30.6 ± 1.61	27.7 ± 2.85	**55.1 ± 6.22**	15.9 ± 1.18	33.1 ± 3.53	34.6 ± 4.31	38.4 ± 2.80	6.17 ± 0.51	23.8 ± 3.29	**44.0 ± 5.95**	27.6 ± 6.73	6.12 ± 0.73	18.9 ± 3.92	24.8 ± 7.73	22.8 ± 6.20
Relationships & Peer	5.41 ± 1.88	11.4 ± 0.54	14.3 ± 0.63	11.7 ± 4.50	11.1 ± 1.29	16.1 ± 1.40	**20.2 ± 1.44**	16.8 ± 1.76	6.39 ± 0.62	13.7 ± 3.88	**20.5 ± 2.75**	2.86 ± 6.39	6.28 ± 1.07	12.8 ± 1.45	16.7 ± 2.68	12.2; ± 21.3
Holidays & Travel	14.7 ± 3.48	28.3 ± 0.72	26.9 ± 2.02	**39.8 ± 2.98**	22.1 ± 0.87	29.6 ± 0.91	32.6 ± 1.03	32.0 ± 1.58	11.4 ± 1.26	26.8 ± 0.68	30.3 ± 1.56	**41.6 ± 1.69**	12.6 ± 2.00	26.4 ± 0.73	28.3 ± 1.81	40.4 ± 2.71
Parties & Celebrations	8.73 ± 2.42	20.0 ± 1.05	21.3 ± 0.59	20.6 ± 3.25	17.1 ± 2.16	28.2 ± 3.22	**32.3 ± 2.13**	28.5 ± 3.74	7.39 ± 0.24	16.3 ± 1.02	**22.1 ± 1.77**	17.3 ± 5.48	6.48 ± 0.57	16.9 ± 0.87	22.0 ± 4.31	16.9 ± 0.87
Style & Self Imagery	21.4 ± 5.48	30.3 ± 0.57	33.3 ± 1.51	37.4 ± 2.10	28.3 ± 2.59	37.8 ± 2.49	**41.9 ± 3.42**	40.6 ± 2.49	19.2 ± 0.99	23.1 ± 0.85	28.2 ± 0.62	**30.4 ± 2.32**	19.0 ± 1.00	23.4 ± 2.20	26.5 ± 2.29	29.8 ± 3.08
Sports	12.2 ± 2.43	20.0 ± 1.36	23.0 ± 1.00	25.0 ± 4.32	18.5 ± 1.20	24.8 ± 2.10	**29.9 ± 2.11**	26.0 ± 2.49	10.3 ± 1.89	16.6 ± 1.19	**22.0 ± 2.60**	16.6 ± 1.19	9.71 ± 1.30	0.00 ± 0.0	20.9 ± 2.27	0.00 ± 0.0
Mean	12.2 ± 2.67	23.0 ± 1.03	24.4 ± 1.25	30.8 ± 3.58	19.4 ± 1.48	28.1 ± 2.17	**32.3 ± 2.38**	30.1 ± 2.40	10.7 ± 0.96	19.4 ± 1.57	**27.8 ± 2.65**	26.6 ± 5.28	10.7 ± 0.96	16.9 ± 1.61	23.7 ± 3.81	24.8 ± 7.83

References

[1] J. Collomosse, S. James, A. Durrant, D. Trujillo-pisanty, W. Moncur, K. M. Orzech, and S. Martindale. Enhancing Digital Literacy by Multimodal Data Mining of the Digital Lifespan. In Proceedings of Digital Economy (DE2014), 2014.
[2] Y. Mori, H. Takahashi, and R. Oka. Image-to-word transformation based on dividing and vector quantizing images with words. In First International Workshop on Multimedia Intelligent Storage and Retrieval Management, 1999.
[3] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *arXiv:1409.0575,* 2014.
[4] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the Devil in the Details: Delving Deep into Convolutional Nets. Proceedings of the British Machine Vision Conference (BMVC), pages 1–11, 2014.
[5] J. Holland. Adaptation in Natural and Artificial Systems. An introductory analysis with applications to biology, control, and artificial intelligence. Univ. Michigan Press, 1st edition, 1975. ISBN: 0-472-08460-7.

Acknowledgements

This work was funded by EPSRC grant EP/L00383X. Thanks to the Charting the Digital Lifespan project partners, who performed the anthropological study in prior work.