Boosting k-NN for Image Classification

Paolo Piro, *PhD student*

University of Nice-Sophia Antipolis

(co-tutorship with University "Campus Bio-medico" of Rome)

Thesis' supervisor: **Prof. Michel Barlaud**

Joint work with **Richard Nock** (University Antilles-Guyane), **Frank Nielsen** (Ecole Polytechnique)

Introduction

Image **Categorization**: to assign one or more **labels** to a given image, based on its content

1. **Representing** visual information:
   - Global descriptors (e.g. GIST)
   - Local descriptors (e.g. SIFT)
2. **Classifying** image representations:
   - Supervised approaches (e.g. SVM)
   - Simple **voting** strategies among neighbors (*k*-NN classification)

***k*-NN** is a gold standard for image retrieval and classification [1,2,3]

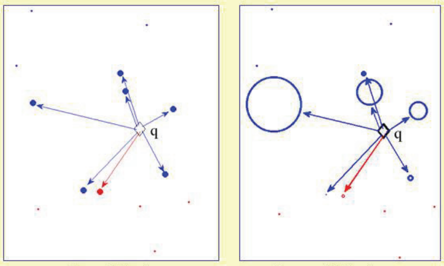
Shortcomings:
- Computational cost of NN search
- Classification errors due to "noisy" voting neighbors

Aim

- Alleviate the uniform voting constraint
- Boost **multiclass** *k*-NN classification

We propose the first boosting algorithm (**UNN**) that induces **leveraged k-NN**

Task: **multilabel**, **multiclass** image categorization



Classic k-NN rule *Leveraged k-NN rule*

Method

Supervised learning:
- Replace *empirical risk* minimization by the minimization of a *surrogate loss* function [4] (e.g. exponential l $\psi = \exp -x$)
- **Multiclass Boosting-like** algorithm [5]:
  - Examples and weak classifiers match altogether
  - Repeatedly leveraging weak classifiers
  - Updating a weight **w** over the examples
- Proved convergence to the **global minimum** of the *surrogate risk*:

$$\varepsilon^{\exp}(\mathbf{h}, \mathcal{S}) \doteq \frac{1}{m}\sum_{i=1}^{m}\exp\left(-\frac{1}{C}\mathbf{y_i}^\top\mathbf{h_i}\right)$$

under very mild assumptions:

1. *Weak learning assumption* (like AdaBoost)
2. *Weak coverage assumption*

UNN algorithm

Induction of **leveraged** ***k*-NN** by minimizing a classification-calibrated surrogate risk

Notation:
- $\mathcal{S}$ training dataset of *m* examples
- (o, y) generic example, with y class vector
- $y_c \in \{-\frac{1}{C_y}, \frac{1}{C-C_y}\}$ membership confidence
- C_y number of negative entries in y
- $i \sim_k i'$: $(\mathrm{o_i}, \mathrm{y_i})$ belongs to the ***k*-NN** of $(\mathrm{o}_{i'}, \mathrm{y}_{i'})$
- $\textsc{Wic}(\{1,2,...,m\},t)$ *weak index chooser* oracle
- $w_j^+,\ w_j^-$ sum of positive (negative) w_j among k-NN
- $\mathbf{h_i}$ strong classifier

Input: $\mathcal{S} = \{(\mathbf{o}_i, \mathbf{y}_i), i = 1, 2, ..., m\}$

Let $\alpha_j \leftarrow 0,\ w_i \leftarrow 1,\ \forall j = 1, 2, ..., m$

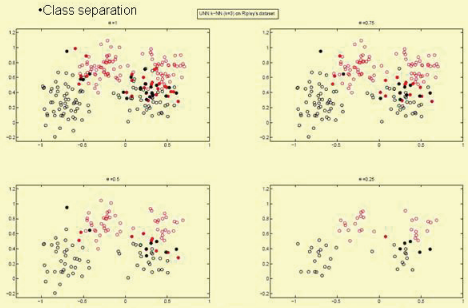
for $t = 1, 2, ..., T$

1. Let $j \leftarrow \textsc{Wic}(\{1,2,...,m\},t)$
2. Let $\delta_j \leftarrow \frac{(C-1)^2}{C}\log\left(\frac{(C-1)w_j^+}{w_j^-}\right)$
3. Let $w_i \leftarrow w_i \exp\left(-\frac{1}{C}\delta_j \mathbf{y_i}^\top \mathbf{y_j}\right)$
4. Let $\alpha_j \leftarrow \alpha_j + \delta_j$

Output: $\mathbf{h_i} = \sum_{i' \sim_k i} \alpha_{i'} \mathrm{diag}(\mathbf{1y_{i'}}^\top)$

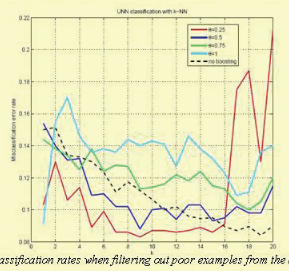
Filtering the database

- Reduce the computational cost of *k*-NN search
- Retain a proportion θ of examples with the largest $|\alpha_j|$
- Cross-validation on the 2-class Ripley's synthetic dataset:
  - Margin maximization effect
  - Class separation



Mixture of two bivariate normal samples: 250 training data, 1000 test data

- Classification performance (Bayes error rate ≈ 8%)



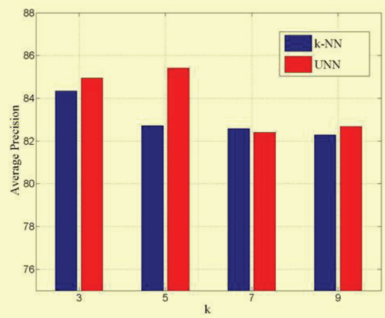
Misclassification rates when filtering out poor examples from the database

Categorization of SIFT descriptors

Holidays-40 database [2]:
- 40 specific image classes
- 250,000 database descriptors
- 1 query image per class

Criteria for database filtering (up to 50%):
- remove examples with the smallest leveraging coefficients (absolute value)
- remove examples with negative lev. coefficients

Examples of images from the Holidays-40 dataset [2].



Average Precision on 40-Holidays database for different values of k.

References

[1] A. Torralba, R. Fergus, W. T. Freeman, 80 million tiny images: a large dataset for non-parametric object and scene recognition, PAMI, vol.30(11), pp. 1958-1970, 2008.

[2] H. Jegou, M. Douze and C. Schmid, Hamming Embedding and Weak Geometric consistency for large-scale image search, Proc. ECCV'08, October, 2008.

[3] A. Bosch, A. Zisserman and X. Munoz, Scene Classification Using a Hybrid Generative/Discriminative Approach, IEEE Transactions on Pattern Analysis and Machine Intelligence (2008).

[4] R. Nock and F. Nielsen, On the Efficient Minimization of Classification-Calibrated Surrogates, NIPS*21 - Advances in Neural Information Processing Systems 2008.

[5] J. Zhu, S. Rosset, H. Zhou and T. Hastie, Multiclass Adaboost. A multiclass generalization of the Adaboost algorithm, based on a generalization of the exponential loss, Stanford University Research Report, 2005.

Acknowledgements

We thank ANR Icos-HD Project for supporting this work.