Multiclass Multilabel Classification with More Classes than Examples

Ofer Dekel[1] and Ohad Shamir[2]

[1]Microsoft Research, Redmond WA, USA
[2]The Hebrew University, Israel

May 13, 2010

flickr from YAHOO!

You aren't signed in Sign In Help

Home The Tour Sign Up Explore Search

Ocean Mist - Grey Whale Cove, California

Uploaded on January 21, 2010 by PatrickSmithPhotography

PatrickSmithPhotography's photostream

92 uploads

This photo also belongs to:

San Francisco Area (Set) 39 items

100 Club (All pics must have been 'Favorited' at least 100 times (Pool)

San Mateo County Group (Pool)

"GreatPicsGallery 300 Faves+ (Pool)

300 Big Ones (Pool)

400+ Faves 1-2-3 (Post one,Comment and Fave two) (Pool)

100 Faves and less than 2000 Views(when posted) (Pool)

World100F – The QualityGroup (Pool)

Chapeaux! (invited only) (Pool)

Favorites: 300 (Pool)

~ Flickr's Finest (100+ Faves Only) ~ (Pool)

25 galleries contain this photo

430 people call this photo a favorite

Tags

landscape
seascape
ocean
sea
mist
fog
pacific
grey whale cove'
pacifica

This past Monday, a big pacific storm this week churned up the ocean into a sea of spray and mist. Then, it all turned a glowing yellow orange, just as the sun set. The RAW file looks just like this. No HDR.

See the 1200 pixel version!
www.flickr.com/photos/patrick-smith-photography/429476936...

Go to my Flickr profile for workshop information and to sign up for my free bi-weekly newsletter. I will answer questions and talk about theories and techniques. No spam will be sent! Also, follow me on Twitter.

Settings etc.:

Canon 5D Mark II
Canon 17-40L @19
1/5-second exposure @F11
LEE soft ND grad (100x150mm) 0.9 + 0.75
Lee foundation Kit filter holder
No polarizer.
ISO 50
Small Slik tripod with Manfrotto pistol grip ball head
RAW file processed with Capture One by Phase One
TIFF file processed with Photoshop
Bare feet and shorts for the inevitable drenching

Story:

The ocean turned into a white frothing fury all of this week as squalls of heavy rain passed through the area. I waited it out behind a cliff under my umbrella and then went out during the sunny openings in the clouds. I was being careful about my back, which is getting better but I made sure to carefully plan my exits for when the water surged up the beach.

And it really surged. This was different than when you get the occasional 'sneaker wave' when 3 waves may occasionally combine at the last moment. This looked more like the tsunami videos I've seen. There were big waves all the time, but sometimes you could look out and see the entire ocean become frothing white and rise high above the horizon until the horizon was blocked from my view. You can see it starting on the left horizon. When that happens, I knew that I had about 30 seconds to get my shot and get out of there. It was not

Tags

- landscape
- seascape
- ocean
- sea
- mist
- fog
- pacific
- grey whale cove'
- pacifica
- san mateo county
- granite
- sunset
- cloud
- sky mountain
- foam
- light
- vacation
- travel
- storm
- lunacy
- swimming
- bay area

Categories: 1452 births | 1519 deaths | People from the Province of Florence | History of anatomy | Italian anatomists | Italian civil engineers | Italian Renaissance humanists | Italian Renaissance XV century painters | Italian vegetarians | Leonardo da Vinci | Mathematics and culture | Ballistics experts | Italian military engineers | People prosecuted under anti-homosexuality laws | Physiognomists | Italian physiologists | Italian inventors | Renaissance architects | Renaissance artists | Renaissance painters | Tuscan painters | Fabulists | 15th century in science | 16th century in science | 15th-century scientists | 16th-century scientists

Wikipedia.org: "Leonardo da Vinci"

Categories: "1452 births" "1519 deaths" "people from the province of Florence" "history of anatomy" "Italian anatomists" "mathematics and culture" "Italian vegetarians" "people prosecuted under anti-homosexuality laws" "Italian inventors" "Renaissance artists" "Tuscan painters"

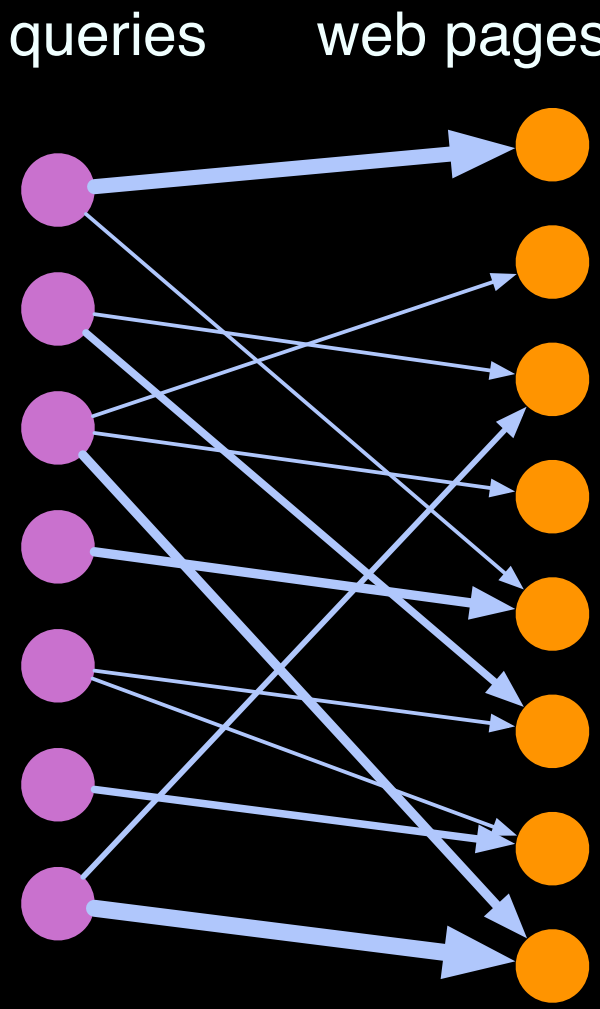
Problem Definition

- Multiclass multilabel classification
- The label set is a *folksonomy* (a.k.a. *collaborative tagging* or *social tagging*)
- We have a large labeled training set (e.g. all Wikipedia pages)
- Goal: categorize unseen instances (e.g. categorize the entire web)

Properties

- m labeled examples, k categories
- $m, k \to \infty$ together (in 2009 Wikipedia had 2.9M articles, 1.5M categories)
- Possibly $k > m$
- Statistical Problem: often can't get an infinite sample from a given class
- Computational Problem: most classification algorithms will choke on millions of labels

Propagating Labels on the Click-Graph



- A bipartite graph derived from search engine logs: clicks encoded as weighted edges
- Wikipedia pages are labeled web pages
- Labels propagate along edges to other pages

An Example

- http://en.wikipedia.org/wiki/Leonardo_da_Vinci passes multiple labels to http://www.greatItalians.com
- Among them
  - "Renaissance artists" - good
  - "1452 births" - bad
- Observation: "1452 births" induces many false-positives (FP): best to remove it altogether from the classifier output (FP $\Rightarrow$ TN, TP $\Rightarrow$ FN)

Notation

- $\mathcal{X}$ is an *instance space*, $\mathcal{Y} = \{0,1\}^k$
- $\mathcal{D}$ is a *distribution* on $\mathcal{X} \times \mathcal{Y}$
- $S = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{2m}$ is an i.i.d. *sample* from $\mathcal{D}$
- A *classifier* $h : \mathcal{X} \to \mathcal{Y}$ suffers *γ-weighted loss*

$$\ell\big(h(\mathbf{x}), \mathbf{y}\big) = \sum_{j=1}^{k} \Big[\gamma \; \mathbb{1}\Big(h_j(\mathbf{x}) = 1; y_j = 0\Big) + (1-\gamma) \; \mathbb{1}\Big(h_j(\mathbf{x}) = 0; y_j = 1\Big)\Big]$$

- *Risk:* $\mathcal{R}(h) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}}\big[\ell\big(h(\mathbf{x}), \mathbf{y}\big)\big]$
- *Empirical risk:* $\hat{\mathcal{R}}(h, S) = \frac{1}{|S|} \sum_{(\mathbf{x}_i,\mathbf{y}_i)\in S} \ell\big(h(\mathbf{x}_i), \mathbf{y}_i\big)$

The Label-Pruning Approach

1. Split data into two halves S_1, S_2
2. Use S_1 to train an initial classifier h_{pre} (e.g. by propagating labels over the click-graph)
3. Apply h_{pre} to S_2, count FP and TP
4. $\forall j \in \{1, \ldots, k\}$, remove label j if

$$\frac{\text{FP}_j}{\text{TP}_j} > \frac{1-\gamma}{\gamma}$$

5. Obtain new "pruned" classifier h_{post}. Note that h_{post} explicitly minimizes $\hat{\mathcal{R}}(h, S_2)$

Analysis

Setting: think of h_{pre} as fixed, S_2 as random

Goal: Prove that w.h.p. our sample S_2 is such that $\mathcal{R}(h_{\text{post}}|S_2) < \mathcal{R}(h_{\text{pre}})$ – positive

Attempt 1: uniform convergence of labels

Prove that

$$\frac{\text{empirical-FP}_j}{\text{empirical-TP}_j} \xrightarrow[m,k\to\infty]{} \frac{\text{expected-FP}_j}{\text{expected-TP}_j}$$

uniformly for all j

Problem: many classes only have a handful of examples

Analysis

Attempt 2: standard empirical estimation tricks

1. By construction: $\hat{\mathcal{R}}(h_{\text{post}}, S_2) < \hat{\mathcal{R}}(h_{\text{pre}}, S_2)$
2. Prove: $\hat{\mathcal{R}}(h_{\text{pre}}, S_2) \xrightarrow[m,k\to\infty]{} \mathcal{R}(h_{\text{pre}})$
3. Prove: $\hat{\mathcal{R}}(h_{\text{post}}, S_2) \xrightarrow[m,k\to\infty]{} \mathcal{R}(h_{\text{post}}|S_2)$

Problem: we can construct cases where $k = \Theta(m)$ and $\hat{\mathcal{R}}(h_{\text{post}}, S_2) - \mathcal{R}(h_{\text{post}}|S_2) \leq c < 0$ for all m

Analysis

Attempt 3: less obvious

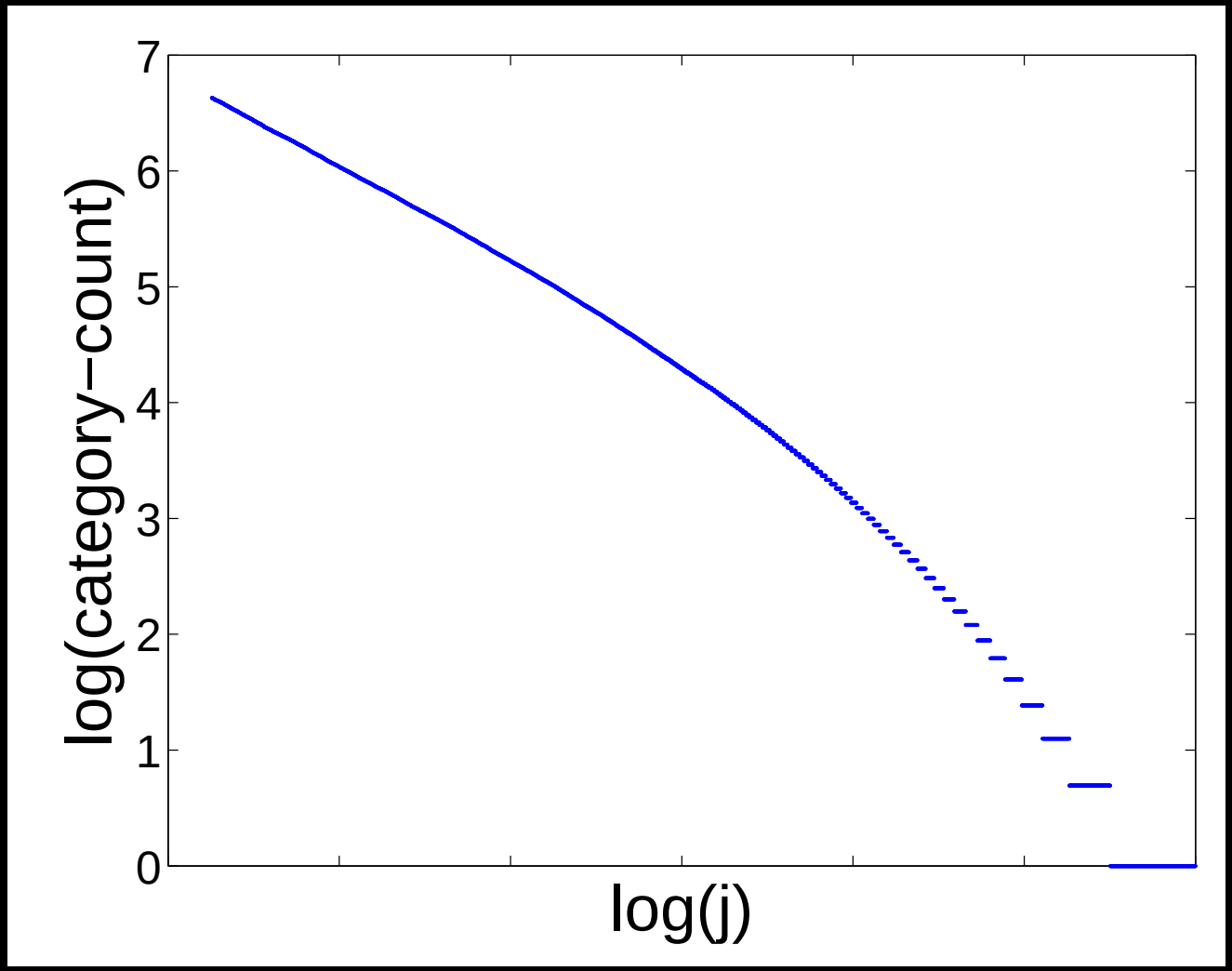
1. If there exists a small s such that $\Pr(\|h_{\text{pre}}(\mathbf{x})\|_1 \leq s) = 1$, then $\mathcal{R}(h_{\text{post}}|S_2) \xrightarrow[m,k\to\infty]{P} \mathbb{E}[\mathcal{R}(h_{\text{post}}|S_2)]$
2. (simplified here for $\gamma = \frac{1}{2}$): Assume labels can be sorted s.t.

$$\Pr\left(h_{\text{pre}}(\mathbf{x})_j = 1\right) = \mathcal{O}(j^{-r})$$

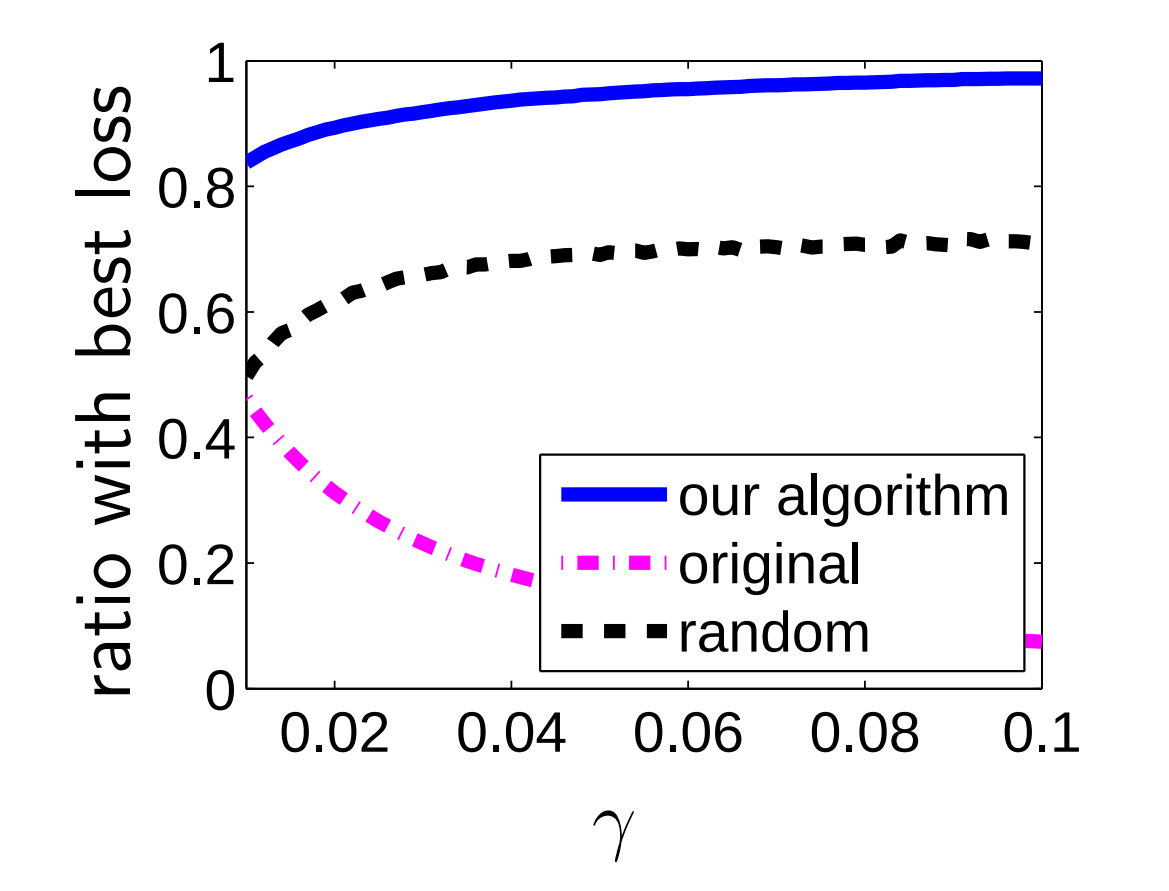
for some $r \in [0, 2)$. Then

$$\mathcal{R}(h_{\text{pre}}) - \mathbb{E}[\mathcal{R}(h_{\text{post}}|S_2)] \geq \text{pos} - \mathcal{O}\left(\sqrt{k^{2-r}/m}\right)$$

Wikipedia Power-Law: $r = 1.6$



Wikipedia Experiment



Conclusion

- The obvious thing is correct, but not for the obvious reason
- $k \to \infty$ violates assumptions of most multiclass analyses
- Our analysis is not an extension of a binary classification analysis
- Future work: more complex label transformation rules (e.g. substitution)