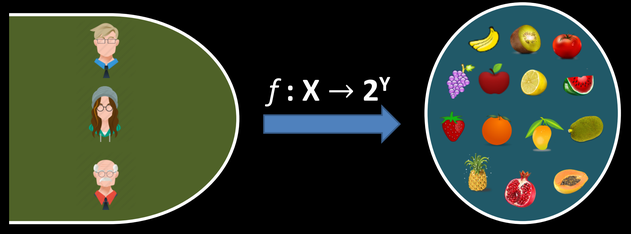
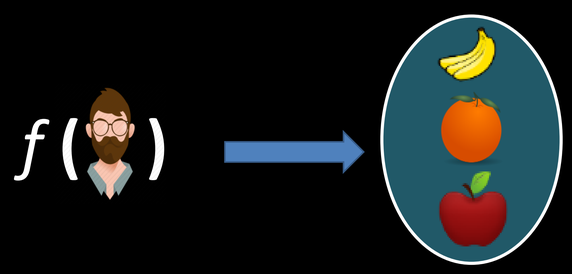
FastXML: A Fast, Accurate and Stable Tree-classifier for eXtreme Multi-label Learning

Yashoteja Prabhu (IIT Delhi) Manik Varma (Microsoft Research)

Extreme Multi-Label Learning

- Learning with millions of labels
- New paradigm for ranking and recommendation





Our Contributions

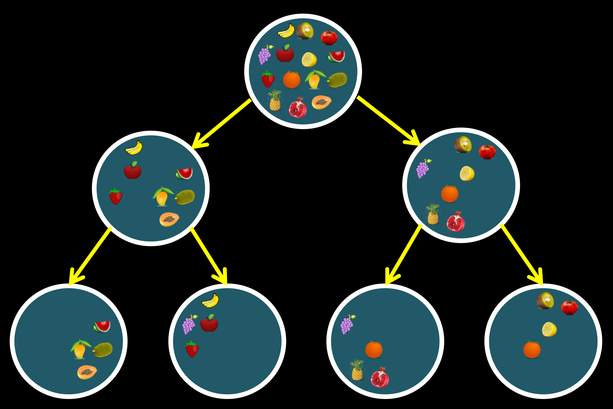
• We formulate a novel node partitioning objective which directly optimizes an nDCG based ranking loss.

• We propose an efficient optimization algorithm for the novel formulation.

• We train on 4 M points, 1 M categories and 160 K features data set in 8 hours on a single core desktop.

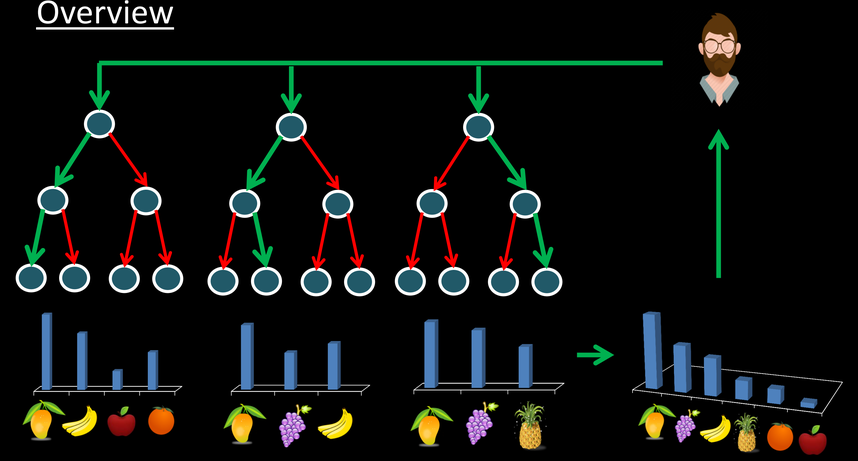
• We improve prediction accuracy significantly over the state of the art Extreme Multi-Label learning algorithms.

Prediction in Logarithmic Time



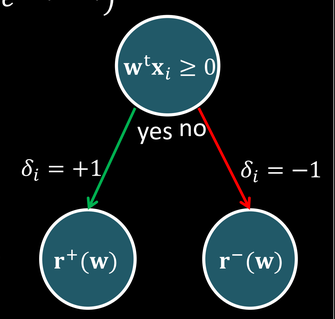
FastXML

Overview



Formulation

$$\text{Min}_{\mathbf{w},\boldsymbol{\delta},\mathbf{r}^{\pm}} \ \|\mathbf{w}\|_1 + \sum_i C_\delta(\delta_i)\log\left(1+e^{-\delta_i \mathbf{w}^t \mathbf{x}_i}\right) + C_r \sum_i L_i^+(\delta_i,\mathbf{r}^+) + L_i^-(\delta_i,\mathbf{r}^-)$$



where

$$L_i^{\pm}(\delta_i,\mathbf{r}^{\pm}) = -\frac{1\pm\delta_i}{2}\sum_j \frac{N_{\mathbf{y}_i y_{ij}}}{\log\left(1+r_j^{\pm}\right)}$$

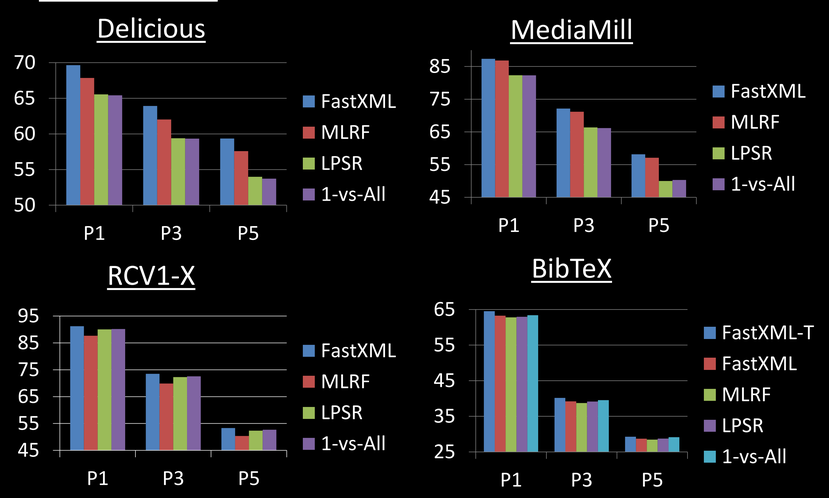
Alternating Minimization

- Initialize $\mathbf{w}$ to 0 and δ_i to +1 or – 1 at random
- Repeat until convergence
  - $\mathbf{r}^{\pm} = \text{rank}\left(\sum_{i:\,\delta_i=\pm 1} N_{\mathbf{y}_i}\mathbf{y}_i\right)$
  - $\delta_i = \text{sign}(v_i^- - v_i^+)$

    $v_i^{\pm} = C_\delta(\pm 1)\log\left(1+e^{\mp \mathbf{w}^t\mathbf{x}_i}\right) - C_r \sum_j \left(\frac{N_{\mathbf{y}_i y_{ij}}}{\log\left(1+r_j^{\pm}\right)}\right)$
- $\mathbf{w} =$ Solve ℓ_1 Logistic Regression$(\{\mathbf{x}_i, \delta_i\})\mathbf{z}$

Results – Small Data Sets

Precision at K

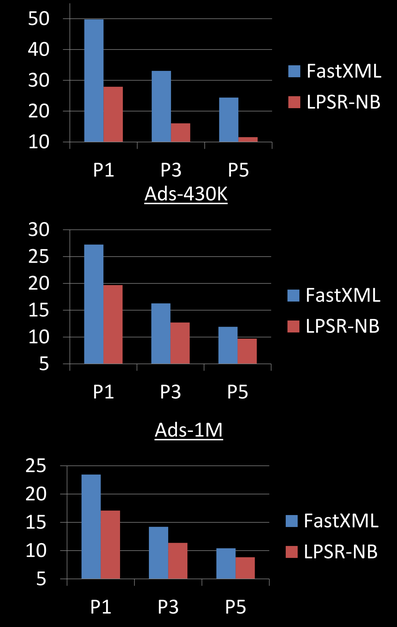


Results – Large Data Sets

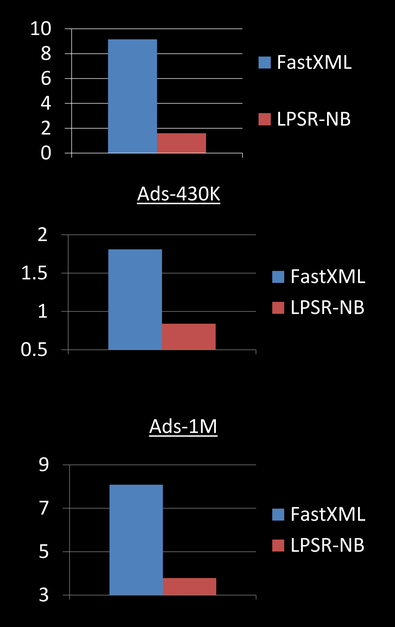
Data Set Statistics

Data Set	# of Training Points (M)	# of Test Points (M)	# of Dimensions (M)	# of Labels (M)
Wikipedia	1.89	0.47	1.62	0.33
Ads-430K	1.12	0.50	0.088	0.43
Ads-1M	3.92	1.56	0.16	1.08

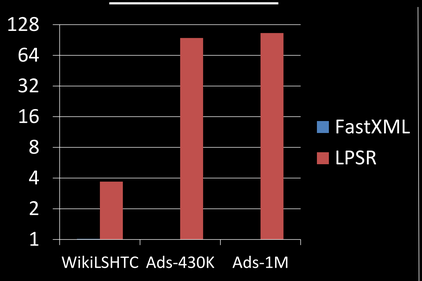
Precision at K



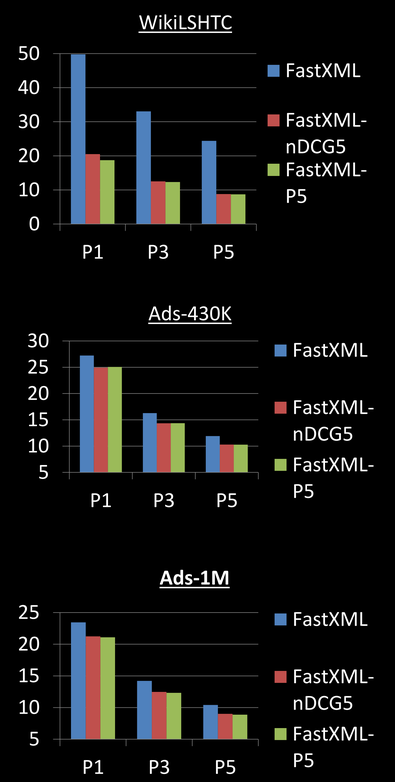
Training Time (Hr)



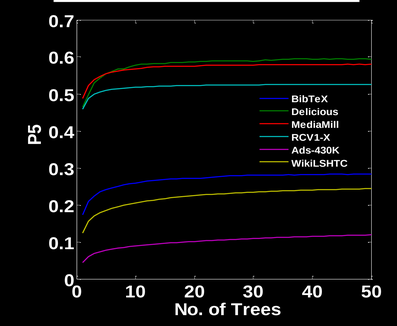
Tree Balance



Variants of FastXML



Random Tree Selection



Speedup vs Cores (Ads-1M)