Extreme Regression for Ranking & Recommendation

Microsoft Research

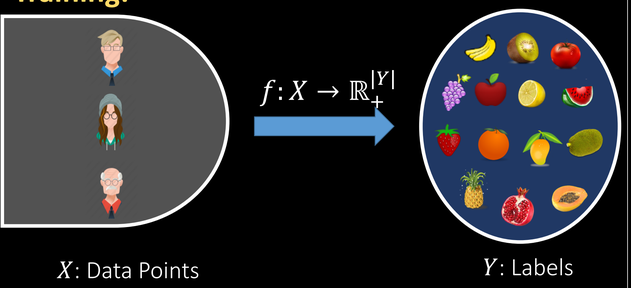
Yashoteja Prabhu*# Aditya Kusupati† Nilesh Gupta* Manik Varma*#

*Microsoft Research India #IIT Delhi †University of Washington

Extreme Regression

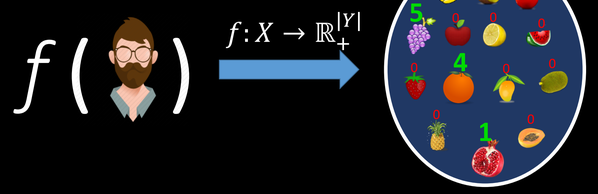
- Predict **relevance scores** of millions of labels towards a given data point
- Reduces to Extreme Classification if relevance scores are binary

Training:



X: Data Points Y: Labels

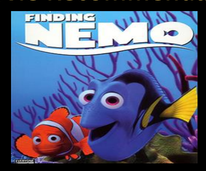
Prediction:



Applications

- A **new paradigm** for reformulating ranking & recommendation
- Predict the most relevant label shortlist & their relevance scores for further reranking

Movie Recommendation:



User	Rating
	5
	4
	1
Others	0

Computational Advertising:

Mens sweaters at Amazon.in - www.amazon.in/Clothing/Mens Sweaters
Ad Shop for **Men's**, Women's & Kid's **Appare** Free Shipping* · Huge Selection · 100% Purch Types: T-Shirts & Polos, Shirts, Trousers, Jean "Most Trusted Online Shopping Brand" – The E

Search Query	PClick
Sweaters for men	100%
Clothing for men	50%
Soft toy	10%
Others	0%

Document Tagging:

Tag	IP Score
Turing awardees	1.8
AI researchers	1.5
Living people	1.0
Others	0

Limitations of Existing Approaches

Extreme Classification:

- Predicts less relevant labels due to binary relevance assumption
- Does not generate useful relevance scores for further filtering or re-ranking

Conventional Regression:

- High accuracy and low latency predictions required in real-world recommendation
- 1-vs-All regressors scale linearly in number of labels
- Scalable tree-based regressors suffer from low accuracy

Extreme Regression Metrics

- Measure the regression errors of millions of labels
- Provide a good proxy for ranking quality
- Irrelevant labels dominate traditional regression metrics

Extreme Mean Absolute Deviation @ k:

$$\text{XMAD@}k(\hat{\mathbf{r}}, \mathbf{r}) = \frac{1}{k} \sum_{l \in S_k} |\hat{r}_l - r_l|$$

where S_k contains k labels with largest errors

Properties:

- $\text{XMAD} @ 1(\hat{\mathbf{r}}, \mathbf{r}) = \|\hat{\mathbf{r}} - \mathbf{r}\|_\infty$
- $\text{XMAD} @ L = \text{MAD}$
- Ranking-regret @ $k \leq 2\,\text{XMAD} @ 2k$

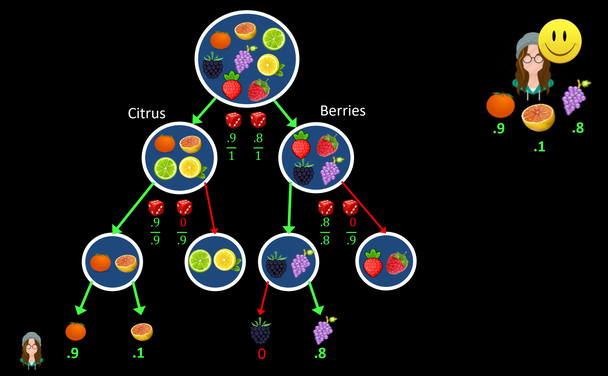
XMAD is a better indicator of filtering & re-ranking qualities than purely ranking or regression metrics

Method	AUPRC	WP-rerank-p @5 (%)	XMAD-p@5	MAD	WP-p @5 (%)
EURLex-4K					
Parabel	0.092	49.67	0.4227	3.96	48.29
XReg	**0.117**	**50.39**	**0.1849**	1.22	**49.72**
XReg-zero	0.085	50.12	0.2255	**1.21**	**49.72**

XReg: Extreme Regressor

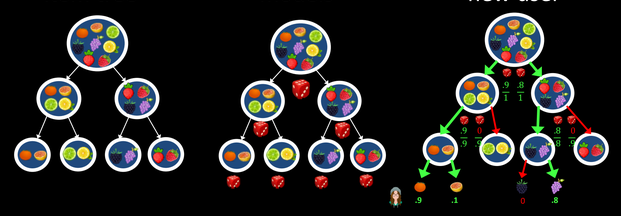
Probabilistic Model:

- How XReg makes recommendations to a user who likes oranges, grapefruits and blueberries ?

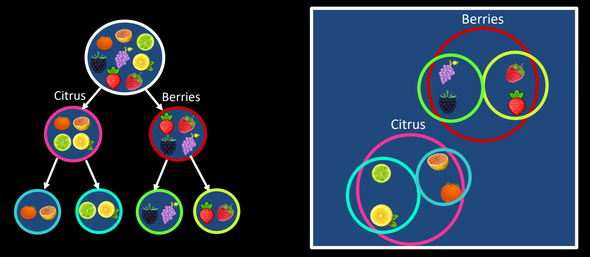


Three stages of Xreg:

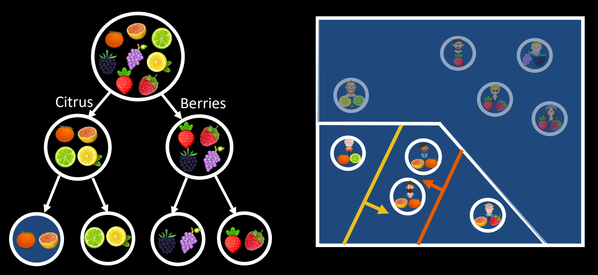
1) Learn item tree 2) Train probabilistic models 3) Predict for a new user



- Learn item tree by hierarchical clustering of items
- Similar items end up in the same leaf node

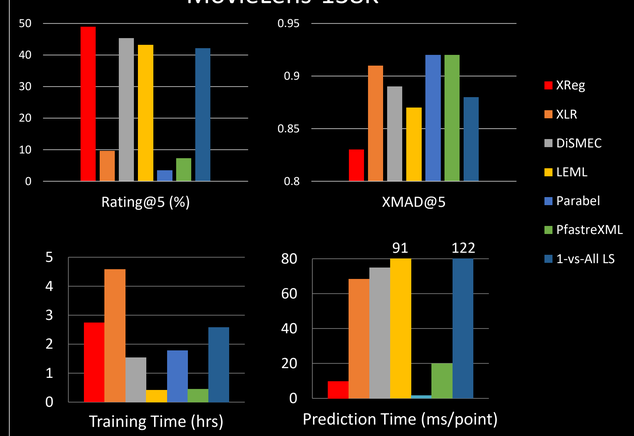


- Train a separate linear regressor for each item in a leaf node
- Recommend items with high regression scores

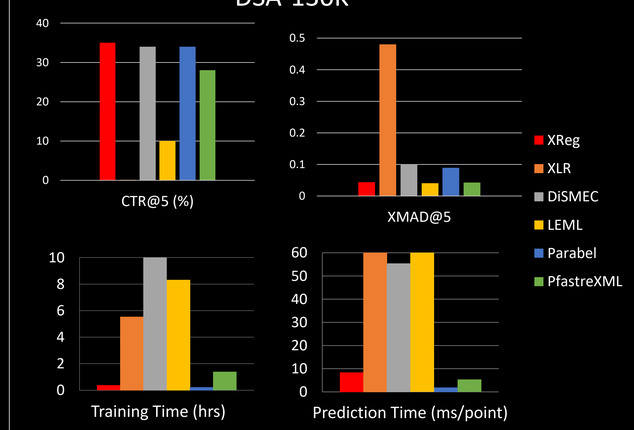


Results

Movie Recommendation:



Computational Advertising:



Document Tagging: