Extreme Classification

Multi-label classification

- When the number of labels reaches the order of hundreds of thousands or millions, using a naive approach that scales linear with a number of labels (i.e., one-vs-all) may not be feasible.

Conditional marginal probability

- Conditional marginal probability of a label $q_l(x) = \sum_{y \in \mathcal{L}} P(y|x) \times P_l(y)$

- Bayes classifiers for popular MLC losses are directly expressed as conditional label probabilities:
  - Hamming loss $\sum_{l=1}^{L} (y_l \neq \hat{y}_l) \times P_l(y)$
  - Precision at $k$: $\sum_{\sum_{l=1}^{k} (y_l \neq \hat{y}_l) \times P_l(y) = 0}$
  - macro F-Measure: $2 \times \frac{P_l(y) \times \text{Recall}_l(y) \times \text{Precision}_l(y)}{P_l(y) \times \text{Recall}_l(y) + \text{Precision}_l(y)}$
  - and others like micro, FM, DC, GCP

- Hence accurate estimation of $q_l(x)$ is crucial for solving XML problems.

Probabilistic Label Trees (PLTs) [6]

- PLTs follows the learning reductions framework: the original problem is decomposed into a set of base problems organized in a tree structure.

- Path from the root to a leaf corresponds to one and only one label and factorizes conditional probabilities:
  \[
  q_l(x) = \prod_{t \in \text{path}} P_l(x_t | x_{\text{parent}(t)})
  \]

- Efficient learning, requires updating at most $|\mathcal{Y}|$ paths per example.

- Different tree-search algorithms can be applied for prediction:
  - beam search (approx. top-$k$ prediction, O(\log n) for balanced trees)
  - uniform-cost search (exact top-$k$ prediction, O(kn) under additional assumptions)
  - threshold-based search (O(\log n) under additional assumptions)

- PLTs has been recently implemented in several state-of-the-art algorithms: FastXML [7], extremeText [1], R poco [8], AttentionXML [9].

Theoretical guarantees of PLTs

- Theorem [6]: For any distribution $P$ and internal node classifiers $f_t$, the following holds:
  \[
  (1 - \epsilon) \sum_{l=1}^{L} P_l(y) \leq \hat{P}_t(y) \leq (1 + \epsilon) \sum_{l=1}^{L} P_l(y)
  \]

- Where $y_l \neq \hat{y}_l$ is binary classification regret for a strongly proper composite loss $l$ (e.g., logistic loss) and $\lambda$ is a constant specific for loss $l$.

- This theorem leads to guarantees for such metrics as Hamming loss, generalized performance metrics, and precision [6].

napkinXC

- Other state-of-the-art methods for XML:
  - It is hard to implement most of the XC algorithms efficiently in high-level languages like Python and Java.
  - Lack of easy to use XC software/libraries.
  - napkinXC:
    - Simple and fast library for extreme multi-class and multi-label classification for Python with C++ back-end.
    - Implements SOTA version of PLs and few other algorithms.
    - Allows training a classifier for very large datasets in few lines of code with minimal resources.
    - Easy to use, follows Scikit-Learn API and supports NumPy and SciPy data types.
    - Implements various prediction algorithms for optimal prediction for different measures (e.g., Hamming loss, precision at k, macro and micro F-measures).

Asterisk paper [5] Science/advances/2020/212218
Source code: https://github.com/mwydmuch/napkinXC
pip: https://pypi.org/project/napkin/

Results: Recommendation

Results: Extreme Classification

Table 1: napkinXC compared to state-of-the-art algorithms on XML repository datasets.

| Dataset | napkinXC | FastXML | PfastreXML | PLtrees | FastXML-M | ExtremeText | T=3 | Offline
|---------|----------|----------|------------|----------|-----------|------------|-----|-------
| Amazon Books | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 |
| Amazon Video | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 |
| Amazon Music | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 |
| Amazon Games | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 |
| Amazon Audible | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 |
| Amazon Appstore | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 |
| Amazon Kindle Store | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 |
| Amazon Associates | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 | 15.95 | 12.05 | 8.57 |