Extreme Classification

BY MANIK VARMA

What would you do if you had the superpower to accurately answer, in a few milliseconds, a multiple-choice question with a billion choices? Would you design the next generation of Web search engines, which could predict which of the billions of documents might be relevant to a given query? Would you build the next generation of retail recommender systems that have things delivered to your doorstep just as you need them? Or would you try and predict the next word about to be uttered by U.S. President Donald Trump?

The objective in extreme classification, a new research area in machine learning, is to develop algorithms with such capabilities. The difficulty of the task can be judged from the fact that, even if it were to take you just a second to read out a choice, it would take you more than 30 years to go through a billion choices. In 2012, state-of-the-art multi-label classification algorithms were struggling to pick the correct subset of options in questions involving thousands of choices. Then, in 2013, a team from Microsoft Research India and IIT Delhi developed a classifier[1] that could scale to 10 million choices, thereby laying the foundations of the area. The approach was based on the realization that only a handful of choices would be relevant for any given question on average. The trick was therefore to quickly eliminate the millions of irrelevant choices. The classifier could then accurately and efficiently choose from the remaining hundred or so options.

Since 2013, extreme classification has come to be a thriving area of research in both academia and industry. Groups from Amazon, CMU, Columbia, Facebook, Fudan University, Google, Humboldt University, IIT Delhi, IIT Kanpur, Max Planck, Microsoft, MIT, Montreal, NEC, NUS, NYU, Stanford, Technion, TU Poznan, UC Davis, UT Austin, Yahoo, and others have developed a plethora of algorithms with varying trade-offs between the prediction accuracy, the prediction time, the training time of the classifier and its size. Most of these algorithms are either based on:

Extreme classification has found applications in diverse areas ranging from information retrieval to recommender systems to computational advertising to natural language processing and even computer vision.

trees that learn a hierarchy over the space of choices so that approximately half the choices are eliminated at each node; embeddings that compress the number of choices by hashing them into a low-dimensional vector space; gating techniques that only consider the handful of relevant options for similar questions seen during training; and deep learning methods that learn the feature representation as well as the classifier. As a result, the community has made remarkable progress over the last six years with training times having been reduced by 10,000x, model sizes having reduced from terabytes to gigabytes, and prediction accuracies on benchmark tasks increasing from 19% in 2013 to 64% today. For instance, the Slice algorithm from MSR India and IIT Delhi, which won the best paper award at WSDM 2019, scales efficiently to problems involving 100-million choices and can be run on a laptop for small problems. Benchmark datasets as well as the source code for many of these algorithms are publicly available at The

Extreme Classification Repository,[3] maintained by IIT Delhi and MSR India, which has become a vital resource for the community.

Extreme classification has found applications in diverse areas ranging from information retrieval to recommender systems to computational advertising to natural language processing and even computer vision. Many papers on extreme classification have been published in top-tier conferences in these areas including AAAI, AISTATS, CVPR, ECCV, IJCAI, ICML, KDD, NAACL-HLY, NeurIPS, SIGIR, WSDM, and WWW. Extreme classification has also opened a new paradigm for key industrial applications such as large-scale ranking and recommendation. For instance, extreme classification can be used to predict which of the top 100 million queries might lead to a click on a given ad or document. Similarly, extreme classification could also be used to predict which of the top 100 million videos you might wish to watch next. In certain cases, reformulating such problems as

Extreme classifiers are helping millions of small and medium enterprises by significantly increasing their sales and revenue, dramatically reducing the costs to reach relevant customers and enabling market growth by the discovery of new customers.

extreme classification tasks might increase revenue by millions of dollars as well as lead to performance improvements over traditional collaborative filtering, learning-to-rank, and content-based approaches. As a result, extreme classification has been deployed in various search and advertising products on the Microsoft Bing platform where it has significantly increased the ability of millions of people around the world to discover the products and services they are looking for. At the same time, extreme

classifiers are also helping millions of small and medium enterprises by significantly increasing their sales and revenue, dramatically reducing the costs to reach relevant customers and enabling market growth by the discovery of new customers.

Extreme classification has brought in many new research questions and technical challenges. A number of workshops have been organized at Dagstuhl,[2] ECML, ICML, NeurIPS, and WWW to discuss these questions. Watch the online videos from these workshops or check out The Extreme Classification Repository[3] if you are looking for an extreme challenge and want to help the community build the next generation of search engines and recommender systems.

References
1. Agrawal, R. et al. Multi-label learning with millions of labels: Recommending advertiser bid phrases for Web pages. In *Proceedings of the Intern. World Wide Web Conference* (Rio de Janeiro, Brazil, May 2013).
2. Bengio, S. et al. Extreme classification (Dagstuhl Seminar 18291). *v*, 7 (2019), 62–80.
3. The Extreme Classification Repository; https://bit.ly/2IDtQbS

Manik Varma (manik@microsoft.com) is a Principal Researcher at Microsoft Research India and an adjunct professor at the Indian Institute of Technology Delhi.