

# Extreme Multi-label Classification: a new paradigm for search and relevance

Marek Wydmuch

OLX Group & Poznań University of Technology, Poland



Prosus Global AI Marketplace Event, Internet, September 8, 2020

# Agenda

- ① Extreme classification
- ② Applications in search and relevance
- ③ Algorithms
- ④ napkinXC demo

# Agenda

- ① Extreme classification
- ② Applications in search and relevance
- ③ Algorithms
- ④ napkinXC demo

## Setting

- Multi-class classification:

$$\boldsymbol{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d \xrightarrow{\boldsymbol{h}(\boldsymbol{x})} \boldsymbol{y} \in \{1, \dots, m\}$$

	$x_1$	$x_2$	$\dots$	$x_d$	$y$
$\boldsymbol{x}$	4.0	2.5		-1.5	5

## Setting

- Multi-class classification:

$$\boldsymbol{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d \xrightarrow{\boldsymbol{h}(\boldsymbol{x})} \boldsymbol{y} \in \{1, \dots, m\}$$

	$x_1$	$x_2$	$\dots$	$x_d$	$y$
$\boldsymbol{x}$	4.0	2.5		-1.5	5

- Multi-label classification:

$$\boldsymbol{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d \xrightarrow{\boldsymbol{h}(\boldsymbol{x})} \boldsymbol{y} = (y_1, y_2, \dots, y_m) \in \{0, 1\}^m$$

	$x_1$	$x_2$	$\dots$	$x_d$	$y_1$	$y_2$	$\dots$	$y_m$
$\boldsymbol{x}$	4.0	2.5		-1.5	1	1		0

## Extreme classification

**Extreme classification**  $\Rightarrow$  a **large** number of **labels**  $m$  ( $\geq 10^5$ )

## Extreme classification

Extreme classification  $\Rightarrow$  a large number of labels  $m$  ( $\geq 10^5$ )

$\Rightarrow$  a large number of features  $d$  ( $\geq 10^5$ )

$\Rightarrow$  a large number of examples  $N$  ( $\geq 10^5$ )

## Extreme classification

**Extreme classification**  $\Rightarrow$  a **large** number of **labels**  $m$  ( $\geq 10^5$ )

$\Rightarrow$  a **large** number of **features**  $d$  ( $\geq 10^5$ )

$\Rightarrow$  a **large** number of **examples**  $N$  ( $\geq 10^5$ )

Basic approach: One-Vs-All with linear models:  $\hat{y} = W^\top x$

## Extreme classification

Extreme classification  $\Rightarrow$  a large number of labels  $m$  ( $\geq 10^5$ )

$\Rightarrow$  a large number of features  $d$  ( $\geq 10^5$ )

$\Rightarrow$  a large number of examples  $N$  ( $\geq 10^5$ )

Basic approach: One-Vs-All with linear models:  $\hat{y} = \mathbf{W}^\top \mathbf{x}$

- Train time complexity  $> 10^{15}$
- Test time complexity  $> 10^{10}$
- Space complexity  $> 10^{10}$

## Computational challenge

### Is it really that hard?

- High performance computing resources available

## Computational challenge

### Is it really that hard?

- High performance computing resources available
- Large data → sparse data (sparse features and labels)

## Computational challenge

### Is it really that hard?

- High performance computing resources available
- Large data → sparse data (sparse features and labels)
- Fast learning algorithms for standard learning problems exist, GPU acceleration is common.

## Computational challenge

### Is it really that hard?

- High performance computing resources available
- Large data → sparse data (sparse features and labels)
- Fast learning algorithms for standard learning problems exist, GPU acceleration is common.
- Learning time for binary model and  $N = 10^5$ :  $\sim 1s$

## Computational challenge

### Is it really that hard?

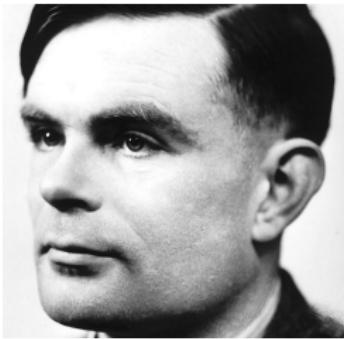
- High performance computing resources available
- Large data → sparse data (sparse features and labels)
- Fast learning algorithms for standard learning problems exist, GPU acceleration is common.
- Learning time for binary model and  $N = 10^5$ :  $\sim 1s$
- Learning time  $m = 10^5$ :  $10^5 \times 2s > 1$  days

# Agenda

- ① Extreme classification
- ② Applications in search and relevance
- ③ Algorithms
- ④ napkinXC demo

## Natural Language Processing: Document tagging

WIKIPEDIA  
The Free Encyclopedia

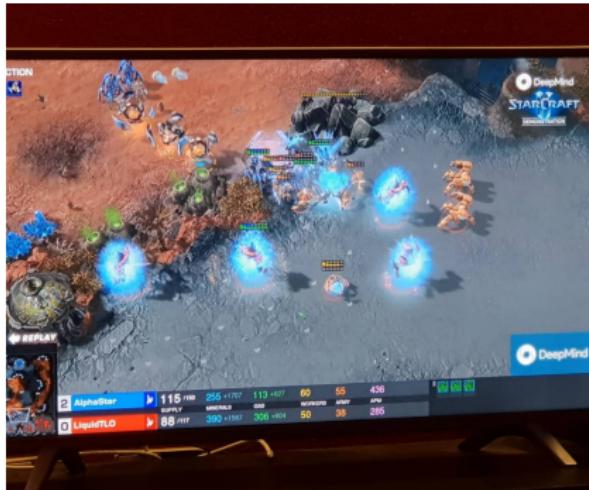


# Natural Language Processing: Document tagging



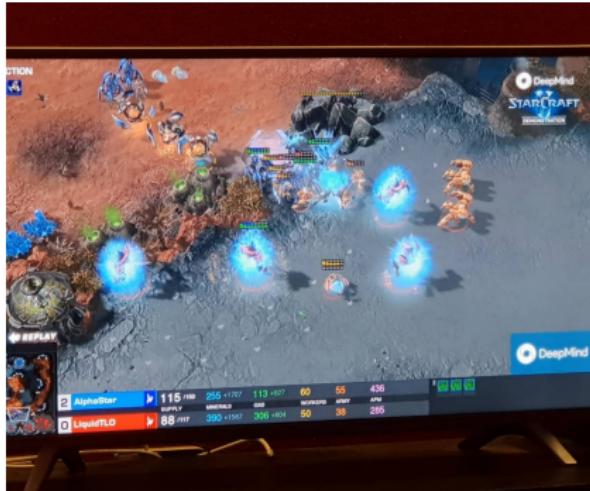
Alan Turing | 1912 births | 1954 deaths  
20th-century mathematicians | 20th-century philosophers  
Academics of the University of Manchester Institute of Science and Technology  
Alumni of King's College | Cambridge Artificial intelligence researchers  
Atheist philosophers | Bayesian statisticians | British cryptographers | British logicians  
British long-distance runners | British people of World War II  
Computability theorists | Computer designers | English atheists  
English computer scientists | English inventors | English logicians  
English long-distance runners | English mathematicians  
English people of Scottish descent | English philosophers | Former Protestants  
Fellows of the Royal Society | Gay men  
Government Communications Headquarters people | History of artificial intelligence  
Inventors who committed suicide | Male long-distance runners  
Mathematicians who committed suicide | Officers of the Order of the British Empire  
Bletchley Park people | People educated at Sherborne School  
People from Maida Vale | People from Wilmslow  
People prosecuted under anti-homosexuality laws | Princeton University alumni  
Programmers who committed suicide | People who have received posthumous pardons  
Recipients of British royal pardons | Academics of the University of Manchester  
Suicides by cyanide poisoning | Suicides in England | Theoretical computer scientists  
Foreign Office personnel of World War II | British anti-fascists | LGBT and suicide  
LGBT scientists

## Post tagging/topic prediction for multimedia search



“AI agents developed by Google’s DeepMind subsidiary have beaten human pros at Starcraft II”

## Post tagging/topic prediction for multimedia search



“AI agents developed by Google’s DeepMind subsidiary have beaten human pros at Starcraft II”

#AI  
#artificialintelligence  
#deeplearning  
#machinelearning  
#videogame  
#DeepMind  
#SC2  
#StarCraft2  
#AlphaStar  
#insane  
#GG  
#LiquidTLD  
#computerscience  
#nerds

# Information Retrieval for search and advertising

Tagging ads with relevant queries<sup>1</sup>:

Ad:



<sup>1</sup> Rahul Agrawal, Archit Gupta, Yashoteja Prabhu, and Manik Varma. Multi-label learning with millions of labels: Recommending advertiser bid phrases for web pages. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013*, pages 13–24. International World Wide Web Conferences Steering Committee / ACM, 2013

# Information Retrieval for search and advertising

Tagging ads with relevant queries<sup>1</sup>:

Ad:



Results of MLRF:

- "geico car insurance"
- "geico auto insurance"
- "geico insurance"
- "www geico com"
- "care geicos"
- "geico com"
- "need cheep auto insurance"
- "wisconsin car insurance quotes"
- "cheap auto insurance florida"
- "all state auto insurance coupon code"

<sup>1</sup> Rahul Agrawal, Archit Gupta, Yashoteja Prabhu, and Manik Varma. Multi-label learning with millions of labels: Recommending advertiser bid phrases for web pages. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013*, pages 13–24. International World Wide Web Conferences Steering Committee / ACM, 2013

# Information Retrieval for search and advertising

Predicting relevant products for queries<sup>2</sup>:

The screenshot shows the Amazon search interface for 'headphones'. The search bar at the top has 'headphones' entered. Below the search bar, there are filters for 'Avg. Customer Review' (4 & Up, 5 & Up, 4 & Up, 5 & Up), 'Shop by feature' (DJ Style, Foldable, Lightweight, Microphone, Noise-Canceling, Noise-Isolating, Phone Control, Sports & Exercise, Tangle-Free Cord, Volume Control), and 'Amazon's Choice' (Sony MDRZX110/BLK ZX Series Stereo Headphones). The results section displays two products: the Sony headphones (Black) and the COWIN E7 Active Noise Cancelling Headphones (Black).

Avg. Customer Review

- 4 & Up
- 5 & Up
- 4 & Up
- 5 & Up
- 4 & Up

Shop by feature

- DJ Style
- Foldable
- Lightweight
- Microphone
- Noise-Canceling
- Noise-Isolating
- Phone Control
- Sports & Exercise
- Tangle-Free Cord
- Volume Control

Amazon's Choice

Sony MDRZX110/BLK ZX Series Stereo Headphones (Black)

★★★★★ - 23,799

Ships to Poland

COWIN E7 Active Noise Cancelling Headphones Bluetooth Headphones with Microphone Deep Bass Wireless Headphones Over Ear, Comfortable Protein Earpads, 30 Hours Playtime for Travel/Work, Black

★★★★★ - 41,338

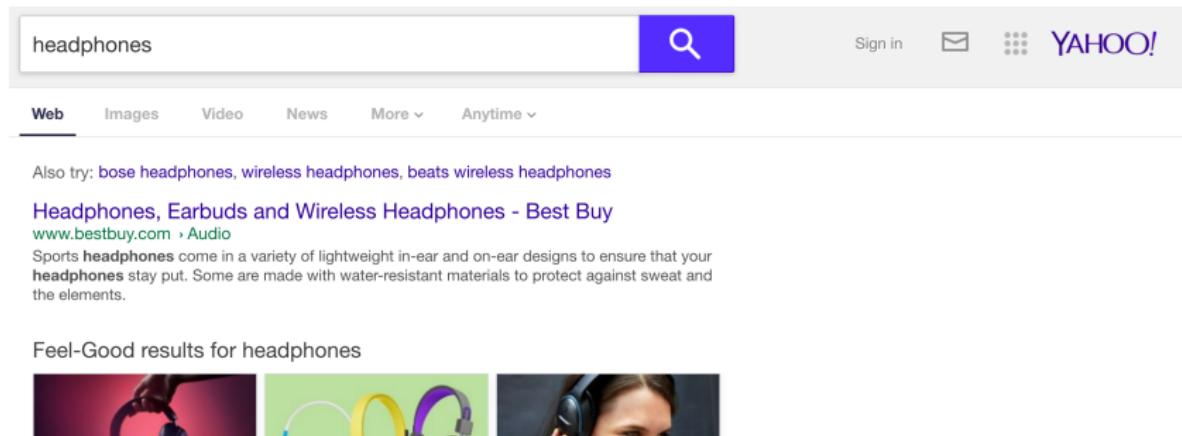
\$59.99

Save 20% with coupon

<sup>2</sup> Tharun Kumar Reddy Medini, Qixuan Huang, Yiqiu Wang, Vijai Mohan, and Anshumali Shrivastava. Extreme classification in log memory using count-min sketch: A case study of amazon search with 50m products. In *Advances in Neural Information Processing Systems* 32, pages 13244–13254, 2019

# Information Retrieval for search and advertising

“Also try” :



A screenshot of a web browser displaying a search results page for "headphones" on the Yahoo! search engine. The search bar at the top contains the query "headphones". To the right of the search bar is a purple search button with a white magnifying glass icon. Further to the right are links for "Sign in", an envelope icon, a grid icon, and the "YAHOO!" logo. Below the search bar is a navigation menu with tabs for "Web", "Images", "Video", "News", "More", and "Anytime". The main content area displays search results. The first result is a link to "Headphones, Earbuds and Wireless Headphones - Best Buy" with the URL "www.bestbuy.com > Audio". Below this link, a snippet of text reads: "Sports headphones come in a variety of lightweight in-ear and on-ear designs to ensure that your headphones stay put. Some are made with water-resistant materials to protect against sweat and the elements." Below this snippet is a section titled "Feel-Good results for headphones" followed by three small thumbnail images: one showing a hand holding a pair of headphones, another showing a colorful pair of headphones (yellow, green, blue), and a third showing a person wearing headphones.

headphones

Sign in YAHOO!

Web Images Video News More ▾ Anytime ▾

Also try: [bose headphones](#), [wireless headphones](#), [beats wireless headphones](#)

**Headphones, Earbuds and Wireless Headphones - Best Buy**  
[www.bestbuy.com > Audio](http://www.bestbuy.com > Audio)

Sports headphones come in a variety of lightweight in-ear and on-ear designs to ensure that your headphones stay put. Some are made with water-resistant materials to protect against sweat and the elements.

Feel-Good results for headphones



# Recommendation and ranking systems

## User to item, item to item recommendation:

Your recently viewed items and featured recommendations

Recommendations & Popular Items

Page 1 of 3

Anker Power Port II  
★★★★★ 37  
EUR 39.99 ✓prime

USB C to DisplayPort Cable – 4 K @ 60Hz, CHOETECH (4ft/1.2 m)  
USB 3.1 Type C...  
★★★★★ 157  
EUR 13.99 ✓prime

Mini Display Port to Display Port Cable  
Posugear Gold Plated Thunderbolt Mini DP to...  
★★★★★ 23  
EUR 8.99 ✓prime

Apple Lightning USB Cable 1 m, MQGJ2ZM/A  
★★★★★ 3  
EUR 21.00 ✓prime

Anker USB C Charger PowerPort PD 2 Wall Charger 30W Dual Port with 18W Power...  
★★★★★ 5  
EUR 18.99 ✓prime

Tech Mat Pencil Lightning Cable Charge Adapter for Apple iPad Pro Female to Female (.75 Inch)  
★★★★★ 45  
EUR 7.99 ✓prime

Sponsored products related to this item

Page 1 of 27

Wslihcure Bluetooth Kopfhörer Kabellos V5.0 Touch Bluetooth Headset Sport Ohrhörer ...  
★★★★★ 310 EUR 51,99 ✓prime

TriLink USB Typ C auf 3,5mm Audio Kopfhörer Jack Adapter(Hi-Resolution Audio & DAC ...  
★★★★★ 6 EUR 12,99 ✓prime

AUSDOM M06 Bluetooth Kopfhörer auf Ohr, VOGEK Faltbare kabelgebundene Kopfhörer mit Mikrofon Over-Ear HiFi Kopfhörer, Bluetoo...  
★★★★★ 23 EUR 31,99 ✓prime

Teufel BAMSTER Schwarz Bluetooth NFC Android iOS Spotify aptX Bluetooth Lautsprecher...  
★★★★★ 28 EUR 13,99 ✓prime

Bluetooth Kopfhörer on Ear Wireless HiFi Stereo Headset mit Mikrofon Freisprecheinr...  
★★★★★ 14 EUR 23,99 ✓prime

Ad feedback

## Problems with embedding-based methods

- The optimization is a pairwise loss function leading to a huge number of training instances.
- The loss functions of embedding-based methods are based on handcrafted similarity thresholds which are not well understood.
- Cold start problem.

# Agenda

- ① Extreme classification
- ② Applications in search and relevance
- ③ Algorithms
- ④ napkinXC demo

# Extreme classification: Algorithms

- Label tree-based methods:
  - ▶ PROBABILISTIC LABEL TREES (PLTs)<sup>3</sup>,
  - ▶ PARABEL<sup>4</sup>,
  - ▶ EXTREMETEXT<sup>5</sup>,
  - ▶ ATTENTIONXML<sup>6</sup>
- Hashing based methods: MERGED-AVERAGED CLASSIFIERS VIA HASHING (MACH)<sup>7</sup>

<sup>3</sup> Kalina Jasinska, Krzysztof Dembczynski, Róbert Busa-Fekete, Karlson Pfannschmidt, Timo Klerx, and Eyke Hüllermeier. Extreme F-measure maximization using sparse probability estimates. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 1435–1444. JMLR.org, 2016

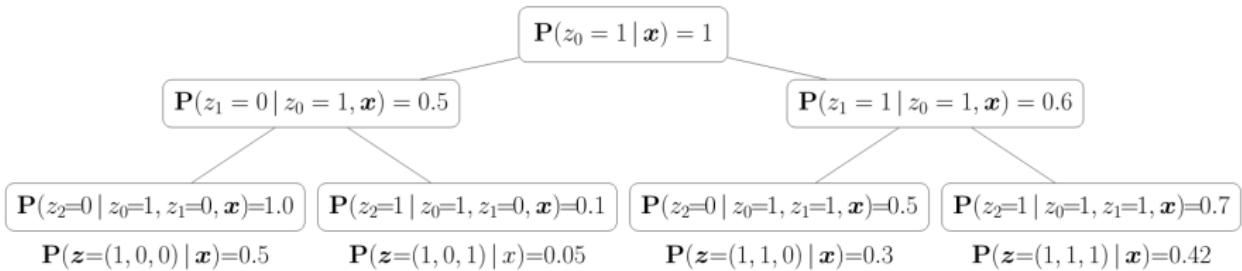
<sup>4</sup> Yashoteja Prabhu, Anil Kag, Shrutiendra Harsola, Rahul Agrawal, and Manik Varma. Parabel: Partitioned label trees for extreme classification with application to dynamic search advertising. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018*, pages 993–1002. ACM, 2018

<sup>5</sup> Marek Wydmuch, Kalina Jasinska, Mikhail Kuznetsov, Róbert Busa-Fekete, and Krzysztof Dembczynski. A no-regret generalization of hierarchical softmax to extreme multi-label classification. In *Advances in Neural Information Processing Systems 31*, pages 6355–6366. Curran Associates, Inc., 2018

<sup>6</sup> Ronghui You, Suyang Dai, Zihan Zhang, Hiroshi Mamitsuka, and Shanfeng Zhu. Attentionxml: Extreme multi-label text classification with multi-label attention based recurrent neural networks. *CoRR*, abs/1811.01727, 2018

<sup>7</sup> Tharun Kumar Reddy Medini, Qixuan Huang, Yiqiu Wang, Vijai Mohan, and Anshumali Shrivastava. Extreme classification in log memory using count-min sketch: A case study of amazon search with 50m products. In *Advances in Neural Information Processing Systems 32*, pages 13244–13254, 2019

## Probabilistic Label Trees<sup>8</sup>



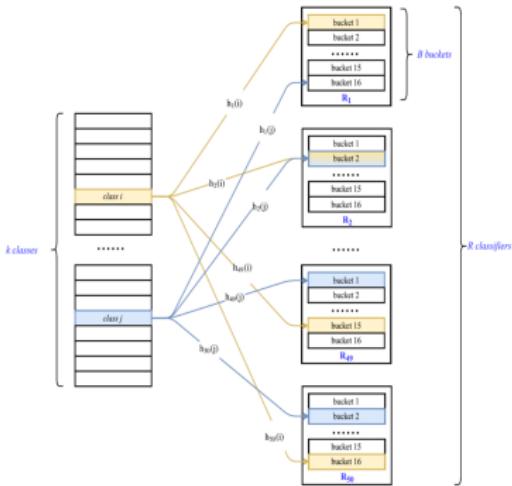
- Each label  $y$  **coded** by  $\mathbf{z} = (1, z_1, \dots, z_l) \in \mathbf{C}$ .
- An internal node identified by a **partial** code  $\mathbf{z}^j = (1, z_1, \dots, z_j)$ .
- **Factorize** the marginal probabilities of labels using a **chain rule**.

$$P(y_i = 1 | \mathbf{x}) = P(\mathbf{z} | \mathbf{x}) = \prod_{i=0}^l \mathbf{P}(z_i | \mathbf{z}^{i-1}, \mathbf{x}).$$

<sup>8</sup>

Kalina Jasinska, Krzysztof Dembczynski, Róbert Busa-Fekete, Karlson Pfannschmidt, Timo Klerx, and Eyke Hüllermeier. Extreme F-measure maximization using sparse probability estimates. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 1435–1444. JMLR.org, 2016

# Merged-Averaged Classifiers via Hashing<sup>9</sup>



- Randomly merges  $K$  labels into  $B$  random-meta-buckets.
- Repeats the process  $R$  times using independent hashing schemes  $h_i, i \in 1, \dots, R$  for labels aggregation.

$$P(y_i = 1 \mid \mathbf{x}) = \frac{\sum_{j=1}^R \mathbf{P}(h_j(i) \mid \mathbf{x})}{R}$$

<sup>9</sup> Tharun Kumar Reddy Medini, Qixuan Huang, Yiqiu Wang, Vijai Mohan, and Anshumali Shrivastava. Extreme classification in log memory using count-min sketch: A case study of amazon search with 50m products. In *Advances in Neural Information Processing Systems 32*, pages 13244–13254, 2019

# Agenda

- ① Extreme classification
- ② Applications in search and relevance
- ③ Algorithms
- ④ napkinXC demo

## Problem with Extreme Classification

Lack of easy to use software for developers...  
...especially for Python...

## napkinXC demo