Extreme Classification

COV 878: Special Topics in Machine Learning

Manik Varma
Microsoft Research & IIT Delhi
Binary Classification

• Answer yes/no questions involving uncertainty

Is this George Washington or not?
Multi-class Classification

- Answer multiple choice questions

Which US President is present in this image?
Multi-label Classification

• Pick multiple answers in a multiple choice question

Which US Presidents are present in this image?
Traditional Classification

- Classification with a small number of choices

Outlook
- Spam or not?

Microsoft Cognitive Services
- < 25K objects

KINECT
- < 100 gestures

Windows Hello
- User or not?

Microsoft
- < 25K objects

Surface Pen
- < 100 characters

MSN
- < 1000 topics

Windows Defender
- Virus or not?

Outlook
- < 1000 tags

Cortana
- ‘Hey Cortana’ or not?
Extreme Classification

• Classification with millions of labels

**Ad**

**Predictions**

<table>
<thead>
<tr>
<th>geico auto insurance</th>
<th>care geicos</th>
</tr>
</thead>
<tbody>
<tr>
<td>geico car insurance</td>
<td>geico insurance</td>
</tr>
<tr>
<td>geico insurance</td>
<td>www geico com</td>
</tr>
<tr>
<td>need cheap auto insurance</td>
<td>geico com</td>
</tr>
<tr>
<td>wisconsin cheap car insurance quotes</td>
<td></td>
</tr>
<tr>
<td>cheap auto insurance florida</td>
<td>all state car insurance coupon code</td>
</tr>
</tbody>
</table>
Extreme Classification

- Publications at AAAI, AISTATS, ECCV, ICML, KDD, NIPS, SIGIR, WSDM, WWW, etc.

- 8 popular workshops organized in 5 years at Dagstuhl, ECML, ICML, NIPS, WWW, etc.

- Code, datasets & benchmarks released on The Extreme Classification Repository

- Wikipedia results have improved from 20% in 2013 to 65% in 2017
Applications

• Information Retrieval
  • Ranking for web search & advertising

• Recommender Systems
  • Item to item recommendation

• Natural Language Processing
  • Language modelling
  • Document tagging

• Computer Vision
  • Person recognition
  • Learning universal feature representations

• Bioinformatics
  • Gene function prediction
Extreme Multi-Label Classification

- Problem formulation

\[ f : X \to 2^Y \]

- X: Users
- Y: Items
Extreme Multi-Label Learning

- Problem formulation

\[ f(\text{ }) \]
Bing Ads – Tesco’s Distilled Water

1-16 of 3,179 results for "distilled water"

Sponsored
Distilled Water - 5 Litre (5L) - Pure Steam Distilled - 0ppm
by Distilled Water UK
£24.95 (£4.99/l)

Adamol 1896 51635139 Distilled Water 5 L
by ADAMOL 1896
£7.20

Bidded Query: distilled water 5 litres
# Predictions: Bing Ads vs Extreme Classification

<table>
<thead>
<tr>
<th>Bing Ads</th>
<th>Extreme Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>water 5</td>
<td>distilled water tesco</td>
</tr>
<tr>
<td></td>
<td>where buy distilled water</td>
</tr>
<tr>
<td></td>
<td>distilled water</td>
</tr>
<tr>
<td></td>
<td>buy distilled water</td>
</tr>
<tr>
<td></td>
<td>distilled water amazon</td>
</tr>
<tr>
<td></td>
<td>distilled water vs purified water</td>
</tr>
<tr>
<td></td>
<td>distilled water uk</td>
</tr>
<tr>
<td></td>
<td>distilled water delivery</td>
</tr>
<tr>
<td></td>
<td>where can I buy distilled water</td>
</tr>
<tr>
<td></td>
<td>distilled water uk supermarket</td>
</tr>
</tbody>
</table>
Traditional Approach

- Reduction to binary classification

\[ h : (\text{Ad}, \text{Phrase}) \rightarrow \{\times, \checkmark\} \]

\[ h(\text{, buy distilled water}) \rightarrow \times \]

\[ h(\text{, water 5}) \rightarrow \checkmark \]
Extreme Classification Approach

- Efficient & accurate prediction via a learnt hierarchy

Extreme Classification for Bing Ads

UK Dynamic Search Ads

French Text Ads

Bided Keywords: la vie assurance, assurance auto, assurance moto

German Product Ads
Kentucky’s rich soil and wonderfully diverse landscapes have for centuries made the state a welcoming habitat for a dazzling variety of wildflowers. From the delicate Kentucky glade cress to the fiery royal catchfly, flowers, grasses, and sedges nestle on rocky hillsides, bask in open woodlands and prairies, and even manage to thrive in busy rights of way and roadsides. As demand for natural resources and land for...
# Predictions: Amazon vs Extreme Classification

<table>
<thead>
<tr>
<th>Amazon</th>
<th>Extreme Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kentucky's Last Great Places</td>
<td>Trees and Shrubs of Kentucky</td>
</tr>
<tr>
<td>Trees and Shrubs of Kentucky</td>
<td>Kentucky's Last Great Places</td>
</tr>
<tr>
<td>Wildflowers and Ferns of Kentucky</td>
<td>Kentucky Trees &amp; Wildflowers: A Folding Pocket Guide to Familiar Species</td>
</tr>
<tr>
<td></td>
<td>Birds of Kentucky Field Guide</td>
</tr>
<tr>
<td></td>
<td>Woody Plants of Kentucky &amp; Tennessee: The Complete Winter Guide to Their Identification &amp; Use</td>
</tr>
<tr>
<td></td>
<td>Native Trees of Kentucky: A Handbook</td>
</tr>
<tr>
<td></td>
<td>Kentucky Wildlife: A Folding Pocket Guide to Familiar Species</td>
</tr>
<tr>
<td></td>
<td>Wildflowers and Ferns of Kentucky</td>
</tr>
<tr>
<td></td>
<td>Kentucky's Natural Heritage: An Illustrated Guide to Biodiversity</td>
</tr>
<tr>
<td></td>
<td>Kentucky Birds: A Folding Pocket Guide to Familiar Species</td>
</tr>
</tbody>
</table>
### Traditional Approach

- Collaborative filtering & matrix factorization

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Banana" /></td>
<td><img src="image" alt="Apple" /></td>
<td><img src="image" alt="Orange" /></td>
<td><img src="image" alt="Pineapple" /></td>
</tr>
<tr>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="✔" /></td>
<td><img src="image" alt="?" /></td>
<td><img src="image" alt="?" /></td>
</tr>
<tr>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="?" /></td>
<td><img src="image" alt="✔" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td><img src="image" alt="?" /></td>
<td><img src="image" alt="X" /></td>
<td><img src="image" alt="?" /></td>
<td><img src="image" alt="X" /></td>
</tr>
</tbody>
</table>

\[
\text{Ratings Matrix} = \text{User Traits} \times \text{Item Attributes}
\]
Extreme Classification Approach

- Recommendation based on user and item features

SwiftXML [Prabhu, Kag, Gopinath, Harsola, Agrawal, Varma WSDM 18]
Bing RS – “cam procedure shoulder”

CAM Procedure | Shoulder Joint Preservation | Osteoarthritis

Dr. Peter J. Millett developed the CAM (Comprehensive Arthroscopic Management) procedure for osteoarthritis of the shoulder.

Comprehensive Arthroscopic Management - CAM Procedure ...

Vail, Colorado (PRWEB) September 20, 2013 -- Newly published research, shows that the innovative CAM procedure is a viable surgical option to delay shoulder surgery.

[VIDEO] CAM Procedure | Shoulder Osteoarthritis Surgery | Vail and ...
### Predictions: Bing vs Extreme Classification

<table>
<thead>
<tr>
<th>Bing</th>
<th>Extreme Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>cam newton shoulder surgery</td>
<td>types of shoulder surgical procedures</td>
</tr>
<tr>
<td></td>
<td>shoulder replacement lawsuits</td>
</tr>
<tr>
<td></td>
<td>how long off work for shoulder surgery</td>
</tr>
<tr>
<td></td>
<td>common shoulder surgeries</td>
</tr>
<tr>
<td></td>
<td>stem cell therapy for rotator cuff tear</td>
</tr>
<tr>
<td></td>
<td>what to wear after shoulder surgery</td>
</tr>
<tr>
<td></td>
<td>cost of arthroscopic shoulder surgery</td>
</tr>
<tr>
<td></td>
<td>arthroscopic shoulder surgery</td>
</tr>
</tbody>
</table>
Divine Comedy

From Wikipedia, the free encyclopedia

"The Divine Comedy" redirects here. For other uses, see The Divine Comedy (disambiguation).
"La commedia" redirects here. For other uses, see Commedia (disambiguation).

The Divine Comedy (Italian: Divina Commedia [diˈvina komˈmɛdja]) is a long narrative poem by Dante Alighieri, begun c. 1308 and completed in 1320, a year before his death in 1321. It is widely considered the preeminent work in Italian literature,[1] and is seen as one of the greatest works of world literature.[2] The poem's imaginative vision of the afterlife is representative of the medieval world-view as it had developed in the Western Church by the 14th century. It helped establish the Tuscan language, in which it is written, as the standardized Italian language.[3] It is divided into three parts: Inferno, Purgatorio, and Paradiso.

The narrative describes Dante's travels through Hell, Purgatory, and Paradise or Heaven,[4] while allegorically the poem represents the soul's journey towards God.[5] Dante draws on medieval Christian theology and philosophy, especially Thomistic philosophy and the Summa Theologica of Thomas Aquinas.[6] Consequently, the Divine Comedy has been called "the Summa in verse".[7]
# Predictions: Wiki vs Extreme Classification

<table>
<thead>
<tr>
<th>Wikipedia</th>
<th>Extreme Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Works by Dante Alighieri</strong></td>
<td><strong>Works by Dante Alighieri</strong></td>
</tr>
<tr>
<td>Divine Comedy</td>
<td>Divine Comedy</td>
</tr>
<tr>
<td>1321 books</td>
<td>1321 books</td>
</tr>
<tr>
<td>1300 in Italy</td>
<td>1300 in Italy</td>
</tr>
<tr>
<td>Visionary poems</td>
<td>Visionary poems</td>
</tr>
<tr>
<td>Epic poems in Italian</td>
<td>Epic poems in Italian</td>
</tr>
<tr>
<td>14th-century Christian texts</td>
<td>14th-century Christian texts</td>
</tr>
<tr>
<td>14th-century books</td>
<td>14th-century books</td>
</tr>
<tr>
<td>Virgil</td>
<td>Virgil</td>
</tr>
<tr>
<td>Afterlife</td>
<td>Dante Alighieri</td>
</tr>
</tbody>
</table>
Recognizing People on Facebook

Choices: Bradley Cooper, Ellen DeGeneres, Meryl Streep, Jennifer Lawrence, Channing Tatum, Julia Roberts, Kevin Spacey, Brad Pitt, Angelina Jolie, Lupita Nyong'o, Peter Nyong'o
Language Modelling

Brevity is the soul of ...
Conclusions

• Extreme classification
  • Tackle applications with millions of choices
  • A new paradigm for ranking & recommendation

• Algorithms & papers
  • MLRF [WWW 2013], FastXML [KDD 2014]
  • SLEEC [NIPS 2015], PfastreXML [KDD 2016]
  • SwiftXML [WSDM 2018], Parabel [WWW 2018]

• The Extreme Classification Repository
  • Code & datasets
  • Benchmark results
  • Papers
Research Questions

- Applications
- Obtaining good quality training data
- Log time and space training and prediction
- Obtaining discriminative features at scale
- Extreme loss functions
- Performance evaluation
- Dealing with tail labels and label correlations
- Dealing with missing and noisy labels
- Explore/exploit for tail labels
- Statistical guarantees
- Fine-grained classification