Multi-label classification via joint space embeddings

Andrija Đurišić

Multi-label classification

Instances

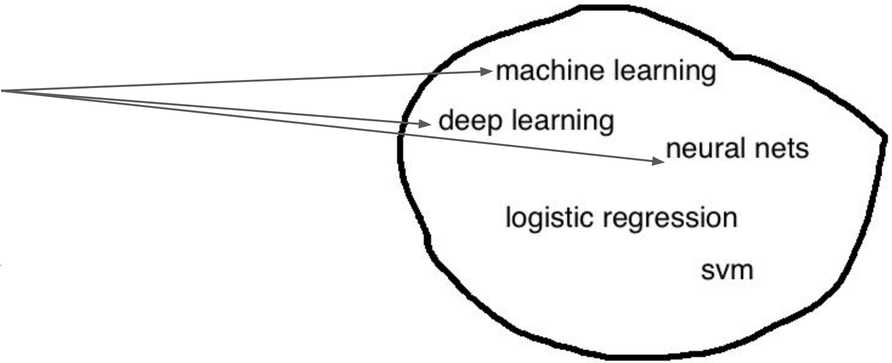
Labels


OPEN ACCESS Freely available online PLoS BIOLOGY

A Role for Parasites in Stabilising the Fig-Pollinator Mutualism


OPEN ACCESS Freely available online PLoS BIOLOGY

A Role for Parasites in Stabilising the Fig-Pollinator Mutualism




OPEN ACCESS Freely available online PLoS BIOLOGY

A Role for Parasites in Stabilising the Fig-Pollinator Mutualism

Not to be confused with multi-class classification

Instances

Labels


A Role for Parasites in Stabilising the Fig-Pollinator Mutualism


A Role for Parasites in Stabilising the Fig-Pollinator Mutualism


A Role for Parasites in Stabilising the Fig-Pollinator Mutualism

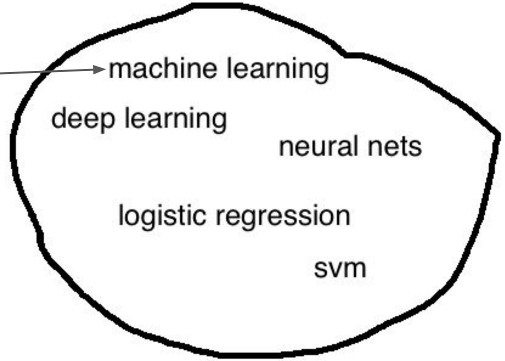


Image annotations problem

Instance

Labels

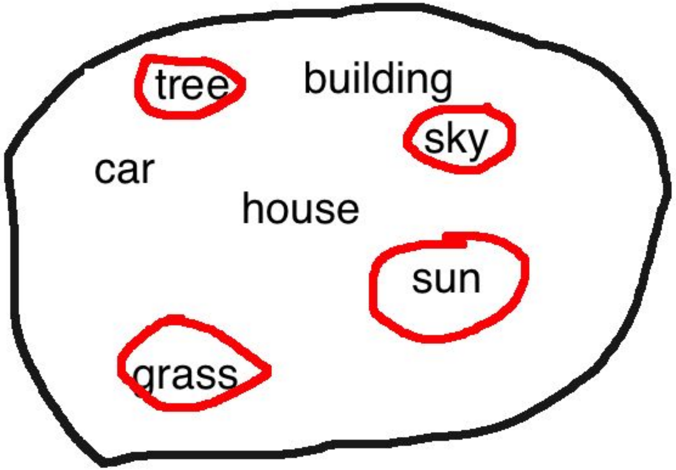


Image annotations problem

Datasets:

- Caltech-256
- Pascal-VOC
- ImageNet 2010 (4m imgs / 16k labels)
- Web (16m imgs / 109k labels)

Requirements:

- Scalable training and testing times
- Scalable memory usage

Ideally, we would like a fast algorithm that fits on a laptop.

Joint space embeddings

Embeddings

- Embedding of an object is vector representation of that object in that vector space
- Similar objects are mapped to close vector representations
- In case of low-dimensional representations savings in memory and computing time are possible

Joint space embeddings

images $x \in \mathbb{R}^d$

annotations $i \in \mathcal{Y} = \{1, \ldots, Y\}$

$\mathbb{R}^D$

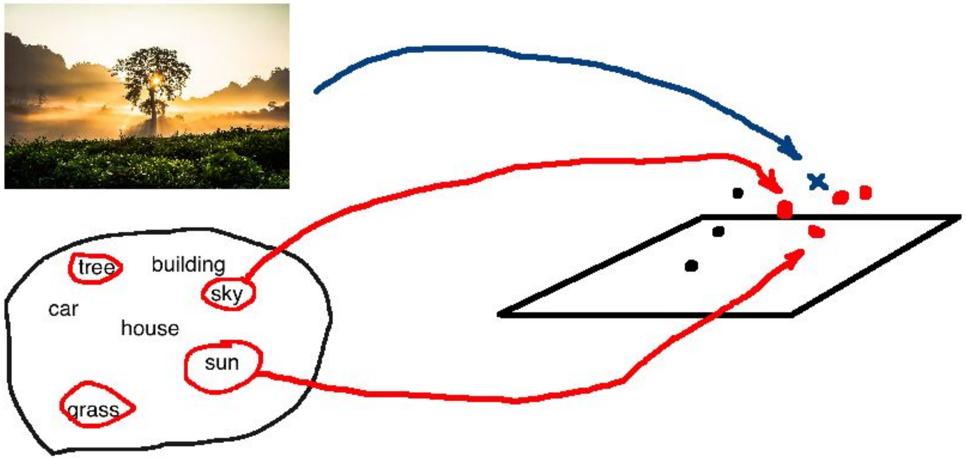
$\Phi_I(x) : \mathbb{R}^d \rightarrow \mathbb{R}^D$

$\Phi_W(i) : \{1, \ldots, Y\} \rightarrow \mathbb{R}^D$

$\Phi_I(x) = Vx$

$\Phi_W(i) = W_i$

$f_i(x) = \Phi_W(i)^\top \Phi_I(x) = W_i^\top Vx$



Joint space embeddings - rank function

$$f(x) = W^{\top} V x$$

$f_{tree}(x) = 13.6$ 0

$f_{buiding}(x) = 2.1$ 4

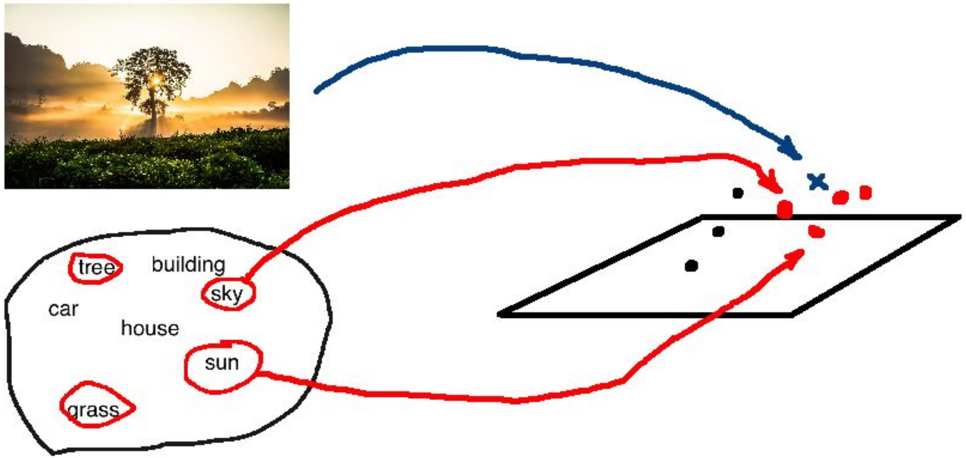
$f_{house}(x) = 1.5$ 5

$f_{car}(x) = 1.2$ 6

$f_{sun}(x) = 11.6$ 3

$f_{grass}(x) = 12.6$ 1

$f_{sky}(x) = 11.7$ 2



$$rank_y(f(x)) = \sum_{i \neq y} I(f_i(x) \geq f_y(x))$$

$$rank_y^1(f(x)) = \sum_{i \neq y} I(1 + f_i(x) > f_y(x)).$$

Joint space embeddings - loss function

$$rank_y(f(x)) = \sum_{i \neq y} I(f_i(x) \geq f_y(x))$$

$$err(f(x), y) = L(rank_y(f(x)))$$

$$L(k) = \sum_{j=1}^{k} \frac{1}{j}$$

$$err(f(x), y) = L\left(rank_y(f(x))\right) \frac{rank_y(f(x))}{rank_y(f(x))}$$

$$err(f(x), y) = \sum_{i \neq y} L\left(rank_y(f(x))\right) \frac{I(f_i(x) \geq f_y(x))}{rank_y(f(x))}$$

$$err(f(x), y) = \sum_{i \neq y} L\left(rank^1_y(f(x))\right) \frac{I(1+f_i(x) \geq f_y(x))}{rank^1_y(f(x))}$$

$$\boxed{\overline{err}(f(x), y) = \sum_{i \neq y} L\left(rank^1_y(f(x))\right) \frac{|1 - f_y(x) + f_i(x)|_+}{rank^1_y(f(x))}}$$

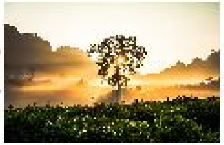
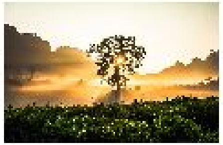
Joint space embeddings - optimization

$$\overline{err}(f(x),y) = \sum_{i\neq y} L\left(rank_y^1(f(x))\right)\frac{|1-f_y(x)+f_i(x)|_+}{rank_y^1(f(x))}$$

$$Risk(f) = \int \overline{err}(f(x),y)dP(x,y).$$

$$\overline{err}_{\bar{y}}(f(x),y,\bar{y}) = L(rank_y^1(f(x)))\,|1-f_y(x)+f_{\bar{y}}(x)|_+$$

$$rank_y^1(f(x)) \approx \left\lfloor \frac{Y-1}{N} \right\rfloor$$

$f_{tree}($  $) > 1 + f_{car}($  $)$

$$\beta_{t+1} = \beta_t - \gamma_t \frac{\partial \overline{err}(f(x),y,\bar{y})}{\partial \beta_t}.$$

Wsabie - Algorithm

Algorithm 1 Online WARP Loss Optimization

Input: labeled data (x_i, y_i), $y_i \in \{1, \ldots, Y\}$.
repeat
  Pick a random labeled example (x_i, y_i)
  Let $f_{y_i}(x_i) = \Phi_W(y_i)^\top \Phi_I(x_i)$
  Set $N = 0$.
  repeat
    Pick a random annotation $\bar{y} \in \{1, \ldots, Y\} \setminus y_i$.
    Let $f_{\bar{y}}(x_i) = \Phi_W(\bar{y})^\top \Phi_I(x_i)$
    $N = N + 1$.
  until $f_{\bar{y}}(x_i) > f_{y_i}(x_i) - 1$ or $N \geq Y - 1$
  if $f_{\bar{y}}(x_i) > f_{y_i}(x_i) - 1$ **then**
    Make a gradient step to minimize:
      $L(\lfloor \frac{Y-1}{N} \rfloor)|1 - f_y(x_i) + f_{\bar{y}}(x_i)|_+$
    Project weights to enforce constraints (2)-(3).
  end if
until validation error does not improve.

Wsabie results

Method	ImageNet 2010		Web	
	prec@1	prec@10	prec@1	prec@10
approx kNN	1.55%	0.41%	0.30%	0.34%
One-vs-Rest	2.27%	1.02%	0.52%	0.29%
Wsabie	4.03%	1.48%	1.03%	0.44%

Wsabie results

Label	Nearest Neighbors
barack obama	barak obama, obama, barack, barrack obama, bow wow
david beckham	beckham, david beckam, alessandro del piero, del piero
santa	santa claus, papa noel, pere noel, santa clause, joyeux noel
dolphin	delphin, dauphin, whale, delfin, delfini, baleine, blue whale
cows	cattle, shire, dairy cows, kuh, horse, cow, shire horse, kone
rose	rosen, hibiscus, rose flower, rosa, roze, pink rose, red rose
eiffel tower	eiffel, tour eiffel, la tour eiffel, big ben, paris, blue mosque
ipod	i pod, ipod nano, apple ipod, ipod apple, new ipod
f18	f 18, eurofighter, f14, fighter jet, tomcat, mig 21, f 16

Wsabie results

	delfini, orca, **dolphin**, mar, delfin, dauphin, whale, cancun, killer whale, sea world
	blue whale, whale shark, great white shark, underwater, white shark, shark, manta ray, **dolphin**, requin, blue shark, diving
	barrack obama, barak obama, barack hussein obama, **barack obama**, james marsden, jay z, obama, nelly, falco, barack
	eiffel, paris by night, la tour eiffel, tour eiffel, **eiffel tower**, las vegas strip, eifel, tokyo tower, eifel tower

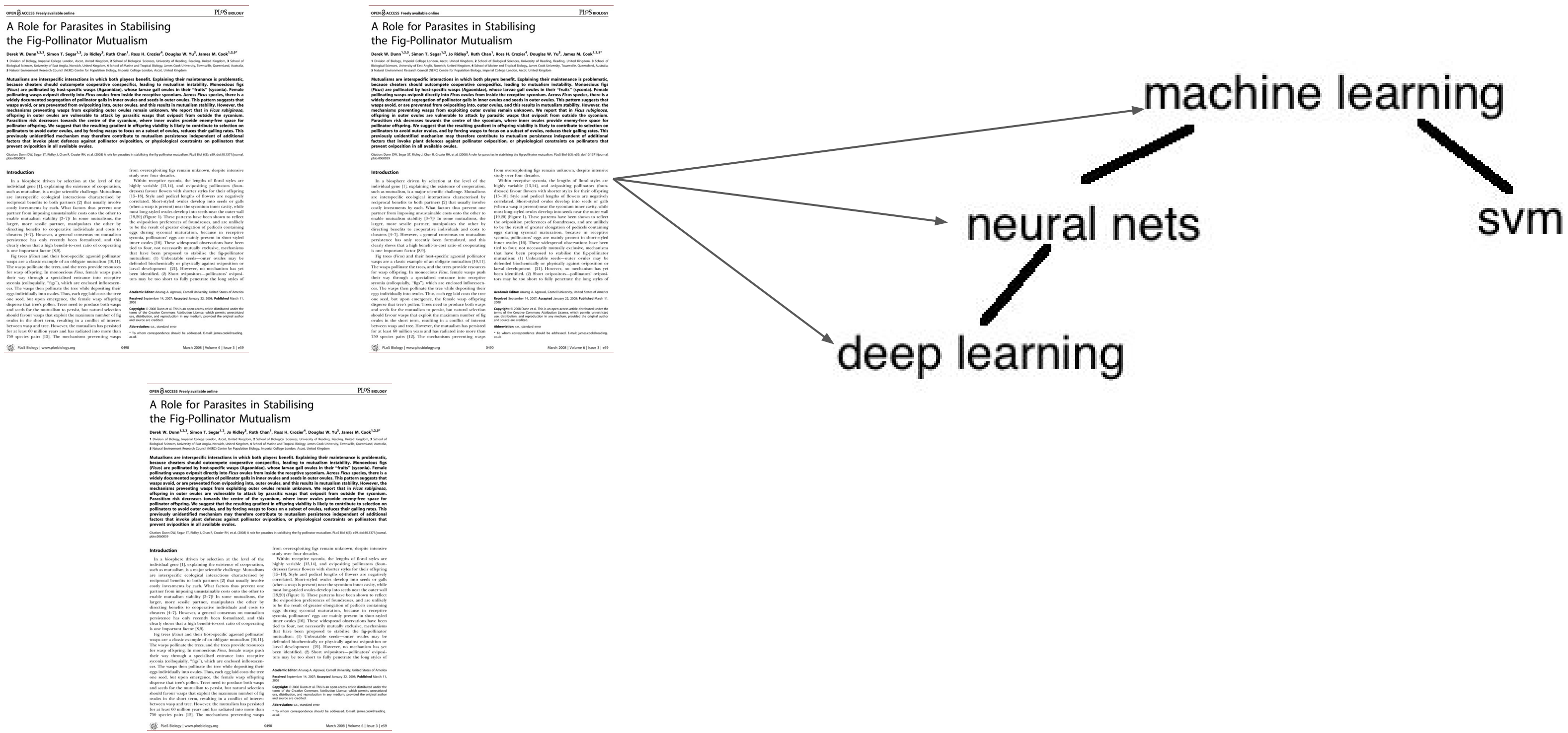
Leveraging label hierarchies

ECML 2016 - Predicting unseen labels using label hierarchies in large-scale multi-label learning - Jinseok Nam @ TU Darmstadt

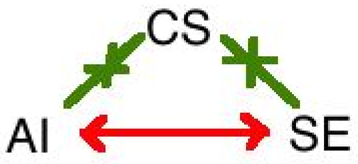
Text classification problem with label hierarchies

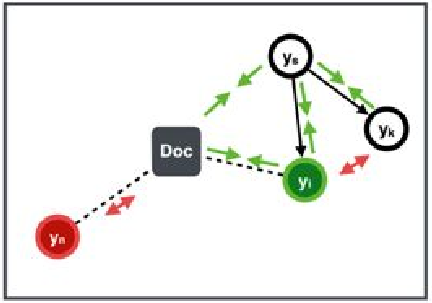
Instances

Labels

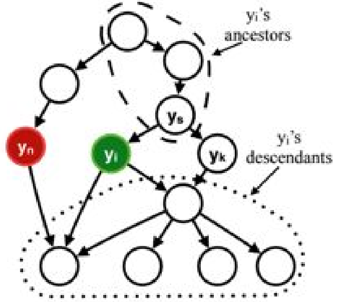


Learning with hierarchical structures over labels





(a) A joint space of instances and labels



(b) A label hierarhcy

$$\Omega(\mathbf{\Theta}_H) = \sum_{n=1}^{N} \frac{1}{\mathcal{Z}_A} \sum_{i \in \mathcal{Y}_n} \sum_{s \in \mathcal{S}_{\mathcal{A}}(i)} - \log p(y_s | y_i, \mathbf{x}_n)$$
$$+ \sum_{l=1}^{Y} \sum_{q \in \mathcal{S}_{\mathcal{P}}(l)} \sum_{\substack{k \in \mathcal{S}_{\mathcal{C}}(q) \\ k \neq l}} L\left(rank_y^1(\mathbf{u}_l)\right)\left[\, 1 - \mathbf{u}_q^T \mathbf{u}_l + \mathbf{u}_k^T \mathbf{u}_l \right]_+$$

$$p(y_s | y_i, \mathbf{x}_n) = \frac{\exp(\mathbf{u}_s^T \hat{\mathbf{u}}_i^{(n)})}{\sum_{v \in L} \exp(\mathbf{u}_v^T \hat{\mathbf{u}}_i^{(n)})},$$

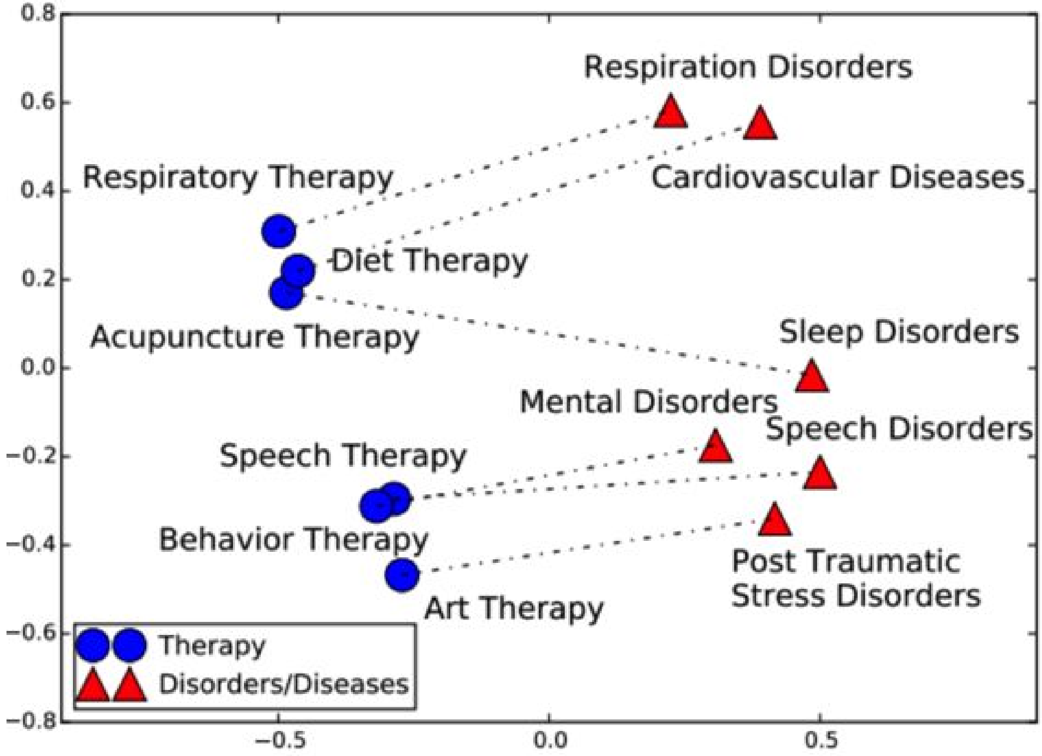
WsabieH results

Datasets

- RCV1-v2 newswire articles - 103 labels organized in a tree - 800k documents
- OHSUMED medical articles - 350k documents

	RCV1-v2		OHSUMED	
	Wsabie	Wsabie_H	Wsabie	Wsabie_H
AvgP	94.34	94.39	45.72	**45.76**
RL	0.44	0.44	4.09	**3.72**

WsabieH results



Predicting unseen labels

Hierarchical relations provide embeddings even of labels which are never seen in data. Therefore, they can appear in predictions in a meaningful way although data don’t provide information about it.

Thanks :)

andrija.m.djurisic@gmail.com

References

Weston, J., Bengio, S., Usunier, N.: Wsabie: Scaling up to large vocabulary image annotations

Jinseok, N., Eneldo, M., Hyunwoo, J., Furnkranz, J.: Predicting unseen labels using label hierarchies in large-scale multi-label learning

https://github.com/xdshang/wsabie