AMA - Green Seminar

Moura Simon

PhD Student
Under supervision of Massih-Reza Amini

Before

Now

ANY QUESTIONS?

Studies

- Master 2 of informatics, option Artificial Intelligence and the Web;
- Bachelor in Applied Mathematics;
- Licence 1 and 2 of fundamental mathematics.

Interests

Academic interests:

- Machine learning;
- Statistics;
- Artificial intelligence in general;

Others:

- Climbing;
- Running;
- Photography;
- ...

Master 2 internship (2015)

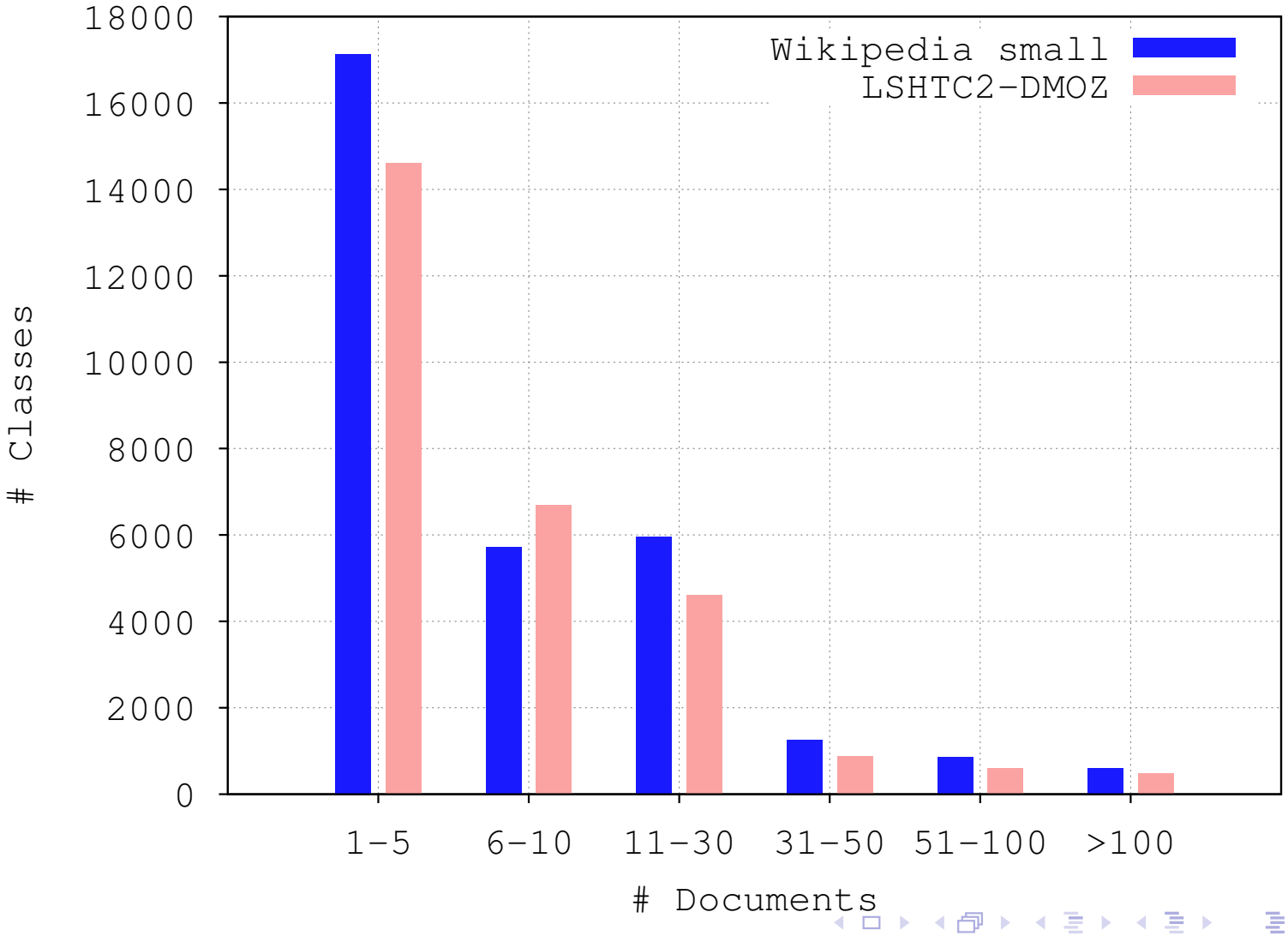
- Supervised by **Ioannis Partalas** (VISEO R&D) and **Massih-Reza Amini**.
- **Sparsification of linear models.**
- **2 main challenges:**
  - Large scale applications;
  - Unbalanced datasets;

Challenge 1/2: Large scale applications

Table: LSHTC datasets and their properties.

	#Categories	#Features	#Documents	Parameters (in GB)
DMOZ-2010	12,294	381,580	128,710	34.9
DMOZ-2011	27,875	594,158	394,756	123.4
DMOZ-2012	11,947	348,548	383,408	31
Wiki Small	36,504	346,299	538,148	94.2
Wiki Large	325,056	1,617,899	2,817,603	3918.3

Challenge 2/2: Unbalanced datasets



Linear model representation & Proposed method

Idea: **Sparsify a posteriori linear models to reduce its size and improve its accuracy**.

$$\begin{array}{c} \\ f1 \\ f2 \\ f3 \\ f4 \\ f5 \end{array} \begin{array}{c} \begin{array}{ccc} C1 & C2 & C3 \end{array} \\ \begin{pmatrix} 1.65 & 0.49 & 0.13 \\ 0.89 & 0.87 & 1.89 \\ 0.56 & 0.55 & 0.36 \\ 3.87 & 1.38 & 0.99 \\ 0.80 & 0.87 & 0.59 \end{pmatrix} \end{array} \rightarrow \begin{array}{c} \\ f1 \\ f2 \\ f3 \\ f4 \\ f5 \end{array} \begin{array}{c} \begin{array}{ccc} C1 & C2 & C3 \end{array} \\ \begin{pmatrix} 1.65 & 0 & 0 \\ 0 & 0 & 1.89 \\ 0 & 0 & 0 \\ 3.87 & 1.38 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{array} \rightarrow \begin{cases} C1 \rightarrow \{1 : 1.65, 4 : 3.87\} \\ C2 \rightarrow \{4 : 1.38\} \\ C3 \rightarrow \{2 : 1.89\} \end{cases}$$

Some results: Memory usage

Table: Sparsity and models size for penalty parameter C= 100.

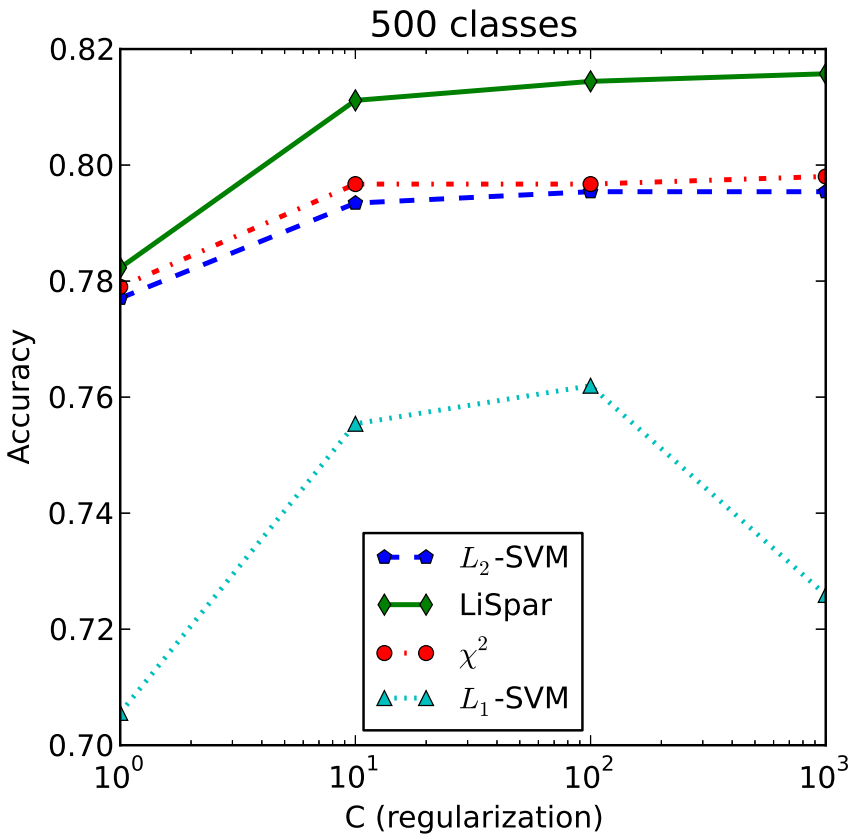
#Classes	L_2-**SVM**	**LiSpar**	L_1-**SVM**	χ^2
500	68.583% (260 Mb)	96.788% (27 Mb)	99.612% (3.1 Mb)	51.708% (72 Mb)
1000	72.479% (723 Mb)	97.062% (77 Mb)	99.701% (7.3 Mb)	57.004% (202 Mb)
2000	75.308% (2200 Mb)	98.228% (148 Mb)	99.769% (18 Mb)	66.072% (674 Mb)
3000	76.415% (4000 Mb)	99.194% (125 Mb)	99.786% (33 Mb)	66.524% (1100 Mb)

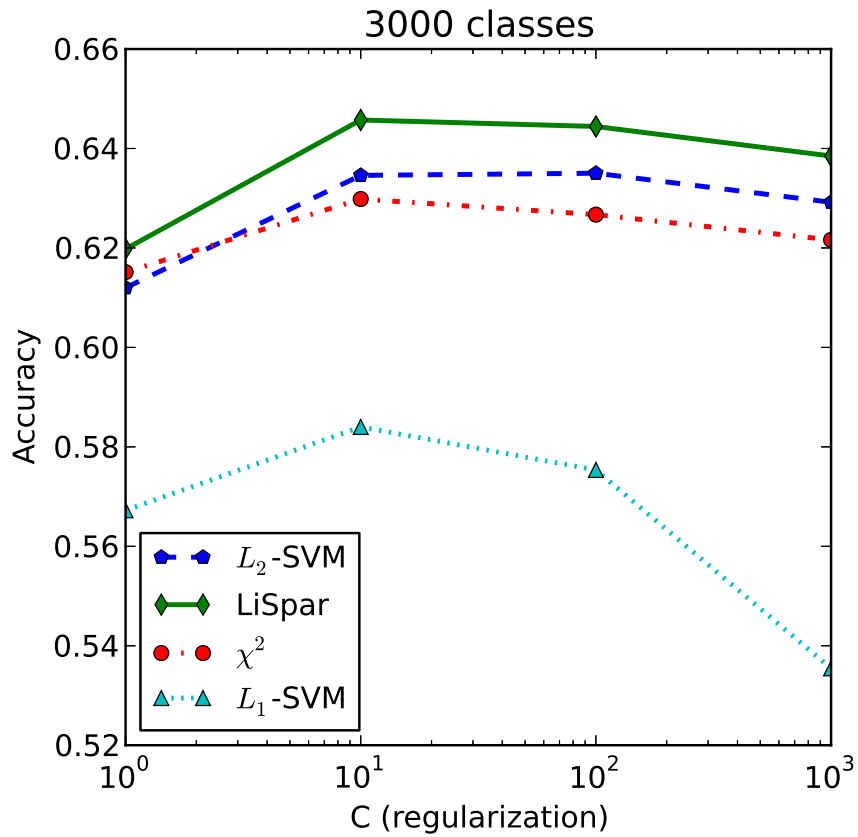
Some results: Time

Table: Prediction time measurements in seconds on sparse representation for L_2-SVM, LisPar, L_1-SVM and χ^2. All models have been trained with a regularization parameter $C = 100$.

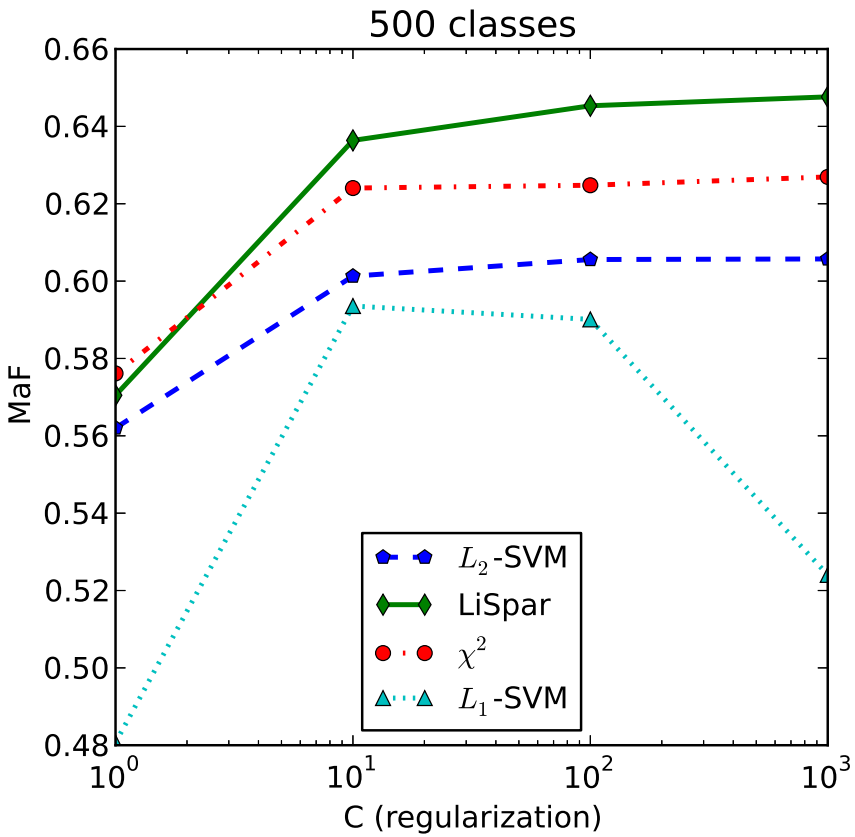
#Classes	L_2-**SVM**	χ^2	**LiSpar**	L_1-**SVM**
500	110.75	20.01	9.05	2.64
1000	439.01	131.94	62.9	12.71
2000	2163.43	584.17	177.74	48.39
3000	5405.64	1508.72	290.31	120.47

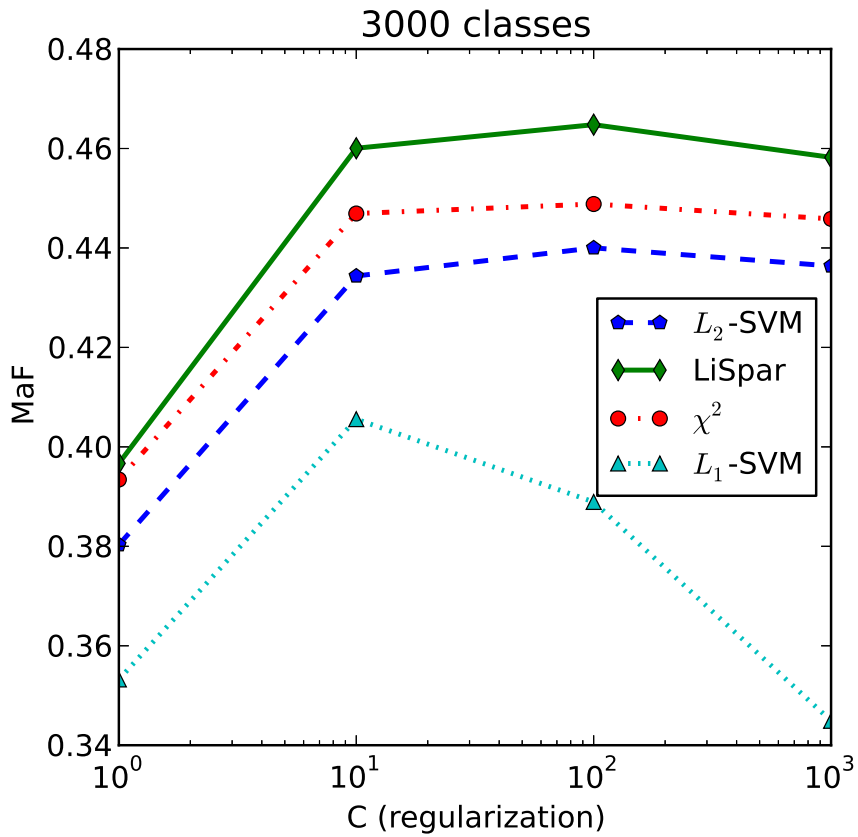
Some results: Accuracy





Some results: MaF





Master 1 internship (2014)

Supervised by **Eric Gaussier**: Hierarchical classification.

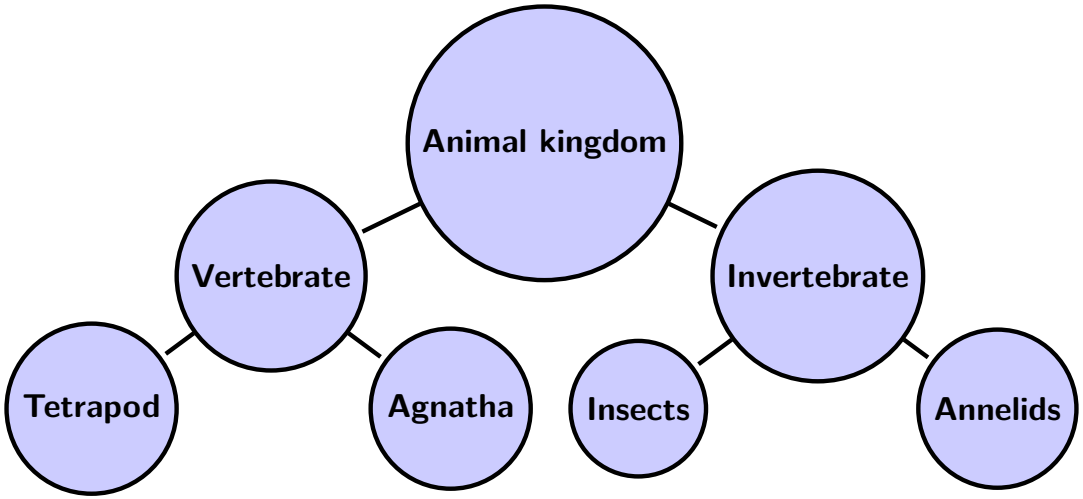


Figure: Animals taxonomy

Generalization error: upper bound

Ideas explored during this internship:

- Use the boundary on generalization error to build an efficient hierarchy for classification [1];
- Grow a hierarchy using a bottom up approach;
- We used a **similarity matrix** to gather "close" classifiers and build a **dendrogram of classifiers**;

Results:

- Close to the original approach in term of precision;
- Faster in term of running time (compared to flat approach);
- But...

Creation of IR function under constraints (2013)

Bachelor internship with **Eric Gaussier**, **Parantapa Goswami** and **Francis Maes**.

Rank documents by relevance regarding a query using:

$$RSV(q,d) = \sum_{w \in q \cap d} x_w^q f(x,y)$$

- q represents a query and d a document.
- $x = x_w^d \log\left(1 + c * \frac{m}{y_d}\right), y = \frac{\#\text{Docs containing w}}{\#\text{Docs in the collection}}$.
- y_d represents the length of document d and m the average length of documents in the collection.

IR constraints

Generate exhaustively functions respecting the following constraints:

- Derivative constraints :

$$\frac{\partial f(x,y)}{\partial x} > 0 \tag{1}$$

$$\frac{\partial f(x,y)}{\partial y} < 0 \tag{2}$$

$$\frac{\partial^2 f(x,y)}{\partial x^2} \leq 0 \tag{3}$$

Using the following operators:

- $+, -, *, /, **$
- $log, sqrt, exp$
- x, y, K

Results

Score comparison between our formulas and BM25 and LM-Dir on set Clef-3 :

CLEF3 :

Functions	**MAP**	**P@10**
BM25	0.4474	0.3352
LM-Dir	0.3968	0.3019
$(x/y)^{\frac{1}{2}}$	0.3425	0.2741
$\sqrt{\frac{\sqrt{x}}{y}}$	**0.4311**	**0.3296**
$\sqrt{\frac{\sqrt{x}}{\sqrt{y}}}$	**0.4412**	**0.3296**
$\frac{\sqrt{x\sqrt{y}}}{y}$	**0.4431**	**0.3093**
$\sqrt{\frac{\sqrt{x*y}}{y}}$	**0.4412**	**0.3296**

Results

Score comparison between our formulas and BM25 and LM-Dir on set TREC-7 :

TREC-7 :

Functions	**MAP**	**P@10**
BM25	0.1828	0.4180
LM-Dir	0.1863	0.3920
$\sqrt{x} - \sqrt{y}$	0.1824	**0.4220**
$\log \frac{x+y}{y}$	**0.1882**	**0.4280**
$\sqrt{\frac{\sqrt{x}}{\sqrt{y}}}$	**0.1944**	**0.4280**
$\sqrt{\frac{\sqrt{x*y}}{y}}$	**0.1944**	**0.4220**
$\sqrt{(\frac{x}{y})^{\frac{k}{2}}}$	**0.1944**	**0.4220**

END: ANY QUESTIONS?

Bibliography

Rohit Babbar, Ioannis Partalas, Eric Gaussier, and Massih-Reza Amini.
On flat versus hierarchical classification in large-scale taxonomies.
In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 1824–1832. Curran Associates, Inc., 2013.