

Results of the Sequence Prediction Competition (SPiCe)

Borja Balle ¹ Rémi Eyraud ² Franco Luque ³ Ariadna
Quattoni ⁴ Sicco Verwer ⁵

¹Lancaster University

²Aix-Marseille Université

³Universidad Nacional de Córdoba and CONICET

⁴Xerox Research Centre Europe

⁵Delft University of Technology

ICGI 2016 (Delft)

Introduction

Goal

Organize a competition relevant to ICGI but inclusive to other communities

The Sequence Prediction Challenge

Observe a string (prefix) of symbols in a finite alphabet, then predict a ranking of possible symbols to be the next element of the sequence

Key Difference with previous competition (PAutomaC)

- ▶ Task more amenable to a wide variety of methods (statistics, neural networks, multi-class classification)
- ▶ Evaluation based on public (on-competition) and private (end of competition) test sets to prevent leaderboard climbing

Some Statistics

- ▶ Number of registered participants: 82
- ▶ Number of active participants: 26
- ▶ Avg. number of submissions per team: 142
- ▶ Number of tasks: 15
- ▶ Range of alphabet sizes: 4 - 6 722 (most around 20)
- ▶ Range of training and test set sizes: 5 000 - 200 000 (train), 748 - 5 000 (test)

Competition Mechanics

- ▶ Once registered, participants get access to data :
 - ▶ a training set of whole sequences
 - ▶ a public and a private test set (prefixes)
- ▶ Submissions are evaluated on-line with one prediction at a time from the public test set
- ▶ Prediction is ranking of 5 most probable symbols to be the next one
- ▶ Evaluation with NCDG-5 metric:

$$NDCG_5(\hat{a}_1^i, \dots, \hat{a}_5^i) = \frac{\sum_{k=1}^5 \frac{p(\hat{a}_k^i | y_i)}{\log_2(k+1)}}{\sum_{k=1}^5 \frac{p_k}{\log_2(k+1)}} ,$$

- ▶ After the training phase is over, entries are evaluated (once!) on a private test set to get final ranking

Type of Data

Problem Nb	Type
1 (20)(20k)(5k)	synthetic (non-stationary HMM with 2 states)
2 (10)(20k)(5k)	synthetic (non-stationary HMM with 2 states)
3 (10)(20k)(5k)	synthetic (non-stationary HMM with 4 states)
4 (33)(6k)(749)	NLP (English verbs, character level, Penn Treebank)
5 (49)(33k+)(4k+)	NLP (character level language modeling, Penn Treebank)
6 (60)(5k)(5k)	partly synthetic, software engineering (RERS 2013 problem 34)
7 (20)(65k+)(5k)	Biology (protein family PF13855, full set, Pfam)
8 (48)(14+)(1.7k)	NLP (Spanish simplified POS sentences, Ancora)
9 (11)(5k)(5k)	partly synthetic, software engineering (RERS 2013 problem 42)
10 (20)(55k)(4.8k)	Biology (protein family PF00400, RP15 subset, Pfam)
11 (6722)(32k+)(4k)	NLP (English lemmas from Flickr-8000)
12 (21)(200k)(3k)	synthetic (PAUTOMAC generator)
13 (702)(26k+)(3.3k)	NLP (English spelling correction from Twitter Typos Corpus)
14 (27)(10k)(5k)	partly synthetic (ALERGIA, DPFA based on problem 4)
15 (32)(50k)(5k)	partly synthetic (ALERGIA, DPFA based on problem 5)

Final Results

Position	Team leader	Global Score	Method
1	Chihiro Shibata	10.4498481	RNN using strictly piecewise features
2	Shanbo Chu	10.0198711	MLP, CNN & LSTM (& weighted n-gram)
3	Ichinari Sato	9.9562550	n-gram, spectral, RNN & tree boosting
4	Farhana Ferdousi Liza	9.4082437	spectral learning of WA & n-gram
5	Du Xi	9.1841280	parameters tuning of spectral learning
6	Martine De Cock	9.0242665	n-gram ($n = 3$ or 4)
7	Benjamin Loos	8.9515982	DPFA (state merging)
8	Quang Vinh Dang	8.9494857	RNN using LSTM cells
9	Rafael Ktistakis	6.7270460	
10	Alok Kumar	6.0221453	

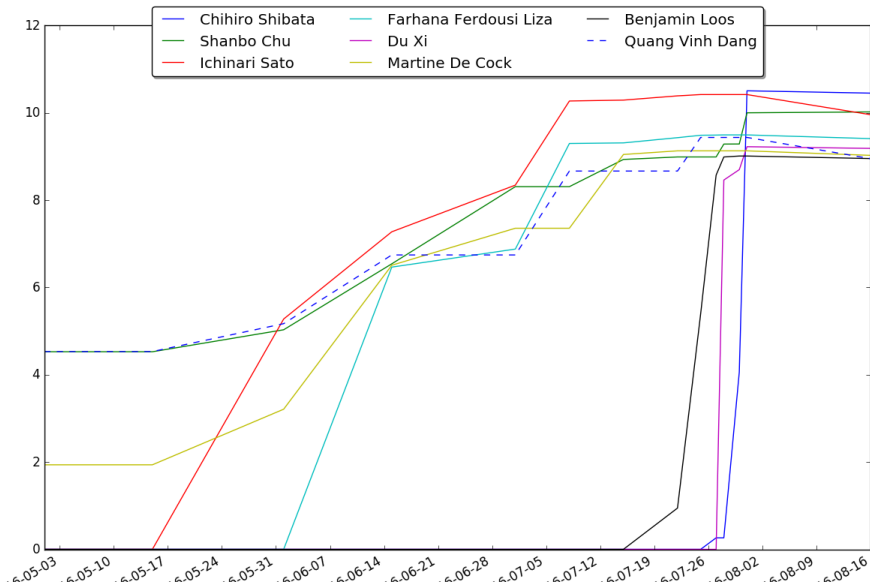
Analysis of Results

Comparison between private and public test sets

Pos.	Team leader	Score	Public Sc.	Public Pos.
1	Chihiro Shibata	10.4498481	10.5036139	1
2	Shanbo Chu	10.0198711	9.9987526	3
3	Ichinari Sato	9.9562550	10.4169258	2
4	Farhana Ferdousi Liza	9.4082437	9.4928019	4
5	Du Xi	9.1841280	9.2227905	6
6	Martine De Cock	9.0242665	9.1279415	7
7	Benjamin Loos	8.9515982	9.0090818	8
8	Quang Vinh Dang	8.9494857	9.4351277	5
-	3-gram baseline	8.6250907	8.2074738	-
9	Rafael Ktistakis	6.7270460	7.5950857	9
10	Alok Kumar	6.0221453	1.7499863	18
11	Ilya Verenich	1.9467110	6.5513241	10

Analysis of Results

Score evolution on the public test sets



Analysis of Results

On synthetic data

Results on problems numbered 1, 2, 3, 12.

Positions indicated in parenthesis are computed without taking into account Problem 12 (very large training sample).

Position	Team leader	Score
1 (1)	Chihiro Shibata	3.5280928
2 (2)	Ichinari Sato	3.4708998
3 (3)	Shanbo Chu	3.4321920
4 (6)	Farhana Ferdousi Liza	3.3883254
5 (4)	Du Xi	3.3385820
6 (5)	Benjamin Loos	3.2766924
7 (7)	Martine De Cock	3.2046625
8 (9)	Quang Vinh Dang	2.9504653
9 (10)	Rafael Ktistakis	2.8351546
10 (8)	Alok Kumar	2.8001947

Analysis of Results

On NLP data

Results on problems numbered 4, 5, 8, 11, 13

Positions indicated in parenthesis are computed without taking into account Problem 11 (very large alphabet).

Position	Team leader	Score
1 (1)	Chihiro Shibata	3.2072033
2 (2)	Shanbo Chu	2.9951987
3 (3)	Ichinari Sato	2.7480545
4 (4)	Quang Vinh Dang	2.6916002
5 (5)	Farhana Ferdousi Liza	2.6789179
6 (8)	Du Xi	2.6670005
7 (6)	Martine De Cock	2.5616705
8 (7)	Benjamin Loos	2.5081655
9 (9)	Rafael Ktistakis	1.4328179
10 (10)	Alok Kumar	1.0457101

Analysis of Results

On other data

Results on Biology data (problems numbered 7 and 10), Software Engineering ones (6,9), and DPFA from NLP ones (14, 15).

Team leader	Biology	Software E.	DPFA from NLP
Chihiro Shibata	1.2879198 (2)	1.8223056 (1)	0.6043264 (10)
Shanbo Chu	1.1959674 (3)	1.6838970 (5)	0.7126157 (4)
Ichinari Sato	1.3365225 (1)	1.7716953 (2)	0.6290825 (7)
Farhana Ferdousi Liza	0.8439897 (6)	1.7420696 (3)	0.7549409 (1)
Du Xi	0.8363763 (7)	1.6295532 (7)	0.7126157 (4)
Martine De Cock	0.9262389 (5)	1.7271268 (4)	0.6045675 (9)
Benjamin Loos	0.8363763 (8)	1.6724305 (6)	0.6579333 (6)
Quang Vinh Dang	1.1323993 (4)	1.5698342 (8)	0.6051864 (8)
Rafael Ktistakis	0.5336884 (10)	1.187472 (9)	0.7379122 (2)
Alok Kumar	0.6025545 (9)	0.857234 (10)	0.7164515 (3)

Conclusion

- ▶ Lot of energy but lot of fun
- ▶ Pushing forward of several methods
- ▶ Clear comparison of different approaches
- ▶ New publicly available Benchmark

Thanks to all participants and
Congratulations to the winners!