2010 Approximate Inference Evaluation

Co-Chairs: Gal Elidan and Amir Globerson **Student organizer:** Uri Heineman The Hebrew University

Advisory board: Jeff Bilmes (Univ. of Washington), Rina Dechter (UCI), Peter Grunwald (CWI), Isabelle Guyon (Clopinet), Peter Spirtes (CMU)

Overview

Tasks:

- MAR: univariate marginals
- PR: partition function (probability of evidence)
- MPE: most probably explanation

Time frames:

- Lightning speed: 20 seconds
- Coffee break: 20 minutes
- Lunch break: 1 hour (reduced from 2 hours)

Anonymous solvers (on request)



Performance Measures

MAR: relative to exact solution (0-2 range)

$$\frac{1}{N} \sum_{i=1}^{N} \sum_{x_i} |P^*(X_i = x_i) - P^s(X_i = x_i)|$$

PR: relative to exact solution (no natural range)

$$|\log Z^* - \log Z^s|$$

- MPE: relative to asynchronous belief propagation $\frac{E^{s}-\min(E^{bp},E^{naive})}{\left|\min(E^{bp},E^{naive})\right|}$
- > Average over networks within domains
- Sum across domains

How it worked

- Fully automated after identity verification
- "Live" leader-board

Time 20 seconds

Group Name	PR	MAR	MPE
Stephen Gould	-	-	-0.7224
Radu Marinescu	-	-	-1.3466
AsynchBP	70.7919	1.1072	0.0073
CogitoFrgoTafor	2 2775	0 3065	-1 0252
Time 20 seconds			

And details

11110 20 50001105					
Categories	PR	MAR	MPE		
SP	8.4864(0/8)	0.0393(0/8)	0.0000(0/21)		
rids	59.2590(0/20)	0.6338(0/20)	0.0000(0/32)		
image Alignment	-	-	0.0000(0/5)		
Medical diagnosis	1.1527(0/26)	0.1722(0/26)	-		
Not see the Deck see to be see	0 1577 (0 (0 ()	0.0005 (0.000)	0.0070(0/201)		

- Multiple solvers allowed (1-5 in practice)
- Resubmission allowed (>10 for some solvers)
- Blind networks only domain name was given
- Few networks revealed to help debugging

Network: The Big Picture

	PR	MAR	MPE	Total
20 sec	204	204	461	869
1200 sec	204	204	532	940
3600 sec	204	204	287	695
Total	612	612	1280	2504

Networks – by domain (1 hour)

Network	PR	MAR	MPE
CSP	8	8	55
Grids	20	20	40
Image Alignment			10
Medical Diagnosis	26	26	
Object Detection	96	96	92
Pedigree	4	4	
Protein Folding			21
Protein-Protein Interaction			8
Segmentation	50	50	50

Many thanks to: Kristian Kersting, Stephen Gould, Menachem Fromer, Ben Packer, Dan Geiger, Ariel Jaimovich, Farshid Moussavi

CPU Hours Per Solver

	PR	MAR	MPE	Total
20 sec	1+	1+	3+	6+
1200 sec	74+	74+	195+	345+
3600 sec	224+	224+	315+	675+
Total	300+	300+	514+	1115+

All runs on Intel XEON 5550 with 4GB RAM per job

Competition Timeline



Competition Timeline



Leader-board Summary

Seconds	PR	MAR	MPE
20	Arthur Choi	Arthur Choi	Joris Mooii
20	(UCLA)	(UCLA)	(Max Planck)
1200	Vibhav	Vibhav	Thomas
	Gogate	Gogate	Schiex
	(UW+UCI)	(UCI)	(INRA)
3600	Vibhav	Vibhav	Joris
	Gogate	Gogate	Mooij
	(UW+UCI)	(UCI)	(Max Planck)

Neck-to-neck Competition

Group Name	PR	MAR	MPE
AsynchBP	75.1875	1.1501	0.6807
CogitoErgoInfer	2.0707	0.2894	-1.2272
CogitoErgoInfer2	2.0707	0.2826	-1.2191
CogitoErgoInfer3	2.0707	0.2810	2.5925
daoopt.mpe	-	_	-1.0059
daoopt.mpe.anytime	-	_	-0.9803
edbp	1.3063	0.1762	0.0711
edbq	1.3134	0.1742	0.0701
edbr	1.3259	0.1841	4.0749
FandS	_		-0.2611
ijgp	-	0.1722	-

Winning Teams

- (MAR) IJGP by Vibahv Gogate, Andrew Gefland, Natasha Flerova and Rina Dechter (UCI): Anytime iterative GBP based algorithm
- (PR) Vgogate by Vibahv Gogate, Pedro Domingos (UW),, Andrew Gefland and Rina Dechter (UCI): Formula based importance sampling
- (PR+MAR) EDBP by Arthur Choi, Adnan Darwiche, with support from Glen Lenker and Knot Pipatsrisawat (UCLA): Anytime BP based anytime thickening of structure
- (MAP) **libDAI** by Joris Mooij (Max Planck): junction tree, LBP/MP, double-loop GBP, Gibbs, decimation
- (MAP) toulbar2 by Thomas Schiex et al (INRA)
 Anytime branch and bound weighted CSP solver

The good and the Bad

- ✓ Active participation: 9 teams, 26 solvers
- "Live" leader board encouraged aggressive competition (over 10 re-submissions for some solvers)
- Solvers from varied communities: propagation based inference, anytime algorithms, CSP,...
- X Little response up to two weeks before deadline to
 - Upper/lower bound track
 - Learning track (inference within estimation)
- X Sensitivity to specific failures

Looking Ahead

To appear in the website in the coming weeks:

- Detailed summary of competition and results
- Solvers descriptions and links (if provided)

Plans for the future:

- Looking into feasibility of continuing service to community (anyone wants to support this?)
- Expand variety of networks / evaluations
- Creating a standardized benchmark for inference

The Team



Gal Elidan



Amir Globerson

