

Reflections on a Diagnosis Competition – DXC'09

First Success and Future Challenges

presented by Tolga Kurtoglu, Ph.D.
Mission Critical Technologies @ NASA Ames Research Center

What is diagnosis?

from Wikipedia, the free encyclopedia

“As a subfield in artificial intelligence, *diagnosis* is concerned with the development of algorithms and techniques that are able to determine whether the behavior of a system is correct. If the system is not functioning correctly, the algorithm should be able to determine, as accurately as possible, which part of the system is failing, and which kind of fault it is facing. The computation is based on observations, which provide information on the current behavior.”

Why is diagnosis a challenge?

incorrect and/or insufficient knowledge about system behavior

limited observability

presence of many different types of faults

(system/supervisor/actuator/sensor faults,
additive/multiplicative faults,
abrupt/incipient faults,
persistent/intermittent faults)

non-local and delayed effect of faults due to dynamic system behavior

presence of other phenomena that mask the symptoms of faults

(unknown inputs acting on system,
noise that affects the output of sensors, etc.)

What lead to a competition in diagnosis?

“Despite the development of a variety of notations, techniques, and algorithms, efforts to evaluate and compare the different diagnosis algorithms have been minimal. One of the major deterrents is the lack of a common framework for evaluating and comparing diagnosis algorithms.”

so we have developed a framework...

the framework accomplishes the following objectives:

accelerate research in theories, principles, modeling and computational techniques for diagnosis of physical systems

encourage the development of software platforms that promise more rapid, accessible, and effective maturation of diagnosis technologies

provide a forum that can be utilized by algorithm developers to test and validate their technologies

systematically evaluate different diagnosis technologies by producing comparable performance assessments

What lead to a competition in diagnosis?

we have the framework...

what better way to have its first concrete implementation then by running a competition...

the birth of a new AI competition...

First International Diagnosis Competition (DXC'09)

url: <http://www.dx-competition.org>

DXC'09 Organizing Committee

Johan de Kleer, Lukas Kuhn PARC

Alexander Feldman Delft University of Technology & PARC

Tolga Kurtoglu Mission Critical Technologies @ NASA Ames

Sriram Narasimhan UC Santa Cruz @ NASA Ames

Scott Poll NASA Ames Research Center

David Garcia Stinger Ghaffarian Technologies @ NASA Ames



DXC Objectives



~~the framework accomplishes the following objectives:~~

accelerate research in theories, principles, modeling and computational techniques for diagnosis of physical systems

encourage the development of software platforms that promise more rapid, accessible, and effective maturation of diagnosis technologies

provide a forum that can be utilized by algorithm developers to test and validate their technologies

systematically evaluate different diagnosis technologies by producing comparable performance assessments

analyze trade-offs among diagnostic performance metrics

DXC'09 Format

12 diagnosis algorithms competed in 3 tracks that included industrial and synthetic systems

initially the participants were provided with system descriptions and a sample data set that included nominal and fault scenarios

the participants had to provide algorithms that communicated with the run-time architecture to receive scenario data and return diagnostic results

for the competition we ran all the algorithms on extended scenario data sets (different from sample set) to compute a set of predefined metrics and a final ranking based on weighted metric scores

monetary prize for winners in each track

DXC'09 Overview: Systems

Track	Tier	Systems	Description
Industrial	1	ADAPT-Lite	Basic faults injected into a simplified EPS (Electrical Power System) testbed
	2	ADAPT	More complex faults injected into the full EPS distribution system
Synthetic	1	ISCAS85	Multiple faults injected into the circuits from the ISCAS85 benchmarks

DXC'09 Overview: Participants

The participating diagnosis algorithms used a variety of reasoning schemes (not just model-based). These include model-based, Bayesian network-based, rule-based, statistical, and optimization approaches.



Submissions came from all over the world including Australia, Italy, USA, The Netherlands, and Sweden.

DXC'09 Overview: Results

The results of the competition indicate that no algorithm was able to dominate all metric categories which illustrates that selecting a diagnosis algorithm involves many trade-offs.

Tier 1 (ADAPT-Lite)

	Weight	RODON	Wizards Of Oz	Fault Buster	ProADAPT	HyDE-A	HyDE-S	RulesRule	FACT	RacerX
FP Rate	1.3	0.0645	0.0000	0.1333	0.0333	0.0000	0.2000	0.8246	0.2813	0.0645
FN Rate	1.3	0.0968	0.5000	0.3438	0.0313	0.4688	0.0741	0.0000	0.0667	0.1613
Det Acc	0.3	0.9194	0.7419	0.7581	0.9677	0.7581	0.8548	0.2419	0.8226	0.8871
Class Errors	2.2	10.000	24.000	32.000	2.000	26.649	26.000	76.000	25.000	32.000
T_det (ms)	2.2	218	11530	1893	1392	13223	130	1000	373	126
T_iso (ms)	1.5	7205	11626	9259	4084	13840	653	282	9796	999999
CPU (ms)	0.6	11766	1039	2039	1601	24795	513	117	1767	139
Mem (kb)	0.6	26679	1781	2539	1680	5447	5795	3788	4340	3572
FINAL SCORES:		59.850	46.300	35.750	72.800	31.750	59.500	51.800	50.400	51.850

Tier 2 (ADAPT)

	Weight	RODON	Wizards Of Oz	Fault Buster	ProADAPT	HyDE	Stanford
FP Rate	1.3	0.5417	0.5106	0.8143	0.0732	0.0000	0.3256
FN Rate	1.3	0.0972	0.0959	0.2400	0.1392	0.3000	0.0519
Det Acc	0.3	0.7250	0.7417	0.4250	0.8833	0.8000	0.8500
Class Errors	2.2	84.067	159.248	130.000	76.000	121.569	110.547
T_det (ms)	2.2	3490	30742	14099	5981	17610	3946
T_iso (ms)	1.5	36331	47625	37808	12486	21982	14103
CPU (ms)	0.6	80261	23387	5798	3416	29612	963
Mem (kb)	0.6	29878	7498	10261	6539	20515	5912
FINAL SCORES:		70.500	51.400	52.400	83.200	61.000	81.500

DXC'09 Technical Challenges, Logistical Considerations

framework development (engineering issues)

how to develop a standard representation format?

how to define a set of *useful* metrics?

how to develop a software run-time architecture?

considerations for the competition (logistics)

where to host the competition?

what are the pros and cons of systems selected?

what modeling information to provide to participants?

how to aggregate individual metric scores into a final score?

DXC'09 Reflections and Lessons Learned

limited lessons learned based on 1 year of experience

it is extremely difficult to come up with *useful, generic* metrics
metrics without context does not mean much

one needs a top-down approach (systems engineering)
to defining the metrics based on user requirements

“industry relevance” is a key ingredient for success

impact of DXC on research directions in diagnosis

feedback from the community: great start, keep it up!

no real influence on research directions yet!

strategizing for winning a competition can be an
enemy for research!

DXC'09 Future Outlook

we intend to continue on a yearly basis...

already started planning DXC'10

sustainability (?)

future objectives of DXC (envisioned impact on diagnosis community)

establish a set of common benchmarks

standardized representation language (similar to PDDL)

compatibility with industry standards

web-based accessibility

scalability

Questions?

