

AUTOMATED DIAGNOSIS OF INBORN ERRORS OF METABOLISM USING HIGH-THROUGHPUT NMR SPECTROSCOPY

Ajit Singh¹, Russell Greiner¹, Victor Dorian² and Brent Lefebvre²

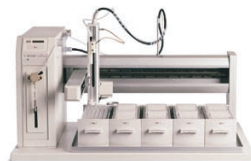
¹Department of Computing Science, University of Alberta

²Chenomx Inc.

CCMA : CONCURRENT CHEMICAL MIXTURE ANALYSIS



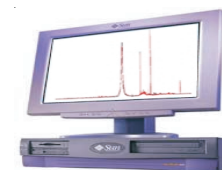
Inborn errors of metabolism (IEMs) are an extremely rare class of genetic diseases.



The Versatile Automated Sampler Transport (VAST) rapidly injects samples into the NMR spectroscope.



High-throughput Nuclear Magnetic Resonance (NMR) spectroscopy.

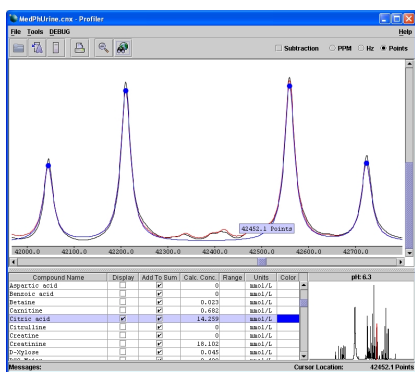


Chemomx's CCMA software automatically identifies 75 small metabolites by their spectral signature.

The sample type for these experiments is urine.

Accurate identification is normally a time-consuming process for experts.

DxTOOL : DIAGNOSTIC TOOL



1-Methylhistidine	0.0687
Acetamide	0.0353
Adipic Acid	0.0000
Leucine	0.0449
Sucrose	0.0144
...	...

Inputs

- Compound-concentration pairs
- Noisy concentrations
- Occasionally incomplete data



DxTool

DxTool uses

- Bayesian networks
- Bayesian error bars for confidence of diagnosis.

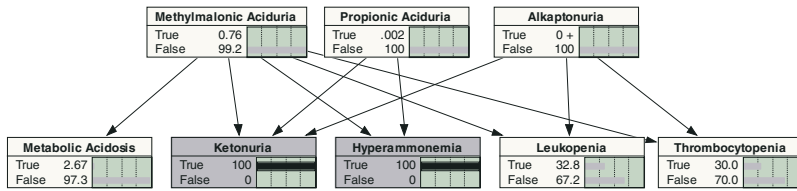


IEM	Pr(IEM E)
Methylmalonic Aciduria	0.0051 ± 0.0023
Hartnup's Disorder	0.0001 ± 0.0012
...	...

DxTool produces

- Probability of each IEM
- Confidence intervals
 - Variance of the response
 - Linear-normal assumption

Design



BN2O (Binary Node 2-Layer Noisy-OR)

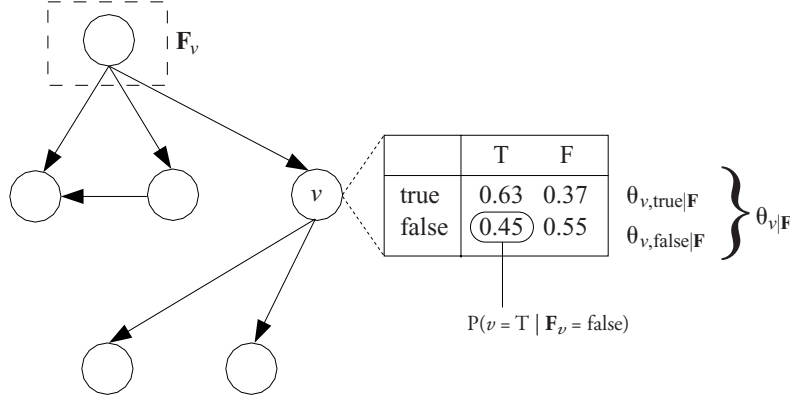
- One layer of disease states
- One layer of symptom States
- Encodes $P(\text{Symptom}|\text{Disease})$

Pros / Cons

- ✓ Small Number of Parameters
 $O(n)$ instead of $O(2^n)$ per CPTable
- ✓ Parameters have clear interpretation
Use of medical literature and experts
- ✓ Discretization to normal/abnormal works well for most symptoms
- ✓ Inference can be done in $O(2^{|\text{positive findings}|})$
Quickscore Algorithm [1]
✗ Not yet possible w/ Bayesian error bars
- ✓ Allows modelling of a patient with two similar IEMs
(e.g. Both 3-Methylglutaconic aciduria and 3-Methylglutaric aciduria)

[1] Heckerman, D. *A tractable inference algorithm for diagnosing multiple diseases*. UAI-89.

Error Bars



Let

H : Query (Hypothesis) Variables
E : Evidence Variables
 $\Theta : \{\theta_{v|F}\}_v$

Then for $Q = P(H|E, \Theta)$

$$\sigma_Q^2 = \sum_{v \in V} \sum_{f \in F_v} \frac{A_{vf} - B_{vf}}{1 + \sum_x \alpha_{v,x|f}}$$

$$A_{vf} = \sum_{v \in V} \frac{\{p_v(h, x, f | e) - p(h | e)p(x, f | e)\}^2}{\mu_{v,x|f}}$$

$$B_{vf} = \{p_v(h, f | e) - p(h | e)p_v(f | e)\}^2$$

In the limit of large sample sizes. [2]

- ◆ Computed using modified bucket elimination (BE) algorithm
- ◆ Effectively another pass of BE

[2] T. van Allen, R. Greiner, P. Hooper. *Bayesian Error Bars for Belief Net Inference*. UAI-01.

Example

(all numbers in mmol/L)

3-D-Hydroxybutyrate	0.2477
3-Furoic acid	0.0000
5-Hydroxylysine	1.0444
Alanine	0.4411
Methylmalonic acid	0.0000
...	

3-D-Hydroxybutyrate:	abnormal
3-Furoic acid:	normal
5-Hydroxylysine:	abnormal
Alanine:	normal
Methylmalonic acid:	normal
...	

Pr(Methylmalonic aciduria) = 0.0220 ± 0.0319
Pr(Glutaric aciduria) = 0.0000 ± 0.0176
Pr(Zellweger) = 0.0210 ± 0.0118

Conclusions

Broad classes of IEMs identified by 1 to 3 metabolites.

Learning structure and parameters nearly impossible here
e.g. IEMs with less than 5 cases.

Confidence of diagnosis allows meaningful discrimination of nearly equiprobable diagnoses