Query-Based Diagnostics

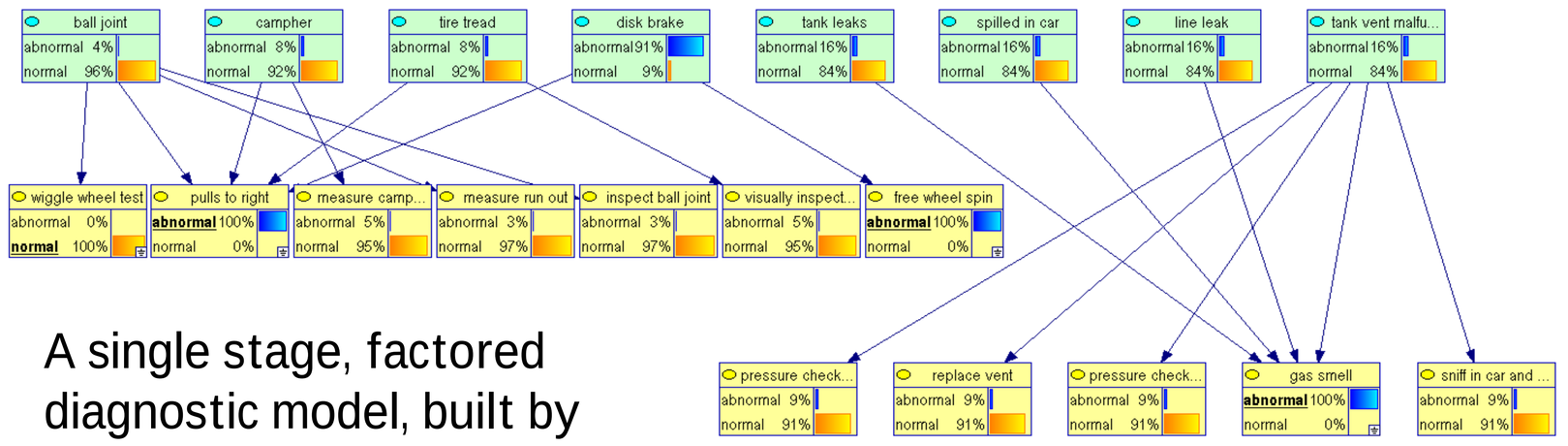
John Mark Agosta*, Thomas H. Gardos *, Marek J. Druzdzel §

* Intel Corporation, Santa Clara, CA, USA § Faculty of Computer Science, Bialystok Technical University, Poland & School of Information Sciences, Univesity of Pittsburgh, Pennsylvania

Modifying the model while problem-solving

We model diagnostic problems based on inferring causes by passive observation of a diagnostician's work-flow. Then by recording their findings and final diagnosis, the model can be modified directly, or improved by learning from cases so acquired.

The hidden probabilistic model of the system under diagnosis is necessarily simplified – based on three-layer Bayesian networks with canonical interactions among the network variables – and we are able to reduce greatly the knowledge engineering effort that goes into model building.



A single stage, factored diagnostic model, built by user-interaction *during a diagnostic session*

- Its is hard for engineers to anticipate all conditions where a model will be used. Model formulation is best done when an actual problem arises, when knowledge about it is cognitively accessible.
- The problem that starts e.g. a diagnostic session, forms a "query."

  Query-based inference means the model is conditioned, all or in part, on the query that initiated the session.
- *Query-based Inference* can help diagnosis, for example by taking hints for formulating the model from the sequence of actions during a user session.
- There are other advantages to "lazy" model building: smaller models, more focused inference.

Learning from just a few cases

Each case specifies the result, as a ranking of faults, f_i by probability, of applying evidence j to the model M. A case is best thought of as a fragment of the correct model, expressing *epistemic* uncertainty rather than stochastic variation in the appearance of faults.

Unlike modifications made by elicitation of causal links, there is no direct way to modify the network by inspection of a set of cases. The *case consistency learning* problem is to determine if the case set is consistent, and if so, to learn, or modify an existing model to be consistent with the set.

Definition: Case Consistency

A model M is consistent with case j to level k if the list of ordered fault marginals given the evidence $e^{(j)}$ agrees with the case.

$$P(f_1 \mid e^{(j)} M) \geq \mathbf{L} \geq P(f_k \mid e^{(j)} M)$$

Then the learning problem is to find a model M^* consistent with all j cases.

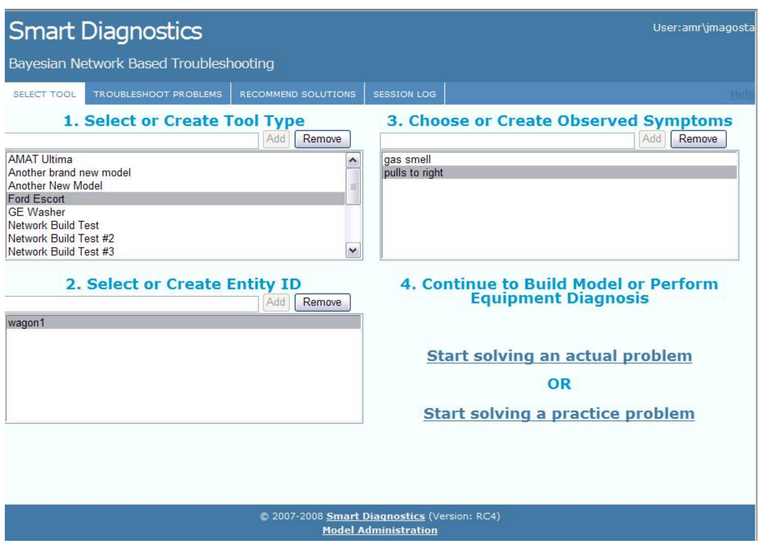
Two web-based applications

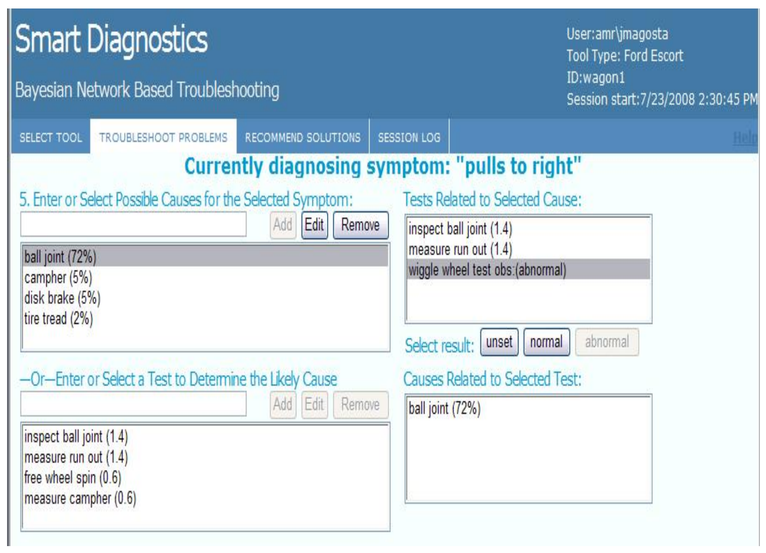
Marilyn is a public website that collects observables, context, and the resulting diagnosis in dialog with the user. It suggests a set of possible diagnoses from among those entered.

Smart Diagnostics is an interactive troubleshooting tool, built in cooperation with our client, by putting a Bayes network behind the workflow-based design they suggested.

Both consist of a web application, the SMILE engine and a database that persists the model.

- *Smart Diagnostics* is a test-bed for exploring Query-Based Inference.
- It hides the network structure and probabilities from the user.
- It logs all actions and outcomes of user sessions.
- The user-generated networks form a prior for learning from the collected session logs.





Query-Based Diagnostics

John Mark Agosta*, Thomas H. Gardos *, Marek J. Druzdzel §

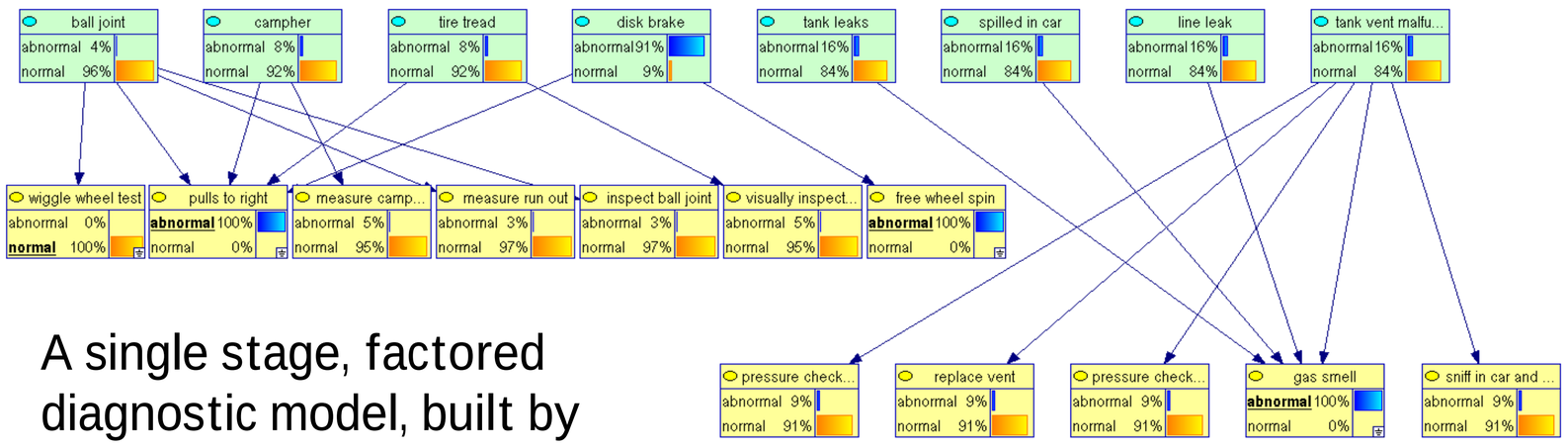
* Intel Corporation, Santa Clara, CA, USA

§ Faculty of Computer Science, Bialystok Technical University, Poland & School of Information Sciences, Univesity of Pittsburgh, Pennsylvania

Modifying the model while problem-solving

We model diagnostic problems based on inferring causes by passive observation of a diagnostician's work-flow. Then by recording their findings and final diagnosis, the model can be modified directly, or improved by learning from cases so acquired.

The hidden probabilistic model of the system under diagnosis is necessarily simplified – based on three-layer Bayesian networks with canonical interactions among the network variables – and we are able to reduce greatly the knowledge engineering effort that goes into model building.



A single stage, factored diagnostic model, built by user-interaction *during a diagnostic session*

- Its is hard for engineers to anticipate all conditions where a model will be used. Model formulation is best done when an actual problem arises, when knowledge about it is cognitively accessible.
- The problem that starts e.g. a diagnostic session, forms a "query."

  Query-based inference means the model is conditioned, all or in part, on the query that initiated the session.
- *Query-based Inference* can help diagnosis, for example by taking hints for formulating the model from the sequence of actions during a user session.
- There are other advantages to "lazy" model building: smaller models, more focused inference.

Learning from just a few cases

Each case specifies the result, as a ranking of faults, f_i by probability, of applying evidence j to the model M. A case is best thought of as a fragment of the correct model, expressing *epistemic* uncertainty rather than stochastic variation in the appearance of faults.

Unlike modifications made by elicitation of causal links, there is no direct way to modify the network by inspection of a set of cases. The *case consistency learning* problem is to determine if the case set is consistent, and if so, to learn, or modify an existing model to be consistent with the set.

Definition: Case Consistency

A model M is consistent with case j to level k if the list of ordered fault marginals given the evidence $e^{(j)}$ agrees with the case.

$$P(f_1 \mid e^{(j)}\, M) \geq \mathbf{L} \geq P(f_k \mid e^{(j)}\, M)$$

Then the learning problem is to find a model M^* consistent with all j cases.

Two web-based applications

Marilyn is a public website that collects observables, context, and the resulting diagnosis in dialog with the user. It suggests a set of possible diagnoses from among those entered.

Smart Diagnostics is an interactive troubleshooting tool, built in cooperation with our client, by putting a Bayes network behind the workflow-based design they suggested.

Both consist of a web application, the SMILE engine and a database that persists the model.

- *Smart Diagnostics* is a test-bed for exploring Query-Based Inference.
- It hides the network structure and probabilities from the user.
- It logs all actions and outcomes of user sessions.
- The user-generated networks form a prior for learning from the collected session logs.