# Rapid Diagnostics based on Data

**Pramod P. Khargonekar**

**Department of Electrical and Computer Engineering**
**University of Florida, Gainesville, FL**

**and**

**Ubiquiti Inc.**
**Ann Arbor, MI**



Science, 2/11

**Data → Information →** ***Knowledge***

# Diagnostics are Important

- A major issue:
  - Obviously drive “Fixed Right the First Time”
  - Reduce issue of “No Trouble Found”
  - Significant fraction repair & maintenance costs
- The market includes:
  - OEMs, Fleets, and indirectly also Suppliers
  - Independent repairshops
- An opportunity waiting to happen:
  - Techniques are available
  - Data is available
  - Computing power & connectivity are available

# Rapid Diagnostics

- **Goal: Rapidly deliver diagnostics**

  ***Diagnoses***, ***Best fixes***, ***Differentiators***, ***Repair procedures***

- **Repairs & Maintenance data** = *Knowledge Base*

  **All** technicians' experiences, **All** regions, **All** points in time

- **Natural Language Query**

  Commonplace description of symptoms and vehicle type

- **Automatically adapt to New Conditions** with new data

  ***Data itself*** is the knowledge base

- **Simple interface** – anytime & anywhere

# Rapid Diagnostics

## Demonstration

# Past Technical Barriers

- **Repair data often has the usable info in unstructured text.**

  Extracting useful info automatically needed the more recent advances in ML and StatisticalNLP.

- **Significant difficulties in modeling systems**

  Approaches using math models, expert systems, simulation, “case-based reasoning” have limits

- **Costs of data storage, computing power, connectivity**

  Capturing, storing, accessing, computing, transmitting info were all major challenges in the past.  Not so anymore.

# Key Elements of Our Approach

- **Use an Ontology.**

  This allows extracting text information**,** and helps to capture domain information organization.

- **Use a Repository of Repairs Data.**

  This may be warranty, repairs, or maintenance datasets.

- **For a New Repair case, find a set of "similar" data points.**

  Use the similar dataset to figure out the needed diagnostics.

- **Instance-Based Learning** (IBL)**.**

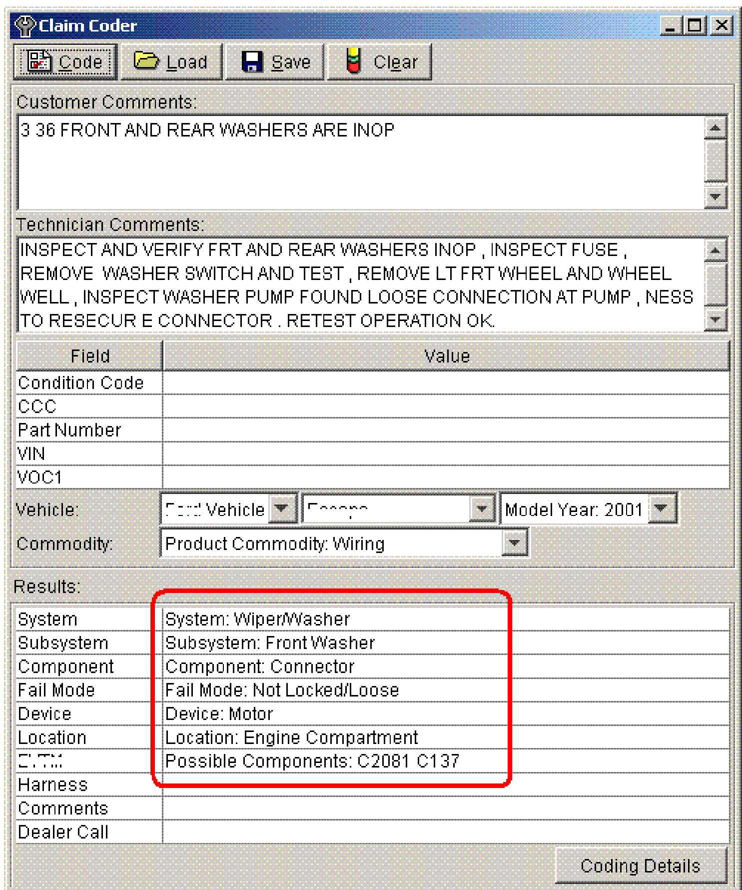
- **Simple for end-user.**

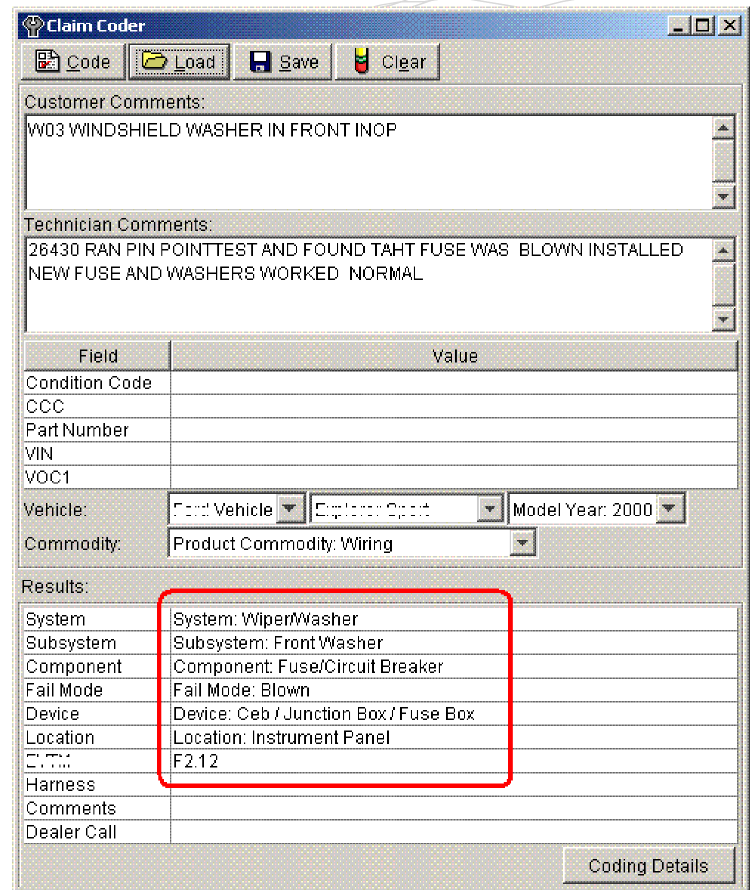
  Similar to Google – uses Search, and usable anywhere & anytime.  "Smarts" in backend server.

# Ubiquiti: Use of "Ontology Maps"

Info (automatically) extracted from Record – with ref to Ontology

Two Examples:

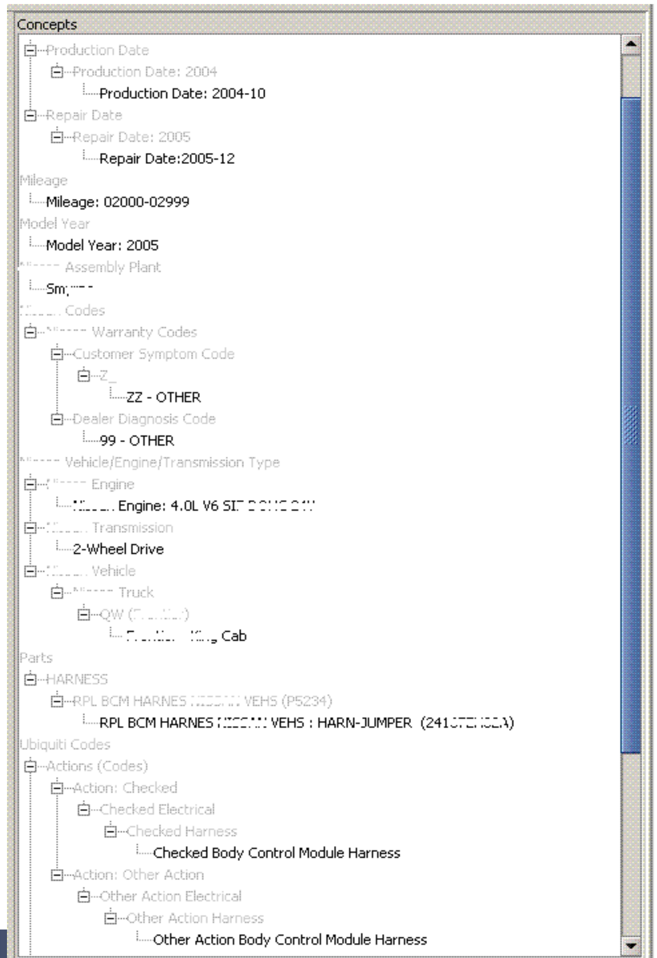


# Similarity-based Search: Ontology-based approach

Use concepts (instead of words), leverage the ontology.

*Dispense with traditional Euclidean geometry*. But more complex and compute-intensive.




# Machine Learning Basics

*Emulate human intelligence*

A long term goal in computer science and engineering

- Unsupervised Learning
- Supervised Learning … (Ubiquiti approach)
- Examples of Success:
  - Automatic Zip code recognition
  - IBM Watson … "Jeopardy" champ
  - Epilepsy – seizure detection

# ROADs Technical Issues

- **Finding "Similar" Data is Critical.**

  Consider a new repair instance - 10K miles and with "Check Engine" light on.

  Is this case *more similar* to a repair with 10K miles with symptoms of engine power loss, or

  to a repair with 20K miles and "Check Engine" light on?

- **Finding <u>too few</u> or <u>too many</u> records.**

  Too few records may lead to "over-fitting" or indicate not having enough information. Too many records would "dilute" the diagnostics, and only give the general distribution of results.

- **Figures of Merit.**

  How should system recognize good result?

- **How Much Data is Needed?**

# Search Execution Strategies

- **Top down or Bottom up approaches.**

  Start with tight similarity and weaken systematically. Or, start with weak constraints, and tighten systematically.

  How to weaken/tighten? When to stop?

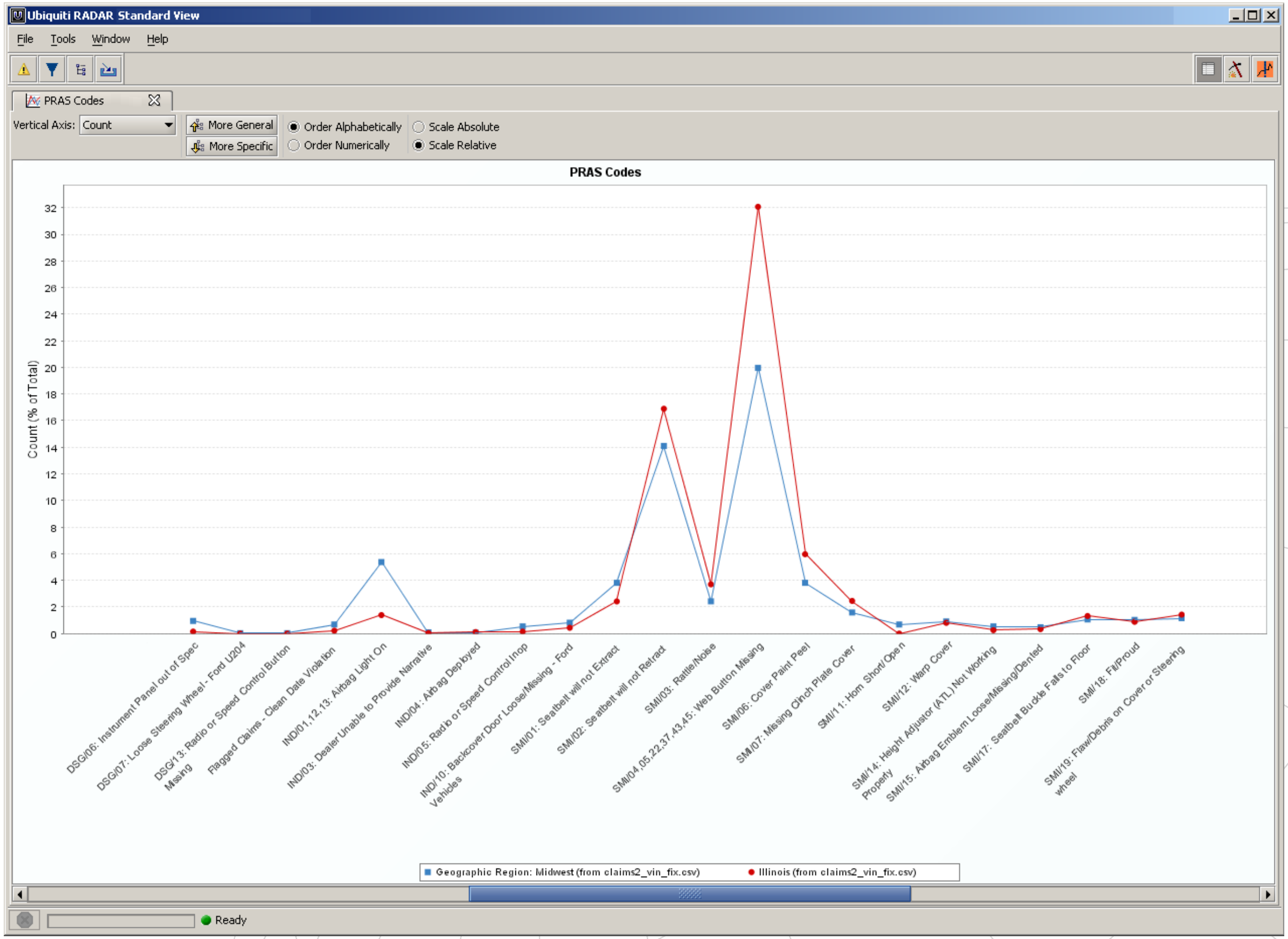
- Ubiquiti ROADs Approach: **Adjust parameters empirically.** Use available data to run many tests and note results – allows gathering info on similarity and stopping criteria

***Find the right similarity neighborhood for the query record***

# Examples of Right Similarity Neighborhood

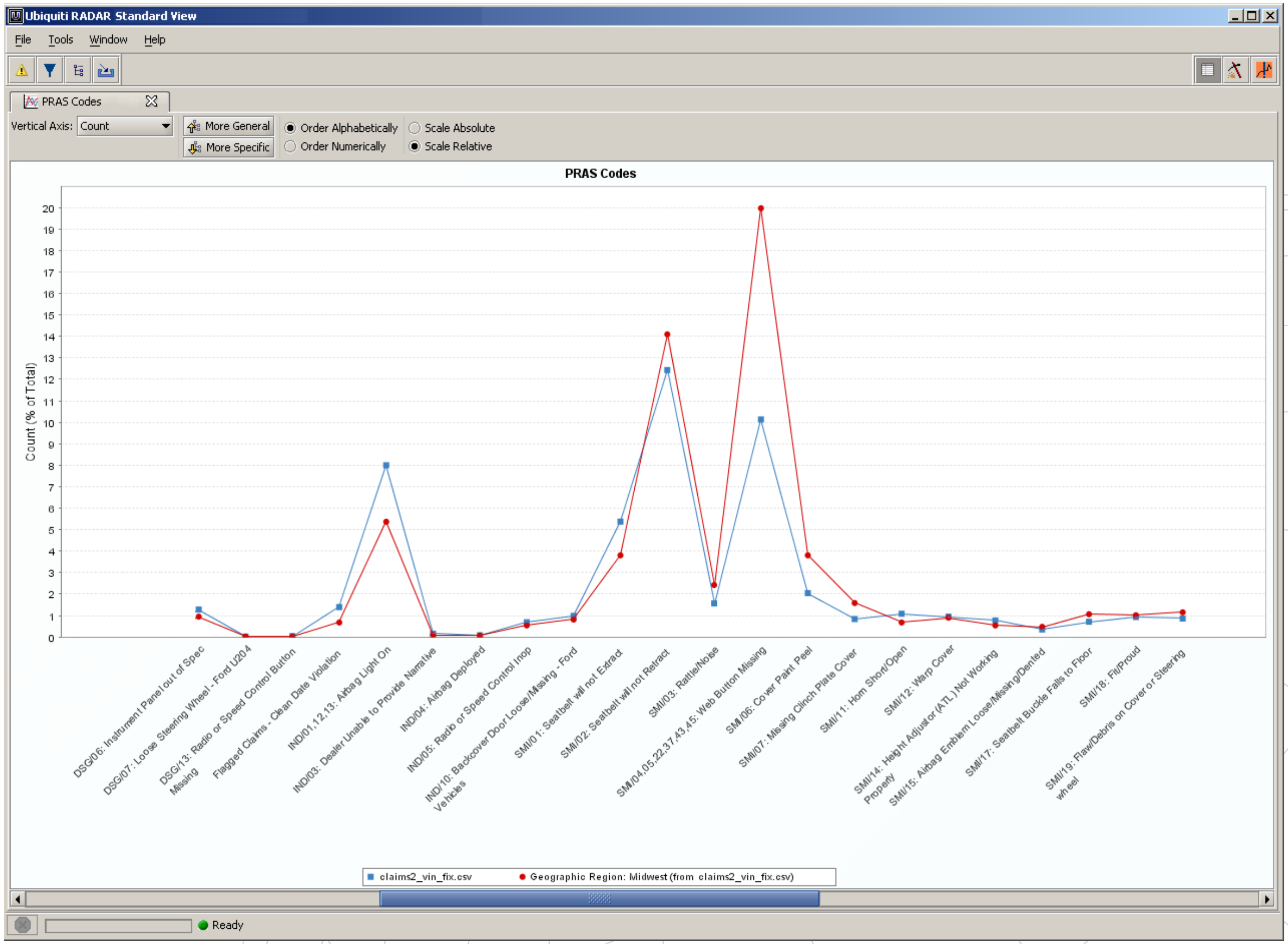
- Minimum or maximum number of result-set records
- Minimum or Maximum Distance from the Repair Instance
- Stop enlarging (or shrinking) the neighborhood at "Phase Change"
- And others

# Example:  Failures IL vs. Midwest




**Difference measured by K-L or LP norms**

# Example: Failures Midwest vs. All



# Figures of Merit

- How should we measure the performance of the diagnostic system? What is “good”? What is “better”?
- Ideal Truth: Expert?
- Data contains past expert results …
- Challenging question

# Other ROADs Parameters to Consider

**Several Parameters for Empirical Study**

- Using weighted nearest-neighbors.
- Using “reputation system” weighting.
- Changing Training (and Scoring) sets.
- Changing Training Set size.

*Meta-level optimization to adjust approach parameters*

# Conclusions

- Rapid diagnostics – a great target for the "data rich future"
- Taken a practical approach
- Leverage:
  - Domain specific knowledge via ontologies
  - Data is the model
  - Machine learning
- Natural adoption path from current Ubiquiti offerings!

***Decision Support with Text***

Web: www.ubiquiti.com
Email: info@ubiquiti.com
Phone: (734) 997-8800
Fax: (413) 622-3123

Ubiquiti Inc.
303 Detroit Street
Suite 202
Ann Arbor, MI 48104