Dempster-Shafer Theory:
Fault Diagnosis Application

Mohammad Shahab, (227598)
Systems Engineering dept.
Feb 2009
Outline

• Classification, Data (Sensor) Fusion
• What is Dempster-Shafer?
  – D-S vs. Bayes
  – D-S Rules
• Fault Diagnosis?
• Literature Review
• Case Study: Engine Faults
  – Overview
  – Solution
• Comments & Conclusion
Data (Sensor) Fusion...

• “... is the combining of sensory data (or data derived from sensory data) from disparate sources such that the resulting information is 'better' than would be possible when these sources were used individually.”

• To classify:
  - Medicine: is it Hepatitis? Cancer?
  - Surveillance: is it the enemy? Friend?
  - Manufacturing: is the machine OK? Not OK?
What is Dempster-Shafer?

  - New concept of ‘probability’... Here it is called ‘Belief’
- Ability to model narrowing hypothesis set as evidence accumulates
- Arthur Dempster: Dempster’s Combination Rule
  - Vs. Bayes’ combination rule

- Robot location (example):
  - Possibilities = { yes, no, no clue}
D-S vs. Bayes

• Normal Probability:
  – Hypotheses = \{ @ cell ‘x’ \}
  – 1 = yes, 0 = no
  – So 1st cell, \( p(1) = 0.4 \), \( p(1') = 0.6 \)

• To combine different ‘masses’:
  – Bayes Rule: say ‘y’ is different sensors

\[
P(x|y) = \frac{P(y|x) \cdot P(x)}{P(y)} = \text{Likelihood} \cdot \text{Prior} \]

\[
P(x|y, z) = \frac{P(y|x, z) \cdot P(x|z)}{P(y|z)}
\]

  – Total Probability:

\[
P(x) = \sum_{\forall z} P(x|z)P(z)
\]
D-S vs. Bayes

• Dempster-Shafer:

• Probability:
D-S vs. Bayes

• Dempster-Shafer:

**Shafer**: “Bayesian theory cannot distinguish between lack of belief and disbelief. It does not allow one to withhold belief from a proposition without according that belief to the negation of the proposition.”
D-S Rules

- **Frame of Discernment** $\Theta=\{\theta_1, \theta_2, \ldots\}$
  - Features, Classes, diseases, faults,
  - Medicine: $\Theta=\{\text{disease } 1, \text{disease } 2, \text{disease } 3, \text{disease } 4\}$

- **Power Set** $2^\Theta = \text{subsets} = \{\text{disease } 1, \text{disease } 2, \text{disease } 3, \text{disease } 4, \{\text{disease } 1, \text{disease } 2\}, \{\text{disease } 1, \text{disease } 3\}, \ldots \{\text{disease } 1, \text{disease } 2, \text{disease } 3, \text{disease } 4\}\}$
  - *I don’t know* $= \{\text{disease } 1, \text{disease } 2, \text{disease } 3, \text{disease } 4\}$

- **Basic Belief Assignment (BBA):** $m(.)$, ‘mass’
  - $m(A)$ represents the belief assigned to an individual element $A$
  - $m(\Theta) = m(\{\text{disease } 1, \text{disease } 2, \text{disease } 3, \text{disease } 4\}) \neq 1$
  - $m(\text{disease } 1) + m(\text{disease } 1') < 1$

- **Dempster’s Combination Rule (e.g. Fusion for sensor (evidence) 1 and 2)**

  $$m^{1,2}(C) = \frac{\sum_{A \cap B = C} m^1(A) m^2(B)}{\sum_{A \cap B \neq \emptyset} m^1(A) m^2(B)} = \frac{\sum_{A \cap B = C} m^1(A) m^2(B)}{1 - \sum_{A \cap B = \emptyset} m^1(A) m^2(B)}$$

  - Look if $C=\{\text{disease } 1\}$
Fault Diagnosis

- In Industry,
- Condition Monitoring, Health Monitoring, Fault Detection, etc.

- “... monitoring a parameter of condition in machinery, such that a significant change is indicative of a failure “

- Mostly done on Rotating Machines
- Vibration Analysis, ‘features’ are extracted via multiple sensors

- So, the questions are:
  1. Which fault is it?
  2. Decision Making: to what degree we are certain about decision?
Literature Review

• Dempster-Shafer Theory:

   - Reviews the D-S Theory
   - with proposing new combination technique with learning & adaptation

   - Review data fusion for Bayes’ Rule (e.g. Kalman) and Dempster-Shafer Theory

   - Comparison, Bayes & D-S:
   - Both have same results, however different in representation
   - Bayes need prior ‘probabilities’, unlike D-S
Literature Review

• Dempster-Shafer Theory:

   – Book chapter, early presentations of the theory
   – Takes the standard problem of Medical Diagnosis

   – Review the D-S Theory with different ‘algorithmic’ study

   – Software sensors
Literature Review

• D-S in Fault Diagnosis:

   – One of two papers, study application of D-S decision making from raw sensor data
   – Proposes new improved D-S

   – Application to diesel engine cooling system

   – Induction motors, current & vibration (electrical & mechanical)
   – Feature in time domain & frequency domain

   – Develops a good approach of transforming sensory readings into information to be used in D-S
   – Vibration, acoustic, pressure, tempareture
Machine Fault Diagnosis

• Problem Overview
  • Let the machine have states or sensor outputs or features of
    \[ X = [x_1 \ x_2 \ \cdots \ x_n] \]
  • Let us have a ‘table’ of faults:
    \[
    \begin{bmatrix}
    F_1 \\
    F_2 \\
    \vdots \\
    F_K \\
    \end{bmatrix}
    = \begin{bmatrix}
    x_{11}^f & x_{21}^f & \cdots & x_{n1}^f \\
    x_{12}^f & \cdots \\
    \vdots \\
    x_{1k}^f & \cdots & \cdots & x_{nk}^f \\
    \end{bmatrix}
    \]
  • So, one can have a measure for ‘distance’ from readings to the faults set
    \[
    d_{ik} = \sqrt{\sum_{n} (x_n - x_{nk}^f)^2}
    \]
  • So the lower the distance, the most likely that fault ‘k’ is occurring
  • So a measure is to have
    \[
    p_{ik} = \frac{1}{d_{ik}}
    \]
  • At the end we have for each sensor, \( K \) measures
Applying Dempster-Shafer

- Let us assume we have 2 faults, and 2 sensors
- So, \( \Theta = \{ F_0, F_1, F_2 \} \), ‘0’ means no fault
- So, the power set would be: \( \{ F_0, F_1, F_2, \{ F_1, F_2 \} \} \) (we can not have fault and no fault, but we can have no clue)

- so we have masses \( m_1, m_2 \) for each sensor that cover the power set
- Example:
  - Let have the conflicting situation:

    - But after using Dempster Combination rule

\[
\begin{array}{|c|c|c|}
\hline
\text{Fault} & m_1 & m_2 \\
\hline
F_0 & 0.09 & 0.02 \\
F_1 & 0.2 & 0.85 \\
F_2 & 0.7 & 0.1 \\
F_{12} & 0.01 & 0.03 \\
\hline
\end{array}
\]

\[
m_{F_u}(F_0) = \frac{m_1(F_0) \times m_2(F_0)}{\mu}
\]

\[
m_{F_u}(F_1) = \frac{m_1(F_1) \times m_2(F_1) + m_1(F_1) \times m_2(\{F_1, F_1\}) + m_2(F_1) \times m_1(\{F_1, F_1\})}{\mu}
\]

\[
m_{F_u}(F_0) = 6.4 \times 10^{-3}
\]

\[
m_{F_u}(F_1) = 0.66
\]

\[
m_{F_u}(F_2) = 0.33
\]

\[
m_{F_u}(\{F_1, F_2\}) = 1.07 \times 10^{-3}
\]
Simulation

• For some feature ‘x’ [0, 2]:
  • No fault: 0 → 0.5 | Fault 1: 0.5 → 1 | Fault 2: 1 → 1.5 | Don’t know if: 1.5 → 2
  • If we have distance for sensor 1:
    \[ d_i = \left[ \frac{1}{|0.25 - s_1|}, \frac{1}{|0.75 - s_1|}, \frac{1}{|1.25 - s_1|}, \frac{1}{|1.75 - s_1|} \right] \]
    
    • Then we can have a normalized
    \[ P_{ik} \]

• In my simulation:
  – I forced “Fault 1”
  – So for two sensors: I added noise to the feature reading
Simulation

- Let actual $x=0.7$ (fault 1)
Simulation

- Sensor 1 decision

- Sensor 2 Decision
Simulation

- Fused Sensor decision
Comments & Conclusion

• Concern is about the computational complexity: imagine if 3, 4, ... features, if more sensors

\[ m_{F_1}(F_1) = \frac{m_1(F_1) \times m_2(F_1) + m_1(F_1) \times m_2(\{F_1, F_1\}) + m_2(F_1) \times m_1(\{F_1, F_1\})}{\mu} \]

• “don’t know” information reflects actual knowledge (or say not knowing!!)

• **Future Directions:**
  – Study Effect of different noise
  – to develop better algorithms

• Conclusion:

• One new view about “belief”

• In my opinion, it should replace probability (conceptual, not yet mathematically)
Q & A

• *Thanks*....