Semantic Web Knowledge Fusion

Jennifer Sleeman
University of Maryland, Baltimore County

Motivation

Data facts associated with knowledge model instances: How does one decide which facts should be associated with which entities?

Data from multiple sources: How to combine and update data over time? Resolve conflicts?

In other domains: "Information Integration" and "Data Fusion"

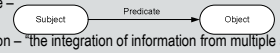
In Semantic web domain: "Knowledge Fusion"
Uncertainty not consistently represented (Bayesian, Fuzzy Logic, Dempster-Shafer and others)

Data Fusion + Uncertainty + Semantic Web

Dempster-Shafer shows promise

Definitions

RDF Triple –



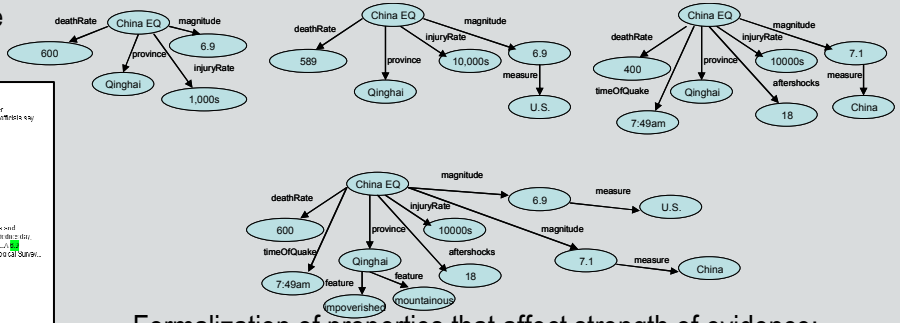
Data Fusion – "the integration of information from multiple sources to produce specific and comprehensive unified data about an entity" [4]

JDL Revised Data Fusion Model [4]

Uncertainty –"a variety of forms of incomplete knowledge, including incompleteness, vagueness, ambiguity, and others." [3]

Methodology

Conduct a study using Dempster-Shafer theory of evidence. Assign a belief and plausibility to each piece of evidence based on a formalization of properties. Use the combining evidence function to combine evidence. Build an ontological structure that can represent the values calculated in DS so evidence can be updated/changed as the knowledge model evolves.



Formalization of properties that affect strength of evidence:

Metadata:

Source – Where is the data retrieved from? Is the source reliable?
Properties such as timestamp can influence confidence

Rule Based:

If numeric value increasing and fact happens later in time likely more accurate

Dempster-Shafer theory of evidence:

Representation of ignorance – general argument is one cannot represent ignorance using a probabilistic method such as Bayesian

Example: Coin toss

With DS there is a concept of a belief: $Bel(A) = \sum_{B|B \subseteq A} m(B)$

And a concept of a plausibility: $Pl(A) = \sum_{B|B \cap A \neq \emptyset} m(B)$

This is based on the universal set and mass functions:

$m: \mathbb{P}(X) \rightarrow [0,1] \qquad m(\emptyset) = 0 \qquad \sum_{A \in \mathbb{P}(X)} m(A) = 1$

Combining Evidence:

$$m_{1,2}(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K} \quad \text{when } A \neq \emptyset \qquad m_{1,2}(\emptyset) = 0 \quad \text{where } K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C)$$

Evaluation

In the area of data fusion there were multiple papers written regarding using Dempster-Shafer and Bayesian. There is also work that uses Dempster-Shafer to combine multi-classifier results. Critics show a marginal increase in accuracy using Dempster-Shafer. Supporters offer experiments that show Dempster-Shafer's advantage.

A criticism often mentioned is related to combining evidence. The denominator of the combining evidence function 1-K is a normalizer and the effect of this is completely ignoring conflict which can produce unexpected results. Dempster's paper states that conflicts should be ignored and hence the normalization but this can produce unexpected results. Recent work shows improvements to Dempster-Shafer which attempt to resolve this issue [2].

The amount of conflict between beliefs can be measured:

$$Con(Bel_1, Bel_2) = \log(\frac{1}{1-\kappa}); \text{ where } \kappa = \sum_{A \cap B_i = \emptyset} m_1(A_i) m_2(B_j)$$

Related to representing this within an ontological structure, there were two papers where the author presented approaches to do this but in one paper the aspect of change over time was completely addressed and the second paper was attempting to solve a different problem.

Future Work

Based on the known flaw with the Dempster-Shafer normalization, apply one of the advancements that correct this problem.
Yager's Modified Dempster Rule – quasi-associative operator

Build the ontological structure and a small experiment which can be used to compare Dempster-Shafer and Bayesian.

References

[1] Resource Description Framework (RDF): Concepts and Abstract Syntax, http://www.w3.org/TR/rdf-concepts/
[2] Sentz, K. and S. Ferson (2002). Combination of Evidence in Dempster-Shafer Theory, SAND2002-0835 Technical Report. Sandia National Laboratories, Albuquerque, NM
[3] "Uncertainty in Ontologies: Dempster-Shafer Theory for Data Fusion Applications", A. Bellenger1 and S. Gatepaille, Defence and Security Information Processing, Control and Cognition department, France
[4] "An Introduction to Multisensor Data Fusion", D. L. Hall and J. Llinas, editors. Handbook of Multisensor Data Fusion. CRC Press, 2001.
[5] "An Introduction to Bayesian and Dempster-Shafer Data Fusion", D. Koks and S. Challa, DSTO Systems Sciences Laboratory, November 2005