Wigmore: A constraint-based language for reasoning about evidence and uncertainty

David Burke, Galois Inc.
HCSS 2016

galois

Overview

|galois|

1. Context
2. Dempster-Shafer (D-S) Theory and its critics
3. Designing an evidential assertion language
4. Illustrative Examples
5. Future enhancements

Context for this work

|galois|

- Challenge: Designing a decision support system for analyzing cyber threats.
- Salient Features:
  1. Analysts are operating in a world of "*irreducible uncertainty*": decisions will have to be made before uncertainty can be resolved.
  2. Analysts are working with evidence that is missing, fragmentary, contradictory, dynamic…
  3. Multiple analysts and workflows.

The Concept of Uncertainty

|galois|

- Plenty of terms with varied meanings: chance, risk, “Knightian” uncertainty, ambiguity, probabilities…
- This key distinction is the core of our approach:
  - ***Aleatory*** uncertainty: uncertainty due to not knowing the chances involved; uncertainty due to noise; stochastic.
  - ***Epistemic*** uncertainty: uncertainty due to ignorance; uncertainty due to lack of knowledge.
- In the cyber domain, so-called ‘zero-day’ attacks are an exemplar of epistemic uncertainty.

Dempster Shafer Theory & Evidence

|galois|

Dempster-Shafer (D-S) theory (and the field of Belief Functions) is intriguing because the fundamental semantics of evidence combination looks very appropriate to the problem space.

Suppose we have a very simple universe, $\Omega = \{A, \sim A\}$:

Probabilistic semantics: If $p(A) = x$, then $p(\sim A) = 1 - x$

Evidential semantics: If $m(A) = x$, then $m(\Omega) = 1 - x$

Two useful (complementary) views:

- *Updating* beliefs if we think of beliefs as generalized probabilities.
- *Combining* beliefs if we think of beliefs as evidence.

Toy Dempster-Shafer Example

|galois|

Is the berry poisonous?

There are two possible worlds: a berry you want to eat is either safe (S) or poisonous (P). Hence Ω = {P,S}.

Suppose we have two pieces of evidence to consider. Let:

bba_1 = ({S} = .7, {P} = .2, {S,P} = .1)

bba_2 = ({S} = .6, {P,S} = .4)

terminology:

bba stands for “basic belief assignment”

Ω stands for “frame of discernment”, or f.o.d.

Toy Dempster-Shafer Example: Calculations

- Combine evidence through set intersection and multiplication:

	{S} = .6	{S,P} = .4
{S} = .7	{S} = .42	{S } = .28
{P} = .2	{ } = .12	{P} = .08
{S,P} = .1	{S} = .06	{S,P} = .04

Collect terms:

{S} = .42 + .28 + .06 = .76

{P} = .08

{} = .12

{S,P} = .04

Two variants for reduction to decision

{S} = .42 + .28 + .06 = .76
{P} = .08
{} = .12
{S,P} = .04

"Pignistic" transform

1. Throw out empty set & normalize:

{S} = .76 + .02 = .78

{P} = .08 + .02 = .10

p(S) = .78 / .88 = .89

p(P) = .10 / 88 = .11

2. Keep empty set as a measure of inconsistency:

p(S) = .78

p(P) = .10

p({}) = .12

Infamous D-S Example (Zadeh)

- Two doctors making a diagnosis:
  - Doctor A: Brain Tumor = .99, Meningitis = .01
  - Doctor B: Flu = .99, Meningitis = .01

Doctor A

Doctor B		{F} = .99	{M} = .01
	{T} = .99	{ } = .98	{ } = .01
	{M} = .01	{ } = .01	{M} = .0001

- Meningitis is the only non-empty result; after normalization, confidence becomes 1.0

Zadeh Example: Two Interpretations

|galois|

	{F} = .99	{M} = .01
{B} = .99	{ } = .98	{ } = .01
{M} = .01	{ } = .01	{M} = .0001

#1: Ignore empty sets & normalize:

{M} = .0001

p(M) = .0001 / .0001 = 1.0

#2: Use empty set as a measure of inconsistency:

p(M) = .001

p({}) = .999

Our takeaway: Not to discard D-S approaches, but rather to ask:
Why not make explicit the concept of evidential consistency?

Design Principles: Perspectives and Constraints

|galois|

Particularly in domains containing epistemic uncertainty, there is not necessarily a definitive or objective means of combining all of the potential evidence.

Therefore, how evidence is combined reflects a particular analyst's point of view – a ***perspective***.

We also don't want to force analysts to blindly combine all their available evidence; instead, give them the ability to flexibly express ***constraints*** about evidence combination.

We borrowed from, and extended existing D-S theory to design a language to incorporate these principles.

Evidence Trees

|galois|

An evidence tree is a recursive datatype:

EV = Simple-Evidence bba (base case)

| ∧ *Ev1 Ev2* (“and” constraint)

| ∨ *Ev1 Ev2* (“or” constraint)

| ¬ *Ev* (“not” constraint; uses set complements)

| Discount *Ev* factor (discount evidence by a given factor)

| Named-Evidence name (lookup of existing assertion tree)

Evaluating Evidence Trees

|galois|

- We need rules and the associated machinery to handle details of evidence tree reduction:
  1. *Frame of discernment:* how two different f.o.d.'s to combine. (say, {a,b,c} and {c,e,f}
  2. *World types:* defining how to operate under a 'closed' universe (empty sets normalized away) or an 'open' one (allowing for the possibility that our enumeration of possible worlds is incomplete).
  3. *Combination mode:* In combining two pieces of evidence, one from a closed universe, and one from an open universe, either the open universe dominate ('inclusive' mode) or does the closed one dominate ('conservative' mode).

Evidence Forests and Their Reduction

|galois|

- Analysts can make multiple assertions, each assertion represented by an evidence tree.
- The set of analyst assertions is an evidence forest.
- An analyst defines a weight vector to the evidence forest and a reduction mode; reducing an forest is simply a linear weighting of the reduced evidence trees

Illustrative Example 1

|galois|

Start with the two doctors from the Zadeh example:

bba1 =([(t,.99),(m,.01)], Ω = {t,f,m}, closed)

bba2 =([(f,.99),(m,.01)], Ω = {t,f,m}, closed)

The first analyst makes two assertions

1. The diagnosis should be consistent with the two doctors:

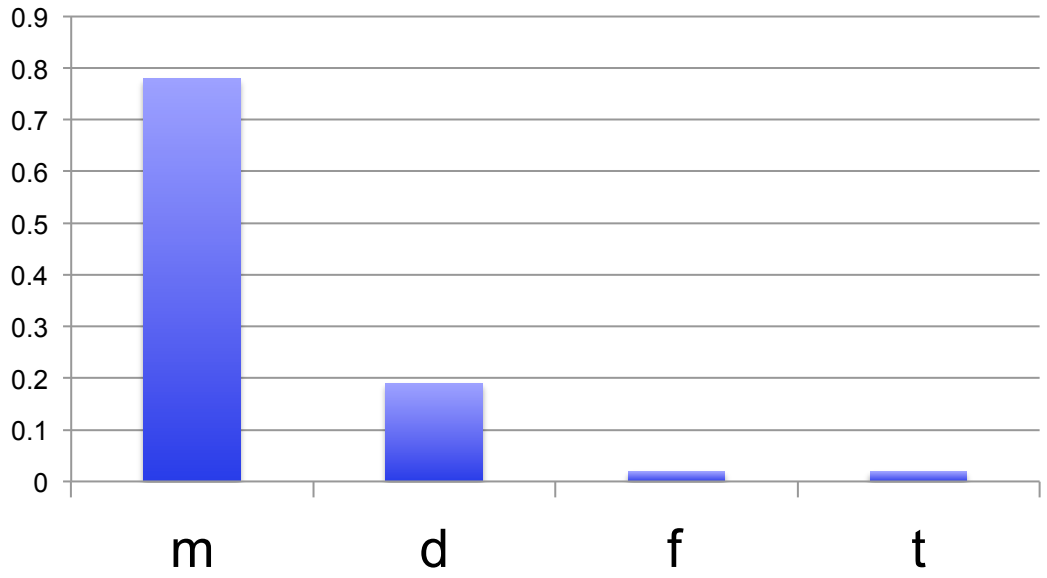
ev3 = (bba1 ∧ bba2)

2. There is evidence that dengue fever (which the doctors didn't run tests for) is the true culprit:

ev4 = ([(d, .7), (Ω, .3)], Ω = {t,f,m,d}, closed)

Illustrative Example 1 (cont.)

Analyst 1 evaluates the evidence in conservative mode with a weight vector of [95,30]:



Meningitis is the clear winner, with dengue fever getting some support as well.

Illustrative Example 2

|galois|

Again, start with the two doctors from the Zadeh example. A second analyst makes four assertions:

Doctor 1 is a fool; believe the opposite!

ev5 = ¬ bba1

Discount Doctor 2's evidence by 25%:

ev6 = $\mathcal{D}$(bba2, .25)

There is evidence that the disease is emphysema:

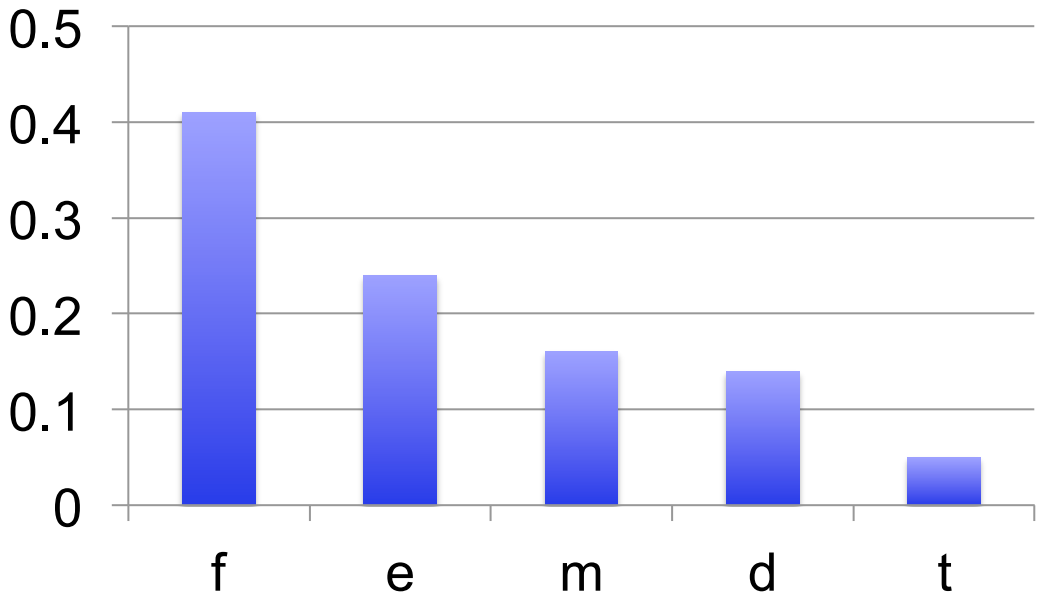
ev7 = ([(e, .6), ({d,e}, .3), (Ω, .1)], Ω = {m,t,f,d,e}, open)

The truth is consistent with assertion 4 (from Analyst 1) or my assertion 7:

ev8 = (ev4 ∨ ev7)

Illustrative Example 2 (cont.)

Analyst 2 uses the inclusive evaluation mode, with a weight vector of [.55,.80,.40,.70]



The aggregated evidence is more evenly spread out than in the previous example, but flu and emphysema are the most likely candidates.

Summary thus far...

- Designed extensions of Dempster-Shafer Theory for more expressive evidence modeling by multiple analysts. Since assertions are explicit, they can be shared and interrogated.
- "*All models are wrong, but some are useful*"
- Initial implementations in Python and Haskell.
- Although the impetus for this work was the cyber domain, our approach is not domain-specific, and could be applied in other contexts where multiple users are making and sharing evidential assessments.

Further Language Extensions

|galois|

1. *Correlation Operator*: A key requirement in belief function analysis is that we have evidential independence, but in practice evidence can be correlated (in the sense that higher confidence in one makes belief in the other more plausible).
2. *Decision Transforms*: We used a default transform rule (based on set cardinality) to go from set masses to singletons; other rules should be part of the language, and operator configurable.
3. *Fixed Point Analysis*: How to deal with the situation where there is mutual recursion between analyst references.
4. *Additional binary operators*: many more operators can (and have) been defined by the belief function community.

|galois|

Questions? Comments? Protests?

David Burke
davidb@galois.com
(503) 330-9512

|galois|

Backup Slides

Dempster-Shafer Example

- Suppose that you have a very simple universe: only two possible worlds, where, in this case, a particular berry you want to eat is either safe (S) or poisonous (P). Hence Ω = {P,S},
- Suppose we have two pieces of evidence to combine. Terminology again: a piece of evidence is often referred to in the literature as a "basic belief assignment (bba)".
- Let:
  - bba_1 = ({S} = .7, {P} = .3)
  - bba_2 = ({S} = .6, {P} = .4)

  (note in this example that nobody gave weights to Ω)

Simplest D-S Problem: Calculations

- The standard D-S operation is set intersection and multiplication. Set intersection can be thought of as consistency of evidence: Ev1 **and** Ev2

	{S} = .6	{P} = .4
{S} = .7	{S} = .42	{ } = .28
{P} = .3	{ } = .18	{P} = .12

- Notice that the off-diagonals are empty sets – we can't simultaneously be in world S and world P. So using set intersection is a way of looking for consistency between pieces of evidence.

D-S Example - two variants

- D-S Theory proposed two ways of handling empty sets:

1. Count up the total masses of the non-empty sets – use this as a normalization factor:
   - Normalization factor = .42 + .12 = .54
   - {S} = .42 / .54 = .78
   - {P} = .12 / .54 = .22
   - Not too surprisingly, this is called “normalized D-S”
2. Treat the total of the empty sets as a special element that has the semantics: “the true state of the world lies outside the current definition of Ω”
   - ({S} = .42, {P} = .12, {} = .42)
   - Called “non-normalized D-S”
   - You can imagine using the mass of {} as a threshold for “the problem as originally specified is, well, ill-specified”

A motivating example

- A robbery has been committed, and there are only three possible suspects.
  - Jack (J) a 70-year-old man
  - Tom (T), a 20-year-old man
  - Sally (S), a 65-year-old woman
- We have two witnesses; we represent their beliefs for each set in a "frame of discernment" (Ω) that contains all atomic possibilities:
  - Witness 1 believes that the robbery was probably committed by a male, and is reasonably confident that Tom is the most likely criminal.

$$bba_1 = (\{T\} = .6, \{J\} = .2, \{J,T\} = .1, \{J,T,S\} = \Omega = .1)$$

  - Witness 2 believes that the robbery was committed by an elderly person, and thinks that Sally is slightly more likely the guilty party.

$$bba_2 = (\{S\} = .4, \{J\} = .3, \{J,S\} = .2, \{J,T,S\} = \Omega = .1)$$

Robbery problem -- calculations

	{S} = .4	**{J} = .3**	**{J,S} = .2**	**Ω = .1**
{T} = .6	{} = .24	{} = .18	{} = .12	{T} = .06
{J} = .2	{} = .08	{J} = .06	{J} = .04	{J} = .02
{J,T} = .1	{} = .04	{J} = .03	{J} = .02	{J,T} = .01
Ω = .1	{S} = .04	{J} = .03	{J,S} = .02	{J,T,S} = .01

- We still get some empty sets (in green), giving us a normalization factor of 1 - .66 = .34
- Notice that we end up with a wide variety of other sets: {J}, {T}, {S}, {J,S}, {J,T}, {J,T,S}.
- If we want to know which suspect to arrest, we need to apply the pignistic transformation after normalization.

Robbery problem - Interpretation

- First, apply normalization (with a touch of rounding error)
  - {J} = .20 / .34 → {J} = .59
  - {S} = .04 / .34 → {S} = .12
  - {T} = .06 / .34 → {T} = .18
  - {J,S} = .02 / .34 → {J,S} = .06
  - {J,T} = .01 / .34 → {J,T} = .03
  - {J,T,S} = .01 / .34 → {J,T,S} = .03
- Finally, do the sums after the pignistic transform of each set above:
  - {J} = .59 + (.06 / 2) + (.03 / 2) + (.03 / 3) = .645 ← robber!
  - {S} = .12 + (.03 / 2) + (.03 / 3) = .145
  - {T} = .18 + (.03 / 2) + (.03 / 3) = .205