SIE 565
(ISE 405)

Reasoning with uncertainty

Mike Worboys
worboys@spatial.maine.edu
or

How we learned to accept human frailty and love vagueness, uncertainty and granularity
Course objectives

To understand what uncertainty is.
To understand that there are many different kinds and degrees of uncertainty.
To learn of the techniques used in uncertainty handling.
To match techniques with uncertainty types.
To relate uncertainty handling to data mining.
To relate uncertainty handling to the spatial domain.
Course structure

Preamble
1. Reasoning
2. Information and uncertainty
3. Quantitative approaches
4. Qualitative approaches
Useful resources

Witten and Frank, Data Mining, 2nd edition, Elsevier, 2005. WF


SEE SIE 565 WEB SITE
Section 1:

Formal aspects of reasoning
Facts and rules

**Facts**: Simple, particular statements.
- Paris is the capital of France
- It is snowing outside.

**Rules**: General principles, often in conditional form.
- All oak trees are broadleaved.
- No US state name begins with the letter “B”.

**Facts can be “interesting”**.
- It will snow outside tomorrow
- 'It is snowing outside’ is a sentence of four words.
- 'It is snowing outside’ is true.
- The sentence I am now speaking is true.
- The sentence that I am now speaking is false.
Knowledge and observations

Facts may often be verified or contradicted by a single observation, if it is the correct observation; this depends on its schema (see below).

Existential knowledge can be verified by a single appropriate observation, but would require many (sometimes an infinite number of) observations to falsify.

Universal knowledge can be contradicted by a single appropriate observation, but require many observations to verify.
Inferences

If it is snowing then John is skiing

It is snowing

John is skiing

All men are mortal

Socrates is a man

Socrates is mortal

Every day in the past the universe existed

The universe existed last Friday

Every day in the past the universe existed

The universe will exist next Friday
Proof and truth

Statement and the world
Inference and model/interpretation
Program and execution
Theory and practice
Proof and truth

With uncertainty we don’t have crisp notions of truth, and inference becomes more difficult.
Models: the link between proof and truth

A theory is a set of assumptions and inferences. An interpretation of a statement (proposition) allows an evaluation of the truth value of the statement. An interpretation of a theory is an interpretation of all the statements in the theory.

A model of a statement is an interpretation in which the statement is assigned truth value true. A model of a theory is an interpretation in which the all the statements are assigned truth value true.
Theory and model

ontable(A).
ontable(B).
on(C,B).
cube(B).
cube(A).
pyramid(C).

A blocks world

This theory has many models.
A model for these facts?

ontable(A).
ontable(B).
on(A,B).

Interpretations can be as we define them.
A model for this theory?

ontable(A).
ontable(B).
red(A).
∀X. ontable(X) → blue(X).
∀X. ¬(red(X) ∧ blue(X)).

This theory has no models.
It is inconsistent.
World - model - theory

∀XYR. on(X,R) ∧ on(Y,R) → connect(R,X,Y).

road(A34)
on(Newbury,A34)
connect(A34,Newbury, Redbridge)

proof

truth falsity

occurrents
continuants
model-world relationships

The model-world relationship can be:

precise
measure of how fully the model specifies the world

accurate
the model matches the world (truth in the model, reflects situations that obtain in the world)

vague
containing representations that cannot be determined as either true or false (borderline cases)

current
temporally up-to-date

...
Theory-interpretation relationships

A theory can be:

consistent
there are no contradictions derivable in the interpretation
there is a interpretation

sound (with respect to a interpretation )
formulas provable in the theory must be interpreted as true in the interpretation

complete (with respect to a interpretation )
formulas interpreted as true in the interpretation must be provable in the theory

...
Consistency

Inconsistent

\[
\text{ontable}(A). \\
\text{ontable}(B). \\
\text{red}(A). \\
\forall X. \text{ontable}(X) \rightarrow \text{blue}(X). \\
\forall X. \neg (\text{red}(X) \land \text{blue}(X)).
\]

Consistent

\[
\text{ontable}(A). \\
\text{ontable}(B). \\
\text{red}(A). \\
\forall X. \text{ontable}(X) \rightarrow \text{bright}(X). \\
\forall X. \text{bright}(X) \rightarrow (\text{yellow}(X) \lor \text{red}(X)).
\]

*find two different models*
**Completeness**

**Theories**

Not complete

- left(A, B).
- left(B, C).

Complete

- left(A, B).
- left(B, C).
- left(A, C).
- left(A, B).
- left(B, C).
- \( \forall xyz. \ (\text{left}(x,y) \land \text{left}(y,z)) \rightarrow \text{left}(x,z). \)

**Model**

3 objects, named A, B, C and relationship 'left-of' between them.
Reasoning example

Suppose a KB contains the following facts:

1. Aland, Bland, Cland, and Dland are countries.
2. Eye, Jay, Cay, and Ell are cities.
3. Exe and Wye are rivers.
4. City Eye belongs to Aland.
5. City Jay belongs to Bland.
6. City Cay belongs to Cland.
7. City Ell belongs to Dland.
8. Cities Eye, Ell, and Cay lie on the river Exe.

and rule:

10. Each river passes through all countries to which the cities that lie on it belong.
Consider the following model

Assume that this representation is accurate.

There are truths expressed by the map but not deducible from the KB.

   e.g. ALand and BLand share a common boundary.

But, restrict attention to facts about countries, cities, rivers, cities in countries, cities on rivers, rivers through countries.

The KB is sound (all the statements in the KB are true in the map).
The KB is not complete: e.g. "River Exe passes through countries Aland, Bland, Dland, Cland", is true but not deducible in the KB.
However, if we add a further city Em, and facts to the KB:

13. Em is a city.
14. Em belongs to the country Bland.
15. The river Exe passes through city Em.

Then the revised KB is sound and complete with respect to map, because we can now deduce:
River Exe passes through the country Bland.
Formal notation

The first ten statements in the KB can be formally written as:

1. \(\text{country}(\text{Aland}) \land \text{country}(\text{Bland}) \land \text{country}(\text{Cland}) \land \text{country}(\text{Dland})\)
2. \(\text{city}(\text{Eye}) \land \text{city}(\text{Jay}) \land \text{city}(\text{Cay}) \land \text{city}(\text{Ell})\)
3. \(\text{river}(\text{Exe}) \land \text{river}(\text{Wye})\)
4. \(\text{belongs to}(\text{Eye, Aland})\)
5. \(\text{belongs to}(\text{Jay, Bland})\)
6. \(\text{belongs to}(\text{Cay, Cland})\)
7. \(\text{belongs to}(\text{Ell, Dland})\)
8. \(\text{lies on}(\text{Eye, Exe}) \land \text{lies on}(\text{Ell, Exe}) \land \text{lies on}(\text{Cay, Exe})\)
9. \(\text{lies on}(\text{Jay, Wye})\)
10. \(\forall r. \forall x. \forall c. (\text{river}(r) \land \text{lies on}(c, r) \land \text{belongs to}(c, x)) \rightarrow \text{passes through}(r, x)\)
Section 2:

Information and uncertainty
Shannon's theory of information communication

- Source
- Receiver
- Channel
- Noise
- Equivocation

Information quantity measured by entropy
Communications system

information source produces a message consisting of an arrangement of symbols.

transmitter operates on message to produce a suitable signal to transmit.

channel the medium used to transmit the signal from transmitter to receiver.

receiver reconstructs the message from the signal.

destination for whom the message is intended.
Entropy as a measure of uncertainty

Entropy has been proposed by Shannon (1948) as a suitable measure of uncertainty of information in the context of communications.

The fundamental problem of communications is that of reproducing at one point a message transmitted at another point.

The actual message is one selected from a set of possible messages.
If we make an observation, the information that we expect to gain from the observation, before we actually make it, is called the entropy $H$. Entropy is about the amount of uncertainty that can on average be eliminated by a piece of information.

Shannon (1948) provides a formula for $H$ as

$$H = - \sum_{i=1}^{n} p_i \log_2 p_i$$

where $p_i$ is the probability of a particular answer being elicited. This allows entropies of components of compound events to be added.

For example, if we know the prior probabilities of the use of a piece of land are:

- residential: 0.4
- industrial: 0.6

then $H = -(0.4 \log_2 0.4 + 0.6 \log_2 0.6) \approx 0.7$ bits
Simple map example

Our map is of a collection of only two kinds of symbols:
dots represent towns or villages
lines represent roads.
Let \( p \) be the probability of any particular symbol chosen at random being a dot.
The average information per symbol is
\[
H = -(p \log_2 p + (1-p) \log_2 (1-p))
\]
So, if there are 9 times as many lines as there are dots,
\[
H = -(0.9 \log_2 0.9 + 0.1 \log_2 0.1)
= 0.47 \text{ bits per symbol.}
\]
Critique of communication theory as a theory of information

The entropy formula as a measure of information is rather arbitrary. Communication theory measures quantities of information, but does not consider at all information content. Shannon has said that the semantic aspects of information are irrelevant to the engineering aspects.
Section 3:

Imperfection, uncertainty, data quality
The components of data quality
Example

CIA - Intel

A report came in a week ago from a new agent that in the next year there will be a new ice cream factory opening somewhere in the Arctic Circle.

(Later intel indicates that this report is incorrect. For “ice cream” read “chemical weapons.”)
Quality dimensions

accuracy  format  comparability
reliability  interpretability  conciseness
Timeliness  content  freedom from bias
Relevance  efficiency  informativeness
Completeness  importance  level of detail
currency  sufficiency  quantitativeness
Consistency  usability  scope
Flexibility  usefulness  understandability
Precision  clarity

accuracy
bias
informativeness
application
data
data collector
data user
completeness
precision
timeliness
Error, imprecision, vagueness, and indeterminacy

Error
  lack of correlation with the actual state of affairs

Imprecision
  lack of specificity in the representation

Vagueness
  existence of borderline cases

Conflict
  existence of multiple representations of the same situation

Indeterminacy
  more than one outcome possible in a given situation
Representations and the world

Error
  lack of correlation with the actual state of affairs

Imprecision
  lack of specificity in the representation

Vagueness
  existence of borderline cases

Conflict
  existence of multiple representations of the same situation

Indeterminacy
  more than one outcome possible in a given situation
Types of imperfection

- Imperfection
  - Error
    - Lack of correlation between domains
    - “The time is 1103 and 26.4 sec”
  - Imprecision
    - Lack of specificity
    - “This talk is taking place in France”
  - Vagueness
    - Existence of borderline cases
    - “La Londe is near Toulon”
Categories of imperfection

**Inherent**: Independent of observations, observational methods, measuring devices.

**Contingent (Representational)**: Dependent on the inadequacy of observations, observational methods, measuring devices.
Vagueness

Vague predicates and objects admit borderline cases for which it is not clear whether the predicate is true of false.

Vagueness is all pervasive in representations of the real world.

Vagueness is not easy to handle using classical reasoning approaches.
Working with vagueness

Context dependence, subjectivism
what counts as being tall?

Non-Boolean semantics
'Fred is tall' true or false?

... but still needs reasoning with
If Mary is short then she is not tall
If George is tall and Lynne is taller than George, then Lynne is tall
Section 3:
Quantitative approaches to uncertainty
Basics of chance and probability
Preliminaries on chance (probability)

Chance arises when observing the outcome of an aleatory (random) experiment, such as the throw of a die.

If $X$ denotes the set of possible outcomes, we can specify a chance function $\text{ch} : X \rightarrow [0,1]$. $\text{ch}(x)$ gives the proportion of times that a particular outcome occurs.

$\text{ch}$ must satisfy the constraint that the sum of chances of all possible outcomes is 1.

For $S \subseteq X$, $\text{ch}(S)$ is the chance of an outcome from set $S$. 
Basic rules of chance

1. \( \text{ch}(\emptyset) = 0 \)
2. \( \text{ch}(X) = 1 \)
3. If \( A \cap B = \emptyset \), then \( \text{ch}(A \cup B) = \text{ch}(A) + \text{ch}(B) \)
Justifying probabilities (JH17-24)

If belief is to be quantified using probability, then it is important to explain where the numbers come from.

Principle of indifference: All elementary outcomes are equally likely.
- Great for card players, many basic situations.
- Problem often to determine what the elementary outcomes are.

Relative frequency approach
- But what about things that happen only once (nuclear power station melt-down)?

Subjective assessment
- What rules must these assessments obey?
Put your money where your mouth is!

Let $W$ be a set of possible events, and $U \subseteq W$. What level of belief do we assign $U$?

Consider the following family of bets:
- If $U$ happens then we win $100(1-p)$,
- but if $U$ doesn’t happen then we lose $100p$.

For what values of $p$ would we take the bet.

Note: if we take the bet for value $p$, and $p' < p$ then we should take it for $p'$ (because we win more and lose less).

So we are interested in the highest value of $p$ we are prepared to risk.

Cases:
- $U$ is certain: $p=1$
- $U$ cannot happen: $p=0$
- $U$ 50-50, $p=0.5$

Note connection with levels of chance, belief, etc.
Probability and choice

Probabilities may not always be the best way of deciding what we do.

Example:

You are shown a box containing 10 balls. You are told that 3 are red, and that the others are either yellow or green. A ball will be drawn from the box. You must predict what the color of the ball will be. You get $100 if you are correct. You lose $100 if you are wrong. What do you predict?
Probability and choice -2

Experiments show that most people choose Red. Probability theory, at least assuming the principle of indifference, will assign probabilities as follows:

\[
\begin{align*}
ch(\text{Red}) &= 0.3 \\
ch(\text{Yellow}) &= 0.35 \\
ch(\text{Green}) &= 0.35
\end{align*}
\]

Maybe what is needed are lower and upper probabilities:

\[
\begin{align*}
\text{lowch}(\text{Red}) &= 0.3 & \text{highch}(\text{Red}) &= 0.3 \\
\text{lowch}(\text{Yellow}) &= 0.0 & \text{highch}(\text{Yellow}) &= 0.7 \\
\text{lowch}(\text{Green}) &= 0.0 & \text{highchance}(\text{Green}) &= 0.7
\end{align*}
\]

depending on which distribution of Y and G is in the bag.

Risk needs to be balanced against payoff.
Compound and conditional outcomes

Given \( n \) independent trials of a single aleatory experiment. The chance of the compound outcome \((x_1, \ldots, x_n)\) is given by:

\[
ch^n(x_1, \ldots, x_n) = ch(x_1) \times \cdots \times ch(x_n)
\]

Suppose an aleatory experiment has been only partly completed and the outcome partly determined to be in the set \( U \subseteq X \). If \( V \subseteq X \) is the outcome set under consideration, we write the chance of \( V \) following the partial determination as \( ch(V | U) \). We have the formula:

\[
ch(V | U) = \frac{ch(U \cap V)}{ch(U)}
\]
Example of conditional chance

A tree seed has been planted in 1 of 9 cells. We have no other initial knowledge. Is it in region V?

We determine that the seed will only grow in region U, as the rest is infertile.

We revise the chance that the seed will grow in V to $ch(V|U)$.

$$ch(V|U) = \frac{ch(U \cap V)}{ch(U)}$$
$$= \frac{2/9}{6/9}$$
$$= 1/3$$
Another example

Suppose J. Doe is a randomly chosen American who was alive on January 1, 2000.

According to the United States Center for Disease Control:
   Roughly 2.4 million of the 275 million Americans alive on that date died during the 2000 calendar year.
   Among the approximately 16.6 million senior citizens (age 75 or greater) about 1.36 million died in 2000.

Let U be the event of J. Doe dying in 2000.
Let V be the event of J. Doe being a senior.

Calculate \( \text{ch}(U) \)
Calculate \( \text{ch}(U|V) \)
Another example

Suppose J. Doe is a randomly chosen American who was alive on January 1, 2000.

According to the United States Center for Disease Control:
- Roughly 2.4 million of the 275 million Americans alive on that date died during the 2000 calendar year.
- Among the approximately 16.6 million senior citizens (age 75 or greater) about 1.36 million died in 2000.

Let U be the event of J. Doe dying in 2000.
Let V be the event of J. Doe being a senior.

Calculate \( \text{ch}(U) \), \( \text{ch}(U|V) \), \( \text{ch}(V|U) \)

\[
\text{ch}(U) = \text{the population-wide mortality rate} = \frac{2.4M}{275M} = 0.00873.
\]
\[
\text{ch}(U\&V) = \frac{1.36M}{275M} = 0.00495,
\]
\[
\text{ch}(V) = \frac{16.6M}{275M} = 0.06036.
\]
\[
\text{ch}(U|V) = \frac{\text{ch}(U\&V)}{\text{ch}(V)} = \frac{0.00495}{0.06036} = 0.082 \text{ (note, this is } \frac{1.36}{16.6}).
\]
\[
\text{ch}(V|U) = \frac{\text{ch}(U\&V)}{\text{ch}(U)} = \frac{0.00495}{0.00873} = 0.57
\]
Beware: The second ace puzzle (JH1)

A deck has 4 cards: AH, 2H, AS, 2S. Cards are shuffled and Alice is dealt 2 cards.

\[ \text{ch(2 aces)} = \frac{1}{6} \]

Alice declares she has at least one ace in her hand.

\[ \text{ch(2 aces | at least 1 ace)} = \frac{\text{ch(2 aces)}}{\text{ch(at least 1 ace)}} \]
\[ = \frac{(1/6)}{(5/6)} = \frac{1}{5}. \]

Alice now declares that she has AH.

\[ \text{ch(2 aces | AH)} = \frac{\text{ch(2 aces)}}{\text{ch(AH)}} \]
\[ = \frac{(1/6)}{(1/2)} = \frac{1}{3}. \]

But, if Alice had said that she had AS, ch would be 1/3.

Why does the chance go up from 1/5 to 1/3 whichever ace Alice says she has? After all, we know that she must have one or the other when she tells us that she has at least one ace!
Bayesian Theory
Bayesian theory of evidence

Developed by English clergyman Thomas Bayes (1702-1761).

A degree of belief with respect to a set $P$ of possibilities (outcomes) is given by a belief function $\text{Bel}: \mathcal{P}(P) \rightarrow [0,1]$, subject to:

1. $\text{Bel}(\emptyset) = 0$
2. $\text{Bel}(P) = 1$
3. If $A \cap B = \emptyset$, then $\text{Bel}(A \cup B) = \text{Bel}(A) + \text{Bel}(B)$

(same as the rules for chance)
Note on Bayesian belief

Rules of chance have been transferred to rules of belief.
Note that the additive rule implies that belief must sum to one.
Consider:
b₁: there is life in the Sirius system
b₂: there is not life in the Sirius system

\[ \text{Bel}(b₁)+\text{Bel}(b₂)=1, \text{ even though there is little evidence either way.} \]
Assign \( \text{Bel}(b₁)=\text{Bel}(b₂)=0.5? \)
Quite high levels based on little evidence.
We return later to this.
Bayesian formula

Suppose we begin with a Bayesian belief function Bel: \( \mathcal{P}(P) \rightarrow [0,1] \), and then learn that \( A \subseteq P \) is the case.

We replace Bel with a new Bayesian belief function \( \text{Bel}_A : \mathcal{P}(P) \rightarrow [0,1] \), given by:

\[
\text{Bel}_A(B) = \frac{\text{Bel}(A \cap B)}{\text{Bel}(A)}
\]

Or in the language of conditioned belief:

\[
\text{Bel}(B|A) = \frac{\text{Bel}(A \cap B)}{\text{Bel}(A)} \quad (\text{BF1})
\]

(note similarity with rule for conditioned chance)
Bayesian formula

Applying the formula BF1

\[ \text{Bel} \ (B | A) = \frac{\text{Bel} (A \cap B)}{\text{Bel} (A)} \]

symmetrically gives:

\[ \text{Bel} \ (A | B) = \frac{\text{Bel} (A \cap B)}{\text{Bel} (B)} \]

Eliminating \( \text{Bel} (A \cap B) \) gives:

\[ \text{Bel} \ (B | A) = \frac{\text{Bel} (B) \times \text{Bel} (A | B)}{\text{Bel} (A)} \quad (BF2) \]

and

\[ \text{Bel} \ (A | B) = \frac{\text{Bel} (A) \times \text{Bel} (B | A)}{\text{Bel} (B)} \quad (BF3) \]
Bayesian formula

\[ \text{Bel}(A|B) = \frac{\text{Bel}(A) \times \text{Bel}(B|A)}{\text{Bel}(B)} \] (BF3)

This is a widely used form of the Bayesian formula. In BF3, our posterior belief \( \text{Bel}(A|B) \) is calculated by multiplying our prior belief \( \text{Bel}(A) \) by the likelihood that \( B \) will occur if \( A \) is true.

The denominator is a normalizing constant.

We now see how it is calculated.
Calculation of \( \text{Bel}(B) \) - marginalization

Note that:
\[
B = (A \cap B) \cup (A' \cap B)
\]
and \((A \cap B) \cap (A' \cap B) = \emptyset\).

Apply Bayesian rule 3 (additivity):
\[
\text{Bel} (B) = \text{Bel}((A \cap B) \cup (A' \cap B)) = \text{Bel}(A \cap B) + \text{Bel}(A' \cap B)
\]

Now apply the Bayesian formula BF1:
\[
\text{Bel} (B) = \text{Bel}(B|A)\text{Bel}(A) + \text{Bel}(B|A')\text{Bel}(A') \quad (BF4)
\]

This is another widely used form of the Bayesian formula.
Bayesian example

Suppose there is a certain disease found randomly in 0.005 of the population (determined by epidemiological experiment).

A certain clinical test is 0.99 effective in detecting the presence of the disease (i.e. it yields a positive result in 99% of the cases where the disease is present (determined by clinical trials).

However, the test also yields false positives in 5% of the cases where the disease is not present (determined by clinical trials).

You take the test and the result is positive. How likely are you to have the disease?
Bayesian example continued

Let A be the outcome that the disease is present. Let B be the outcome that the test is positive.

We have:

\[ \text{Bel}(A)=.005, \ \text{Bel}(B|A)=.99, \ \text{Bel}(B|A')=.05. \]

We wish to know Bel(A|B).

Using BF4:

\[ \text{Bel}(B) = \text{Bel}(B|A)\text{Bel}(A) + \text{Bel}(B|A')\text{Bel}(A') \]
\[ = .99*.005 + .05*.995 \]
\[ = .0547 \]

Using BF3:

\[ \text{Bel}(A|B) = \frac{\text{Bel}(A) \times \text{Bel}(B|A)}{\text{Bel}(B)} \]
\[ = .005*.99/.0547 \]
\[ = .0905 \]

This should be the level of our belief that we have the disease following a positive test.
Bayesian example continued

Determine the following:

1. Chance of a negative test, irrespective of whether the disease is present or not.

2. Following a positive test, the level of belief that the disease is not present (false positive)

3. Following a negative test, the level of belief that the disease is not present (true negative).

4. Following a negative test, the level of belief that the disease is present (false negative).
Exercise on Bayes' Rule

Suppose that Bob tests positive on an Aids test that is 99% reliable.

Suppose, in this case, that a test is \( x \)% reliable if during clinical trials:

- \( x \)% of subjects with Aids tested positive
- \( x \)% of subjects that did not have Aids tested negative

How likely is it that Bob has Aids? Write an expression for this level of belief in terms of the proportion of the general population that has Aids. Comment on this expression. (E.g., plot the results using Excel)
Inconsistency checking

Suppose we have propositions:
A: I am in Maine
B: It is snowing

and we hold the following beliefs:
Bel(A) = 0.5
Bel(A|B) = 0.3
Bel(B|A) = 0.7

Then Bel(B) = Bel(B|A)Bel(A)/Bel(A|B)
= (0.7 * 0.5) / 0.3
> 1
Chain rule

We can rearrange BF1 to get the so-called **product rule**:
\[ b(A,B) = b(A|B) \cdot b(B) \]

We can extend this for three variables:
\[ b(A,B,C) = b(A|B,C) \cdot b(B,C) = b(A|B,C) \cdot b(B|C) \cdot b(C) \]

and in general to \( n \) variables:
\[ b(A_1, A_2, \ldots, A_n) = b(A_1|A_2, \ldots, A_n) \cdot b(A_2|A_3, \ldots, A_n) \ldots b(A_{n-1}|A_n) \cdot b(A_n) \]

In general we refer to this as the **chain rule**.

**Note use of comma for intersection**
Bayesian Belief Nets

Used for determining levels of causality between many events.

- cloudy
- sprinkler
- mudpath
- rain
Background to Bayesian networks

A Bayesian network $B = (V,E)$ is a directed acyclic graph in which each node in $V$ is annotated with quantitative probability/belief information. A set $V$ of events label the nodes of the network.

If there is an edge from node $X$ to node $Y$ in $B$, then $X$ is said to be the parent of $Y$. Each node $X$ in $V$ has a conditional probability distribution $P(X|\text{Parents}(X))$ associated with it.
b(C) = 0.5
b(C') = 0.5

b(R|C) = 0.8
b(R|C') = 0.2

b(R'|C) = 0.2
b(R'|C') = 0.8

b(S|C) = 0.1
b(S|C') = 0.5

b(S'|C) = 0.9
b(S'|C') = 0.5

b(M|S,R) = 0.99
b(M|S,R') = 0.9
b(M|S',R) = 0.9
b(M|S',R') = 0.0

Observe the path is muddy. What is the likelier cause, rain or the sprinkler?
Calculation (adding the tears)

Chain rule gives \( b(A,B,C,D) = b(A|B,C,D) b(B|C,D) b(C|D) b(D) \)
So \( b(M,R,S,C) \)
\[ = b(M|R,S,C) b(R|S,C) b(S|C) b(C) \]
\[ = b(M|R,S) b(R|C) b(S|C) b(C) \quad \text{because of dependencies in BBN} \]
Similarly with \( b(M',R,S,C), \ldots \)
These joint beliefs can be calculated from the conditional beliefs given.
We want to compare \( b(R|M) \) with \( b(S|M) \).
\( b(R|M) = \frac{b(M,R)}{b(M)} \) and \( b(S|M) = \frac{b(M,S)}{b(M)} \) by BF1
So we must compare \( b(M,R) \) with \( b(M,S) \).
\[ b(M,R) = b(M,R,S,C) + b(M,R,S',C) + b(M,R,S,C') + b(M,R,S',C') \]
\[ b(M,S) = b(M,R,S,C) + b(M,R',S,C) + b(M,R,S,C') + b(M,R',S,C') \]
\[ b(M,R) - b(M,S) = (b(M,R,S',C) + b(M,R,S',C')) - (b(M,R',S,C) + b(M,R',S,C')) \]
\[ = + 0.9*0.8*0.9*0.5 + 0.9*0.2*0.5*0.5 \]
\[ - 0.9*0.2*0.1*0.5 - 0.9*0.8*0.5*0.5 \]
\[ = 0.18 \]
\[ > 0 \]
So rain is the likelier cause!
Please check!
BN example (Subramanian)

- Burglary
- Earthquake
- Alarm
- John calls
- Mary calls
What does the network mean?

Burglary and Earthquake directly affect the probability of the alarm going off. Whether or not John or Mary calls depends on the alarm. John and Mary do not directly perceive burglary or minor earthquakes.

\[
b(M|J,A,E,B) = b(M|A) \\
b(J|M,A,E,B) = b(J|A)
\]

JohnCalls is **conditionally independent** of MaryCalls given Alarm.

But note: \( b(J|M) = b(J) \) is not necessarily the case.
Independence and conditional independence

$X$ is independent of $Y$ if
\[ P(X, Y) = P(X)P(Y) \]

$X$ is conditionally independent of $Y$ given $Z$ if
\[ P(X \mid Y, Z) = P(X \mid Z), \text{ or equivalently} \]
\[ P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z) \]

In the example, $\text{JohnCalls}$ is conditionally independent of $\text{MaryCalls}$ given $\text{Alarm}$. 
BN example (Subramanian)

- Burglary
  - $b(B) = 0.001$
  - $b(A | B, E) = 0.95$
  - $b(A | B, E') = 0.95$
  - $b(A | B', E) = 0.29$
  - $b(A | B', E') = 0.001$

- Earthquake
  - $b(E) = 0.002$
  - $b(A | B, E) = 0.95$
  - $b(A | B, E') = 0.95$
  - $b(A | B', E) = 0.29$
  - $b(A | B', E') = 0.001$

- Alarm
  - $b(J | A) = 0.9$
  - $b(J | A') = 0.05$

- John calls

- Mary calls
  - $b(M | A) = 0.7$
  - $b(M | A') = 0.01$
Questions

1. John calls. What is the chance that the alarm went off?
2. Mary calls. What is the chance that the alarm went off.
3. John calls. What is the chance of an earthquake?
Dempster-Shafer Theory of Evidence
Shafer’s (1976) theory of evidence provides a way of pooling the total evidence available.
A real number between 0 and 1 indicates the degree of support a body of evidence provides for a proposition.
The theory focuses on the combination of degrees of belief or support from distinct bodies of evidence
Dempster’s Rule of Combination provides a method for changing beliefs in the light of new evidence.
Elements of Dempster-Shafer Theory

DS theory can be thought of as an extension of probability theory, and generalization of Bayesian statistics.
Uncertain belief for propositions based on evidence is represented by belief functions.
A frame of discernment $F$ is a set of mutually exhaustive possibilities.
For each piece of evidence, there is an associated mass function (or basic probability assignment, bpa) over $F$ that assigns to each subset of $S \subseteq F$ a value from 0 to 1, subject to:
- $m(\emptyset) = 0$
- sum of $m(S)$ over all $S \subseteq F$ is 1
Elements of Dempster-Shafer Theory

The idea is that $m(S)$ gives the degree of belief assigned to exactly the set $S$ (and not to any proper subset of $S$).

If $m(S) = 0.6$, then 60\% of one's total belief is assigned to (committed to) exactly $S$.

The subsets $S \subseteq F$ for which $m(S) > 0$ are called the focal elements of the piece of evidence.

The union of all focal elements is the focus of the evidence.
Elements of Dempster-Shafer Theory

The belief function $Bel$ induced by mass function $m$ assigns to each subset $S \subseteq F$ a value from 0 to 1.

$Bel(S) = \text{sum of } m(T) \text{ over all } T \subseteq S$

So $Bel(S)$ is a measure of the total belief in $S$, including the belief in more specific propositions.

Credibility or Support ($Bel$)
- amount of evidence for a proposition

Plausibility ($Pl$)
- lack of evidence against a proposition

$Pl(S) = 1 - Bel(S')$

$Pl(S) = \text{sum of } m(T) \text{ over all } T \text{ that overlap with } S$

Mass, belief and plausibility can all be defined in terms of one another
Example

Assume $F$ is \{H,C,P,Q\} and
\[
\begin{align*}
    m(\{H\}) &= 0.3 \\
    m(\{H,C\}) &= 0.2 \\
    m(\{H,C,P\}) &= 0.5
\end{align*}
\]

Then
\[
\begin{align*}
    Bel(\{H\}) &= 0.3 & Pl(\{H\}) &= 1.0 \\
    Bel(\{H,C\}) &= 0.5 & Pl(\{H,C\}) &= 1.0 \\
    Bel(\{H,P\}) &= 0.3 & Pl(\{H,P\}) &= 1.0 \\
    Bel(\{P\}) &= 0 & Pl(\{P\}) &= 0.5 \\
    Bel(\{C,P\}) &= 0 & Pl(\{C,P\}) &= 0.7
\end{align*}
\]
DS in not Bayesian

To describe fully the evidence concerning a proposition $A$, we need to know not only the degree of support there is for $A$, $\text{Bel}(A)$, but also the degree to which the negation of $A$ is not supported, $\text{Pl}(A)$. Note that $\text{Bel}(A) \leq \text{Pl}(A)$.

We can think of $\text{Bel}(A)$ as the lower limit of support/belief and $\text{Pl}(A)$ as the upper limit.

It is not necessarily true that:

Belief for $A$ + Belief against $A$ = 1.

There may be some belief that is uncommitted. We are able to distinguish between disbelieve and lack of belief.

This contrasts with the Bayesian Theory which requires that the beliefs for and against a proposition should sum to one.
Integration of imperfect information

Dempster’s rule of combination
Dempster's Rule of Combination

Suppose $F$ is a frame. Suppose we have two separate pieces of evidence concerning the frame, with mass functions $m_1$ and $m_2$. We want to calculate the mass function $m$ for the combined evidence. Note that the pieces of evidence may support or conflict with each other.

Let $A_1, \ldots, A_n$ be focal elements for $m_1$.

Let $B_1, \ldots, B_p$ be focal elements for $m_2$.

For $A \subseteq F$,

$$m(A) = \frac{\text{sum of products } m_1(A_j) m_2(B_k) \text{ for all } A_j B_k \text{ whose intersection is } A}{\text{sum of products } m_1(A_j) m_2(B_k) \text{ for all } A_j B_k \text{ whose intersection is non-empty}}.$$ 

The denominator ensures the measure of the total belief of the joint evidence is 1.
Example: reinforcement, no normalization

Suppose $F = \{D,D'\}$

\[
\begin{align*}
m_1(\{D\}) &= 0.8 & m_1(\{D'\}) &= 0 & m_1(\{D,D'\}) &= 0.2 \\
m_2(\{D\}) &= 0.9 & m_2(\{D'\}) &= 0 & m_2(\{D,D'\}) &= 0.1
\end{align*}
\]

We create a table with rows and columns named by non-empty subsets of $F$.

<table>
<thead>
<tr>
<th></th>
<th>$m_1({D})$</th>
<th>$m_1({D'})$</th>
<th>$M_1({D,D'})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_2({D})$</td>
<td>0.9</td>
<td>0</td>
<td>0.18</td>
</tr>
<tr>
<td>$m_2({D'})$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$m_2({D,D'})$</td>
<td>0.08</td>
<td>0</td>
<td>0.02</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
m_1+m_2(\{D\}) &= (0.72 + 0.08 + 0.18)/1 = 0.98 \\
m_1+m_2(\{D'\}) &= 0
\end{align*}
\]
Example: normalization

Suppose \( F = \{a,b,c\} \)

\[
\begin{align*}
m1(\{a\}) &= 0.5 & m1(\{a,b\}) &= 0.2 & m1(\{a,c\}) &= 0.3 \\
m2(\{b\}) &= 0.6 & m2(\{b,c\}) &= 0.4
\end{align*}
\]

We create a table with rows and columns named by non-empty subsets of \( F \).

<table>
<thead>
<tr>
<th></th>
<th>( m1({a}) )</th>
<th>( m1({a,b}) )</th>
<th>( m1({a,c}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m2({b}) ) 0.6</td>
<td>0.3</td>
<td>0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>( m2({b,c}) ) 0.4</td>
<td>0.2</td>
<td>0.08</td>
<td>0.12</td>
</tr>
</tbody>
</table>

\[
m1+m2(\{b\}) = \frac{(0.12+0.08)}{(0.12+0.08+0.12)} = 0.63
\]
Simple geographical example 1 of Dempster’s Rule of Combination

Let frame be $F=\{f,g\}$ where:
- $f$ is ‘pixel is of type forest’
- $g$ is ‘pixel is not of type forest’

Sensor 1 provides support that pixel $P$ is of type forest as follows: $m_1(\{f\})=0.6$, $m_1(\{f,g\})=0.4$

Sensor 2 provides support that pixel $Q$ is of type forest as follows: $m_2(\{f\})=0.7$, $m_2(\{f,g\})=0.3$

The combination of two sensors provides support that region $P \cap Q$ is of type forest as follows: $m(\{f\})=0.88$, $m(\{f,g\})=0.12$
Let frame be $\mathbf{F} = \{f, g\}$ where:

- $f$ is 'pixel is of type forest'
- $g$ is 'pixel is not of type forest'

Sensor 1 provides support that pixel $P$ is of type forest as follows: $m_1(\{f\}) = 0.6$, $m_1(\{f, g\}) = 0.4$

Sensor 2 provides support that pixel $Q$ is of type forest as follows: $m_2(\{g\}) = 0.7$, $m_2(\{f, g\}) = 0.3$

The combination of two sensors provides support for the land type of region $P \cap Q$ as follows:

$m(\{f\}) = 0.31$, $m(\{g\}) = 0.48$, $m(\{f, g\}) = 0.21$
Exercise

Alice has a coin, and she knows that it either has a bias towards heads or a bias towards tails (but not both).
Let the heads bias be denoted: BH
Let the tails bias be denoted: BT.
Suppose that Alice tosses the coin and it comes down heads. Let the value of the corresponding mass function \( m_1(BH) = k \) (the exact value of \( k \) does not matter, but assume \( k \) is strictly between 0 and 1.)
1. Calculate Alice's belief that the coin is biased towards heads.
2. Suppose that Alice tosses the coin again and it comes down heads. Calculate Alice's new belief that the coin is biased towards heads.
3. Suppose that Alice tosses the coin \( n \) times, and it always comes down heads. Calculate Alice's belief that the coin is biased towards heads.
4. After the initial toss, suppose the 2\(^{nd}\) toss comes down tails. Calculate all the mass functions and beliefs.
Counter Intuitive?

Zadeh has given an example of a use of Dempster's Rule which has unexpected results.

Two doctors examine a patient and agree that he/she suffers from either meningitis (M), concussion (C), or brain tumor (T).
Thus \( U = \{M,C,T\} \).

The doctors have different diagnosis:
- \( m_1(\{M\}) = 0.99, \quad m_1(\{T\}) = 0.01 \);
- \( m_2(\{C\}) = 0.99, \quad m_2(\{T\}) = 0.01 \);

They agree on their low expectation of a tumor, but disagree on the likely cause.

If we now combine beliefs and compute \( m_1 + m_2(\{T\}) \) we find that it is 1!

Verify, Discuss
Section 4:

Qualitative approaches to uncertainty
Possible worlds
Belief and knowledge
Belief revision
Many-valued logics
Fuzzy sets
Rough sets
Possible worlds
Possible world representation of uncertainty (WD9.3.1)

The worlds or outcomes that an agent considers possible are called
- possible worlds
- states
- elementary outcomes

E.g. toss a die: W1, W2, ..., W6.

The set of possible worlds is sometimes called the sample space.

An event is a set U of possible worlds = the possibility under consideration.

The set U consists of all worlds in which the event can take place.

E.g. the event of a die landing an even number = \{W2, W4, W6\}. 
Representing an agent’s uncertainty using possible worlds

Let $W$ be the set of all states in a specific case.

The uncertainty of an agent may be represented by a subset $W'$ of $W$.

$W'$ is the subset of worlds that the agent considers possible.

The larger $W'$, the greater the agent’s uncertainty.

Let the event under consideration be $U$.

If $U$ and $W'$ are not disjoint, then event $U$ is possible, otherwise $U$ is impossible.

If $W'$ is a subset of $U$, then the event is certain.

Usually, we assume that $W$ is finite.
Example

Toss of a die.

$W = \{W_1, W_2, W_3, W_4, W_5, W_6\}$

$W' = \{W_1, W_2, W_3\}$

(the agent has been told that the die landed with a 1, 2 or 3.

$U$ is the event of an even landing $= \{W_2, W_4, W_6\}$

$W'$ and $U$ are not disjoint, so $U$ is possible.

$W'$ is not a subset of $U$ so $U$ is not definite.
Belief and knowledge
Belief and knowledge

Belief results if a statement is held to be true by an agent.
Knowledge is often defined as justified, true belief.

I might believe that pigs can fly, but I cannot know this.
I might believe that there is life on Sirius-A, but I cannot currently know this.
Modal formalism for knowledge and belief

Let \( p \) be the proposition “The region is forested”. Then:

- \( Kp \) indicates that I (the agent, the KB) know that the region is forested.
- \( Bp \) indicates that I (the agent, the KB) believe that the region is forested.

Logics of knowledge and belief allow us to use the modal operators \( K \) and \( B \), and contain axioms like:

- \( Kp \rightarrow Bp \)
- \( Kp \rightarrow p \)
- \( \neg Kp \rightarrow \neg p \) (Closed World Assumption)
- \( Kp \rightarrow KKp \) (positive introspection)
- \( \neg Kp \rightarrow K\neg Kp \) (negative introspection)
Inconsistency, uncertainty, and belief revision
Inconsistency may result from:

- measurement error
- specification mismatch (imprecision)
- concept vagueness
- temporal variation
- representation mismatch (semantic heterogeneity)
Inconsistent information may be:

resolved before entry into database
  e.g. statistical averaging
  weighting according to preference
  (scale, currency, source reliability, …)

held in the database and resolved at application run time
  more context-sensitive
  measurement GIS approach
Inconsistency can result in several actions:

**learning**: resulting in revision of information

**information acquisition**: seeking further information to resolve inconsistency

**inconsistency removal**: belief revision, localization of inconsistency, non-monotonic reasoning

**invoking preference**: resolve inconsistency by preferring some information to other

**argumentation**: instituting a dialogue between the agents holding the differing information
But how to reason with an inconsistent knowledge base?

We can maintain consistency or introduce methods of reasoning with inconsistent knowledge.
Revision and consistency

旧知识库（old KB）：一致（consistent）

新知识库（new KB）：一致（consistent）？

新信息（new information）
Flooding application

Aerial photography flows balance flow
water height \langle l, h \rangle

\langle 10, 20 \rangle
\langle 10, 30 \rangle
\langle 30, 40 \rangle
\langle 10, 30 \rangle
\langle 30, 40 \rangle
Belief change example
(Gardenfors and Rott, 1995)

Beliefs
The bird caught in the trap is a swan.
The bird caught in the trap comes from Sweden
Sweden is part of Europe
All European swans are white

Consequences
The bird caught in the trap is white

New information
The bird caught in the trap is black

Which sentence(s) would you give up?
Motivation

Problems arise is several areas
- IS: new entry inconsistent with DB
- Sensors: Sensor info inconsistent with theory/previous sensor/other sensor
- Diagnosis: Device behavior inconsistent with device description

Problem nontrivial
- Choice involved
- The may be unforeseen indirect consequences of revision

Representation
We can think of the issue as 'constraint satisfaction', where the constraints may be 'overdetermining' the system.
Desirable

To follow some rational principles
  minimal change
  maximal consistency
  computational tractability
Revision v update
(Katsuno and Mendelzon)

Revision: agent gets new information about a static world

you believe book is either on the table or on the chair
I tell you that the book is not on the table
you then believe book is on chair

Update: agent gets new information about a changing world

you believe book is either on the table or on the chair
you are told that all books on the table have now been removed from the room
you believe that the book is either on the chair or not in the room
Non-monotonic reasoning systems

Non-monotonic logics more closely model common-sense reasoning, as they allow both the introduction and retraction of beliefs. The field of non-monotonic reasoning includes work on:

- logics of knowledge and belief
- default logic
- truth maintenance
- closed-world databases
- probabilistic reasoning

... Default logic allows us to reason with assumptions that we may retract if we have evidence contradicting them. The reasoning is *defeasible*. 


Default reasoning

Default rules: apply unless they lead to a contradiction

from 'Jack is an accountant' infer 'Jack is neat in appearance' unless we know or can infer otherwise.

accountant(Jack) : neat(Jack) / neat(Jack)

If Jack is an accountant then we infer that Jack is neat, unless and until we have any evidence that definitely refutes the condition(s).
Default Logic

The database $W$. 

$W = \{F_b, \neg F_f, A(a,b), A(b,c), A(c,d), A(d,e), A(e,f), \forall x,y \ A(x,y) \leftrightarrow A(y,x)\}$

$D = \{A(x,y) \land F_x : F_y / F_y, A(x,y) \land \neg F_x : \neg F_y / \neg F_y\}$

An extension to the default theory $(W, D)$, a possible way that the world might be, is found by applying default rules $D$ to the data in $W$. 

SIE 565
Default Logic

The database $W$.

$W = \{F_b, \neg F_f, A(a,b), A(b,c), A(c,d), A(d,e), A(e,f), \forall x,y A(x,y) \leftrightarrow A(y,x)\}$

$D = \{A(x,y) \land F_x : F_y / F_y, A(x,y) \land \neg F_x : \neg F_y / \neg F_y\}$
Extensions in a default theory
Many-valued logics
A 3-valued logic
(Kleene's system)
Kleene’s 3-valued logic

Assume three truth values, T,F,?.
The meaning of ? depends on the system.
In Kleene’s system, ? means there are not sufficient (computing) resources to arrive at an answer.
Kleene’s truth tables provide rules for computing the truth values of composite statements.
p = 'The class is SIE 565'
q = 'The trillionth prime ends in a 3'
What are the truth values of p,q,¬p,¬q, p∧q, p∨q, p→q?
Truth tables for Kleene’s logic

<table>
<thead>
<tr>
<th>∧</th>
<th>T ? F</th>
<th>∨</th>
<th>T ? F</th>
<th>¬</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T ? F</td>
<td>T</td>
<td>T T T</td>
<td>T</td>
<td>T F</td>
</tr>
<tr>
<td>F</td>
<td>F F F</td>
<td>F</td>
<td>T ? F</td>
<td>F</td>
<td>T</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>→</th>
<th>T ? F</th>
<th>↔</th>
<th>T ? F</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T ? F</td>
<td>T</td>
<td>T ? F</td>
</tr>
<tr>
<td>F</td>
<td>T T T</td>
<td>F</td>
<td>F ? T</td>
</tr>
</tbody>
</table>
Logics of fusion of imperfect information

<table>
<thead>
<tr>
<th>fuse</th>
<th>1</th>
<th>?</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>0</td>
<td>?</td>
<td>?</td>
<td>0</td>
</tr>
</tbody>
</table>

Vague region
no errors
conservative

<table>
<thead>
<tr>
<th>fuse</th>
<th>1</th>
<th>?</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>X</td>
</tr>
<tr>
<td>?</td>
<td>1</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>X</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Crisp region
no errors
weighted towards definite
Some significant places of the campus. Which places are near other places?
The nearness experiment

22 subjects (campus staff), divided into 2 equal size groups A and B.

For a given place P:
  - ask A-subjects which places are near P
  - ask B-subjects which places are not near P

Repeat for each place P on the campus.
Example: Near the Library?

<table>
<thead>
<tr>
<th>Place</th>
<th>Near</th>
<th>Not near</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 hour Reception</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Academic Affairs</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Barnes Hall</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Biological Sciences</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Chancellor’s Building</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Chapel</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Chemistry</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Clock House</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Computer Science</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Earth Sciences</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Health Centre</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Holly Cross / The Oaks</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Horwood Hall</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Keele Hall</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Lakes</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Leisure Centre</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Library</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Lindsay Hall</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Observatory</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Physics</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Students Union</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Visual Arts</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

What do we do with the data?
... adding leftness

The leftness experiment was of the same form.
In this case the Library was taken as fixed throughout the questionnaires.
In each questionnaire, subjects were asked to imagine standing at one of the reference places facing the Library, and to identify each place P for which it was true (false) to say that P was on the person’s left.
Weight of evidence

Simple minded approach:

\[ \sigma_v(p, q) = \frac{n_T(p, q) - n_\perp(p, q)}{N} \]

\[ \sigma_\lambda(p, q) = \frac{l_T(p, q) - l_\perp(p, q)}{N} \]

Using Dempster-Shafer:

\[ \sigma_v^+(p, q) = \frac{n_T(p, q)(N - n_\perp(p, q))}{N^2 - n_T(p, q)n_\perp(p, q)} \]

\[ \sigma_\lambda^+(p, q) = \frac{l_T(p, q)(N - l_\perp(p, q))}{N^2 - l_T(p, q)l_\perp(p, q)} \]
Variation of nearness $\sigma_v$ with distance
Variation of leftness $\sigma_\lambda$ with angle
Taking advantage of the true/false information

Scattergram analysis
Belnap’s 4-valued logic
### Four-valued logic

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>n</th>
<th>b</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>or</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>n</td>
<td>1</td>
<td>n</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>1</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>n</td>
<td>b</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>n</th>
<th>b</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>and</td>
<td>1</td>
<td>1</td>
<td>n</td>
<td>b</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>n</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>0</td>
<td>b</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>n</th>
<th>b</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>not</td>
<td>1</td>
<td>n</td>
<td>b</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>n</td>
<td>b</td>
<td>1</td>
</tr>
</tbody>
</table>
de Morgan lattice

Votes for near

Votes for not near

near

not near

conflict
Fuzzy sets
Preamble: Levels of Belief
Thought experiment

We want to estimate levels of belief in two propositions, p and q. Select N people and ask each:
- Check box A if you believe p.
- Check box B if you believe q.

(They can check neither, one, or both).

Suppose P people check box A and Q check box B. Define:
- \( \text{bel}(p) = \frac{P}{N} \)
- \( \text{bel}(q) = \frac{Q}{N} \)

Suppose we move out of the area, but later become interested in the level of belief in \( p \lor q \)? How do we calculate it?
Assumption 1

Rational voters

1. If voter checks for proposition p, and $p \rightarrow r$ is the case, then the voter will check r.
2. If voter does not check for p, and $r \rightarrow p$ is the case, then the voter will not check r.

Then

$$\max (\text{bel}(p), \text{bel}(q)) = \leq \text{bel}(p \lor q) = \leq \text{bel}(p)+\text{bel}(q)$$
Assumption 2

Graded sets of voters

The collection of voters in each group can be ordered so that if v votes for p, and v < w, then w votes for p.

Then:

\[ \text{bel}(p \lor q) = \max (\text{bel}(p), \text{bel}(q)) \]
Assignment

1. Do the same thought experiment for the conjunction of beliefs.
2. Investigate other three-valued logics, and write a page about the possible systems you have found.
Basic concepts in fuzzy sets

Let $U$ be a universe of discourse.

A fuzzy membership function is a function $\mu$ from $U$ to $[0,1]$.

A fuzzy set $A$ in $U$ is a set of ordered pairs $(u, \mu_A(u))$ for all $u \in U$, where $\mu_A$ is a fuzzy membership function.

Fuzzy set $A$ is empty if if $\mu_A(u) = 0$ for all $u$.

Fuzzy set $A$ is contained in $B$ if $\mu_A(u) \leq \mu_B(u)$, for all $u$.

Fuzzy sets $A$ and $B$ are equal if $\mu_A(u) = \mu_B(u)$, for all $u$.

Fuzzy set $B$ is the complement of $A$ if $\mu_B(u) = 1 - \mu_A(u)$, for all $u$. 
Basic concepts in fuzzy sets

The union of fuzzy sets $A$ and $B$ is the set $A \cup B$ with membership function $\max(\mu_A(u), \mu_B(u))$, for $u$ in $U$.

The intersection of fuzzy sets $A$ and $B$ is the set $A \cap B$ with membership function $\min(\mu_A(u), \mu_B(u))$, for $u$ in $U$.

The support of fuzzy set $A$ is the crisp set $\text{support}(A) = \{u \mid \mu_A(u) > 0\}$.

For $0 \leq \alpha \leq 1$, the $\alpha$-cut of fuzzy set $A$ is the crisp set given by $A_\alpha = \{u \mid \mu_A(u) > \alpha\}$.
Generalized fuzzy conjunction

Let $T$ be a function from $[0,1] \times [0,1]$ to $[0,1]$. $T$ is a triangular norm (T-norm) if it satisfies:

- $T(1,1)=1$  
- $T(0,x)=T(x,0)=0$  
- $T(x,y)=T(y,x)$  
- $u \leq v$, $w \leq x$ implies $T(u,w) \leq T(v,x)$  
- $T(T(x,y),z)=T(x,T(y,z))$

boundary  
boundary  
commutativity  
monotonicity  
associativity
Example T-norms

\[ \min \{a, b\} \]

\[ \max \{a + b - 1, 0\} \quad \text{(Weak, Lukasiewicz)} \]

\[ ab \quad \text{(probability)} \]

\[ ab / (k + (1-k)(a + b - ab)), k \in (0, 1) \quad \text{(Hamacher)} \]

\[ ab / \max\{a, b, k\}, k \in (0, 1) \quad \text{(Dubois and Prade)} \]

there are many others
Generalized fuzzy disjunction

Let $S$ be a function from $[0,1] \times [0,1]$ to $[0,1]$. $S$ is a triangular conorm (S-conorm) if it satisfies:

- $S(0,0)=0$  
- $S(1,x)=S(x,1)=1$  
- $S(x,y)=S(y,x)$  
- $u \leq v$, $w \leq x$ implies $S(u,w) \leq S(v,x)$  
- $S(S(x,y),z)=S(x,S(y,z))$

boundary  
boundary  
commutativity  
monotonicity  
associativity
Example S-conorms

\[
\begin{align*}
\max \{a, b\} \\
\min \{a + b, 1\} & \quad \text{(Lukasiewicz)} \\
a + b - ab & \quad \text{(probability)}
\end{align*}
\]

there are many others
Generalized fuzzy complement

Let $C$ be a function from $[0,1]$ to $[0,1]$. $T$ is a generalized fuzzy complement if it satisfies:

- $C(0) = 1$ boundary
- $C(1) = 0$ boundary
- $CC(x) = x$ involution
- $x < y$ implies $C(y) < C(x)$ anti-monotonicity
De Morgan pairs

T-norms and S-norms come in DeMorgan pairs

\[ C(T(x,y)) = S(C(x), C(y)) \]
\[ C(S(x,y)) = T(C(x), C(y)) \]

e.g. min and max

Note, one can be defined in terms of the other

\[ T(x,y) = C(S(C(x), C(y))) \]
\[ S(x,y) = C(T(C(x), C(y))) \]
Three semantics of fuzzy sets

**Similarity**

$\mu_A(u)$ is the degree of proximity of $u$ to prototype elements of $A$.

E.g. cluster analysis.

**Preference**

$A$ represents a set of more or less preferred entities (or values of a decision variable $x$).

$\mu_A(u)$ represents the intensity of preference for object $u$, or the feasibility of selecting $u$ as a value of $x$.

**Uncertainty**

$\mu_A(u)$ is the degree of possibility that a parameter $x$ has value $u$, given that all that is known is that “$x$ is $A$”.

Dubois and Prade, Fuzzy Sets and Systems 90(1997) 141-150
Example

Let $U$ be a universe of discourse. Let $A$ be the fuzzy set of urban areas.

**Similarity**

Construct one or more prototype urban areas with specific characteristics (density of population, amount of roof pixels). $\mu_A(u)$ is the degree of proximity of an observed region $u$ to the prototype element(s).

**Preference**

We may wish to site a store in an urban area. $\mu_A(u)$ represents the intensity of our preference for a specific target site.

**Uncertainty**

Suppose that all we know about where Rick lives is that he lives in an urban area. $\mu_A(London)$ is the degree of possibility that Rick lives in London.
Application of Fuzzy Sets: Image Analysis

Satellite Raster Image

Fuzzy Set Membership
Fuzziness and image analysis

The standard approach to image analysis and recognition begins by segmenting the image into regions and computing various properties of and relationships among these regions. However, the regions are not always 'crisply' defined; it is sometimes more appropriate to regard them as fuzzy subsets of the image... It is not always obvious how to measure geometrical properties of fuzzy sets, but definitions have been given and basic properties established for a variety of such properties and relationships, including connectedness and surroundedness, convexity, area, perimeter and compactness, extent and diameter"

Examples of fuzzy geometry

Area of a fuzzy region

\[ a(\mu) = \sum \mu . \]

Perimeter of a fuzzy region

\[ p(\mu) = \sum_{m=1}^{M} \sum_{n=1}^{N-1} |\mu_{mn} - \mu_{m,n+1}| + \sum_{n=1}^{N} \sum_{m=1}^{M-1} |\mu_{mn} - \mu_{m+1,n}|. \]

Compactness of a fuzzy region

\[ \text{Compactness}(\mu) = \frac{a(\mu)}{p^2(\mu)}. \]
Do these definitions work?

For example, does area have its common properties?

\[ \text{area}(A \cup B) = \text{area}(A) + \text{area}(B) - \text{area}(A \cap B) ? \]

\[ \text{area}(A) = \text{area}(A \cap B) + \text{area}(A \cap B') ? \]
... and are they the right definitions?

Note that the definitions given for area, perimeter and compactness are crisp.

For fuzzy area definitions see IJGIS 2004.
Suppose we have \( n \) feature vectors \( x_1, x_2, \ldots, x_n \) all from the same class, and we know that they fall into \( k \) compact clusters, \( k < n \).

Let \( m_i \) be the mean of the vectors in Cluster \( i \).

If the clusters are well separated, we can use a minimum-distance classifier to separate them.

That is, we can say that \( x \) is in Cluster \( i \) if \( || x - m_i || \) is the minimum of all the \( k \) distances. This suggests the following procedure for finding the \( k \) means:
Cluster analysis K-means 2

Make initial guesses for the means $m_1, m_2, ..., m_k$

Until there are no changes in any mean

{ 
  Use the estimated means to classify the examples into clusters

  For $i$ from 1 to $k$
    Replace $m_i$ with the mean of all of the examples for Cluster $i$

}
Example: use 2-means to cluster into 2 groups
First approximation to 2-means
Cluster 1
Second approximation to 2-means
Cluster 2
Third approximation to 2-means
Cluster 3, and stable
Final cluster
Fuzzy k-means

The clusters produced by the k-means procedure are sometimes called "hard" or "crisp" clusters, since any feature vector $x$ either is or is not a member of a particular cluster.

This is in contrast to "soft" or "fuzzy" clusters, in which a feature vector $x$ can have a degree of membership in each cluster.

The fuzzy-k-means procedure allows each feature vector $x$ to have a degree of membership in Cluster $i$. 
Fuzzy k-means

Make initial guesses for the means \( m_1, m_2, \ldots, m_k \)

Until there are no changes in any mean:

\[
\begin{align*}
\text{Use the estimated means to find the degree of membership } u(j,i) \\
\text{of } x_j \text{ in Cluster } i; \\
\text{for example, if } a(j,i) = \exp(-|x_j - m_i|^2), \\
\text{we might use } u(j,i) = a(j,i) / \text{sum}_j a(j,i) \\
\text{For } i \text{ from 1 to } k \\
\text{Replace } m_i \text{ with the fuzzy mean of all of the examples for Cluster } i - \\
\end{align*}
\]

\[
m_i = \frac{\sum_{j} u(j,i)^2 x_j}{\sum_{j} u(j,i)^2}
\]
Mid-term test 1 (5 mins)

Without looking at your notes, list (with brief definitions) as many components of uncertainty as you can in 10 minutes.
“Go to the first major intersection, take a hard right, merge onto the highway, take any of the next three exits, keep going for 1.5 miles until you are in the mountains”.

(Reality check: The mountains are less than a mile away).

Analyze this sentence in terms of the uncertainty it contains.
We have an area of land, which we have partitioned into 100 quadrants. In each area we have counted the numbers of trees of species Oak, Pine, and Fir.

Suppose that we are interested in classifying each quadrant into one of two land-types (as yet unspecified).

Briefly outline how you would fuzzy cluster analysis to determine possible land types and classify the quadrants.
Rough sets
# Decision under uncertainty

## Land suitability information system

<table>
<thead>
<tr>
<th>SiteID</th>
<th>Land use</th>
<th>Height</th>
<th>Temperature</th>
<th>Cost</th>
<th>Suitability</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>Urban</td>
<td>Low</td>
<td>Hot</td>
<td>Low</td>
<td>Yes</td>
</tr>
<tr>
<td>E2</td>
<td>Urban</td>
<td>High</td>
<td>Hot</td>
<td>Low</td>
<td>Yes</td>
</tr>
<tr>
<td>E3</td>
<td>Urban</td>
<td>Low</td>
<td>Cold</td>
<td>Low</td>
<td>Yes</td>
</tr>
<tr>
<td>E4</td>
<td>Urban</td>
<td>High</td>
<td>Cold</td>
<td>High</td>
<td>No</td>
</tr>
<tr>
<td>E5</td>
<td>Rural</td>
<td>Low</td>
<td>Hot</td>
<td>Low</td>
<td>Yes</td>
</tr>
<tr>
<td>E6</td>
<td>Rural</td>
<td>High</td>
<td>Cold</td>
<td>Low</td>
<td>No</td>
</tr>
<tr>
<td>E7</td>
<td>Rural</td>
<td>High</td>
<td>Cold</td>
<td>High</td>
<td>No</td>
</tr>
</tbody>
</table>
Representative Readings


Introduction to rough set theory

Theory originated at Warsaw by Pawlak in the 70s. The theory deals with the classification and analysis of data drawn from imprecise observations.

Primary notions

- approximation space
- lower and upper approximations to a set

The approximation space corresponds to our knowledge frame.

A set (or proposition) is defined by two approximations:

- lower approximation contains elements known definitely to be in the set
- upper approximation contains elements known possibly to be in the set
Indiscernibility

For any observation, assume an indiscernibility relation $\rho$ on set $S$.

Given relation $\rho$ on set $S$ and $s \in S$, define:

$$R(s) = \{ t \in S \mid t \rho s \}$$

If $\rho$ is an equivalence relation, then the collection of sets of the form $R(s)$ for $s \in S$ forms a partition of $S$.

The set of equivalence classes of $S$ with respect to $\rho$ is denoted $S/\rho$. 
Indiscernibility and partitioning

Elements in the same block of the partition are indiscernible from each other.

$S$

$S/\rho$
Upper and lower sets

Given equivalence (assumed now) relation $\rho$ on set $S$, define:

$L(T) = \{x \in S / \rho \mid x \subseteq T\}$

$U(T) = \{x \in S / \rho \mid x \cap T \neq \emptyset\}$

called the lower and upper approximations to set $T$ with respect to $\rho$ on set $S$.

Sometimes, we view the approximations directly in terms of elements of set $S$.

In this case, we define:

$L^*(T) = \bigcup \{x \in S / \rho \mid x \subseteq T\}$

$U^*(T) = \bigcup \{x \in S / \rho \mid x \cap T \neq \emptyset\}$
Upper and lower sets

Note that $L^*(T) \subseteq T \subseteq U^*(T)$.

$L^*(T) = U^*(T)$ (or $L(T) = U(T)$ )

if and only if

set $T$ can be represented precisely (crisply) with respect to $\rho$

Otherwise:

$T$ is represented approximately (roughly) with respect to $\rho$. 
Indiscernibility results in roughness ...
lower and upper sets ...
Information Systems
(example from Komorowski rough sets tutorial, 1998, CiteSeer)

- Information system (informal definition)
  An information system is a data set, represented as a table.
  - Each row represents a case (entity, object, thing, event, person)
  - Each column represents an attribute (variable, property, measurement)

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>LEMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>16-30</td>
<td>50</td>
</tr>
<tr>
<td>X2</td>
<td>16-30</td>
<td>0</td>
</tr>
<tr>
<td>X3</td>
<td>31-45</td>
<td>1-25</td>
</tr>
<tr>
<td>X4</td>
<td>31-45</td>
<td>1-25</td>
</tr>
<tr>
<td>X5</td>
<td>46-60</td>
<td>26-49</td>
</tr>
<tr>
<td>X6</td>
<td>16-30</td>
<td>26-49</td>
</tr>
<tr>
<td>X7</td>
<td>46-60</td>
<td>26-49</td>
</tr>
</tbody>
</table>

- Information system (formal definition)
  A pair $I = (U,A)$, where:
  $U$: A finite, non-empty set of objects, called the universe
  $A$: A finite, non-empty set of properties

in the above table,

$U = \{X1, \ldots, X7\}$

$A = \{\text{Age, LEMS}\}$
A decision system is an information system in which the values of a special decision attribute classify the cases.

Decision system (formal definition)
A DS of the form $D = (U, A \cup \{d\})$, where:

- $U \rightarrow$ A finite, non-empty set of objects, called the universe
- $A \rightarrow$ A finite, non-empty set of conditional attributes, or conditions
- $d \rightarrow$ The decision attribute

In the above table,

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>LEMS</th>
<th>Walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>16-30</td>
<td>50</td>
<td>Yes</td>
</tr>
<tr>
<td>$X_2$</td>
<td>16-30</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>$X_3$</td>
<td>31-45</td>
<td>1-25</td>
<td>No</td>
</tr>
<tr>
<td>$X_4$</td>
<td>31-45</td>
<td>1-25</td>
<td>Yes</td>
</tr>
<tr>
<td>$X_5$</td>
<td>46-60</td>
<td>26-49</td>
<td>No</td>
</tr>
<tr>
<td>$X_6$</td>
<td>16-30</td>
<td>26-49</td>
<td>Yes</td>
</tr>
<tr>
<td>$X_7$</td>
<td>46-60</td>
<td>26-49</td>
<td>No</td>
</tr>
</tbody>
</table>
Rules and redundancy in Decision Tables

Values of conditional attributes in a decision table may be used to classify cases.

A decision table is the extension of a rule base leading from properties of the cases to the decision.

A decision table may be redundant in at least two ways:

- The same type of case may be represented several times (unnecessary for the creation of new rules)
- Some of the attributes may be superfluous (unnecessary for the classification purposes)
Indiscernibility Relation

Formal definition:

If $I = (U, A)$, is an information system, then with any subset $B$ of $A$ there is associated an indiscernibility relation on $U$:

$$IND_A(B) = \{(x, x') \in U^2 \mid \forall a \in B \ a(x) = a(x')\}$$

This indiscernibility relation is an equivalence relation.
Indiscernibility Example

The equivalence classes of the $B$-indiscernibility relation are:

- if $B=\emptyset$ then:
  \[
  \text{IND (}\emptyset\text{)}=\{[X1, X2, X6, X3, X4, X5, X7]\}
  \]

- if $B=\{\text{Age}\}$ then:
  \[
  \text{IND (}\{\text{Age}\}\text{)}=\{[X1, X2, X6], [X3, X4], [X5, X7]\}
  \]

- if $B=\{\text{LEMS}\}$ then:
  \[
  \text{IND (}\{\text{LEMS}\}\text{)}=\{[X1], [X2], [X3, X4], [X5, X6, X7]\}
  \]

- if $B=\{\text{Age}, \text{LEMS}\}$ then:
  \[
  \text{IND (}\{\text{Age, LEMS}\}\text{)}=\{[X1], [X2], [X3, X4], [X5, X7], [X6]\}
  \]
**Set Approximation**

- Partitions of interest are those where the value of the decision attribute is the same for all cases in the subset.
- Take the set of walking patients.
- It cannot be defined with certainty using the available attributes.

However, we can define the following:
- The set of patients that definitely walk.
- The set of patients that definitely do not walk.
- The set of patients for which we do not know whether they walk or not.
Set Approximation

- The *lower approximation* of the set contains the objects that can be classified with certainty as members of $X$ based on the knowledge of the conditions.

- The *boundary region* contains the objects that we cannot decisively classify into $X$ based on knowledge of the conditions.

- The *upper approximation* is the sum of the lower approximation and the boundary. It contains the certain and possible members of $X$.

- The *outside region* of $X$ consists of those objects that can be with certainty classified as not belonging to $X$ on the basis of knowledge from the conditions.

- A set is *rough* if the boundary region is not empty, otherwise it is *crisp*.
Categories of Rough Sets

- There are four categories of rough sets (four categories of vagueness)

- X is roughly B-definable iff:
  \[ \overline{BX} \neq 0 \text{ and } \overline{BX} \neq U \]

- X is internally B-indefinable iff:
  \[ \overline{BX} = 0 \text{ and } \overline{BX} \neq U \]

- X is externally B-indefinable iff:
  \[ \overline{BX} \neq 0 \text{ and } \overline{BX} = U \]

- X is totally B-indefinable iff:
  \[ \overline{BX} = 0 \text{ and } \overline{BX} = U \]
Accuracy of Approximation

- Rough sets can be characterized numerically by a measure called *accuracy of approximation*

\[ a_B(X) = \frac{|B(X)|}{|\overline{B}(X)|} \]

- If the coefficient is 1, then the set is crisp

- If the coefficient is <1, then the set is rough with respect to B
Next Step in Data Reduction

A decision table may be redundant in at least two ways:

- The same type of case may be represented several times (unnecessary for the creation of new rules)

- Some of the attributes may be superfluous (unnecessary for the classification purposes)
Formal Definition of Reducts

Given an information system $I = (U, A)$, a reduct is a minimal subset $B$ of attributes such that:

$$IND_A(B) = IND_A(A)$$
Discernibility Matrix

- A discernibility matrix of an information system $A=(U,A)$ is a symmetric $n \times n$ matrix with entries $c_{ij}$, such that:

$$c_{ij} = \{ a \in A \mid a(x_i) \neq a(x_j) \}$$

for $i,j$ between 1 and $n$

where $U = \{ x_i \mid i = 1, \ldots, n \}$
## Deriving Reducts

**An example:**

<table>
<thead>
<tr>
<th></th>
<th>Diploma</th>
<th>Experience</th>
<th>French</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>MBA</td>
<td>Medium</td>
<td>Yes</td>
<td>Excellent</td>
</tr>
<tr>
<td>x2</td>
<td>MBA</td>
<td>Low</td>
<td>Yes</td>
<td>Neutral</td>
</tr>
<tr>
<td>x3</td>
<td>MCE</td>
<td>Low</td>
<td>Yes</td>
<td>Good</td>
</tr>
<tr>
<td>x4</td>
<td>MSc</td>
<td>High</td>
<td>Yes</td>
<td>Neutral</td>
</tr>
<tr>
<td>x5</td>
<td>MSc</td>
<td>Medium</td>
<td>Yes</td>
<td>Neutral</td>
</tr>
<tr>
<td>x6</td>
<td>MSc</td>
<td>High</td>
<td>Yes</td>
<td>Excellent</td>
</tr>
<tr>
<td>x7</td>
<td>MBA</td>
<td>High</td>
<td>No</td>
<td>Good</td>
</tr>
<tr>
<td>x8</td>
<td>MCE</td>
<td>Low</td>
<td>No</td>
<td>Excellent</td>
</tr>
</tbody>
</table>
Discernibility Matrix Example

An example:

<table>
<thead>
<tr>
<th></th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
<th>x5</th>
<th>x6</th>
<th>x7</th>
<th>x8</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x2</td>
<td>er</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x3</td>
<td>der</td>
<td>dr</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x4</td>
<td>der</td>
<td>de</td>
<td>der</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x5</td>
<td>dr</td>
<td>de</td>
<td>der</td>
<td>e</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x6</td>
<td>de</td>
<td>der</td>
<td>der</td>
<td>r</td>
<td>er</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x7</td>
<td>efr</td>
<td>efr</td>
<td>def</td>
<td>der</td>
<td>defr</td>
<td>dfr</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>x8</td>
<td>def</td>
<td>dfr</td>
<td>fr</td>
<td>def</td>
<td>defr</td>
<td>defr</td>
<td>def</td>
<td>der</td>
</tr>
</tbody>
</table>
Discernibility Function

- Having generated the discernibility matrix, we can now define the discernibility function.

- The discernibility function allows us to discern all objects from each other and is formally defined as:

$$f_A(a_1^*, \ldots a_m^*) = \bigwedge \{ \bigvee c_{ij}^* \mid 1 \leq j \leq i \leq n, c_{ij}^* \neq 0 \}$$

- Take all the entries in the discernibility matrix as a disjunction and combine them as conjunction.
## Discernibility Function Example

$$f_{A}(d,e,f,r) = (e \lor r) \land (d \lor e \lor r) \land (d \lor e \lor r) \land (d \lor r) \land (e \lor f \lor r) \land (d \lor e \lor f)$$

$$\land (d \lor r) \land (d \lor e) \land (d \lor e \lor r) \land (e \lor f \lor r) \land (d \lor f \lor r)$$

$$\land (d \lor e \lor r) \land (d \lor e \lor r) \land (d \lor e \lor r) \land (d \lor e \lor f) \land (f \lor r)$$

$$\land (e) \land (r) \land (d \lor f \lor r) \land (d \lor e \lor f \lor r)$$

$$\land (e \lor r) \land (d \lor e \lor f \lor r) \land (d \lor e \lor f \lor r)$$

$$\land (d \lor f \lor r) \land (d \lor e \lor f)$$

$$\land (d \lor e \lor r)$$

<table>
<thead>
<tr>
<th></th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
<th>x5</th>
<th>x6</th>
<th>x7</th>
<th>x8</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x2</td>
<td>er</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x3</td>
<td>dr</td>
<td>dr</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x4</td>
<td>dr</td>
<td>de</td>
<td>dr</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x5</td>
<td>dr</td>
<td>de</td>
<td>dr</td>
<td>er</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x6</td>
<td>de</td>
<td>dr</td>
<td>dr</td>
<td>dr</td>
<td>er</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x7</td>
<td>erf</td>
<td>erf</td>
<td>def</td>
<td>der</td>
<td>defr</td>
<td>dfr</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>x8</td>
<td>def</td>
<td>dfr</td>
<td>fr</td>
<td>defr</td>
<td>defr</td>
<td>defr</td>
<td>def</td>
<td>dr</td>
</tr>
</tbody>
</table>
## Deriving the Reduct

After simplification this function gives:

\[ f_A(d, e, f, r) = e \land r \]

<table>
<thead>
<tr>
<th></th>
<th>Diploma</th>
<th>Experience</th>
<th>French</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>MBA</td>
<td>Medium</td>
<td>Yes</td>
<td>Excellent</td>
</tr>
<tr>
<td>x2</td>
<td>MBA</td>
<td>Low</td>
<td>Yes</td>
<td>Neutral</td>
</tr>
<tr>
<td>x3</td>
<td>MCE</td>
<td>Low</td>
<td>Yes</td>
<td>Good</td>
</tr>
<tr>
<td>x4</td>
<td>MSc</td>
<td>High</td>
<td>Yes</td>
<td>Neutral</td>
</tr>
<tr>
<td>x5</td>
<td>MSc</td>
<td>Medium</td>
<td>Yes</td>
<td>Neutral</td>
</tr>
<tr>
<td>x6</td>
<td>MSc</td>
<td>High</td>
<td>Yes</td>
<td>Excellent</td>
</tr>
<tr>
<td>x7</td>
<td>MBA</td>
<td>High</td>
<td>No</td>
<td>Good</td>
</tr>
<tr>
<td>x8</td>
<td>MCE</td>
<td>Low</td>
<td>No</td>
<td>Excellent</td>
</tr>
</tbody>
</table>
Discernibility Matrix in the case of a decision attribute

- A discernibility matrix of an information system $A=(U, A)$ is a symmetric $n \times n$ matrix with entries $c_{ij}$, such that:
  
  $$c_{ij} = \{ a \in A \mid a(x_i) \neq a(x_j) \}$$

for $i, j$ between 1 and $n$, when $x_i$ and $x_j$ have different values on the decision attribute.

- Otherwise $c_{ij} = T$
## Land suitability information system

<table>
<thead>
<tr>
<th>SiteID</th>
<th>Land use</th>
<th>Height</th>
<th>Temperature</th>
<th>Cost</th>
<th>Suitability</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>Urban</td>
<td>Low</td>
<td>Hot</td>
<td>Low</td>
<td>Yes</td>
</tr>
<tr>
<td>E2</td>
<td>Urban</td>
<td>High</td>
<td>Hot</td>
<td>Low</td>
<td>Yes</td>
</tr>
<tr>
<td>E3</td>
<td>Urban</td>
<td>Low</td>
<td>Cold</td>
<td>Low</td>
<td>Yes</td>
</tr>
<tr>
<td>E4</td>
<td>Urban</td>
<td>High</td>
<td>Cold</td>
<td>High</td>
<td>No</td>
</tr>
<tr>
<td>E5</td>
<td>Rural</td>
<td>Low</td>
<td>Hot</td>
<td>Low</td>
<td>Yes</td>
</tr>
<tr>
<td>E6</td>
<td>Rural</td>
<td>High</td>
<td>Cold</td>
<td>Low</td>
<td>No</td>
</tr>
<tr>
<td>E7</td>
<td>Rural</td>
<td>High</td>
<td>Cold</td>
<td>High</td>
<td>No</td>
</tr>
</tbody>
</table>
Properties of rough sets

(1) $\underline{B}(X) \subseteq X \subseteq \overline{B}(X)$,
(2) $\underline{B}(\emptyset) = \overline{B}(\emptyset) = \emptyset$, $\underline{B}(U) = \overline{B}(U) = U$,
(3) $\underline{B}(X \cup Y) = \overline{B}(X) \cup \overline{B}(Y)$,
(4) $\underline{B}(X \cap Y) = \underline{B}(X) \cap \underline{B}(Y)$,
(5) $X \subseteq Y$ implies $\underline{B}(X) \subseteq \underline{B}(Y)$ and $\overline{B}(X) \subseteq \overline{B}(Y)$,
(6) $\underline{B}(X \cup Y) \supseteq \underline{B}(X) \cup \underline{B}(Y)$,
(7) $\overline{B}(X \cap Y) \subseteq \overline{B}(X) \cap \overline{B}(Y)$,
(8) $\underline{B}(-X) = -\overline{B}(X)$,
(9) $\overline{B}(-X) = -\underline{B}(X)$,
(10) $\underline{B}(\underline{B}(X)) = \overline{B}(\underline{B}(X)) = \underline{B}(X)$,
(11) $\overline{B}(\overline{B}(X)) = \underline{B}(\overline{B}(X)) = \overline{B}(X)$,
Rough membership

The rough membership function indicates the degree of membership of an element in a set. Let $X$ be a (classical) set, and $[x]$ the equivalence class of $x$ under the granularity relation $B$. Then:

$$\mu^B_X : U \rightarrow [0, 1] \text{ and } \mu^B_X(x) = \frac{|[x]_B \cap X|}{|[x]_B|}$$
Spatial Roughness

Rough subsets based on the granularity
Properties of rough regions

- Not connected
- Not connected
- Not connected
- Maybe connected
Properties of vague regions

<table>
<thead>
<tr>
<th>not connected</th>
<th>maybe connected</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Vague spatial relationships

How can we speak about the spatial relationship between $<A_1,B_1>$ and $<A_2,B_2>$?
Integrating rough observations: frames
Integrating rough observations: rough sets
Lifestyle generalization

[Diagram]

SIE 565
General applications of rough sets

imprecise knowledge representation
multiresolution data modelling
conflict analysis
non-invasive data analysis
identification of data dependencies
information-preserving data reduction