Belief Functions for the Working Scientist

Joint CCT DSERC and ISERC Meeting

Dr Fabio Cuzzolin

Artificial Intelligence and Vision research group

October 28, 2015

Current and Future Projects

- Mathematics of Uncertainty theme
  - UAI 2015 Tutorial https://www.youtube.com/watch?v=nhGzn0R5TgM
  - The geometry of uncertainty -> upcoming Springer monograph
  - Survey of 50 years of belief functions -> upcoming IEEE Fuzzy Systems survey paper(s) (with Cigdem Sengul)
  - The Total Belief Theorem
  - Pose estimation using Belief Modelling Regression -> IEEE TFS (with Wenjuan Gong)
  - Climate change predictions (with Oxford's Jim Hall?)
  - Generalisation of random variables: Belief variables
  - Belief function logic for the semantic web (with Oxford's Thomas Lukasiewicz)
  - Uncertainty-theoretic learning theory (with Cambridge's Adrian Weller)
  - Random Set Random Trees/Forests for multi-label classification (with MS India?)

Belief Functions as Random Sets

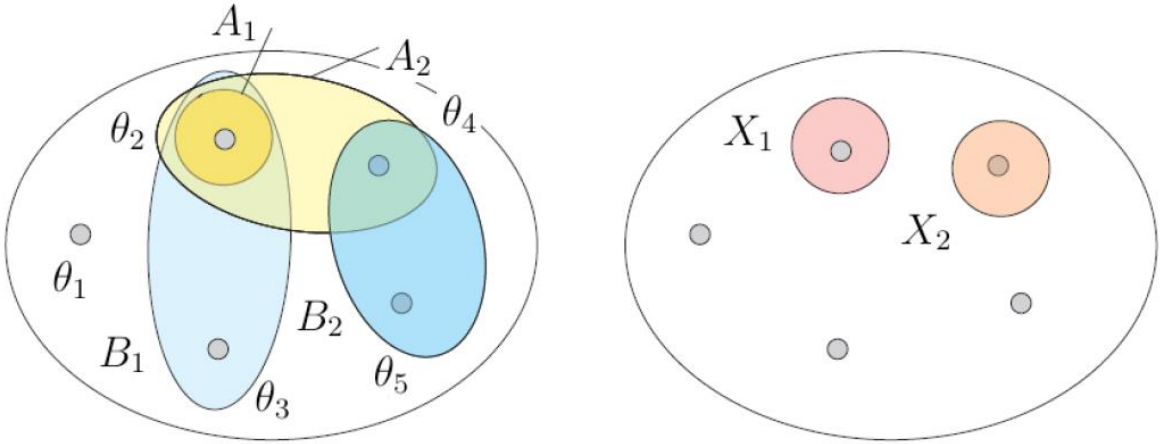
A strong rationale: missing or insufficient data

- mathematically, belief functions can be seen as random sets
- a random set is a set-valued random variable, which for certain outcomes of the probability space (P,F,Ω) produce an entire set of possible values

- example: the cloaked dice: if you cover three faces of a dice and you roll it, when you get a covered face you can only say that one of the three non-visible faces has come up
- hence, you assign prob values (e.g. 1/6) to sets of outcomes

BELIEF FUNCTIONS
AS GENERALISED PROBABILITIES

- Probabilities do not represent well ignorance and lack of data
- Evidence is normally limited, rather than infinite as assumed by probability
  Sometimes expert knowledge needs to be combined with hard evidence
- In extreme cases (rare events or far-future predictions) very little data
- Belief functions proposed by Dempster and Shafer in late Sixties
- Idea: evidence is typically available in support of events/propositions directly, and can be translated into a distribution over all events (a random set)
- Can be fused by Dempster's rule, a generalisation of Bayes' rule



UAI (Uncertainty in Artificial Intelligence) 2015

tutorial: belief functions for the working scientist

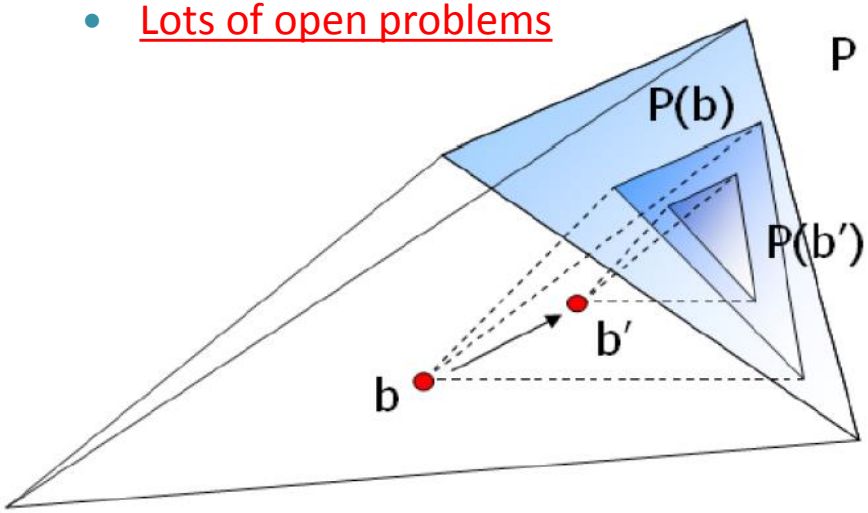
UAI 2015 Amsterdam Tutorial: Belief Functions for the Working Scientist

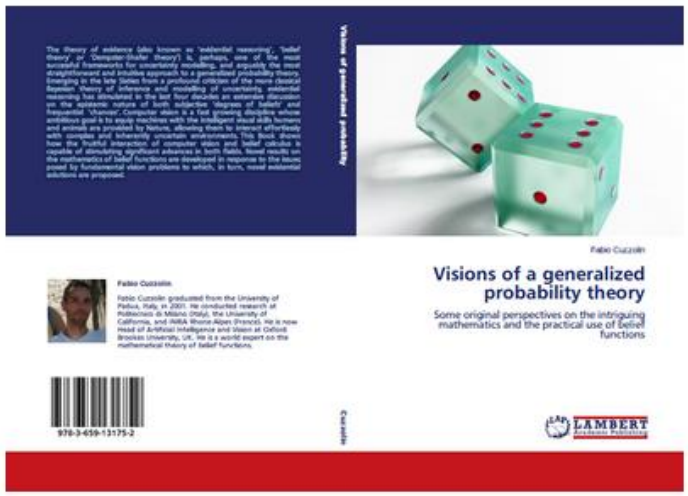
- part of a campaign to embed uncertainty theory and random set theory in the wider Artificial Intelligence community
- similar tutorials proposed at the top upcoming AI conferences, AAAI, IJCAI, KR, ECAI

The Geometry of Uncertainty

Treating uncertainty measures as geometric entities

- Belief functions can be seen as a point of a simplex
- Probabilities are part of the border of this simples
- Possibility and fuzzy measures too -> unified geometric approach to uncertainty
- Problems can be approaches there:
  - Conditioning belief functions (IEEE TFS to revise)
  - Decision making with belief functions (AIJ to revise)
  - Lots of open problems





UPCOMING SURVEY PAPERS ON 50 YEARS OF BFS

THEORY AND APPLICATIONS

takes place on a testing database. Data fusion leads to a significant improvement of classification performances with respect to the actual system.
Cited by 15

Modeling, combining, and discounting mine detection sensors within the Dempster-Shafer framework
N Milisavljevic, I Bloch ... - AeroSense ..., 2000 - proceedings.spiedigitallibrary.org
abstract In this paper, ideas for modeling humanitarian mine detection sensors and their combination within Dempster-Shafer framework are presented. Reasons for choosing this framework are pointed out, taking into account specificity and sensitivity of the problem. ...
Cited by 12

A verified Matlab toolbox for the Dempster-Shafer theory
E Auer, W Luther, G Rebner and P Limbourg - BELIEF 2010
The importance of the Dempster-Shafer theory (DST) for modeling and propagating uncertainty has grown in the recent past. An obstacle for wider application of this theory in industrial practice is the lack of software support for analysts. The few tools available depend on floating point arithmetic and do not consider the inherently interval-based nature of the DST to the full extent. Therefore, the obvious next step is to combine the DST ideas with those from interval arithmetic. An additional advantage of employing interval methods is the guarantee that the results obtained on a computer are mathematically correct.
In this paper, we introduce a new verified DST implementation for MATLAB based on the previously developed IPP TOOLBOX. It extends this software using interval arithmetic and simultaneously takes care of the rounding errors. After giving a short overview of the Dempster-Shafer theory and interval methods, we describe the main features of the new toolbox and show its potential using several examples.
Cited by 13

Using the Dempster–Shafer theory of evidence with a revised lattice structure for activity recognition
J Liao, Y Bi, C Nugent - Information Technology in Biomedicine, ..., 2011 - ieeexplore.ieee.org
Abstract—This paper explores a sensor fusion method applied within smart homes used for the purposes of monitoring human activities in addition to managing uncertainty in sensor-based readings. A three-layer lattice structure has been proposed, which can be used to ...
Cited by 17

A modified Dempster-Shafer theory for multicriteria optimization
L Chen, SS Rao - Engineering optimization, 1998 - Taylor & Francis
Abstract A new methodology, based on a modified Dempster-Shafer (DS) theory, is proposed for solving multicriteria design optimization problems. It is well known that considerable amount of computational information is acquired during the iterative process ...
Cited by 11

Feature extraction using rough set theory and genetic algorithms—an application for the simplification of product quality evaluation
LY Zhai, LP Khoo, SC Fok - Computers & Industrial Engineering, 2002 - Elsevier
... Many theories, techniques and algorithms have been developed to deal with the analysis of imprecise or inconsistent data. The most successful ones are based on fuzzy set theory and the Dempster–Shafer theory of evidence. ...
Cited by 111

The Dempster-Shafer combination rule as a tool to classifier combination
MR Ahmadzadeh, M Petrou ... - Geoscience and Remote ..., 2000 - ieeexplore.ieee.org
ABSTRACT In this paper we present Dempster-Shafer theory as a framework within which the results of a Bayesian and a fuzzy classifier can be combined to produce a better final classification. We deal with the case when the two original classifiers use different classes ...
Cited by 9

Combining fingerprint and voiceprint biometrics for identity verification: an experimental comparison
Y Wang, Y Wang, T Tan - Biometric Authentication, 2004 - Springer

... The experimental results show that Support Vector Machine and the Dempster-Shafer method are superior to other schemes. 1 Introduction The emergency of biometrics helps to solve the problems that the traditional methods such as password and IC cards have faced. ...
Cited by 78

Extended Dempster-Shafer combination rules based on random set theory
Y Zhu, XR Li - Defense and Security, 2004 - proceedings.spiedigitallibrary.org
abstract The Dempster combination rule has been widely discussed and used since it is a convenient and promising method to combine multi-source information with their own confidence degrees/evidences. On the other hand, it has been criticized and debated ...
Cited by 12

Assessing dependability of safety critical systems using diverse evidence
N Fenton, B Littlewood, M Neil, L Strigini ... - Software, IEE ..., 1998 - ieeexplore.ieee.org
... work. 2 modelling uncertainty Our first task was to examine in depth the various methods and technologies for modelling uncertainty, including Bayesian probability, Dempster-Shafer the-ory, fuzzy sets and possibility theory. In ...
Cited by 73

The Application of Dempster Shafer Theory to a Logic-Based Visual Recognition System
GM Provan - Proceedings of the Fifth Annual Conference on ..., 1990 - dl.acm.org
The Application of Dempster Shafer Theory to a Logic-Based Visual Recognition System. Author: Gregory M. Provan, Published in: - Proceeding, UAI '89 Proceedings of the Fifth Annual Conference on Uncertainty in Artificial Intelligence. ...
Cited by 11

A few remarks on measures of uncertainty in Dempster-Shafer theory
J VEJNAROVÁ - International Journal Of General System, 1993 - Taylor & Francis
It is shown, through two simple examples, that neither the measure of discord nor the measure of total uncertainty used in [Klir, 1991] satisfy one of the fundamental requirements generally taken as necessary for a meaningful measure of information (or uncertainty): the ...
Cited by 10

Reliability assessment and optimization under uncertainty in the Dempster-Shafer framework
P Limbourg, D Germann - 27th ESReDA seminar, Glasgow, GB, 2004
Epistemic uncertainty has been a subject heavily discussed in recent years. Where uncertainty of estimations for long time has been disregarded, it is now viewed as an inherent property of the system model. This work demonstrates how different data types like expert estimations and manufacturers data with various degrees of epistemic uncertainty are acquired and aggregated in a coherent framework. Using this model we show how to select between design alternatives using multi-objective optimization.
A Pareto-based evolutionary algorithm is applied to find a choice of non-dominated solutions. A new strategy for biasing the search to desired objectives without loosing diversity is presented. The user can select a posteriori between solutions covering a wide range of the objective space but clustering in the specified area.
Cited by 11

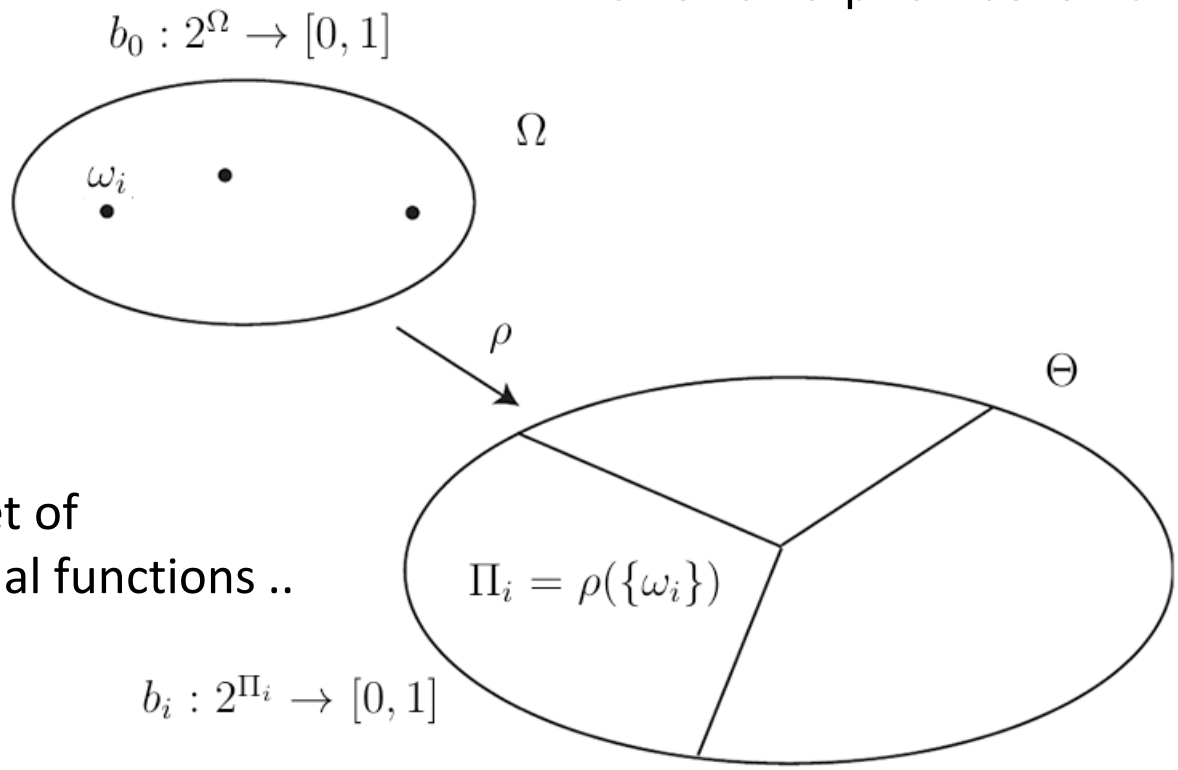
Combining classifiers for word sense disambiguation based on Dempster–Shafer theory and OWA operators
CA Le, VN Huynh, A Shimazu, Y Nakamori - Data & Knowledge Engineering 63, Issue 2, November, 2007, Pages 381-396
In this paper, we discuss a framework for weighted combination of classifiers for word sense disambiguation (WSD). This framework is essentially based on Dempster-Shafer theory of evidence (Shafer, 1976) and ordered weighted averaging (OWA) operators (Yager, 1988). We first determine various kinds of features which could provide complementarily linguistic information for the context, and then combine these sources of information based on Dempsters rule of combination and OWA operators for identifying the meaning of a polysemous word. We experimentally design a set of individual classifiers, each of which corresponds to a distinct representation type of context considered in the WSD literature, and then the discussed combination strategies are tested and compared on English lexical samples of Senseval-2 and Senseval-3.
Cited by 13

- Cigdem has been very helpful with compiling a list of 3000+ abstracts, collecting PDF and assembling two survey papers

The Total Belief Theorem

The generalisation of the total probability theorem

- Given an a-priori belief function..

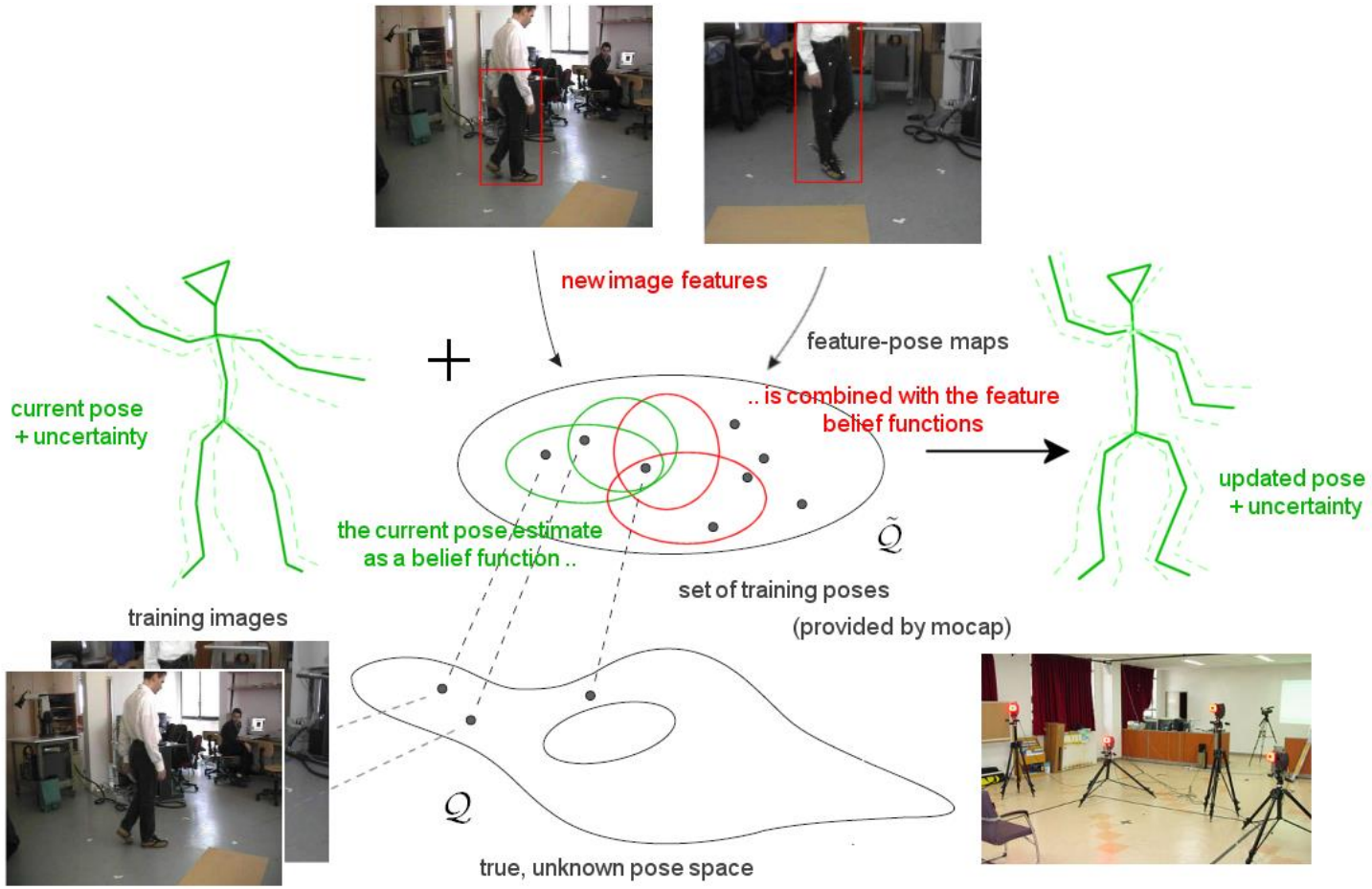


- ..and a set of conditional functions ..

- .. find the total belief function(s) meeting these constraints
- Fundamental methodological contribution

Belief Modelling Regression

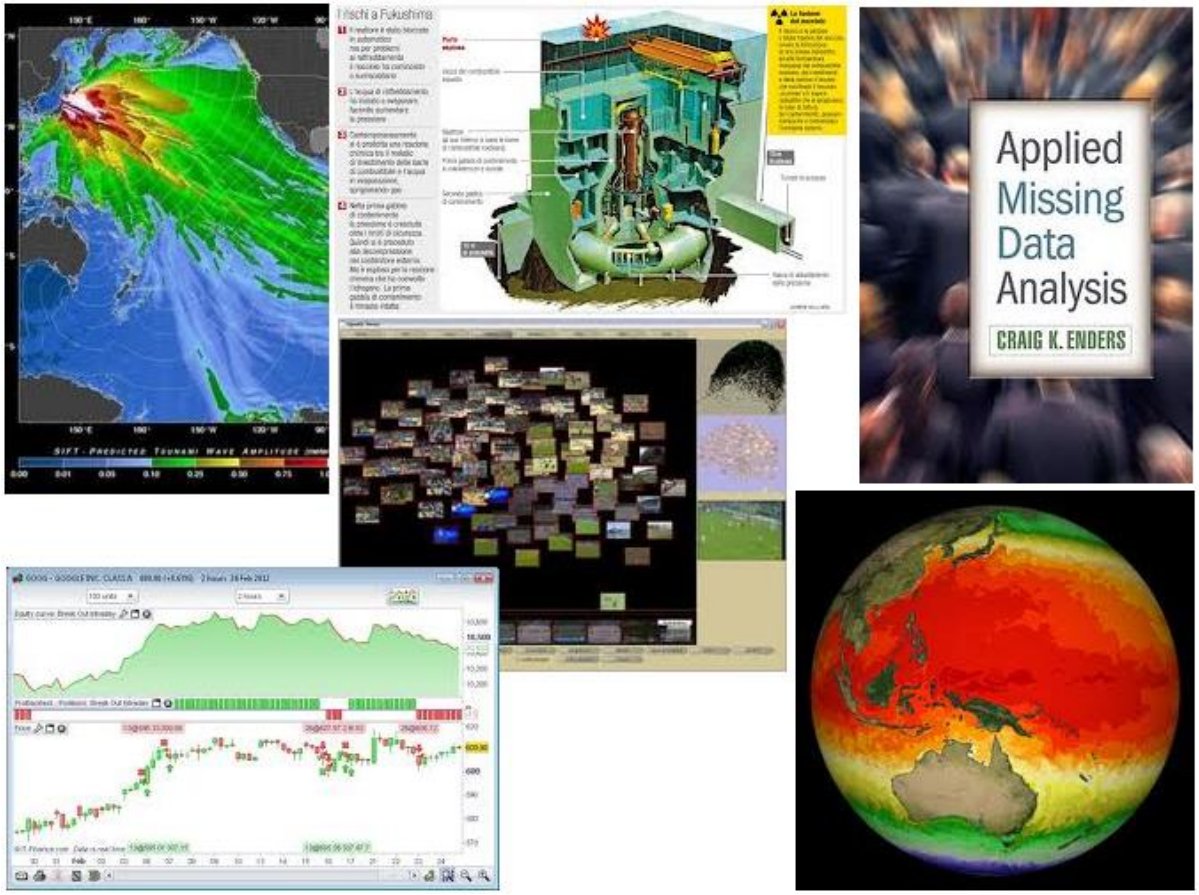
for example-based pose estimation



- IEEE Transaction on Fuzzy Systems paper under revision with W. Gong

Modelling Climatic Change

in the framework of the theory of belief functions



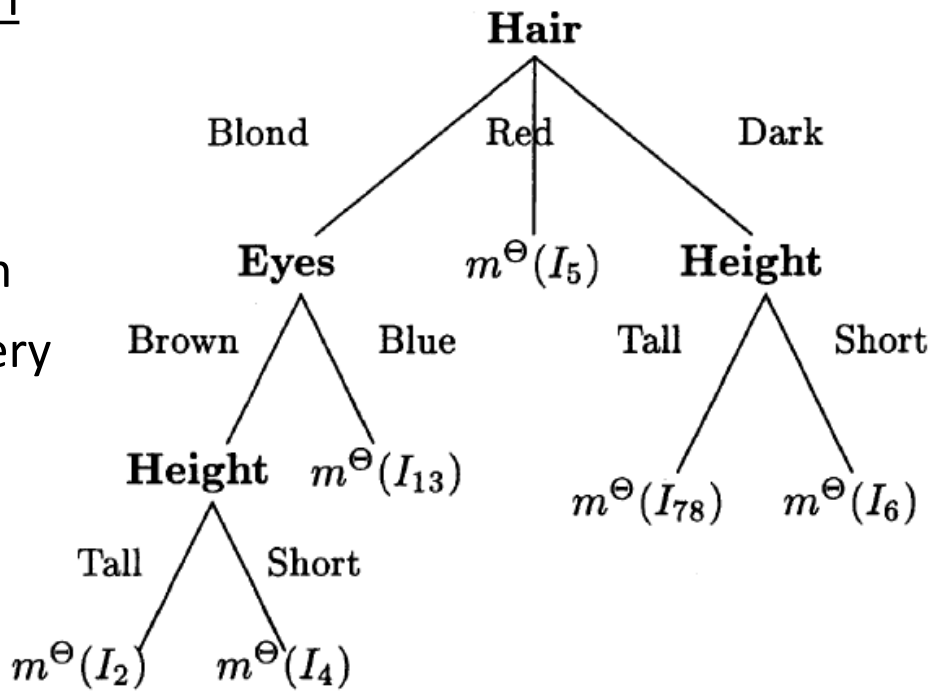
- possible collaboration with Oxford's Jim Hall, lots of work to do

Belief Functions Logic
for the Semantic Web

- Uncertainty: statements are true or false, but due to lack of knowledge we can only estimate to which probability degree they are true or false, e.g., "John wins in the lottery with the probability 0.01"
- Vagueness: statements involve concepts for which there is no exact definition, such as tall, small, close, far, cheap, and expensive; statements are true to some degree, e.g., "Hotel Verdi is close to the train station to degree 0.83" (related to fuzziness)
- uncertainty and vagueness are crucial in the Semantic Web
- many existing proposals for extensions of SW languages (RDF, OWL, DLs, rules) by uncertainty and vagueness (Thomas Lukasiewicz)
  - http://cms.brookes.ac.uk/staff/FabioCuzzolin/BELIEF2014/files/Lukasiewicz.pdf
- as belief theory subsumes both fuzzy and probability theory, why not attach belief values to propositions?
- various belief logic frameworks have been proposed

Random Sets Random Trees
for multi-label classification

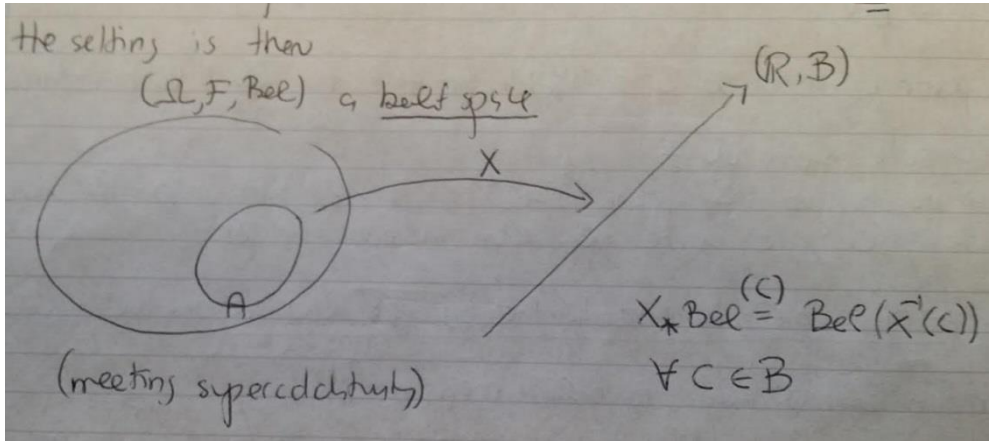
- In many classification problems items are denoted by more than one label
- Example: when browsing for images one want to use more then one label to describe them
- Computational requirements very strict -> random forest are fast and effective
- However, need to split nodes according to entropy of sets of label values
- Belief decision trees
  - Interest from Microsoft Research India on this topic



Belief Random Variables and Processes

Generalising the Radon-Nikodym derivative

- Belief functions/random sets are nothing but set-valued random variables
- (one-to-many maps from a measurable space to another domain)
- can we instead <u>generalise the notion of random variable</u>, as a mapping from a "belief space" (a set with an attached belief measure) to R^N ?



- can we define the <u>analogous of a Cumulative Distribution Function</u>?
- how is this related to p-boxes (lower and upper bounds to CDFs)?
- to define density we need a <u>generalisation of Radon-Nikodym derivative</u> to monotone capacities (related work by Molchanov)

Towards Uncertainty-Theoretic Learning Theory
A new foundation for modern machine learning?

- in mainstream ML, focus on fitting the observable outputs in the training data
- 'overfitting' -> e.g., an autonomous driving system performs well on validation but fails catastrophically in the real world (a radically new setting)
- 'imprecise' probabilities arise when evidence is insufficient to estimate distributions
- training sets constitute a glaring example of data insufficient in quantity and quality
- classical statistical learning theory (Vapnik) contemplates 'generalisation' criteria which are based on a naïve correlation between smoothness and generality …
- ..or makes predictions on the reliability of a training set based on simple quantities such as number of samples (no description of quality of representation)
- uncertainty theory can provide new insights on the matter by providing worst-case, cautious predictions, and produce AI agents aware of their own limitations
- will start from a re-examination of the "free lunch theorem" ("a general-purpose universal optimization strategy is theoretically impossible"), and a generalisation of the concept of "Probably Approximately Correct", i.e. the ability of an algorithm of finding with probability at least $(1-\delta)$ the correct answer with error no more than ε.

Some more speculative ideas
but somewhat fascinating

- an imprecise probabilistic foundation for quantum mechanics
  - there has been very limited work in this direction
  - fundamental imprecision on the wave function?
- representing random sets using exterior algebras
  - Seems to be a more correct geometric representation of set-valued functions
- belief functions and gambles (from imprecise probability)
  - they can be extended in order to assign masses to any continuous functions on the domain
  - (events are just indicator functions)
  - can this generalise imprecise probabilities too?