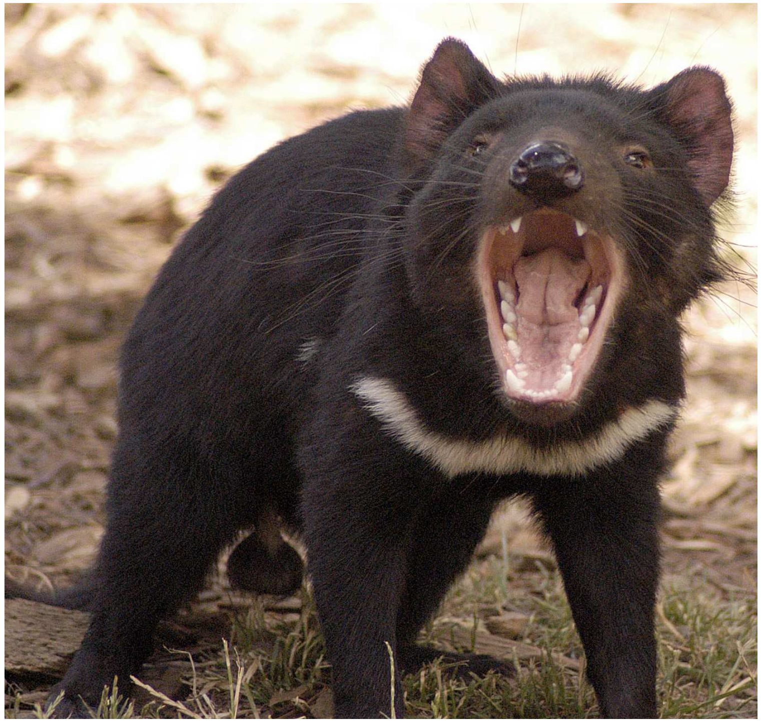
Introduction to fuzzy systems and aggregation operators

Gleb Beliakov

Deakin University, Melbourne, Australia

gleb@deakin.edu.au



Outline

Fuzzy sets

- Characteristic function $\rightarrow$ membership function $\mu : X \rightarrow [0, 1]$ which associates with each object $x \in X$ a degree of membership in a set.
- Example: the set of *tall* people.

$$\mu(x) = \begin{cases} 0 & if\ x \leq 160\ cm \\ 1 & if\ x \geq 180\ cm \\ (x - 160)/20 & otherwise \end{cases}$$

Fuzzy sets

- Fuzzy sets allow us to deal with imprecise concepts, where it is natural to assign membership degrees. it is a different kind of uncertainty to randomness, where probability theory works.
- It is the basis for possibility theory
- Example: Hans eats for breakfast x eggs. Probability that Hans eats 1,2,3, etc eggs today. versus Possibility that hans could eat x eggs.

Fuzzy sets

- Fuzzy sets are useful in modelling situations where it is impossible or pointless to assign precise values.
- Patient came to a clinic. Measuring Body mass index $weight/height^2$. Clinical guideline: if the patient is *overweight* then xxx (treatment, consultation).
- Overweight $= BMI > 25$. Do we make a cutoff at 25? What about 24.9 and 25, the difference is about 200gm = 1 glass of water .

Fuzzy sets

- Operations of set union, intersection, complement are defined using max, min and $1 - \mu(x)$ operations. These are the only ones consistent with mutual distributivity.
- Relaxing these condition leads to various theories of multivalued logic.
- max and min are too restrictive in practice and do not coincide with how people perform these operations.
- Experiments: students assigning degrees of membership in fuzzy sets *metallic*, *container* and *metallic container*: mean works better than min.
- This lead to studies of other operations, now called aggregation functions.

Fuzzy rule based systems

- An attempt to model descriptive rules, understood by people.
- If x *is* A *and* y *is* B *then* z *is* C
- If speed is LOW and distance is HIGH then set engine to HIGH
- Here one can control the membership functions of all fuzzy sets and also the logical connectives AND and implication

Fuzzy rule based systems

- The system has several rules
- If x is A_1 and y is B_1 then z is C_1
- If x is A_2 or y is B_2 then z is C_2
- We need to combine the antecedents (model AND, OR), combine the consequents (what is the value of z - need OR and defuzzification) - Mamdani control.
- Sugeno control:
  If x is A_1 and y is B_1 then $z = f(x, y)$

Fuzzy rule based systems

- Example in structural reliability. Rules could be:
- If Cracks are narrow AND cracks are few then wall is good
- If Cracks are wide AND cracks are few then wall is moderately damaged
- If Cracks are wide AND cracks are many then wall is damaged

Aggregation

Aggregation

- *Aggregation functions* play an important role in several areas, including fuzzy logic, decision making, expert systems, risk analysis and image processing.

Aggregation

- *Aggregation functions* play an important role in several areas, including fuzzy logic, decision making, expert systems, risk analysis and image processing.
- Aggregation functions combine input values into a single output value, which represents all the inputs.

Aggregation

- *Aggregation functions* play an important role in several areas, including fuzzy logic, decision making, expert systems, risk analysis and image processing.
- Aggregation functions combine input values into a single output value, which represents all the inputs.
- Typical examples are *weighted means, medians, OWA functions and t-norms/t-conorms.* But there are many other aggregation functions, and infinitely many members in most families.

Aggregation

- *Aggregation functions* play an important role in several areas, including fuzzy logic, decision making, expert systems, risk analysis and image processing.
- Aggregation functions combine input values into a single output value, which represents all the inputs.
- Typical examples are *weighted means, medians, OWA functions and t-norms/t-conorms.* But there are many other aggregation functions, and infinitely many members in most families.
- One question is which particular aggregation function I use in *my* application?

Definitions and classes of AGOPs

Definition 1 *An **aggregation function** is a function of* $n > 1$ *arguments* $f : [0,1]^n \to [0,1]$, *with the properties*

(i) $f(\underbrace{0,0,\ldots,0}_{n-times}) = 0$ *and* $f(\underbrace{1,1,\ldots,1}_{n-times}) = 1.$

(ii) $\mathbf{x} \leq \mathbf{y}$ *implies* $f(\mathbf{x}) \leq f(\mathbf{y})$ *for all* $\mathbf{x}, \mathbf{y} \in [0,1]^n$.

The vector inequality is understood componentwise.

Aggregation functions may possess various properties, which often classify them into special classes.

Classes of aggregation functions

- **Averaging** if it is bounded by

$$\min(\mathbf{x}) = \min_{i=1,\dots,n} x_i \le f(\mathbf{x}) \le \max_{i=1,\dots,n} x_i = \max(\mathbf{x}).$$

- **Conjunctive** if it is bounded by $f(\mathbf{x}) \le \min(\mathbf{x})$.
- **Disjunctive** if it is bounded by $\max(\mathbf{x}) \le f(\mathbf{x})$.
- **Mixed** if it is neither conjunctive, disjunctive or averaging.
- **Idempotent** if $f(t, t, \dots, t) = t$ for any $t \in [0, 1]$.

  Monotonicity and idempotency implies averaging behaviour.
- **Symmetric** (commutative) if $f(\mathbf{x}) = f(\mathbf{x}_P)$ for any $\mathbf{x} \in [0, 1]^n$ and any permutation P of $\{1, \dots, n\}$.

Main Classes: **Averaging functions**

- Weighted arithmetic means $M_{\mathbf{w}}(\mathbf{x}) = \sum_{i=1}^{n} w_i x_i$
- Weighted quasi-arithmetic $M_{\mathbf{w},g}(\mathbf{x}) = g^{-1}\left(\sum_{i=1}^{n} w_i g(x_i)\right)$
- Ordered weighted averaging $OWA_{\mathbf{w}}(\mathbf{x}) = \sum_{i=1}^{n} w_i x_{(i)}$
- Generalised OWA $OWA_{\mathbf{w},g}(\mathbf{x}) = g^{-1}\left(\sum_{i=1}^{n} w_i g(x_{(i)})\right)$
- Other means (identric, logarithmic, Gini, Lermer etc)
- Median, weighted median, quasi-median
- Fuzzy integrals: Choquet, Sugeno, and special cases (WOWA, weighted max/min)

Main Classes: **Conjunctive/disjunctive functions**

Duality: for a strong negation N, $f_N(\mathbf{x}) = N(f(N(\mathbf{x})))$ is called N-dual of f. In particular, standard negation $f_d(x_1 \ldots, x_n) = 1 - f(1 - x_1, \ldots, 1 - x_n)$.

Duals of conjunctive functions are disjunctive.

- Triangular norms and conorms
- Copulas and their duals
- Semi-copulas and Quasi-copulas
- Weighted t-norms and t-conorms
- Functions defined pointwise

Main Classes: **Mixed functions**

- Uninorms. Representable uninorms
  $U_g(\mathbf{x}) = g^{-1}\left(\sum_{i=1}^{n} g(x_i)\right)$
- Nullnorms
- Generated functions with neutral element
- Weighted generated uninorms
- T-S functions $M_{\gamma,g}(T(\mathbf{x}), S(\mathbf{x}))$ where $M_{\gamma,g}$ is a weighted quasi-arithmetic mean with weights $\gamma, 1-\gamma$.
- Symmetric sums
- T-OWA, S-OWA, ST-OWA functions
  $T - OWA_{T,\mathbf{w}}(\mathbf{x}) = \sum_{i=1}^{n} w_i T(x_{(1)}, \ldots, x_{(i)})$

Choosing aggregation functions

We look at two things:

Choosing aggregation functions

We look at two things:

- What are the requirements of the application? This translates into mathematical properties: idempotency, neutral element, symmetry, and so on. In turn, these properties define a class of aggregation functions.

Choosing aggregation functions

We look at two things:

- What are the requirements of the application? This translates into mathematical properties: idempotency, neutral element, symmetry, and so on. In turn, these properties define a class of aggregation functions.
- What is other information? Do we have desired input-outputs? These data allow us to choose specific members of the family, which best fit the data.

Application requirements: properties and data

Application requirements: properties and data

- We look at the problem of choosing an aggregation function.
- Mathematical properties: idempotency or neutral element, symmetry or weights.
- Continuity (and stronger Lipschitz-continuity) are important in applications.

Application requirements: properties and data

- We look at the problem of choosing an aggregation function.
- Mathematical properties: idempotency or neutral element, symmetry or weights.
- Continuity (and stronger Lipschitz-continuity) are important in applications.
- Typically these are very general properties that define very broad classes of functions. Example: we want averaging behaviour. How many means/OWA/fuzzy integrals there are?
- We need some data to choose a specific function.

Application requirements: properties and data

- Experimental data. Desired values.
- Can be tuples $(x_1, \ldots, x_n, y)$, so that we require $f(x_1, \ldots, x_n) \approx y$, or can be just an ordering of outputs $f(\mathbf{x}_1) \geq f(\mathbf{x}_2)$. Can be interval valued inputs and outputs.
- Can be precise: data $(x_1, \ldots, x_n, y)$ need to be interpolated, or approximated: $f(x_1, \ldots, x_n) \approx y$.

Application requirements: properties and data

Table 1: A data set with inputs of varying dimension.

k	n_k	x_1	x_2	x_3	x_4	x_5	y
1	3	x_{11}	x_{21}	x_{31}			y_1
2	2	x_{12}	x_{22}				y_2
3	3	x_{13}	x_{23}	x_{33}			y_3
4	5	x_{14}	x_{24}	x_{34}	x_{44}	x_{54}	y_4
5	4	x_{15}	x_{25}	x_{35}	x_{45}		y_5
$\vdots$							

Application requirements: properties and data

Generic problem formulation

Find an aggregation function from a given class, such that for all data $k = 1, \ldots, K$ $f(\mathbf{x}_k) \approx y_k$.

Minimization problem:

$$\text{minimize } ||\mathbf{r}||$$

$$\text{subject to } f \text{ satisfies properties} \quad \mathcal{P}_1, \mathcal{P}_2, \ldots,$$

where $||\mathbf{r}||$ is the norm of the residuals, i.e., $\mathbf{r} \in R^K$ is the vector of the differences between the predicted and observed values $r_k = f(\mathbf{x}_k) - y_k$.

Application requirements: properties and data

Choice of the norm

- Least squares $||\mathbf{r}||^2 = \sum_{i=1}^{K} r_i^2$
- Least Abs. deviations $||\mathbf{r}|| = \sum_{i=1}^{K} |r_i|$
- Chebyshev $||\mathbf{r}|| = \max |r_i|$
- M-Estimators
- Least median
- Least trimmed squares $\sum_{i=1}^{h} r_{(i)}^2, h \leq K$
- Least trimmed abs. differences $\sum_{i=1}^{h} |r_{(i)}|, h \leq K$
- Etc

Fitting to data: Parametric approaches

- Fixed algebraic form
- Fitting additive generators
- Fitting weighting vectors
- Fitting fuzzy measures

Fitting to data: Fixed algebraic form

Example: Let us have the data and a family of AGOPs

$$f_p(x_1, \ldots, x_n) = \left(\sum_{i=1}^{n} \frac{1}{n} x_i^p \right)^{1/p}$$

To choose the best value of p we solve (using LS criterion)

$$\text{Minimise}_{\ p} \sum_{k=1}^{K} \left[\left(\sum_{i=1}^{n} \frac{1}{n} x_{ik}^p \right)^{1/p} - y_k \right]^2$$

subject to $p \in \Re$. A nonlinear problem, with possibly many locally optimal solutions. Need a global method.

Fitting to data: Fitting additive generators

Example: suppose we fit a family of Yager t-norms with the generators $g_p(t) = (1-t)^p$. We need to fit the data $g_p(x_1) + \ldots + g_p(x_n) \approx g_p(y)$.

To choose the best value of p we solve (using LS criterion)

$$\text{Minimise}\ _p \sum_{k=1}^{K} \left[\sum_{i=1}^{n} (1 - x_{ik})^p - (1 - y_k)^p \right]^2$$

subject to $p \in [0, \infty]$.

Fitting to data: Fitting weighting vectors

Example: We fit a WAM with unknown weights.

To choose the best weighting vector **w** we solve (using LS criterion)

$$\text{minimize}_{w} \sum_{k=1}^{K} \left(\sum_{i=1}^{n} w_i x_{ik} - y_k\right)^2,$$

subject to $w_i \geq 0$ and $\sum w_i = 1$.

This is a convex quadratic programming problem. LSEI (Least squares with equality and inequality constraints) is a classical algorithm to solve such problem.

Fitting to data: Fitting weighting vectors

Example: We fit a WAM using LAD criterion.

$$\text{minimize}_{w} \sum_{k=1}^{K} \left| \sum_{i=1}^{n} w_i x_{ik} - y_k \right|,$$

subject to $w_i \geq 0$ and $\sum w_i = 1$. Convert to an LP:

$$\begin{aligned} &\text{minimize} && \sum_{k=1}^{K} r_k^+ + r_k^- && (1) \\ &\text{subject to} && r_k^+ - r_k^- - \sum_{i=1}^{n} w_i x_{ik} = -y_k,\ k = 1, \ldots, K \\ & && \sum_{i=1}^{n} w_i = 1,\ r_k^-, r_k^+ \geq 0, w_i \geq 0. \end{aligned}$$

where r_k^+, r_k^- are positive and negative parts of $\sum_{i=1}^{n} w_i x_{ik} - y_k$. Note that $r_k^+ + r_k^- = |\sum_{i=1}^{n} w_i x_{ik} - y_k|$.

Fitting to data: Fitting weighting vectors

Example: We fit an OWA using LAD criterion.

$$\text{minimize}_{\ w} \sum_{k=1}^{K} \left| \sum_{i=1}^{n} w_i x_{(i)k} - y_k \right|,$$

subject to $w_i \geq 0$ and $\sum w_i = 1$, $orness = \alpha$. Convert to an LP:

$$\text{minimize} \quad \sum_{k=1}^{K} r_k^+ + r_k^- \tag{2}$$

$$\text{subject to} \quad r_k^+ - r_k^- - \sum_{i=1}^{n} w_i x_{(i)k} = -y_k, \; k = 1, \ldots, K$$

$$\sum_{i=1}^{n} w_i = 1, \; \frac{1}{n-1} \sum_{i=1}^{n} w_i (n - i) = \alpha, \; r_k^-, r_k^+ \geq 0, w_i \geq 0.$$

Fitting fuzzy measures

- In addition to assigning weights to each input, we can assign weights to all possible combinations of inputs.
- Some inputs may reinforce each other while others are redundant.
- We model it with the help of fuzzy measures $v : 2^{\mathcal{N}} \to [0, 1]$ and fuzzy integrals.
- Choquet and Sugeno integrals are the most important ones. Choquet integral generalises WAM and OWA, while Sugeno integral generalises the median
- We can explicitly control inputs relations using Shapley values and interaction indices.

Available tools

- There are several techniques to construct aggregation functions from data.
  - Fitting a parameter
  - Fitting a weighting vectors, fuzzy measures
  - Fitting additive generators with splines
  - Tensor product splines
  - Monotone Lipschitz approximation
- Some methods are simple, whereas others are complex.
- It would be nice to have computational tools (packages, libraries).

Available tools

- Such tool exist.
- We consider AOTools, fmtool and kappalab

AOTool is a software package with spreadsheet interface to fit various classes of AGOPs to data.

FMtool is an open source programming library to deal with and fit fuzzy measures.

Kappalab is a package for R to manipulate and fit fuzzy measures.

Summary

- Fuzzy rule based systems provide a model to deal with imprecise concepts, define meaningful rules, and execute these rules in a mathematically consistent way.
- Aggregation functions allow us to model conjunction, disjunction and averaging
- We looked at construction of aggregation functions from data, subject to application specific properties.
- We examined several methods: Fitting parameters, fitting weighting vectors
- There are other methods for: fitting fuzzy measures,

fitting additive generators with splines, tensor product splines, monotone Lipschitz approximation

- There are tools to perform such constructions
- Details in our book Beliakov, Pradera, Calvo, *Aggregation Functions: A Guide for Practitioners*, Springer, 2007

Conclusion. Questions ?