IEEE WCCI 2020
IEEE World Congress on Computational Intelligence
Virtual Conference – July 19-24, 2020

A Weighted Matrix Visualization for Fuzzy Measures and Integrals

Andrew R. Buck, Derek T. Anderson, James M. Keller, Timothy Wilkin, and Muhammad Aminul Islam

University of Missouri

Outline

- Background
  - Fuzzy measures and the Choquet integral
  - Shapley value and interaction index
- Visualizing fuzzy measures
  - Weighted matrix diagram
  - Showing data coverage
- Examples
  - OWA operators
  - MCDM
  - Information fusion
- Conclusion

Fuzzy Measure

- Given a finite set $X = \{x_1, \ldots, x_n\}$
- A fuzzy measure is a function $g: \mathcal{P}(X) \to \mathbb{R}^+$ which satisfies
  (i) $g(\emptyset) = 0$
  (ii) $A \subseteq B \subseteq X$ implies $g(A) \leq g(B)$
- Usually we define $g(X) = 1$.
- $\mathcal{P}(X)$ is the power set of X.
  - E.g. for $X = \{x_1, x_2, x_3\}$, $\mathcal{P}(X) = \{\emptyset, \{x_1\}, \{x_2\}, \{x_3\}, \{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_3\}, \{x_1, x_2, x_3\}\}$.
- We can think of $g(A)$ as representing the value or utility of the subset $A \subseteq X$.

Shapley Value

- The Shapley value of a fuzzy measure g is defined as the vector $[s_1, \dots, s_n]$ where

$$s_i = \sum_{K \subseteq X \setminus i} \frac{(n - |K| - 1)!\,|K|!}{n!} [g(K \cup i) - g(K)], \qquad \text{with } n = |X|$$

- The vector is normalized such that $\sum_{i=1}^{n} x_i = g(X)$.
- Each s_i represents the average contribution that x_i makes to the worth of the set when added to an existing subset.

Interaction Index

- The Shapley value can be generalized to extend to arbitrary subsets of X.
- The interaction index of a subset $A \subseteq X$ for a fuzzy measure g is defined as

$$I(A) = \sum_{B \subseteq X \setminus A} \frac{(n - |B| - |A|)!\,|B|!}{(n - |A| + 1)!} \sum_{C \subseteq A} (-1)^{|A \setminus C|} g(C \cup B), \qquad \text{with } n = |\mathrm{X}|$$

- $I(A)$ gives a sense of the worth of the set A in the context of the fuzzy measure.
- When $I(A) > 0$, there is positive synergy between the elements.
- When $I(A) < 0$, the elements are redundant.

Fuzzy Integral

- Let $h: X \rightarrow [0,1]$ be a function that specifies the value of a single element $x_i \in X$.
- Given h and g, the discrete Choquet integral is defined as

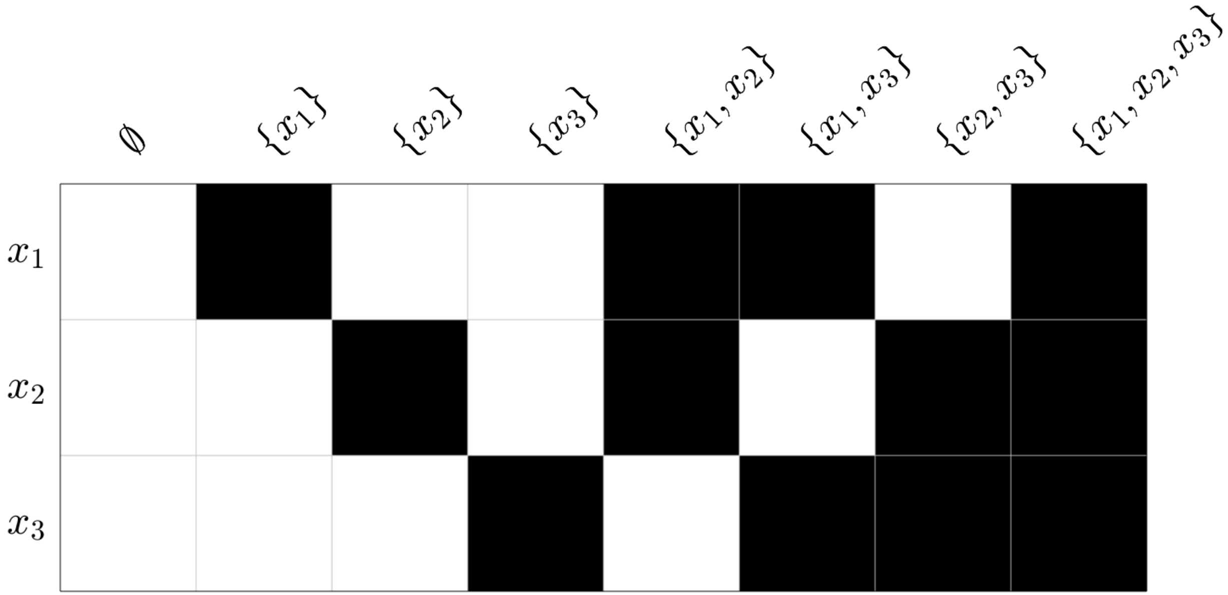
$$C_g(h) = \int_C h \circ g = \sum_{i=1}^{n} h(x_{\pi(i)})[g(A_i) - g(A_{i-1})]$$

where π is a permutation of X such that $h(x_{\pi(1)}) \geq h(x_{\pi(2)}) \geq \cdots \geq h(x_{\pi(n)})$ and $A_i = \{x_{\pi(1)}, \ldots, x_{\pi(i)}\}$ with $g(A_0) = 0$.

- A single data sample h produces an output $C_g(h)$ that uses only n of the 2^n possible subsets of X, implying that data coverage may be important in learning a good measure.

Visualizing a Fuzzy Measure (1)

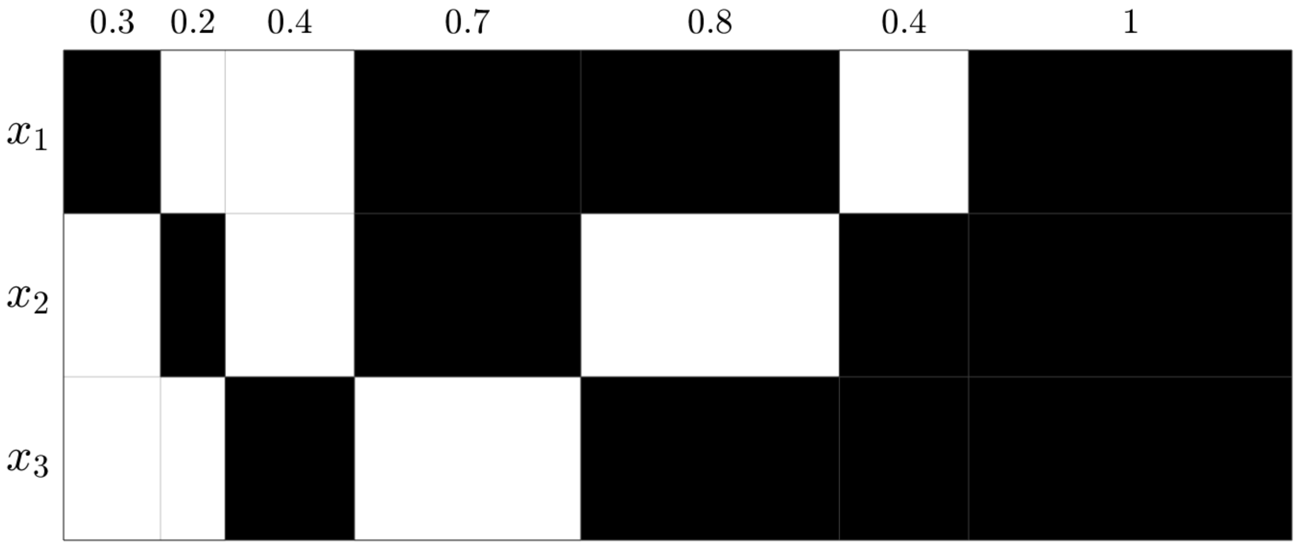
- Suppose we have a fuzzy measure g defined over a set $X = \{x_1, x_2, x_3\}$.
- Begin by constructing a binary indicator matrix.



A	$g(A)$
$\emptyset$	0
$\{x_1\}$	0.3
$\{x_2\}$	0.2
$\{x_3\}$	0.4
$\{x_1, x_2\}$	0.7
$\{x_1, x_3\}$	0.8
$\{x_2, x_3\}$	0.4
$\{x_1, x_2, x_3\}$	1

Visualizing a Fuzzy Measure (2)

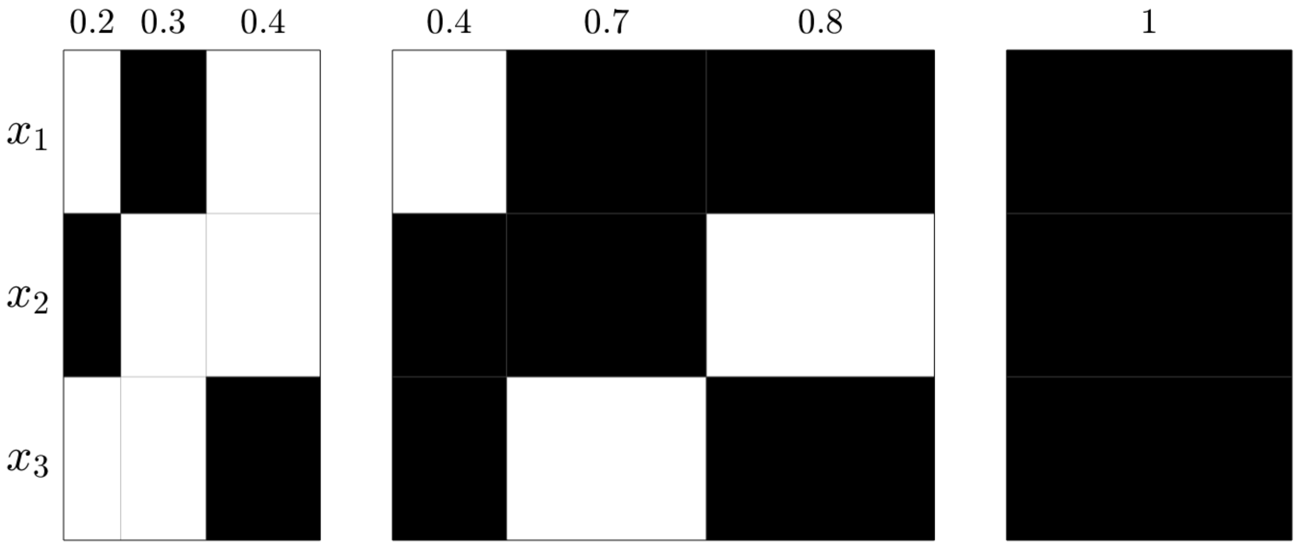
- Next, adjust the width of each column subset A to be proportional to $g(A)$.
- Since $g(\emptyset) = 0$, the first column is not shown.



A	$g(A)$
$\emptyset$	0
$\{x_1\}$	0.3
$\{x_2\}$	0.2
$\{x_3\}$	0.4
$\{x_1, x_2\}$	0.7
$\{x_1, x_3\}$	0.8
$\{x_2, x_3\}$	0.4
$\{x_1, x_2, x_3\}$	1

Visualizing a Fuzzy Measure (3)

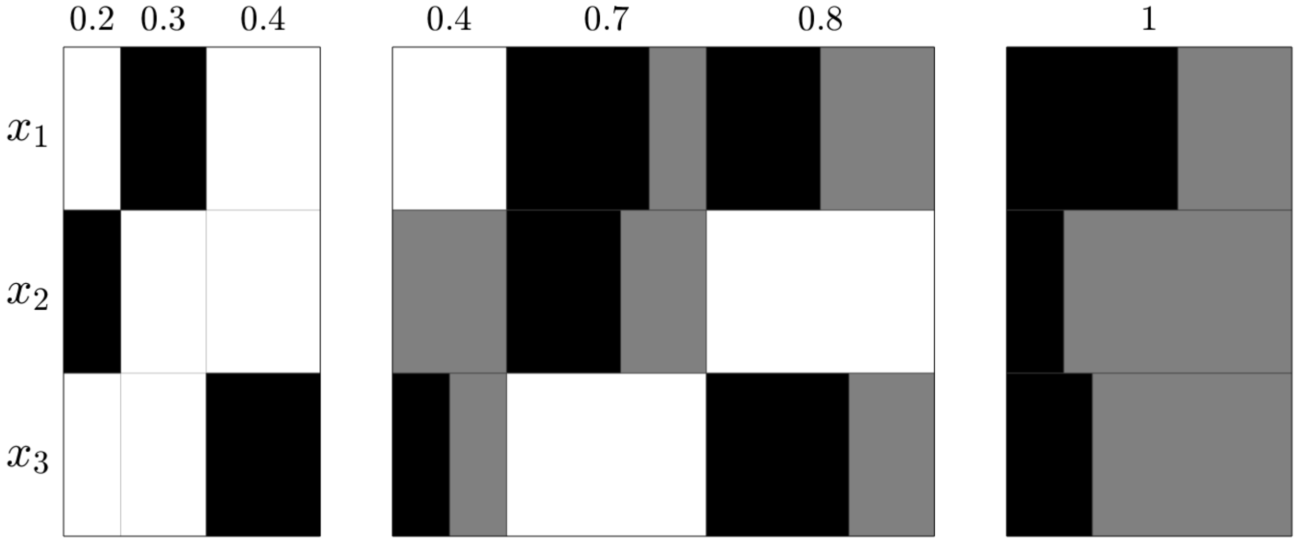
- Then, separate the cardinality sets and sort the columns by increasing value within each set.



A	$g(A)$
$\emptyset$	0
$\{x_1\}$	0.3
$\{x_2\}$	0.2
$\{x_3\}$	0.4
$\{x_1, x_2\}$	0.7
$\{x_1, x_3\}$	0.8
$\{x_2, x_3\}$	0.4
$\{x_1, x_2, x_3\}$	1

Visualizing a Fuzzy Measure (4)

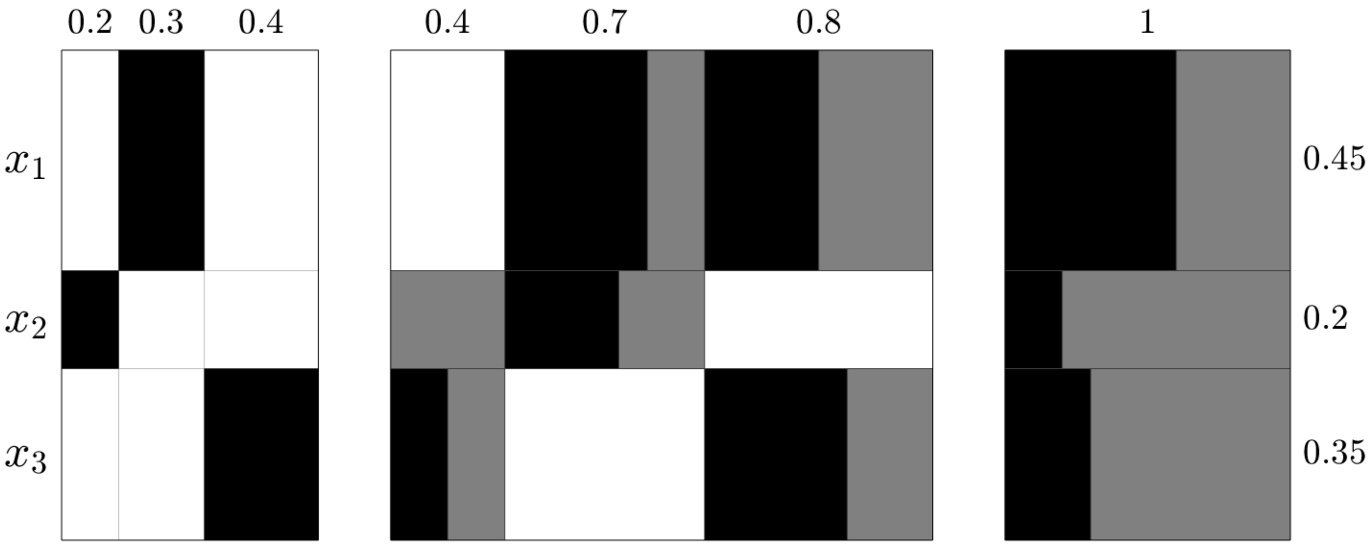
- The incremental contribution of source i in subset A_j is $\Delta g_{ij} = g(A_j) - g(A_j \backslash i)$.
- These are shown as horizontal black bars and the indicator matrix becomes gray.



A	$g(A)$
$\emptyset$	0
$\{x_1\}$	0.3
$\{x_2\}$	0.2
$\{x_3\}$	0.4
$\{x_1, x_2\}$	0.7
$\{x_1, x_3\}$	0.8
$\{x_2, x_3\}$	0.4
$\{x_1, x_2, x_3\}$	1

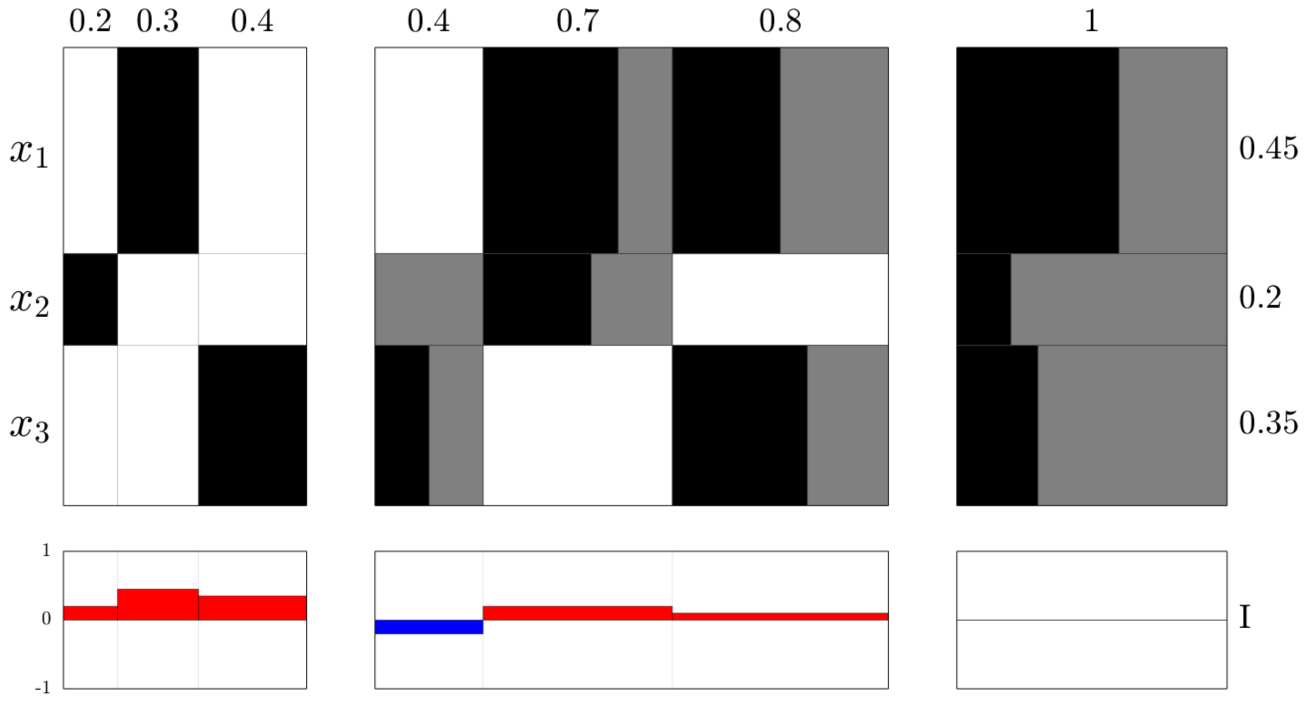
Visualizing a Fuzzy Measure (5)

- The row heights are scaled proportionally to the Shapley index of each source.
- Wider rows indicate more important sources.



A	$g(A)$
$\emptyset$	0
$\{x_1\}$	0.3
$\{x_2\}$	0.2
$\{x_3\}$	0.4
$\{x_1, x_2\}$	0.7
$\{x_1, x_3\}$	0.8
$\{x_2, x_3\}$	0.4
$\{x_1, x_2, x_3\}$	1

Visualizing a Fuzzy Measure (6)



- Finally, we can show the interaction indices for each subset as a bar graph below the diagram.
- The interaction index $I(A)$ for singletons is equal to the Shapley value.
- Positive values (red) show synergy and negative values (blue) show redundancy.

Learning a FM from Data

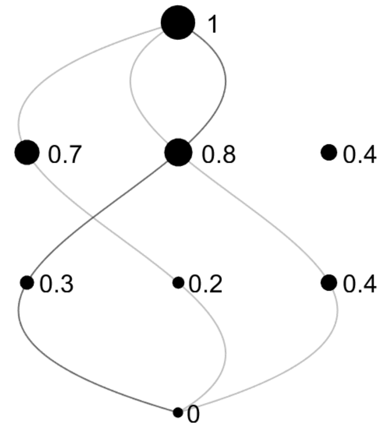
- A FM lattice visualization can show which variables are supported by data [1].
- Each data sample is shown as a walk through the lattice.
- We can see unsupported subsets, but it can be difficult to identify the corresponding source elements.

TABLE I: Fuzzy Measure

A	$g(A)$
$\emptyset$	0
$\{x_1\}$	0.3
$\{x_2\}$	0.2
$\{x_3\}$	0.4
$\{x_1, x_2\}$	0.7
$\{x_1, x_3\}$	0.8
$\{x_2, x_3\}$	0.4
$\{x_1, x_2, x_3\}$	1

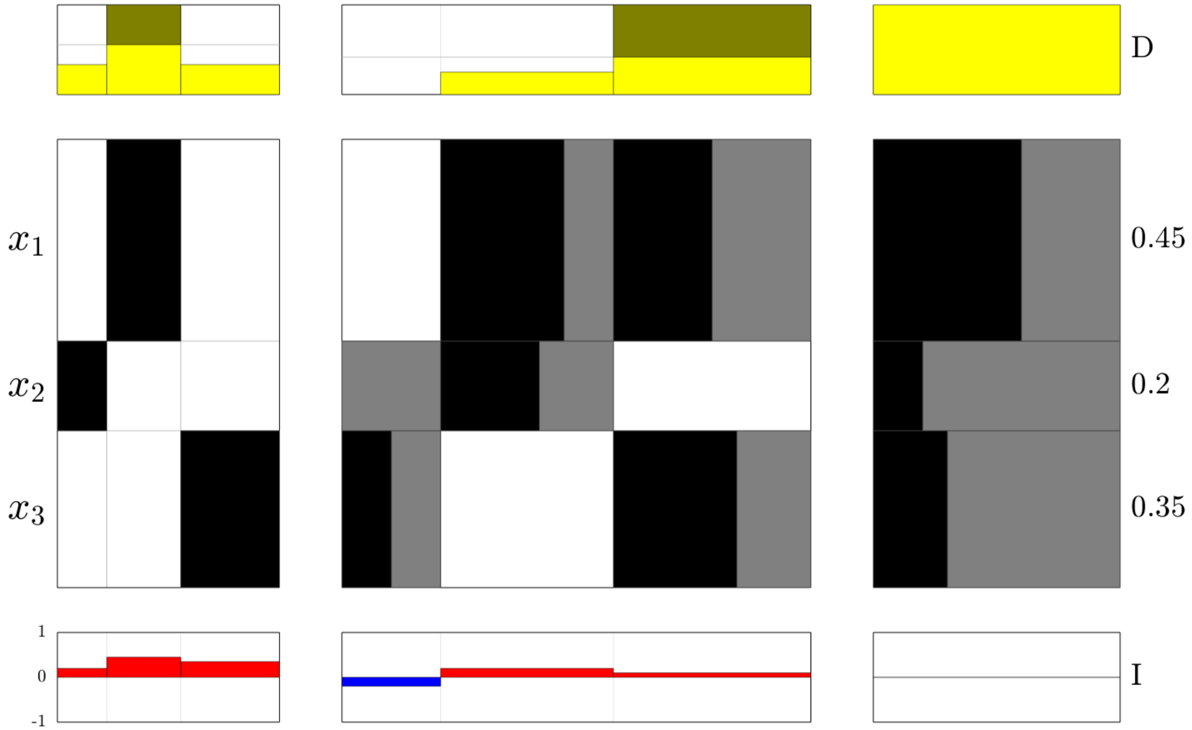
TABLE II: Example Data

$h(x_1)$	$h(x_2)$	$h(x_3)$	$\pi_{(1)}$	$\pi_{(2)}$	$\pi_{(3)}$
0.74	0.13	0.14	1	3	2
0.94	0.09	0.74	1	3	2
0.97	0.13	0.75	1	3	2
0.92	0.96	0.74	2	1	3
0.91	0.20	0.92	3	1	2



[1] A. J. Pinar, T. C. Havens, M. A. Islam, and D. T. Anderson, "Visualization and learning of the Choquet integral with limited training data," in *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Naples, Italy, July 2017.

Looking at Data Coverage



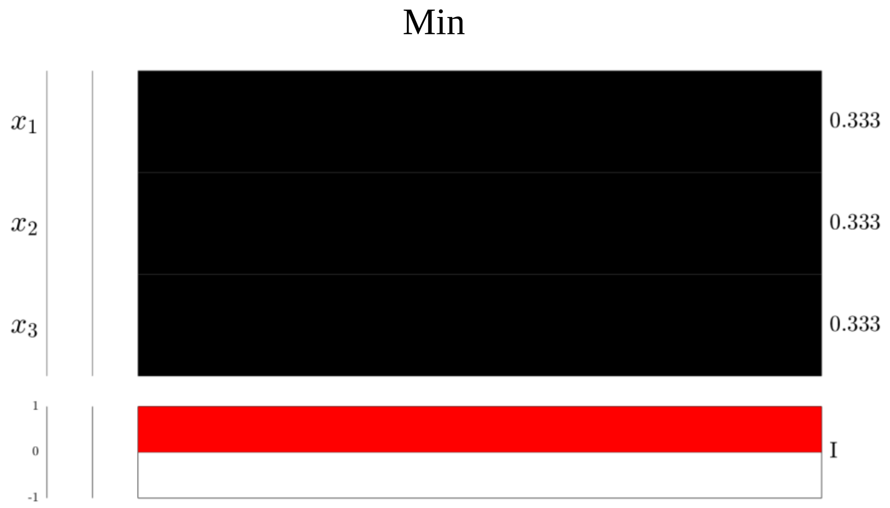
- Add a data coverage histogram above the diagram and normalize within each cardinality set.
- A horizontal line shows the mean visitation value and bars that extend above this are darkened.

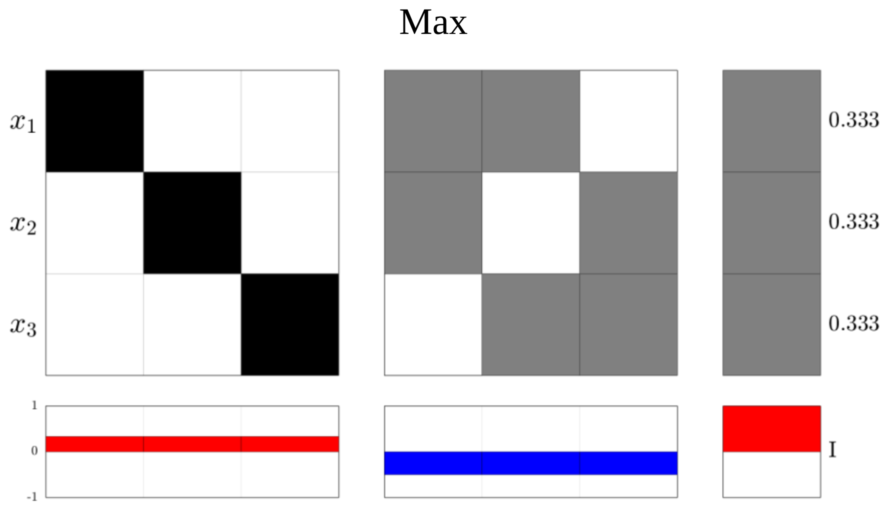
$h(x_1)$	$h(x_2)$	$h(x_3)$	$\pi_{(1)}$	$\pi_{(2)}$	$\pi_{(3)}$
0.74	0.13	0.14	1	3	2
0.94	0.09	0.74	1	3	2
0.97	0.13	0.75	1	3	2
0.92	0.96	0.74	2	1	3
0.91	0.20	0.92	3	1	2

Examples: Min & Max

- A fuzzy measure defines how the Choquet integral behaves.
- The min operator only gives value to the last element added in the sort.
- The max operator gives full value to the first element.

A	$g_{\min(A)}$	$g_{\max(A)}$
$\emptyset$	0	0
$\{x_1\}$	0	1
$\{x_2\}$	0	1
$\{x_3\}$	0	1
$\{x_1, x_2\}$	0	1
$\{x_1, x_3\}$	0	1
$\{x_2, x_3\}$	0	1
$\{x_1, x_2, x_3\}$	1	1

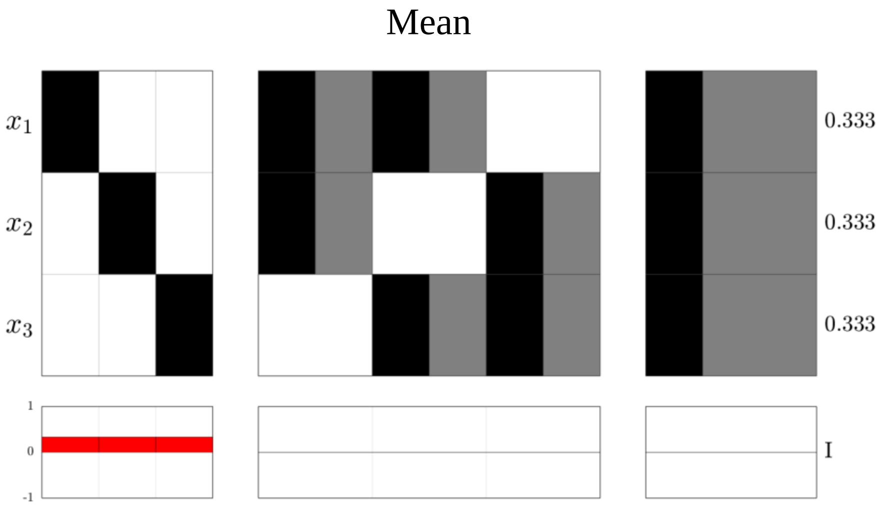


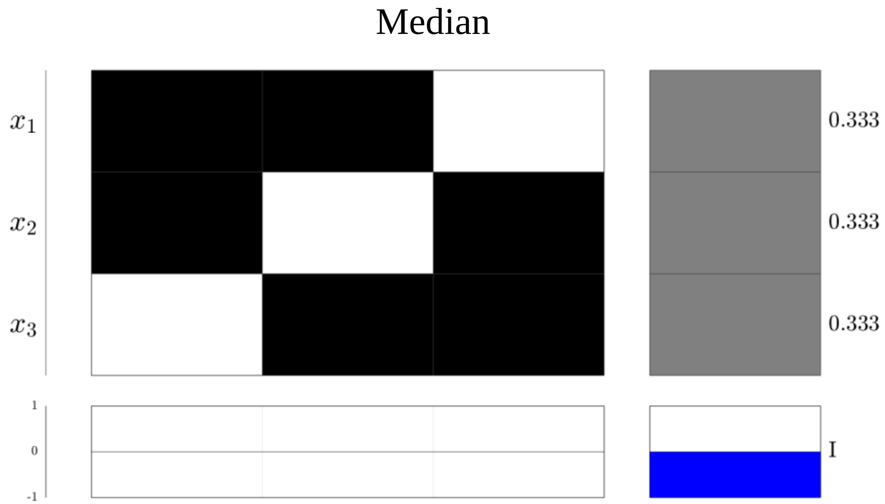


Examples: Mean & Median

- The mean operator shows uniform black bar sizes and a vertical striping pattern.
- The median operator gives all value to the middle cardinality set.

A	$g_{\text{mean}(A)}$	$g_{\text{median}(A)}$
$\emptyset$	0	0
$\{x_1\}$	0.33	0
$\{x_2\}$	0.33	0
$\{x_3\}$	0.33	0
$\{x_1, x_2\}$	0.67	1
$\{x_1, x_3\}$	0.67	1
$\{x_2, x_3\}$	0.67	1
$\{x_1, x_2, x_3\}$	1	1

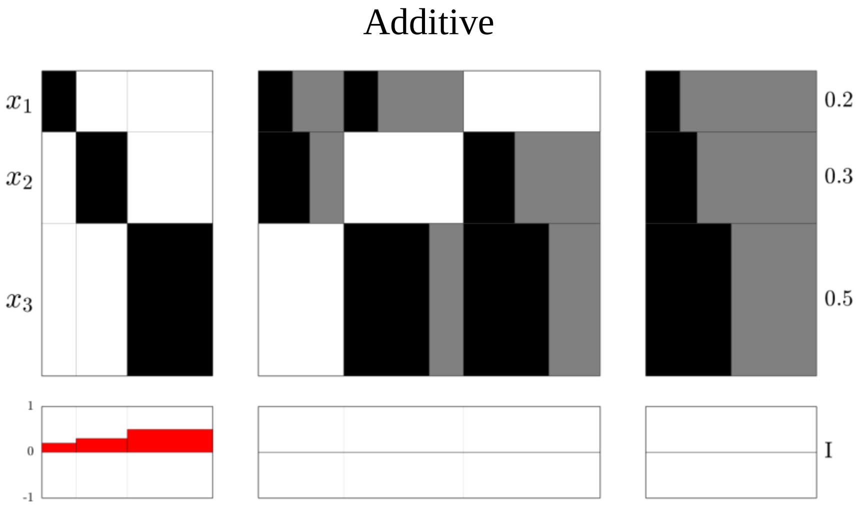


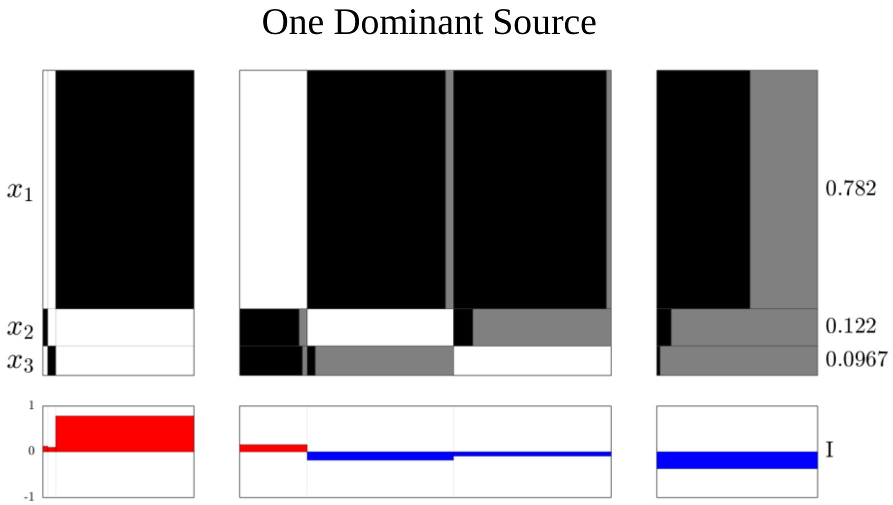


More 3-Source Examples

- In the additive FM, $g(A \cup B) = g(A) + g(B)$, and there are no interactions.
- When one source is dominant, it has a wider row and more black area.

A	$g_{\text{add}(A)}$	$g_{\text{one}(A)}$
$\emptyset$	0	0
$\{x_1\}$	0.2	0.86
$\{x_2\}$	0.3	0.03
$\{x_3\}$	0.5	0.05
$\{x_1, x_2\}$	0.5	0.98
$\{x_1, x_3\}$	0.7	0.91
$\{x_2, x_3\}$	0.8	0.42
$\{x_1, x_2, x_3\}$	1	1

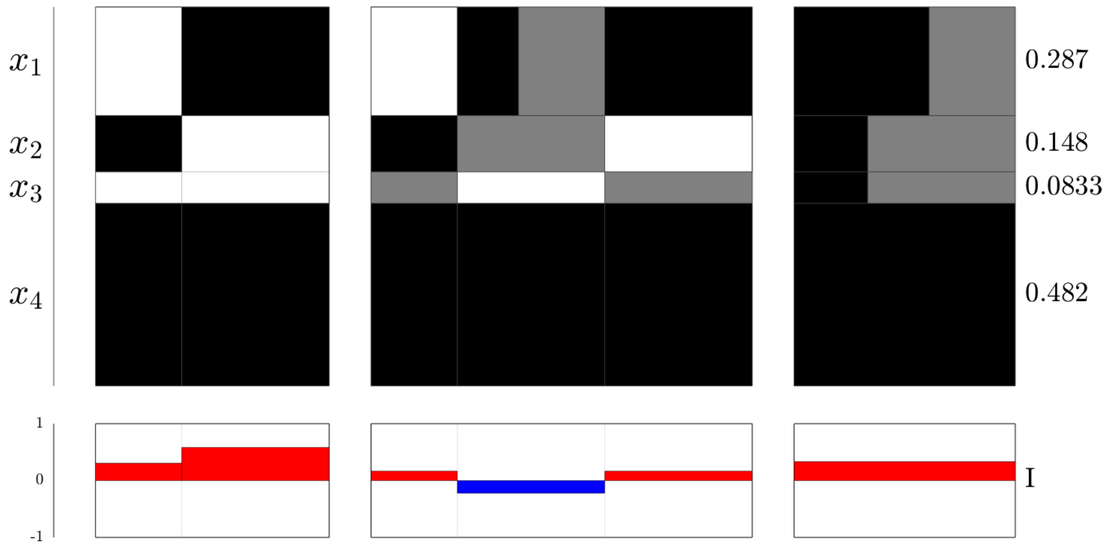




MCDM Example

- Fuzzy integrals can also be used for multicriteria decision-making.
- This example from [2] is used to score individuals on four judging criteria, x_1 to x_4.

A	$g(A)$
$\emptyset$	0
$\{x_1\}$	10^{-6}
$\{x_2\}$	10^{-6}
$\{x_3\}$	10^{-6}
$\{x_4\}$	10^{-6}
$\{x_1, x_2\}$	10^{-6}
$\{x_1, x_3\}$	10^{-6}
$\{x_1, x_4\}$	0.666667
$\{x_2, x_3\}$	10^{-6}
$\{x_2, x_4\}$	0.389743
$\{x_3, x_4\}$	10^{-6}
$\{x_1, x_2, x_3\}$	10^{-6}
$\{x_1, x_2, x_4\}$	0.666667
$\{x_1, x_3, x_4\}$	0.666667
$\{x_2, x_3, x_4\}$	0.389743
$\{x_1, x_2, x_3, x_4\}$	1



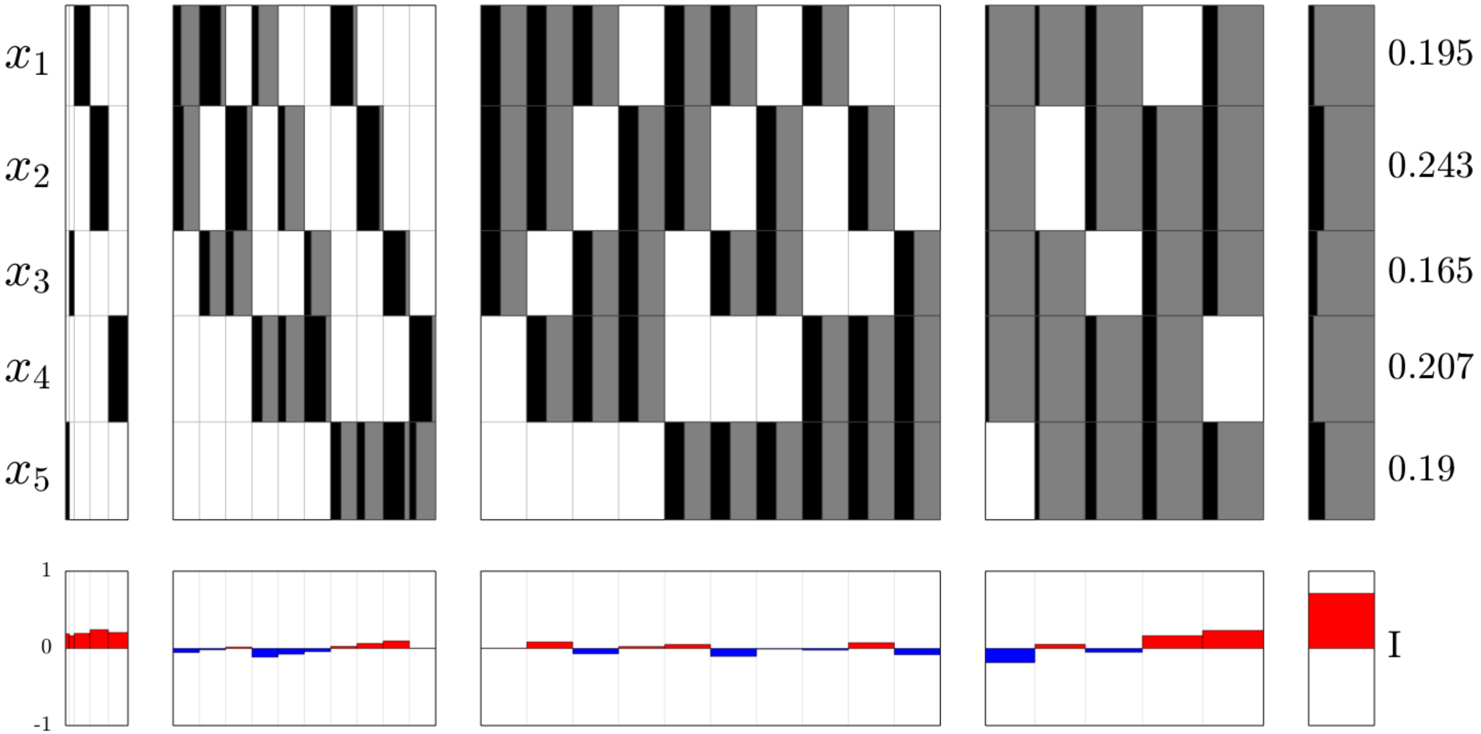
[2] M. Grabisch and M. Roubens, "Application of the Choquet integral in multicriteria decision making," *Fuzzy Measures and Integrals: Theory and Applications*, pp. 348–374, 2000.

Embedded OWA Operators

- A fuzzy measure can have a *embedded OWA operator.*
- In this example,

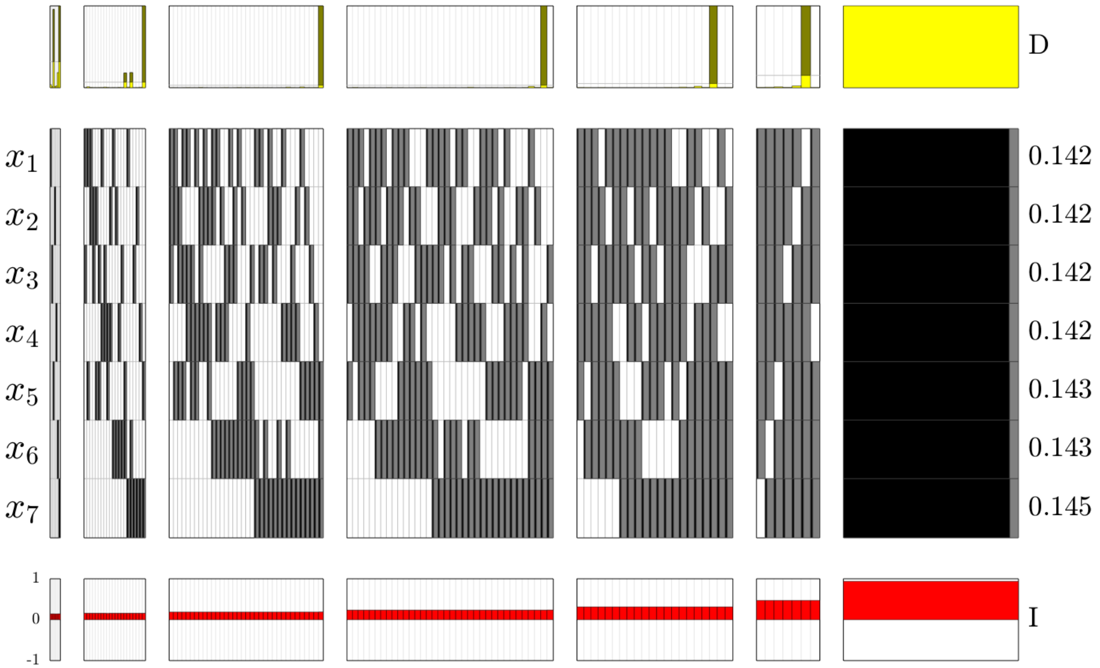
$$g(A) = \begin{cases} 0 & A = \emptyset \\ U(0, 0.4) & |A| = 1 \\ 0.4 & |A| = 2 \\ 0.7 & |A| = 3 \\ U(0.7, 1) & |A| = 4 \\ 1 & A = X \end{cases}$$

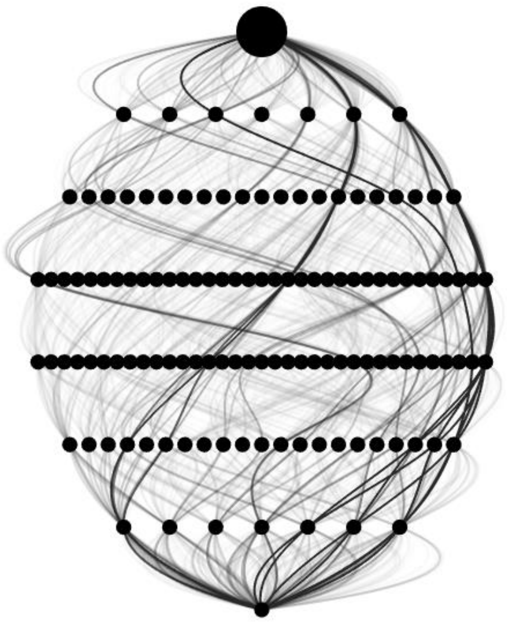
- Vertical striping can be observed in the cardinality 3 and 4 sets.



Measures Learned from Data

- In [3], a fuzzy measure is learned to aggregate the output of 7 neural networks.
- The diagram shows this FM acts mainly as a min operator and could be represented as an OWA.
- All sources are roughly equal in importance.
- The data visitation histogram shows almost all the data used a single walk.





[3] M. A. Islam, D. T. Anderson, A. J. Pinar, T. C. Havens, G. Scott, and J. M. Keller, "Enabling explainable fusion in deep learning with fuzzy integral neural networks," *IEEE Transactions on Fuzzy Systems (accepted)*, 2019, arXiv: 1905.04394.

Conclusion

- We presented a weighted matrix visualization for understanding fuzzy measures.
- This technique provides detail into the interactions between sources and can help determine if the full expressive power of the full fuzzy integral is required.
- Although it's possible to use an arbitrary number of sources, interpretability decreases with many sources.
- An interactive version that renormalizes subsets may be useful for large problems.
- Code is available on Code Ocean at https://codeocean.com/capsule/6663959.