Scaling-Invariant Description of Dependence Between Fuzzy Variables: Towards a Fuzzy Version of Copulas

Gerardo Muela\textsuperscript{1}, Vladik Kreinovich\textsuperscript{1}, and Christian Servin\textsuperscript{2}

\textsuperscript{1}Department of Computer Science
University of Texas at El Paso, El Paso, Texas 79968, USA
gdmuela@miners.utep.edu, vladik@utep.edu

\textsuperscript{2}Computer Science and Information Technology
Systems Department, El Paso Community College
El Paso, TX 79915, USA, cservin@gmail.com
1. Fuzzy Degrees: a Brief Reminder

- In many real-life situations, it is important to incorporate expert knowledge into a computer-based system.
- Experts are often not 100% confident about their statements.
- They may use heuristic rules that they know to be sometimes false.
- Thus, it is important to describe the expert’s degree of confidence in different statements.
- Experts usually describe their degree of confidence by using words from a natural language, such as “usually”.
- However, computers are not very efficient in processing natural language.
- They are more efficient in doing what they were originally designed for – processing numbers.
2. Fuzzy Degrees (cont-d)

- It is therefore reasonable to describe expert’s degrees of confidence by numbers.

- In the computers, “true” is usually represented as 1, and “false” as 0.

- Thus, it makes sense to represent intermediate degrees of confidence by numbers from the interval $[0, 1]$.

- This is one of the main ideas behind fuzzy logic.

- There are many ways to assign a numerical degree to a natural-language term.

- For example, we can ask an expert to mark his/her degree of confidence on a scale from, say, 0 to 10.

- If the expert marks 7, we take the ratio $7/10$ as the desired degree of confidence.
3. Fuzzy Degrees (cont-d)

• Alternatively, we can ask the expert to select between:
  – getting a certain small monetary award when his/her statement is true in a random situation or
  – getting the same award with some probability.

• This measures the expert’s subjective probability that his/her statement is true.

• In general, different methods lead to different numerical degrees.

• In all these cases, the more confident the expert is in a statement, the larger the numerical degree.

• Thus, ideally, a larger degree on one scale corresponds to a larger number on a different scale.
4. Formulation of the Problem

- Often, a term used by an expert depends on two or more real-valued variables.

- Example: healthiness degree depends on blood pressure, body-mass index, etc.

- For each combination \((x_1, \ldots, x_n)\), we have a degree \(\mu(x_1, \ldots, x_n) \in [0, 1]\).

- The membership function \(\mu(x_1, \ldots, x_n)\) described the dependence between \(x_i\).

- However, the numerical values of this function change if we use a different scale.

- In this talk, we describe a scale-invariant way to describing this dependence.
5. Copulas Solve a Similar Prob. Problem

- A random variable $X$ is described by its cumulative distribution function (cdf) $F(x) \overset{\text{def}}{=} \text{Prob}(X \leq x)$.
- A joint distribution is described by a joint cdf
  \[ F(x_1, x_2) \overset{\text{def}}{=} \text{Prob}(X_1 \leq x_1 \& X_2 \leq x_2). \]
- For each $i$, we have a marginal $F_i(x_i) \overset{\text{def}}{=} \text{Prob}(X_i \leq x_i)$:
  \[ F_1(x_1) = F(x_1, +\infty) \text{ and } F_2(x_2) = F(+\infty, x_2). \]
- The joint cdf contains the information:
  - about the marginals and
  - about the relation between $X_i$.
- How can we describe just the information about the dependence between the random variables?
6. Copulas (cont-d)

- E.g., independence means $F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2)$.
- In general, $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$ for some function $C(u_1, u_2)$ called a copula.
- For $n > 2$ variables,
  
  $F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$.

- Copulas do not change if we re-scale the variables $x_i$:
  
  $x_i \rightarrow x'_i = f_i(x_i)$. 
7. What Is the Fuzzy Analog of a Marginal?

- We want to use an analogy with copulas.
- For each pair \((x_1, x_2)\), the value \(\mu(x_1, x_2)\) describes the degree to which this pair is possible.
- A value \(x_1\) is possible if \((x_1, x_2)\) is possible for some \(x_2\):
  - either \((x_1, 0)\) is possible,
  - or \((x_1, 0.01)\) is possible, etc,
- To get a degree to which \(x_1\) is possible, we need to combine the degrees \(\mu(x_1, x_2)\) by some “or”-operation.
- In principle, there are many different “or”-operations: \(f_\lor(a, b) = \max(a, b), f_\lor(a, b) = 1 + a - a \cdot b\), etc.
- However, for most of them, combining infinitely many degrees leads to a useless value 1.
8. Fuzzy Marginal (cont-d)

- The only “or”-operation that makes sense is max, so
  \[ \mu_i(x_i) = \max_{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n} \mu(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n). \]

- Let’s assume that \( \mu_i(x_i) \) is a normal membership function: continuous and attains values 0 and 1.

- In this case, for every \( r \in [0, 1] \), there exists a \( v_i \) for which \( \mu_i(v_i) = r \).

- How do fuzzy marginals change under re-scaling
  \[ \mu'(x_1, x_2, \ldots, x_n) = f(\mu(x_1, x_2, \ldots, x_n))? \]

- One can prove that they change the same way:
  \[ \mu'_i(x_i) = f(\mu_i(x_i)). \]
9. How to Get a Scaling-Invariant Description of Dependence

• We want to find a description of the dependence that does not change if we re-scale all the degrees:

\[ \mu(x_1, x_2, \ldots, x_n) \rightarrow \mu'(x_1, x_2, \ldots, x_n) = f(\mu(x_1, x_2, \ldots, x_n)). \]

• For \( r = \mu(x_1, \ldots, x_n) \), there exists a number \( r_i(x_1, \ldots, x_n) \) for which

\[ \mu(x_1, \ldots, x_n) = \mu_i(r_i(x_1, \ldots, x_n)). \]

• The function \( r_i(x_1, \ldots, x_n) \) describes the dependence in the sense that:

  – if we know this function and the marginals \( \mu_i(x_i) \),
  – then we can reconstruct the original membership function \( \mu(x_1, \ldots, x_n) \).

• One can prove that each function \( r_i(x_1, \ldots, x_n) \) is scale-invariant; so, it is the desired dependence.
10. Examples of $\mu(x_1, \ldots, x_n) = \mu_i(r_i(x_1, \ldots, x_n))$

- For the Gaussian membership function $\mu(x_1, x_2) = \exp(-x_1^2 - x_2^2)$, marginal is $\mu_1(x_1) = \exp(-x_1^2)$.

- The definition of $r_1(x_1, x_2)$ becomes $\exp(-x_1^2 - x_2^2) = \exp(-(r_1(x_1, x_2))^2)$.

- By taking $-\ln$ of both sides, we get $x_1^2 + x_2^2 = (r_1(x_1, x_2))^2$, so $r_1(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$.

- For $\mu(x_1, x_2) = \frac{1}{1 + x_1^2 + x_2^2}$, we have $\mu_1(x_1) = \frac{1}{1 + x_1^2}$.

- Here, $\frac{1}{1 + x_1^2 + x_2^2} = \frac{1}{1 + r_1(x_1, x_2)^2}$, so $1 + r(x_1, x_2)^2 = 1 + x_1^2 + x_2^2$, and $r_1(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$.

- For $\mu_1(x_1, x_2) = \exp(-|x_1| - |x_2|)$, we have $\mu_1(x_1) = \exp(-|x_1|)$, so $\exp(-|x_1| - |x_2|) = \exp(-r_1(x_1, x_2))$ and $r_1(x_1, x_2) = |x_1| + |x_2|$.
11. Examples (cont-d)

- For $\mu(x_1, x_2) = \exp(-x_1^2 - x_1 \cdot x_2 - x_2^2)$, the maximum w.r.t. $x_2$ is when derivative is 0, i.e., when

$$x_1 + 2x_2 = 0 \text{ and } x_2 = -\frac{x_1}{2},$$

so

$$\mu_1(x_1) = \mu(x_1, x_2(x_1)) = \exp \left(-\frac{3}{4} \cdot x_1^2 \right).$$

- Thus, $\mu(x_1, x_2) = \mu_1(r_1(x_1, r_2))$ implies

$$\exp(-x_1^2 - x_1 \cdot x_2 - x_2^2) = \exp \left(-\frac{3}{4} \cdot (r_1(x_1, x_2))^2 \right).$$

- So, $(r_1(x_1, x_2))^2 = \frac{4}{3} \cdot (x_1^2 + x_1 \cdot x_2 + x_2^2)$ and

$$r_1(x_1, x_2) = \frac{2 \cdot \sqrt{3}}{3} \cdot \sqrt{x_1^2 + x_1 \cdot x_2 + x_2^2}.$$
12. Auxiliary Question: What Is the Relation Between Different $r_i(x_1, \ldots, x_n)$?

- To describe the dependence, we can use $r_i(x_1, \ldots, x_n)$ for any $i$.
- If we know $r_i(x_1, \ldots, x_n)$, can we reconstruct $r_j(x_1, \ldots, x_n)$ for $j \neq i$?
- Let us assume that all $\mu_i(x_i)$ increase when $x_i \leq c_i$ and decrease after that.
- As values $r_i(x_1, \ldots, x_n)$, we select values $r_i \geq c_i$.
- Then, we have $r_j(x_1, \ldots, x_n) = s_{ij}^{-1}(r_i(x_1, \ldots, x_n))$, where:
  \[
  s_{ij}(x_j) \overset{\text{def}}{=} \min_{x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n} r_i(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n).
  \]
13. Acknowledgments

- This work was supported in part by the National Science Foundation grants:
  - HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and
  - DUE-0926721, and
- by an award from Prudential Foundation.
14. Proof That $r_i(x_1, \ldots, x_n)$ Is Scaling-Invariant

- We define $r_i(x_1, \ldots, x_n)$ by the formula

  $$\mu(x_1, \ldots, x_n) = \mu_i(r_i(x_1, \ldots, x_n)).$$

- Let’s re-scale membership values, i.e., replace:

  - $\mu(x_1, \ldots, x_n) \rightarrow \mu'(x_1, \ldots, x_n) = f(\mu(x_1, \ldots, x_n))$ and

    - $\mu_i(x_i) \rightarrow \mu'_i(x_i) = f(\mu_i(x_i)).$

  - By applying $f(x)$ to both sides of the above equality, we get $f(\mu(x_1, \ldots, x_n)) = f(\mu_i(r_i(x_1, \ldots, x_n))).$

  - Thus, $\mu'(x_1, \ldots, x_n) = \mu'_i(r_i(x_1, \ldots, x_n)).$

  - So, for the re-scaled membership degrees, we have the exact same function $r_i(x_1, \ldots, x_n).$

  - Thus, $r_i(x_1, \ldots, x_n)$ is indeed scaling-invariant.
15. Proof of the $r_i \rightarrow r_j$ Formula

- $\mu(x_1, \ldots, x_n) = \mu_i(r_i(x_1, \ldots, x_n))$ implies that
  $$
  \mu_j(x_j) = \max_{x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n} \mu(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n) = \max_{x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n} \mu_i(r_i(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n)).
  $$

- $\mu_i(x_i)$ is strictly decreasing for $x_i \geq c_i$.

- Thus, $\mu_i(r_i(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n))$ is max $\iff$ $r_i(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n)$ is min; so, $\mu_j(x_j) = \max_{x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n} \mu_i(r_i(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots, x_n))$.

- Thus, $\mu_j(x_j) = \mu_i(s_{ij}(x_j))$ and $\mu_i(x_i) = \mu_j(s_{ij}^{-1}(x_i))$.

- From $\mu(x_1, \ldots, x_n) = \mu_i(r_i(x_1, \ldots, x_n))$, we get
  $$
  \mu(x_1, \ldots, x_n) = \mu_j(s_{ij}^{-1}(r_i(x_1, \ldots, x_n))).
  $$

- Thus, $\mu(x_1, \ldots, x_n) = \mu_j(r_j(x_1, \ldots, x_n))$ for
  $$
  r_j(x_1, \ldots, x_n) = s_{ij}^{-1}(r_i(x_1, \ldots, x_n)). \text{ Q.E.D.}
  $$