Fuzzy Set Theory in Computer Vision: **Example 5**

Derek T. Anderson and James M. Keller

FUZZ-IEEE, July 2017

Training Data

Let a training data set, T, be

$$T = \{(O_j, \alpha_j) | j = 1, ..., m\}$$

where $\mathbf{O} = \{O_1, ..., O_j, ..., O_m\}$ is a set of "objects" and α_j are their corresponding labels (specifically, $\Re$-valued numbers). For example, O_j could be the strengths in some hypothesis from N different experts, signal inputs at time j, algorithm outputs for input j, kernel inputs or kernel classifier outputs for feature vector j, etc. Subsequently, α_j could be the corresponding function output, class label, membership degree, etc.

ChI with respect to T

Let h_j be the jth integrand, i.e., $h_j(x_i)$ is the input for the ith source with respect to object j. The discrete ChI, for finite X and object O_j is

$$\int h_j \circ \mu = C_\mu(h_j) = \sum_{i=1}^{N} [h_j(x_{\pi_j(i)}) - h_j(x_{\pi_j(i+1)})] \mu(A_{\pi_j(i)}),$$

for $A_{\pi_j(i)} = \{x_{\pi_j(1)}, ..., x_{\pi_j(i)}\}$ and permutation π_j such that $h_j(x_{\pi_j(1)}) \geq ... \geq h_j(x_{\pi_j(N)})$, where $h_j(x_{\pi_j(N+1)}) = 0$.

Sum of Squared Error of ChI and T

Let the SSE between T and the CI be

$$E_1 = \sum_{j=1}^{m} (C_\mu(h_j)) - \alpha_j)^2.$$

$$E_1 = \sum_{j=1}^{m} (\mathbf{A}_{O_j}^t \mathbf{u} - \alpha_j)^2,$$

where $\mathbf{u}$ is the lexicographically encoded capacity vector and

$$\mathbf{A}_{O_j}^t = \left(\ldots, h_j(x_{\pi_j(1)}) - h_j(x_{\pi_j(2)}), \ldots, 0, \ldots, h_j(x_{\pi_j(N)}) \right)^t$$

is of size $1 \times (2^N - 1)$. The function differences correspond to their respective μ locations in $\mathbf{u}$.

Sum of Squared Error of ChI and T

Folding our equation out further, we find

$$E_1 = \sum_{j=1}^{m} (\mathbf{u}^t \mathbf{A}_{O_j} \mathbf{A}_{O_j}^t \mathbf{u} - 2\alpha_j \mathbf{A}_{O_j}^t \mathbf{u} + \alpha_j^2)$$

$$= \mathbf{u}^t \mathbf{D} \mathbf{u} + \mathbf{f}^t \mathbf{u} + \sum_{j=1}^{m} \alpha_j^2,$$

where

$$\mathbf{D} = \sum_{j=1}^{m} \mathbf{A}_{O_j} \mathbf{A}_{O_j}^t \text{ and } \mathbf{f} = \sum_{j=1}^{m} (-2\alpha_j \mathbf{A}_{O_j}).$$

Constraints

In total, the capacity has $(N(2^{N-1}-1))$ monotonicity constraints. These constraints can be represented in a compact linear algebra (aka matrix) form. The following is the minimum number of constraints needed to represent the FM. Let $\mathbf{Cu}+\mathbf{b}\leq\mathbf{0}$, where

$$\mathbf{C}^t=\left(\Psi_1^t,\Psi_2^t,...,\Psi_{N+1}^t,...,\Psi_{N(2^{N-1}-1)}^t\right)^t,$$

and Ψ_1 is a vector representation of constraint 1, $\mu_1-\mu_{12}\leq 0$. Specifically, $\Psi_1^t\mathbf{u}$ recovers $\mathbf{u}_1-\mathbf{u}_{N+1}$. Thus, C is simply a matrix of $\{0,1,-1\}$ values (next slide)

Constraints

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & \dots & -1 & 0 & \dots & \dots & 0 \\ 1 & 0 & \dots & 0 & -1 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 1 & -1 \end{bmatrix}, \quad (1)$$

which is of size $(N(2^{N-1} - 1)) \times (2^N - 1)$.
Also, $\mathbf{b} = \mathbf{0}$, a vector of all zeroes. Note, in some works, $\mathbf{u}$ is of size $(2^N - 2)$, as $\mu(\phi) = 0$ and $\mu(X) = 1$ are explicitly encoded. In such a case, $\mathbf{b}$ is a vector of 0s and the last N entries are of value -1. Herein, we use the $(2^N - 1)$ format.

Quadratic Program

Given T, the search for FM μ reduces to a QP of the form

$$\min_{\mathbf{u}} \frac{1}{2}\mathbf{u}^t\hat{\mathbf{D}}\mathbf{u} + \mathbf{f}^t\mathbf{u}, \tag{2}$$

subject to $\mathbf{C}\mathbf{u} + \mathbf{b} \geq \mathbf{0}$ and $(\mathbf{0}, 1)^t \leq \mathbf{u} \leq \mathbf{1}$. The difference between the two equations is $\hat{\mathbf{D}} = 2\mathbf{D}$ and the inequality need only be multiplied by -1.

ℓ_1-Norm of a Lexicographically Encoded FM

Let $\mathbf{u} \in \mathbb{R}^{2^N-1}$ be $\mathbf{u} = (\mu_1, \mu_2, ..., \mu_{12}, \mu_{13} ..., \mu_{12...N})^t$. Note that we define this ordering such that it is also sorted by cardinality. A relatively simple index of the complexity of μ is

$$v_{\ell 1}(\mu) = \sum_{j=1}^{2^N-1} |\mathbf{u}_j| = \sum_{j=1}^{2^N-1} \mathbf{u}_j.$$

As an example, consider the case of $N = 3$ where the vector $\mathbf{u}$ is $\mathbf{u} = (\mu_1, \mu_2, \mu_3, \mu_{12}, \mu_{13}, \mu_{23}, \mu_{123})^t$.

Regularization

In general, the challenge of QP-based learning of the ChI relative to a regularization term is the optimization of

$$E_2 = \sum_{j=1}^{m} (\mathbf{u}^t \mathbf{A}_{O_j} \mathbf{A}_{O_j}^t \mathbf{u} - 2\alpha_j \mathbf{A}_{O_j}^t \mathbf{u} + \alpha_j^2) + \lambda v_*(\mu), \qquad (3)$$

where $v_*(\mu)$ is the regularizer term. In order for our equation to be suitable for the QP, v_* must be linear or quadratic. See our paper "A. Pinar, D. T. Anderson, T. Havens, A. Zare, T. Adeyeba, Measure of the Shapley Index for Learning Lower Complexity Fuzzy Integrals, Granular Computing, 2017" for different fuzzy measure information theoretic regularizers.