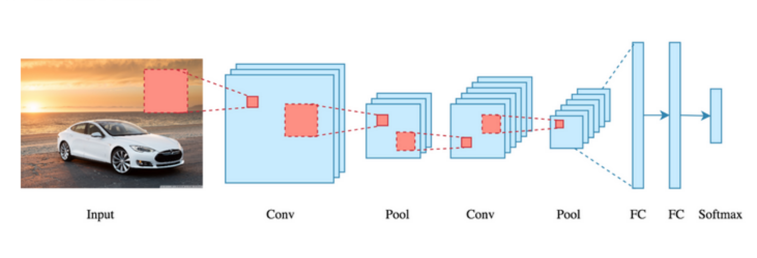
Aggregation and pre-aggregation functions. Extensions of fuzzy integrals and their applications to the computational brain, fuzzy rule systems and decision making

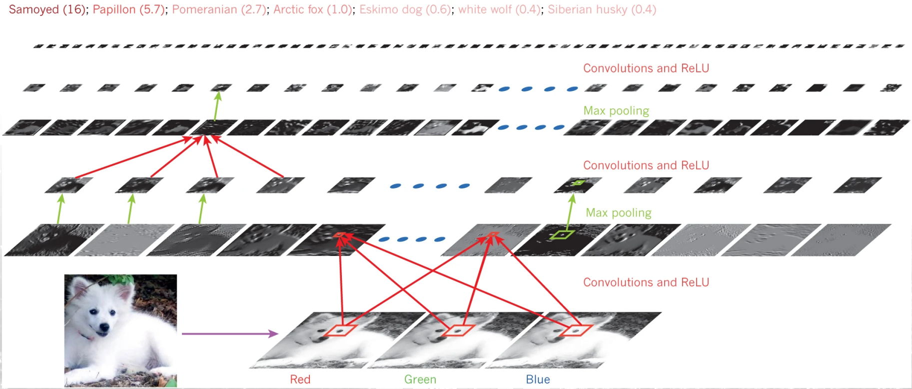
H. Bustince
Public University of Navarra

Pamplona 22 June 2022

Convolutional Neural Network: CNN



Convolutional Neural Network: CNN



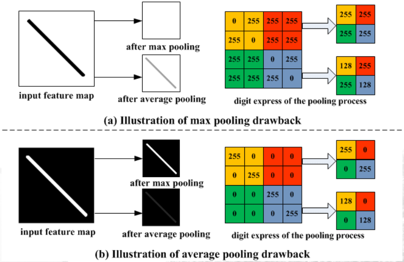
LeCun, Bengio, and Hinton, "Deep learning"

Pooling

It aggregates the features extracted by the convolution layers:

$\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$;
$\mathbf{A}(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$

- Summary of relevant information in a local way.
- Max pooling: $\mathbf{A}(\mathbf{x}) = \max_{i=1}^{n} \mathbf{x}_i$
- Avg pooling: $\mathbf{A}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$



(a) Illustration of max pooling drawback

(b) Illustration of average pooling drawback

Yu et al., "Mixed pooling for convolutional neural networks"

Data fusion functions using numbers in $[0, 1]$

Fusion functions

Definition

Let $n \geq 2$. An (n-ary) fusion function is an arbitrary function $F : [0,1]^n \to [0,1]$.

- The choice of the unit interval is not relevant. Any other interval of real numbers would do.
- No conditions are imposed at all to F.

Monotonicity and related notions

Definition

A function $F : [a, b]^n \to [a, b]$ is increasing if for every $x_1, \dots, x_n$, $y_1, \dots, y_n \in [a, b]$ such that $x_i \leq y_i$ for every $i = 1, \dots, n$ the inequality

$$F(x_1, \dots, x_n) \leq F(y_1, \dots, y_n)$$

holds.

Aggregation functions

Definition

An aggregation function is a function $M : [0,1]^n \to [0,1]$ such that:

1. M is increasing;
2. $M(0, \ldots, 0) = 0$
3. $M(1, \ldots, 1) = 1.$

Definition

An aggregation function is a function $M : [0,1]^n \to [0,1]$ such that:

1. M is increasing;
2. $M(0, \ldots, 0) = 0$
3. $M(1, \ldots, 1) = 1.$

Definition

An aggregation function M is called idempotent if for every $t \in [0,1]$, $M(t, \cdots, t) = t$

Definition

An aggregation function is a function $M : [0,1]^n \to [0,1]$ such that:

1. M is increasing;
2. $M(0,\dots,0) = 0$
3. $M(1,\dots,1) = 1$.

Definition

An aggregation function M is called idempotent if for every $t \in [0,1]$, $M(t,\cdots,t) = t$

Definition

An aggregation function M is called averaging if $\min(\mathbf{x}) \le M(\mathbf{x}) \le \max(\mathbf{x})$

Definition

An aggregation function $T : [0,1]^2 \to [0,1]$ is a triangular norm (t-norm) if it satisfies the following conditions:

- T1 T is commutative;
- T2 T is associative.
- T3 $T(x,1) = x$ for every $x \in [0,1]$.

Definition

A function $O : [0,1]^2 \to [0,1]$ is an overlap function if it satisfies the following conditions:

- O1 O is commutative;
- O2 $O(x,y) = 0$ if and only if $xy = 0$;
- O3 $O(x,y) = 1$ if and only if $xy = 1$;
- O4 O is increasing;
- O5 O is continuous.

Every continuous t-norm without divisor of zero is an overlap function

Some relevant aggregation functions: copulas

Definition

A function $C : [0,1]^2 \to [0,1]$ is a copula if, for all $x, x', y, y' \in [0,1]$ such that $x \leq `x'$ and $y \leq y'$, it satisfies the following conditions:

C1 $C(x,y) + C(x',y') \geq C(x,y') + C(x'y)$;

C2 $C(x,0) = C(0,x) = 0$;

C3 $C(x,1) = C(1,x) = x$.

Choquet and Sugeno Integrals

Fuzzy measures: the idea

- Fuzzy measures are used for evaluating the relationship between the elements to be aggregated.
- They allow to represent the importance of the different coalitions that may be constructed with the different inputs.

Fuzzy measures

Definition

Let $N = \{1, \cdots, n\}$. A function $\mathfrak{m} : 2^N \to [0,1]$ is a discrete fuzzy measure if, for all $X, Y \subseteq N$, it satisfies the following properties:

($\mathfrak{m}1$) Increasingness: if $X \subseteq Y$, then $\mathfrak{m}(X) \leq \mathfrak{m}(Y)$;

($\mathfrak{m}2$) Boundary conditions: $\mathfrak{m}(\emptyset) = 0$ and $\mathfrak{m}(N) = 1$.

- *Power measure*:

$$\mathfrak{m}_{PM}(A) = \left(\frac{|A|}{n}\right)^q, \text{ with } q > 0.$$

Utility of fuzzy measures: The Choquet integral

Definition

Let $\mathfrak{m} : 2^N \to [0,1]$ be a fuzzy measure. The discrete Choquet integral of $\mathbf{x} = (x_1, \ldots, x_n) \in [0,1]^n$ with respect to $\mathfrak{m}$ is defined as a function $C_{\mathfrak{m}} : [0,1]^n \to [0,1]$, given by

$$C_{\mathfrak{m}}(\mathbf{x}) = \sum_{i=1}^{n} \left(x_{(i)} - x_{(i-1)}\right) \cdot \mathfrak{m}\left(A_{(i)}\right),$$

where $\left(x_{(1)}, \ldots, x_{(n)}\right)$ is an increasing permutation on the input $\mathbf{x}$, that is, $0 \leq x_{(1)} \leq \ldots \leq x_{(n)}$, with the convention that $x_{(0)} = 0$, and $A_{(i)} = \{(i), \ldots, (n)\}$ is the subset of indices of the $n - i + 1$ largest components of $\mathbf{x}$.

The Choquet integral is a continuous piecewise linear idempotent aggregation function

Utility of fuzzy measures: The Sugeno integral

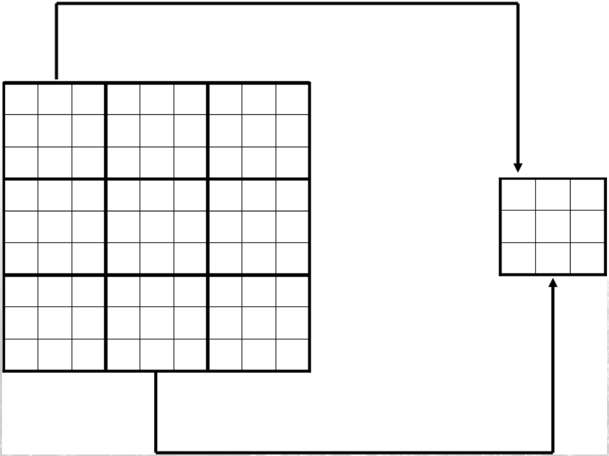
Definition

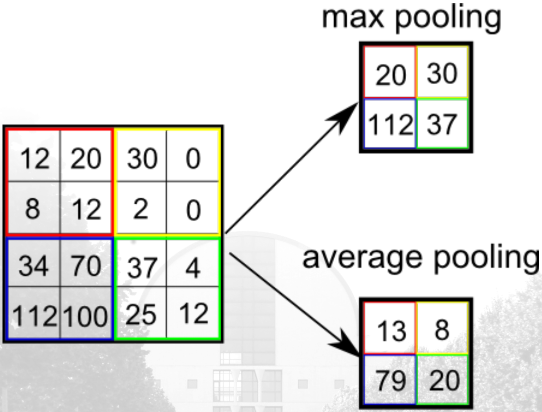
Let $\mathfrak{m} : 2^N \to [0,1]$ be a fuzzy measure. The discrete Sugeno integral of $\mathbf{x} = (x_1, \ldots, x_n) \in [0,1]^n$ with respect to $\mathfrak{m}$ is defined as a function $S_\mathfrak{m} : [0,1]^n \to [0,1]$, given by

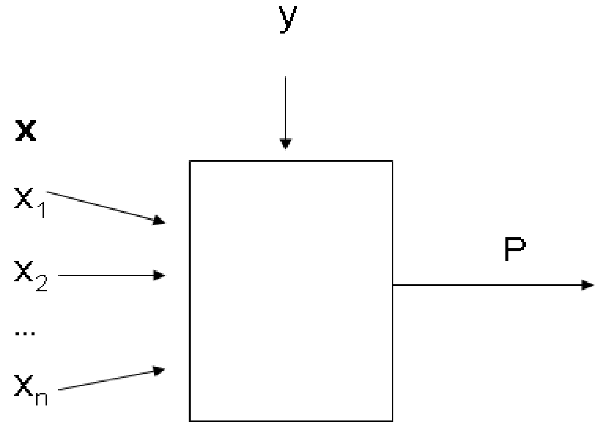
$$S_\mathfrak{m}(\mathbf{x}) = \bigvee_{i=1}^{n} \min\left\{x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right\}.$$

where $\left(x_{(1)}, \ldots, x_{(n)}\right)$ is an increasing permutation on the input $\mathbf{x}$, that is, $0 \leq x_{(1)} \leq \ldots \leq x_{(n)}$, with the convention that $x_{(0)} = 0$, and $A_{(i)} = \{(i), \ldots, (n)\}$ is the subset of indices of the $n - i + 1$ largest components of $\mathbf{x}$.

The problem of choosing the best fusion function




max pooling
12
20
30
0
8
12
2
0
34
70
37
4
112
100
25
12
20
30
112
37
average pooling
13
8
79
20



$$P(x_1, \ldots, x_n, y) = \sum_{i=1}^{n} (x_i - y)^4$$

Penalty functions

Definition

A penalty function is a mapping

$$P : [a,b]^{n+1} \to \mathbb{R}^+ = [0,\infty]$$

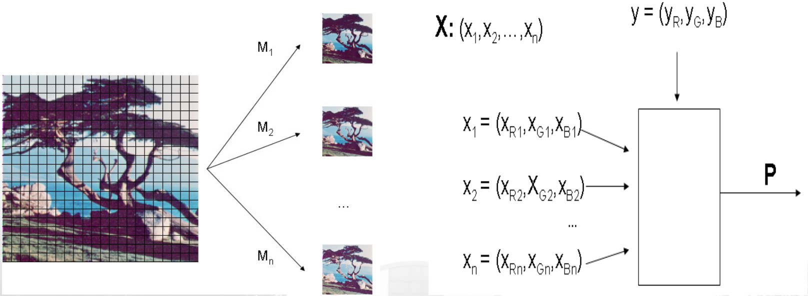
such that:

1. $P(\mathbf{x}, y) = 0$ if $x_i = y$ for every $i = 1, \cdots, n$;
2. $P(\mathbf{x}, y)$ is quasi-convex in y for every $\mathbf{x}$; that is,

$$P(\mathbf{x}, \lambda \cdot y_1 + (1-\lambda) \cdot y_2) \leq \max(P(\mathbf{x}, y_1), P(\mathbf{x}, y_2))$$

Aggregation functions based on penalties. Tomasa Calvo, Gleb Beliakov, Fuzzy Sets and Systems, 161 (10), 1420-1436 (2010)

On the definition of penalty functions in data aggregation. Humberto Bustince, Gleb Beliakov, Gracaliz Pereira Dimuro, Benjamin Bedregal, Radko Mesiar, Fuzzy Sets and Systems, 323 (15), 1-18 (2017)



.

- *Construction of image reduction operators using averaging aggregation functions.* D. Paternain, J. Fernandez, H. Bustince, R. Mesiar, G. Beliakov Fuzzy Sets and Systems, 261, 87-111 (2015)
- *Consensus in multi-expert decision making problems using penalty functions defined over a Cartesian product of lattices.* H. Bustince, E. Barrenechea, T. Calvo, S. James, G. Beliakov Information Fusion 17, 56–64 (2014)

Pre-aggregation functions

A different problem

The monotonicity problem

We are asking for **monotonicity**

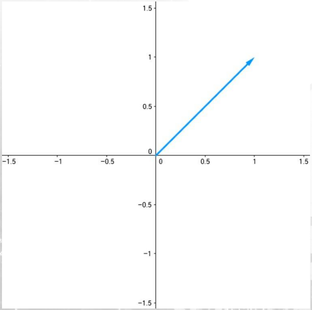
But some fusion methods are not monotone:

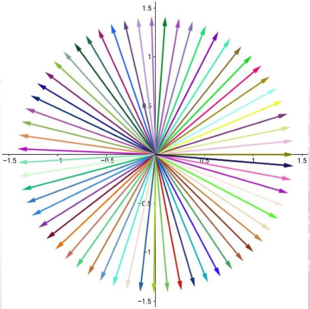
- Statistical operators (the mode)
- Implication functions
- Similarity measures
- Distances

So then?

One step ahead: directional monotonicity

- Weak monotonicity along the direction $(1, \ldots, 1)$ (2015, T. Wilkin, G. Beliakov)
- Generalization: Let's consider any direction $\vec{r} \in \mathbb{R}^n$





The formal definition

Definition

Let $\vec{r}$ be a real vector ($\vec{r} \neq 0$). A fusion function $F : [0,1]^n \to [0,1]$ is $\vec{r}$-increasing if for every $\mathbf{x} \in [0,1]^n$ and for every $c > 0$ such that $\mathbf{x} + c\vec{r} \in [0,1]^n$ it holds that:

$$F(\mathbf{x} + c\vec{r}) \geq F(x)$$

Some examples:

- Every implication function $I : [0,1]^2 \to [0,1]$ is $(-1,1)$-increasing.
- $F(x,y) = x - \max(0,(x-y)^2)$ is $(1,1)$-increasing and $(0,1)$-decreasing, but it is not $(1,0)$-increasing nor $(1,0)$-decreasing.

Directional monotonicity of fusion functions, H. Bustince, J. Fernandez, A. Kolesárová, R. Mesiar, European Journal of Operational Research 244 (1), 300-308 (2015).

Even further: ordered directionally monotone functions

Definition

Let $F : [0,1]^n \to [0,1]$ be a fusion function and let $\vec{r} \neq \vec{0}$ be an n-dimensional vector. F is said to be ordered directionally (OD) $\vec{r}$-increasing if for any $\mathbf{x} \in [0,1]^n$, for any $c > 0$ and for any permutation $\sigma : \{1,\ldots,n\} \to \{1,\ldots,n\}$ with $x_{\sigma(1)} \geq \cdots \geq x_{\sigma(n)}$ and such that

$$1 \geq x_{\sigma(1)} + cr_1 \geq \cdots \geq x_{\sigma(n)} + cr_n \geq 0$$

it holds that

$$F(\mathbf{x} + c\vec{r}_{\sigma^{-1}}) \geq F(\mathbf{x})$$

where $\vec{r}_{\sigma^{-1}} = (r_{\sigma^{-1}(1)}, \ldots, r_{\sigma^{-1}(n)})$

Ordered Directionally Monotone Functions: Justification and Application, H. Bustince, E. Barrenechea, M. Sesma-Sara, J. Lafuente, G. P. Dimuro, R. Mesiar, A. Kolesárová, IEEE Transactions on Fuzzy Systems 26 (4), 2237–2250 (2017).

Definition

An aggregation function is a function $M : [0,1]^n \to [0,1]$ such that:

1. M is increasing;
2. $M(0, \ldots, 0) = 0$
3. $M(1, \ldots, 1) = 1$.

Definition

A function $F : [0,1]^n \to [0,1]$ is said to be an n-ary pre-aggregation function if the following conditions hold:

- (PA1) There exists a real vector $\vec{r} \in [0,1]^n$ $(\vec{r} \neq \vec{0})$ such that F is $\vec{r}$-increasing.
- (PA2) F satisfies the boundary conditions: $F(0,\ldots,0) = 0$ and $F(1,\ldots,1) = 1$.

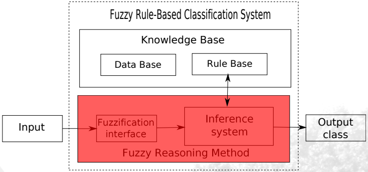
If F is a pre-aggregation function with respect to a vector $\vec{r}$ we just say that F is an $\vec{r}$-pre-aggregation function.

Preaggregation Functions: Construction and an Application. Giancarlo Lucca; José Antonio Sanz; Gracaliz Pereira Dimuro; Benjamín Bedregal; Radko Mesiar; Anna Kolesárová; Humberto Bustince, IEEE Transactions on Fuzzy Systems, 24(2), 260 - 272 (2016).

The generalized Choquet integral and the classification problem

R_j : If x_{p1} is A_{j1} and ... and x_{pn} is A_{jn} then *Class* $= C_j$ with RW_j

- Fuzzy Reasoning Method:
  1. *Matching degree:*
     $\mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn}))$
  2. *Association degree:*
     $$b_j^k = h(\mu_{A_j}(x_p), RW_j^k)$$
  3. *Association degree by classes:*
     $$Y_k = f(b_j^k,\ b_j^k > 0)$$
  4. *Classification:*
     $$C_{best} = \arg \max_{k=1,\cdots,M} (Y_k)$$



- k = 1, ..., M (n. classes).
- j = 1,..., L (n. rules).

O. Cordón, M. J. del Jesús, F. Herrera: A proposal on reasoning methods in fuzzy rule-based classification systems. Int. J. Approx. Reason., 20:1 (1999) 21–45.

Generalized Choquet integral

The state-of-art of the generalizations of the Choquet integral: From aggregation and pre-aggregation to ordered directionally monotone functions. G. P. Dimuro, J. Fernández, B. Bedregal, R. Mesiar, J. A. Sanz, G. Lucca, H. Bustince Information Fusion, 57, 27–43 (2020)

We are going to use the Choquet integral... with a "small" change:

The first idea

$$C_{\mathfrak{m}}(\mathbf{x}) = \sum_{i=1}^{n} \left(x_{(i)} - x_{(i-1)}\right) \cdot \mathfrak{m}\left(A_{(i)}\right)$$

$$\Downarrow \qquad\qquad\qquad \Downarrow$$

$$C_{\mathfrak{m}}^{M}(\mathbf{x}) = \sum_{i=1}^{n} M\left(x_{(i)} - x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)$$

But...

- Testing results

Dataset	WR	Power_GA+Ham
App	**84.89**	82.99
Bal	82.08	**82.72**
Ban	84.30	**85.96**
Bnd	68.56	**72.13**
Bup	61.16	**65.80**
Cle	55.23	**55.58**
Eco	75.61	**80.07**
Gal	**63.11**	63.10
Hab	71.22	**72.21**
Hay	79.46	**79.49**
Iri	**94.67**	93.33
Led	**69.80**	68.60
Mag	79.60	**79.76**
New	94.42	**95.35**
Pag	**94.52**	94.34
Pho	82.01	**83.83**
Pim	**75.38**	73.44
Rin	**90.00**	88.79
Sah	67.31	**70.77**
Sat	**80.40**	**80.40**
Seg	92.99	**93.33**
Tit	**78.87**	**78.87**
Two	84.32	**85.27**
Veh	67.62	**68.20**
Win	94.36	**96.63**
Wis	96.49	**96.78**
Yea	**56.54**	56.53
Mean	78.70	**79.42**

If we take $C_{\mathfrak{m}}^{M}(\mathbf{x})$:

$$\sum_{i=1}^{n} M\left(x_{(i)} - x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right),$$

we overcome the winning rule (the maximum).

We want more: let's go for FURIA and FARC!!!

WHAT ELSE CAN WE DO??

One step more

$$C_{\mathfrak{m}}(\mathbf{x}) = \sum_{i=1}^{n} \left(x_{(i)} - x_{(i-1)}\right) \cdot \mathfrak{m}\left(A_{(i)}\right)$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$C_{\mathfrak{m}}^{M}(\mathbf{x}) = \sum_{i=1}^{n} M\left(x_{(i)} - x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)$$

One step more

$$C_{\mathfrak{m}}(\mathbf{x}) = \sum_{i=1}^{n} \left(x_{(i)} - x_{(i-1)}\right) \cdot \mathfrak{m}\left(A_{(i)}\right)$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$C_{\mathfrak{m}}^{M}(\mathbf{x}) = \sum_{i=1}^{n} M\left(x_{(i)} - x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)$$

The second idea

$$C_{\mathfrak{m}}(\mathbf{x}) = \sum_{i=1}^{n} \left(x_{(i)} \cdot \mathfrak{m}(A_{(i)}) - x_{(i-1)} \cdot \mathfrak{m}\left(A_{(i)}\right)\right)$$

$$\Downarrow \qquad\qquad \Downarrow$$

$$C_{\mathfrak{m}}^{F_1,F_2}(\mathbf{x}) = \sum_{i=1}^{n} F_1(x_{(i)}, \mathfrak{m}(A_{(i)})) - F_2(x_{(i-1)}, \mathfrak{m}(A_{(i)}))$$

To get a value smaller than 1 we do:

$$C_{\mathfrak{m}}^{(F_1,F_2)}(\mathbf{x}) = \min\left\{1, \sum_{i=1}^{n} F_1\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - F_2\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)\right\},$$

Conditions for F_1 and F_2?

Proposition *

Let $F_1, F_2 : [0,1]^2 \to [0,1]$ be two bivariate functions such that, for every $x, y \in [0,1]$, it holds that:

1. F_1 is $(1,0)$-increasing;
2. $F_1(0,x) = F_2(0,x)$;
3. $F_1(0,1) = F_2(0,1) = 0$;
4. $F_1(1,1) = 1$;
5. $F_1(x,y) \geq F_2(x,y)$.

Then, for any fuzzy measure $\mathfrak{m}$, the function $C_{\mathfrak{m}}^{(F_1,F_2)}$ is well-defined and satisfies:

$$0 \leq C_{\mathfrak{m}}^{(F_1,F_2)}(\mathbf{x}) \leq 1$$

for every $\mathbf{x} \in [0,1]^n$.

Does it work?

Proposition

If we take:

- $F_1(x, y) = \sqrt{xy}$
- $F_2(x, y) = \max(x + y - 1, 0)$,

then

$$C_{\mathfrak{m}}^{(F_1,F_2)}(\mathbf{x}) = \min\left\{1, \sum_{i=1}^{n} F_1\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - F_2\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)\right\}$$

is a non-averaging pre-aggregation function.

Results

Table 1: Results achieved in testing considering the $F_1 F_2$ approach

Dataset	FURIA	AC	ProbSum	GM_LK
appendicitis	**87.71**	83.03	85.84	84.89
balance	83.68	85.92	87.20	**89.76**
banana	**88.57**	85.30	84.85	85.23
bands	69.40	68.28	68.82	**70.49**
bupa	**70.14**	67.25	61.74	66.67
cleveland	56.57	56.21	**59.25**	58.57
contraceptive	**54.17**	53.16	52.21	53.50
ecoli	80.06	82.15	80.95	**84.53**
glass	**72.91**	65.44	64.04	64.99
haberman	72.55	**73.18**	69.26	**73.18**
hayes-roth	**81.00**	77.95	77.95	79.43
ion	**89.75**	88.90	88.32	89.75
iris	94.00	94.00	**95.33**	94.67
led7digit	**71.80**	69.60	69.20	69.60
magic	80.65	**80.76**	80.39	80.18
newthyroid	94.88	94.88	94.42	**96.28**
pageblocks	95.25	95.07	94.52	**95.98**
penbased	92.45	92.55	**93.27**	92.64
phoneme	**85.90**	81.70	82.51	82.44
pima	**76.17**	74.74	75.91	75.26
ring	85.54	**90.95**	90.00	90.41
saheart	70.33	68.39	69.69	**70.56**
satimage	**82.27**	79.47	80.40	79.47
segment	**97.32**	93.12	92.94	92.86
shuttle	**99.68**	95.59	94.85	97.33
sonar	78.90	78.36	82.24	**83.23**
spectfheart	77.88	77.88	77.90	**80.12**
titanic	78.51	**78.87**	**78.87**	**78.87**
twonorm	88.11	90.95	90.00	**91.76**
vehicle	**70.21**	68.56	68.09	68.67
wine	93.78	**96.03**	94.92	**96.03**
wisconsin	96.63	96.63	**97.22**	96.34
yeast	58.22	58.96	**59.03**	58.96
Mean	**81.06**	80.12	80.07	80.99

The generalized Sugeno integral and the computational brain

The case of the computational brain

Consider the problem of determining whether a subject is thinking of moving the left or the right hand.
EEG nowadays are not able to determine this

$\Rightarrow$

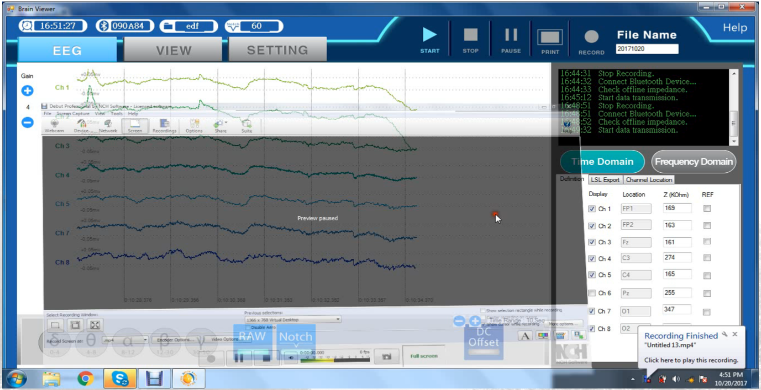
The case of the computational brain

Consider the problem of determining whether a subject is thinking of moving the left or the right hand.

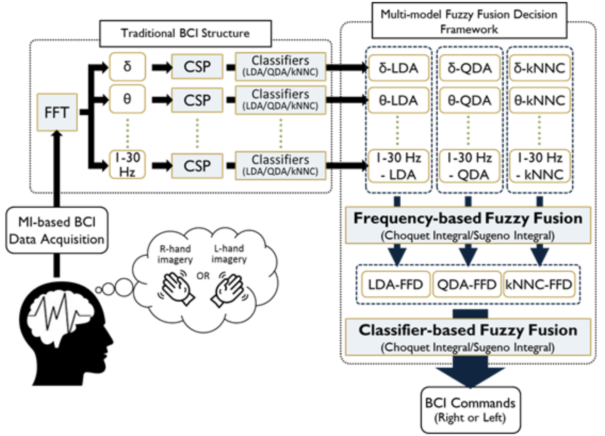
Classification problem with two classes

Not appropriate for deep learning!

Computational brain



STRUCTURE OF THE ALGORITHM:



Multimodal Fuzzy Fusion for Enhancing the Motor-Imagery-based Brain Computer Interface, Li-Wei Ko, Yi-Chen Lu, Humberto Bustince, Yu-Cheng Chang, Yang Chang, Javier Fernandez, Yu-Kai Wang, Jose Antonio Sanz, Gracaliz Pereira Dimuro, Chin-Teng Lin, IEEE Computational Intelligence Magazine,14 (1), 96–106 (2019)

Sugeno-based construction method

Discrete Sugeno integral $S_{\mathfrak{m}}\colon [0,1]^n \to [0,1]$ can be written as

$$S_{\mathfrak{m}}(\mathbf{x}) = \bigvee_{i=1}^{n} \min\left\{x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right\}.$$

What happens if we replace the minimum by another aggregation function?

Discrete Sugeno integral $S_{\mathfrak{m}}\colon [0,1]^n \to [0,1]$ can be written as

$$S_{\mathfrak{m}}(\mathbf{x}) = \bigvee_{i=1}^{n} \min\left\{x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right\}.$$

What happens if we replace the minimum by another aggregation function?

$$S_{\mathfrak{m}}^{M}(\mathbf{x}) = \bigvee_{i=1}^{n} M\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right). \tag{1}$$

Proposition

Let $M\colon [0,1]^2 \to [0,1]$ be a function increasing in the first variable and let for each $y \in [0,1]$, $M(0,y) = 0$ and $M(1,1) = 1$. Then $S_{\mathfrak{m}}^{M}$ is a pre-aggregation function for any fuzzy measure $\mathfrak{m}$.

Sugeno-like construction method of pre-aggregation functions

- Let $M\colon [0,1]^2 \to [0,1]$ be any aggregation function. Then $S_{\mathfrak{m}}^{M}\colon [0,1]^n \to [0,1]$ is also an aggregation function, independently of $\mathfrak{m}$.
- Consider the function F, $F(x,y) = x|2y-1|$. Note that F is a proper pre-aggregation function which satisfies our constraints, and thus, for any $\mathfrak{m}$, the function $S_{\mathfrak{m}}^{F}\colon [0,1]^n \to [0,1]$, $S_{\mathfrak{m}}^{F}(\mathbf{x}) = \bigvee_{i=1}^{n} F\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right)$ is a pre-aggregation function (even an aggregation function thought F is not).

STRUCTURE OF THE ALGORITHM:

We make two steps:

1. Fuse the results for each band and each classifier.
2. Fuse the global result of each classifier.

Pre-aggregations vs. BCI

We use aggregation and pre-aggregation functions to fuse the results of each classifier

- M-S1: Sugeno.
- M-S2: S^M integral with M the Hamacher t-norm:

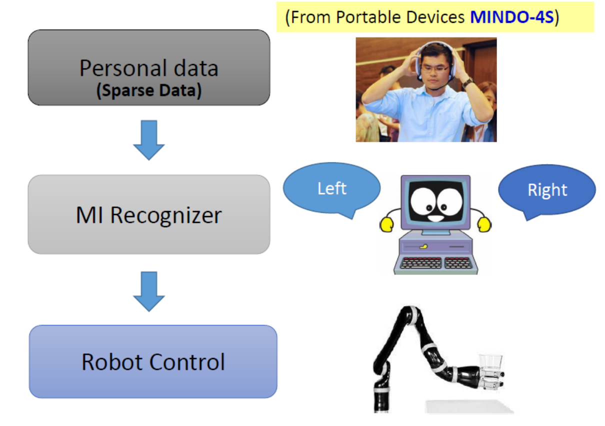
$$F(x,y) = \begin{cases} 0 & \text{if } x = y = 0 \\ \frac{xy}{x+y-xy} & \text{otherwise.} \end{cases}$$

- M-S3: S^M integral with M given by:

$$M(x,y) = x|2y-1|$$

- *Multimodal Fuzzy Fusion for Enhancing the Motor-Imagery-based Brain Computer Interface*, Li-Wei Ko, Yi-Chen Lu, Humberto Bustince, Yu-Cheng Chang, Yang Chang, Javier Fernandez, Yu-Kai Wang, Jose Antonio Sanz, Gracaliz Pereira Dimuro, Chin-Teng Lin, IEEE Computational Intelligence Magazine,14 (1), 96–106 (2019)

The BCI experiment



One video

CT Lin's BCI Lab in Taiwan/Australia

d-Choquet integrals

We can modify the Choquet integral in a different way:

The idea of d-integrals

$$C_{\mathfrak{m}}(\mathbf{x}) = \sum_{i=1}^{n} \left(x_{(i)} - x_{(i-1)}\right) \cdot \mathfrak{m}\left(A_{(i)}\right)$$

$$\Downarrow \qquad\qquad\qquad \Downarrow$$

$$C_{\mathfrak{m}}^{M}(\mathbf{x}) = \sum_{i=1}^{n} d\left(x_{(i)}, x_{(i-1)}\right) \cdot \mathfrak{m}\left(A_{(i)}\right)$$

where d is a dissimilarity.

Dissimilarity functions

Definition

A function $\delta : [0,1]^2 \to [0,1]$ is called a restricted dissimilarity function on $[0,1]$ if it satisfies, for all $x, y, z \in [0,1]$, the following conditions:

1. $\delta(x,y) = \delta(y,x)$;
2. $\delta(x,y) = 1$ if and only if $\{x,y\} = \{0,1\}$;
3. $\delta(x,y) = 0$ if and only if $x = y$;
4. if $x \leq y \leq z$, then $\delta(x,y) \leq \delta(x,z)$ and $\delta(y,z) \leq \delta(x,z)$.

H. Bustince, E. Barrenechea, M. Pagola, Relationship between restricted dissimilarity functions, restricted equivalence functions and normal en-functions: Image thresholding invariant, Pattern Recognition Letters 29 (4) (2008) 525 – 536.

d-Choquet integrals: Choquet integrals based on dissimilarities. H. Bustince, R. Mesiar, J. Fernandez, M. Galar, D. Paternain, A. Altalhi, G.P. Dimuro, B. Bedregal, Z Takáč Fuzzy Sets and Systems, available online

Definition

Let $N = \{1, \ldots, n\}$ be a positive integer and $\mathfrak{m} : 2^N \to [0,1]$ be a fuzzy measure on N. Let $\delta : [0,1]^2 \to [0,1]$ be a restricted dissimilarity function. An n-ary discrete d-Choquet integral on $[0,1]$ with respect to $\mathfrak{m}$ and δ is defined as a mapping $C_{\mathfrak{m},\delta} : [0,1]^n \to [0,n]$ such that

$$C_{\mathfrak{m},\delta}(x_1, \ldots, x_n) = \sum_{i=1}^{n} \delta(x_{\sigma(i)}, x_{\sigma(i-1)})\mathfrak{m}\left(A_{\sigma(i)}\right) \tag{2}$$

where σ is a permutation on N satisfying $x_{\sigma(1)} \leq \ldots \leq x_{\sigma(n)}$, with the convention $x_{\sigma(0)} = 0$ and $A_{\sigma(i)} = \{\sigma(i), \ldots, \sigma(n)\}$.

Observe that, in general, the range of $C_{\mu,\delta}$ is a subset of $[0, n]$. Since, for some applications, it may be desired that the range of $C_{\mu,\delta}$ would be $[0, 1]$, we often impose the following condition:

(P1) $\delta(0, x_1) + \delta(x_1, x_2) + \ldots + \delta(x_{n-1}, x_n) \leq 1$ for all $x_1, \ldots, x_n \in [0, 1]$ where $x_1 \leq \ldots \leq x_n$.

Proposition

Let $C_{\mu,\delta} : [0, 1]^n \to [0, n]$ be an n-ary discrete d-Choquet integral on $[0, 1]$ with respect to μ and δ. If δ satisfies the condition $(P1)$, then

$$C_{\mu,\delta}(x_1, \ldots, x_n) \in [0, 1]$$

for all $x_1, \ldots, x_n \in [0, 1]$ and for any measure μ.

Theorem

Let $\delta : [0,1]^2 \to [0,1]$ be a restricted dissimilarity function. Consider $f_\delta : [0,1] \to [0,1]$, defined, for each $x \in [0,1]$, by

$$f_\delta(x) = \delta(x, 0)$$

and $\delta^* : [0,1]^2 \to [0,1]$, defined, for each $x, y \in [0,1]$, by

$$\delta^*(x, y) = |f_\delta(x) - f_\delta(y)|.$$

Then δ^* is a restricted dissimilarity function which satisfies (P1) if and only if f_δ is injective.

Relation with the standard Choquet integral

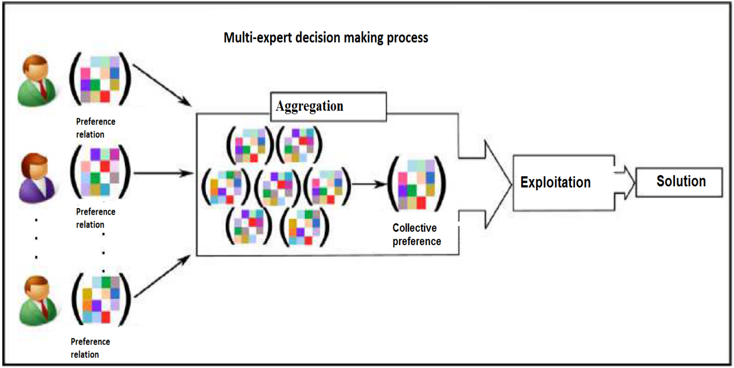
Theorem

Let n be a positive integer, $N = \{1, \ldots, n\}$, $\mathfrak{m} : 2^N \to [0,1]$ be a fuzzy measure on N, $\delta : [0,1]^2 \to [0,1]$ be the function $\delta(x,y) = |x - y|$, $C_{\mathfrak{m},\delta} : [0,1]^n \to [0,1]$ be an n-ary discrete d-Choquet integral on $[0,1]$ with respect to $\mathfrak{m}$ and δ and $C_{\mathfrak{m}} : [0,1]^n \to [0,1]$ be an n-ary discrete Choquet integral on $[0,1]$ with respect to $\mathfrak{m}$. Then

$$C_{\mathfrak{m},\delta}(x_1, \ldots, x_n) = C_{\mathfrak{m}}(x_1, \ldots, x_n)$$

for all $x_1, \ldots, x_n \in [0,1]$.

Decision making with d-integrals

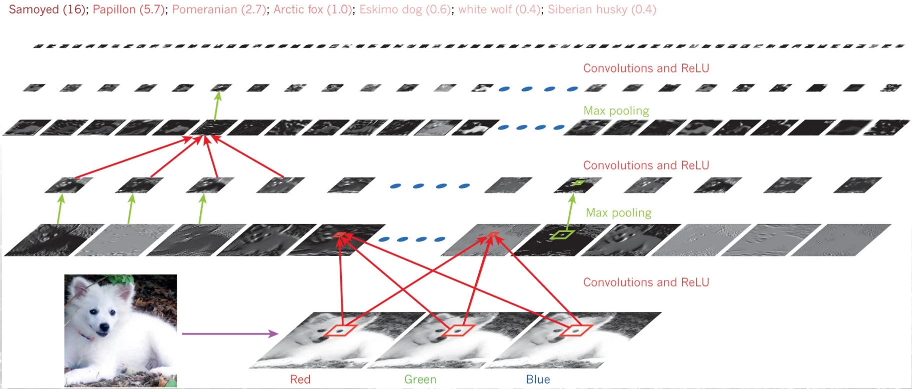


Data fusion functions using real numbers

Objective: To replace subsampling mechanisms in Convolutional Neural Networks (CNNs).

- The features extracted by convolution layers are usually aggregated using the mean or the maximum.
- The chosen pooling function ignores possible coalitions among data.
- The choice of the pooling function acts as an hyperparameter for the model.

Convolutional Neural Network: CNN

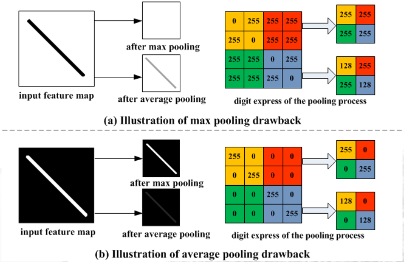


LeCun, Bengio, and Hinton, "Deep learning"

Pooling

It aggregates the features extracted by the convolution layers:
$\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$;
$\mathbf{A}(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$

- Summary of relevant information in a local way.
- Max pooling: $\mathbf{A}(\mathbf{x}) = \max_{i=1}^{n} \mathbf{x}_i$
- Avg pooling: $\mathbf{A}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$



Yu et al., "Mixed pooling for convolutional neural networks"

Increasing functions (I)

From now on, we assume that $2 \leq n \in \mathbb{N}$, $1 \leq r \in \mathbb{N}$.

Notation

We denote by $\mathbf{x}_\sigma = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ a permutation of $\mathbf{x}$.
Si $x_{\sigma(1)} \leq \cdots \leq x_{\sigma(n)}$, decimos que $\mathbf{x}_\sigma = \mathbf{x}_{(\nearrow)}$.

A function $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}$ is increasing if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \mathbf{x} \leq \mathbf{y}$ implies that $\mathbf{A}(x) \leq \mathbf{A}(y)$

Example

Let $r \in 1, \ldots, n$. Denote by $\mathbf{OS}_r$ the r-th order statistics, that is, the function $\mathbf{OS}_r : \mathbb{R}^n \to \mathbb{R}$ given, if $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n : \mathbf{OS}_r(\mathbf{x}) = x_{\sigma(r)}$, by $\mathbf{x}_\sigma = \mathbf{x}_{(\nearrow)}$.

Increasing functions (II)

Example

Denote by $\mathbf{AM}$ the arithmetic mean, that is, the function $\mathbf{AM} : \mathbb{R}^n \to \mathbb{R}$, given by $\mathbf{AM}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$, if $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

A fuuzy measure in N is a mapping $\nu : 2^N \to [0, +\infty)$ such that

1. $\nu(\emptyset) = 0$
2. $S \subseteq T \subseteq N$ implies $\nu(S) \leq \nu(T)$

Example

The Sugeno integral associated to the fuzzy measure ν is the function $\mathbf{S}_\nu : \mathbb{R}^n \to \mathbb{R}$ give, for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$
$\mathbf{S}_\nu(\mathbf{x}) = \max(\min(x_{\sigma(1)}, \nu(\sigma(1), \ldots, \sigma(n)), \ldots, \min(x_{\sigma(n)}, \nu(\sigma(n))),$
donde $\mathbf{x}_\sigma = \mathbf{x}_{(\nearrow)}$

Generalized Sugeno integral (I)

Let $\mathbb{U}$ be a connected subset of $\mathbb{R}$ such that $0 \in \mathbb{U}$. A $\mathbb{U}$-fuzzy measure N is a function $\nu : 2^N \to \mathbb{U}$ such that

1. $\nu(\emptyset) = 0$
2. $S \subseteq T \subseteq N$ implies $\nu(S) \leq \nu(T)$

From the expression of Sugeno integral, we propose a generalization given by:

Definition

Let $\mathbb{U}$ and $\mathbb{I}$ be two connected subsets of $\mathbb{R}$ such that $0 \in \mathbb{U} \subseteq \mathbb{I}$. Let $\nu : 2^N \to \mathbb{U}$ be a $\mathbb{U}$-fuzzy measure. The functions $\mathrm{F} : \mathbb{I} \times \mathbb{U} \to \mathbb{I}$ and $\mathrm{G} : \mathbb{I}^n \to \mathbb{U}$ will be called ν-admissible if the function $\mathbf{A} : \mathbb{I}^n \to \mathbb{I}$ given by $\mathbf{A}(\mathbf{x}) = \mathrm{G}(\mathrm{F}(x_{\sigma(1)}, \nu(\sigma(1), \ldots, \sigma(n)), \ldots, \mathrm{F}(x_{\sigma(n)}, \nu(\sigma(n))), \text{donde}$ $\mathbf{x}_\sigma = \mathbf{x}_{(\nearrow)}$, is well defined.

Generalized Sugeno integral (II)

$$\mathbf{A}(\mathbf{x}) = \mathrm{G}(\mathrm{F}(x_{\sigma(1)}, \nu(\sigma(1), \ldots, \sigma(n)), \ldots, \mathrm{F}(x_{\sigma(n)}, \nu(\sigma(n)))$$

,

In our case, we fix:

- $G(\mathbf{x}) = \sum_{i=1}^{n} x_i$
- $F(x, y) = xy$

And we refer to this version of the generalized Sugeno integral as:

$$\mathbf{D}_\nu(\mathbf{x}) = \sum_{i=1}^{n} ((x_{\sigma(1)}\nu(\sigma(1), \ldots, \sigma(n)), \ldots, (x_{\sigma(n)}\nu(\sigma(n)))$$

,

Combination of increasing functions (I)

Our goal is to build increasing functions $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}$ as follows:

$$\mathbf{A}(\mathbf{x}) = \alpha_1 \mathbf{A}_1(\mathbf{x}) + \cdots + \alpha_n \mathbf{A}_n(\mathbf{x})$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_n : \mathbb{R}^n \to \mathbb{R}$ are increasing functions and $\mathbf{x} \in \mathbb{R}^n$

Notation

Let $\mathbf{A}_1, \ldots, \mathbf{A}_r : \mathbb{R}^n \to \mathbb{R}$ be increasing functions. We denote: $\mathcal{I}(\mathbf{A}_1, \ldots, \mathbf{A}_r) = (\alpha_1, \ldots, \alpha_r) \in \mathbb{R}^n$ such that $\alpha_1 \mathbf{A}_1 + \cdots + \alpha_n \mathbf{A}_n$ is increasing.

Combination of increasing functions (II)

- Combination of order statistics:

Proposition

Take $i_1, \ldots, i_r \in N$, $i_1 < \cdots < i_r$. *Then*
$\mathcal{I}(\mathbf{OS}_{i_1}, \ldots, \mathbf{OS}_{i_r}) = \{(\alpha_1, \ldots, \alpha_r) | \alpha_1, \ldots, \alpha_r \geq 0\}$

- Combination with the arithmetic mean:

Proposition

Take $i_1, \ldots, i_r \in N$, $i_1 < \cdots < i_r, r < n$. *Then*
$\mathcal{I}(\mathbf{AM}, \mathbf{OS}_{i_1}, \ldots, \mathbf{OS}_{i_r}) = \{(\alpha, \beta_1, \ldots, \beta_r) | \alpha, \alpha + n\beta_1, \ldots, \alpha + n\beta_r \geq 0\}$

Combination of increasing functions (III)

- Combination with Sugeno integral:

Definition

A fuzzy measure $\nu : 2^N \to [0, +\infty)$ is strict in $k \in N$ if, either $k = n$ or there is a permutation σ such that $\nu_k^\sigma > \nu_{k+1}^\sigma$. A fuzzy measure is strict if it is strict for every $k \in N$

Proposition

Take $i_1, \ldots, i_r \in N$, $i_1 < \cdots < i_r, r < n$. If there is $k \in N \setminus \{i_1, \ldots, i_r\}$ such that $\nu : 2^N \to [0, +\infty)$ is strict in k, then
$\mathcal{I}(\mathbf{OS}_{i_1}, \ldots, \mathbf{OS}_{i_r}, \mathbf{S}_\nu) = \{(\alpha_1, \ldots, \alpha_r, \beta) | \alpha_1, \ldots, \alpha_r, \beta \geq 0\}$

Proposition

Let $\nu : 2^N \to [0, +\infty)$ be a fuzzy measure. Then
$\mathcal{I}(\mathbf{AM}, \mathbf{S}_\nu) = \{(\alpha, \beta) | \alpha, \alpha + n\beta \geq 0\}$

Combination of increasing functions (IV)

- Combination with the integral $\mathbf{D}_\nu$:

Proposition

Take $M \subseteq N$ *and* $M' = N \setminus M$. *Take* $i_1, \ldots, i_r \in N$, $i_1 < \cdots < i_r$.
$\mathcal{I}(\mathbf{OS}_{i_1}, \ldots, \mathbf{OS}_{i_r}, \mathbf{D}_\nu) = \{(\alpha_1, \ldots, \alpha_r, \beta) | \alpha_i + \beta\nu(i, \ldots, n) \geq 0$ *if* $i \in M$ *y* $\beta\nu(i, \ldots, n) \geq 0$ *if* $i \in M'\}$

Proposition

$\mathcal{I}(\mathbf{AM}, \mathbf{D}_\nu) = \{(\alpha, \beta) | \beta \geq 0, \alpha + n\beta\nu(n) \geq 0$, *o*
$\beta \leq 0, \alpha + n\beta\nu(1, \ldots, n)\}$

Proposition

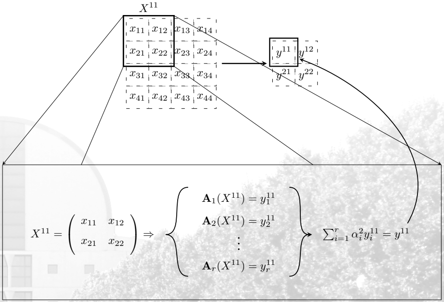
Let $\nu : 2^N \to [0, +\infty)$ *be a strict fuzzy measure. Then*
$\mathcal{I}(\mathbf{D}_\nu, \mathbf{S}_\nu) = \{(\alpha, \beta) | \alpha, \alpha\nu(n) + \beta \geq 0\}$

CombPool layer

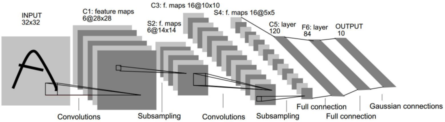
Given r increasing functions $\mathbf{A}_1, \ldots, \mathbf{A}_r$ such that $\mathbf{A}_i : \mathbb{R}^n \to \mathbb{R}$ for every $i = 1, \ldots, r$:

$$y = \sum_{i=1}^{r} \alpha_i^2 \cdot \mathbf{A}_i(\mathbf{x})$$

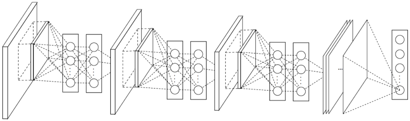
- The contribution of each function is learnt.
- The resulting combinations are always increasing.
- Each function provides different information.



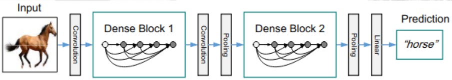
LeNet-5[1]



Network in Network[2][3]



DenseNet[4]

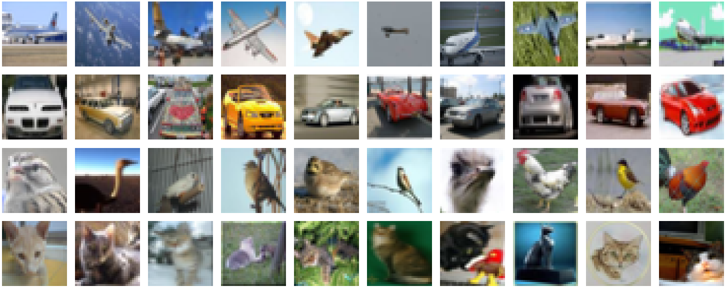


[1] LeCun, Bottou, et al., "Gradient-based learning applied to document recognition".
[2] Lin, Chen, and Yan, "Network in network".
[3] Lee, Gallagher, and Tu, "Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree".
[4] Huang et al., "Densely connected convolutional networks".

Dataset: CIFAR-10[a]

[a] Krizhevsky, Hinton, et al., "Learning multiple layers of features from tiny images".

- Real color images (32×32 pixels)
- 50000 training images
- 10000 test images
- 10 classes



- Experiment 1: Individual functions
  - Arithmetic mean
  - Maximum
  - Minimum
  - Median
  - $\mathbf{S}_\nu$: Sugeno integral
  - $\mathbf{D}_\nu$: Generalized Sugeno integral
- Experiment 2: CombPools layers
  - Combinations of increasing functions

Results: Experiment 1

Table: Accuracy rate for individual functions.

	A1	A2	A3
AM	77.08	87.65	**89.25**
Max	**77.39**	87.85	87.99
Min	70.24	87.61	88.28
Median	70.62	87.07	88.76
$\mathbf{S}_\nu$	72.47	86.79	88.97
$\mathbf{D}_\nu$	73.42	**88.70**	87.20

Table: Accuracy rates for models using CombPool layers.

	A1	A2	A3
Min + Max	**77.02**	87.42	88.55
Min + Max + Median	76.91	**87.43**	**89.77**
AM + Min	75.04	87.23	89.48
AM + Max	**77.23**	**87.78**	86.99
AM + Min + Max	77.17	87.25	88.52
AM + Min + Max + Median	77.04	87.57	**89.83**

Results: Experiment 2 (II)

Table: Accuracy rates for models using CombPool layers.

	A1	A2	A3
$\mathbf{S}_\nu$ + Min	72.41	87.46	88.58
$\mathbf{S}_\nu$ + Max	77.09	87.60	89.48
$\mathbf{S}_\nu$ + Min + Max	**77.30**	87.01	89.42
$\mathbf{S}_\nu$ + Min + Max + Median	77.03	87.29	**89.66**
$\mathbf{S}_\nu$ + AM	76.93	**88.23**	86.99
$\mathbf{D}_\nu$ + Min	72.19	88.51	89.03
$\mathbf{D}_\nu$ + Max	76.80	88.20	89.58
$\mathbf{D}_\nu$ + Min + Max	**77.81**	**88.61**	89.83
$\mathbf{D}_\nu$ + Min + Max + Median	76.15	88.30	89.75
$\mathbf{D}_\nu$ + AM	76.39	88.40	**89.87**
$\mathbf{D}_\nu$ + $\mathbf{S}_\nu$	74.92	88.03	89.68

Conclusions

- More parameterized models get more benefits using CombPool layers
- Individually poor functions are good candidates for combinations.
- Combinations with the generalization D_ν of Sugeno integral usually provide the best results.

Future lines

- To study again pooling functions for CombPool layers
- To consider other aggregation functions, such as Choquet integrals and their generalizations.
- To use increasing combinations in other layers of the network (GAP, activation functions...)

The case of Choquet Integrals in CNNs

Thanks for your attention

Questions?

bustince@unavarra.es
Departamento de Estadística, Informática y Matemáticas (UPNA)
Grupo de Inteligencia Artificial y Razonamiento Aproximado (GIARA)