Tutorial, IEEE WCCI 2016, Vancouver, Canada

Fuzzy Logic & Machine Learning

H.R.Tizhoosh

KIMIA Lab, University of Waterloo, Canada

tizhoosh.uwaterloo.ca :: tizhoosh@uwaterloo.ca

Overview

- Fuzzy Logic
- Machine Learning
- Fuzzy Logic and Machine Learning
- A Critical Review

Fuzzy Logic

A Brief History of Fuzzy Logic

- 1920: Three-Valued Logic (Jan Łukasiewicz)

NOT(A)

A	¬A
F	T
U	U
T	F

AND(A, B)

A ∧ B		B		
		F	U	T
A	F	F	F	F
	U	F	U	U
	T	F	U	T

OR(A, B)

A ∨ B		B		
		F	U	T
A	F	F	U	T
	U	U	U	T
	T	T	T	T

Source: https://en.wikipedia.org/wiki/Three-valued_logic

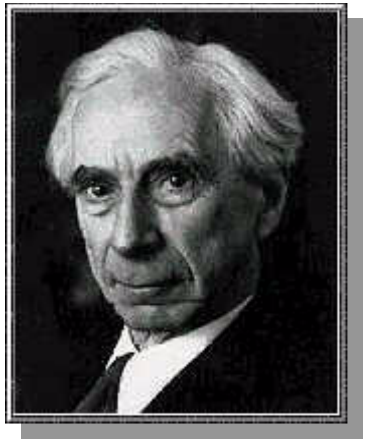
Vagueness

Source: www.uspsoig.gov

A Brief History of Fuzzy Logic

- 1923: Paper on vagueness (Bertrand Russell)

All traditional logic habitually assumes that *precise symbols* are being employed. It is therefore not applicable to this terrestrial life, but only to an imagined celestial existence.



Bertrand Russell

Vagueness

A Brief History of Fuzzy Logic







Vagueness

A Brief History of Fuzzy Logic

The vagueness of the word **chair** is typical of all terms whose application involves the use of the senses. In all such cases "**borderline cases**" and "**doubtful objects**" are easily found to which we are unable to say either that the class name does or does not apply.

Max Black
Vagueness: An exercise in logical analysis, 1937

Vagueness

Probability Theory

Yesterday
Will tomorrow rain?

Fuzzy Logic

Today
It is raining!

But what is the rain intensity?
Drizzle, light, moderate, heavy, extreme?

Event occurs

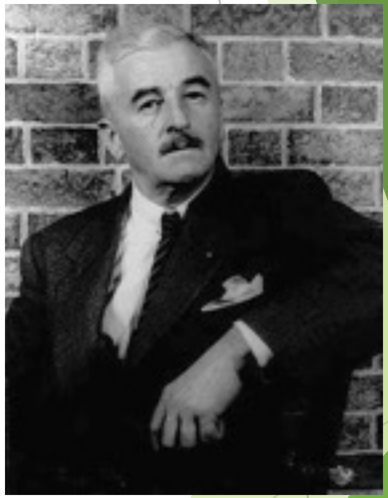
Uncertain vs. Vague

~~On Friday, April 23, 1954, at 2:33 p.m.~~ as Miss Emily Grierson died, ~~100% of population~~ of our town went to her funeral: the men through a in 5% of cases usual respectful affection for a fallen monument, 97.23% of women out of curiosity to see the inside of her house, which no one save an 76 years and 7 months old manservant– 50% gardener, 50% cook--had seen in >= 10 years.

Uncertain vs. Vague

When Miss Emily Grierson died, our **whole town** went to her funeral: the men through **a sort of** respectful affection for a fallen monument, the women **mostly** out of curiosity to see the inside of her house, which no one save an **old** manservant--**a combined** gardener and cook--had seen in **at least ten** years.

William Faulkner
A Rose for Emily



Uncertain vs. Vague

Modeling of Real World

Modeling of Human Reasoning

↓

Set Theory
Logic
Measure Theory

- 1965: Paper of Fuzzy Sets (Lotfi Zadeh)
- 1966: Patern Recognition as interpolation of membership functions (Zadeh et al.)

A *fuzzy set* (*class*) A in X is characterized by a *membership* (*characteristic*) *function* $f_A(x)$ which associates with each point[2] in X a real number in the interval [0, 1],[3] with the value of $f_A(x)$ at x representing the "grade of membership" of x in A. Thus, the nearer the value of $f_A(x)$ to unity, the higher the grade of membership of x in A.

INFORMATION AND CONTROL 8, 338-353 (1965)

Fuzzy Sets*

L. A. Zadeh

Department of Electrical Engineering and Electronics Research Laboratory, University of California, Berkeley, California

A fuzzy set is a class of objects with a continuum of grades of membership. Such a set is characterized by a membership (characteristic) function which assigns to each object a grade of membership ranging between zero and one. The notions of inclusion, union, intersection, complement, relation, convexity, etc., are extended to such sets, and various properties of these notions in the context of fuzzy sets are established. In particular, a separation theorem for convex fuzzy sets is proved without requiring that the fuzzy sets be disjoint.

Fuzzy Sets

- 1965: Paper of Fuzzy Sets (Lotfi Zadeh)
- 1966: Patern Recognition as interpolation of membership functions (Zadeh et al.)

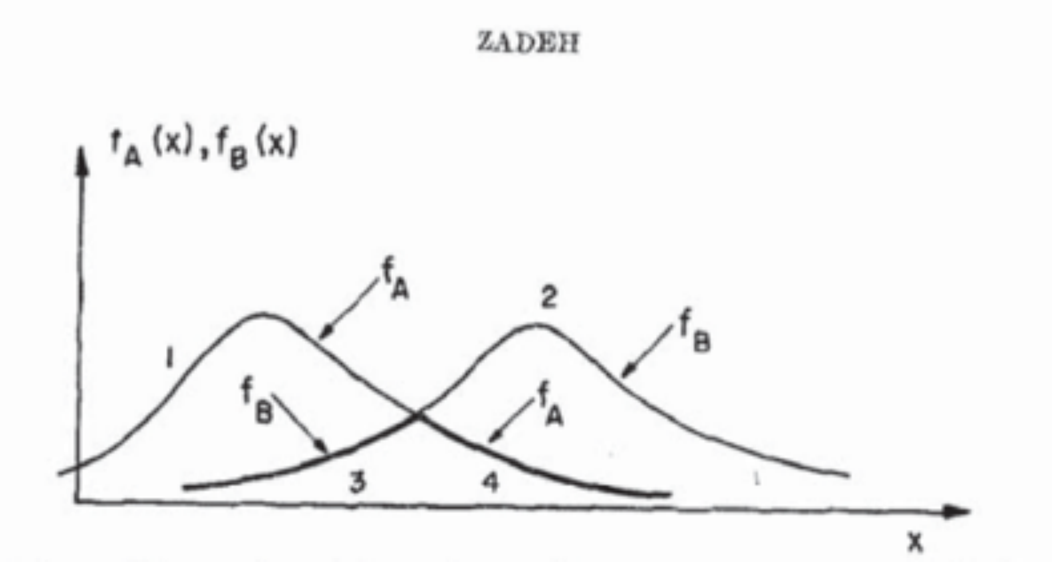
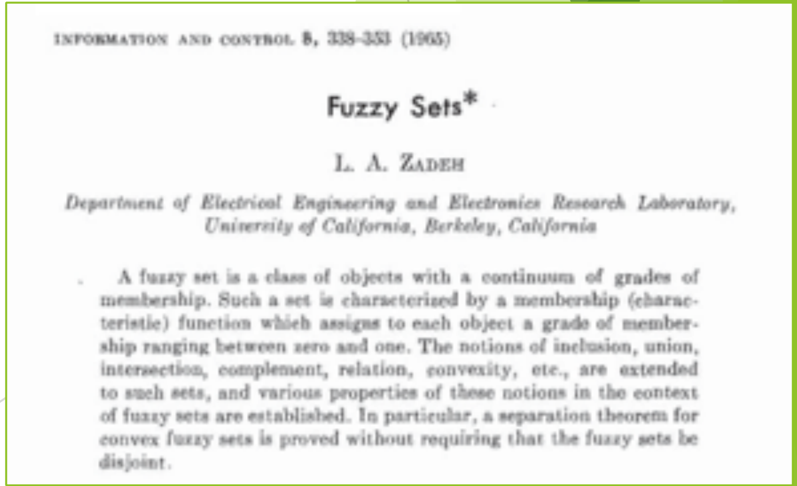


Fig. 1. Illustration of the union and intersection of fuzzy sets in R^1

Fuzzy Sets



INFORMATION AND CONTROL 8, 338-353 (1965)

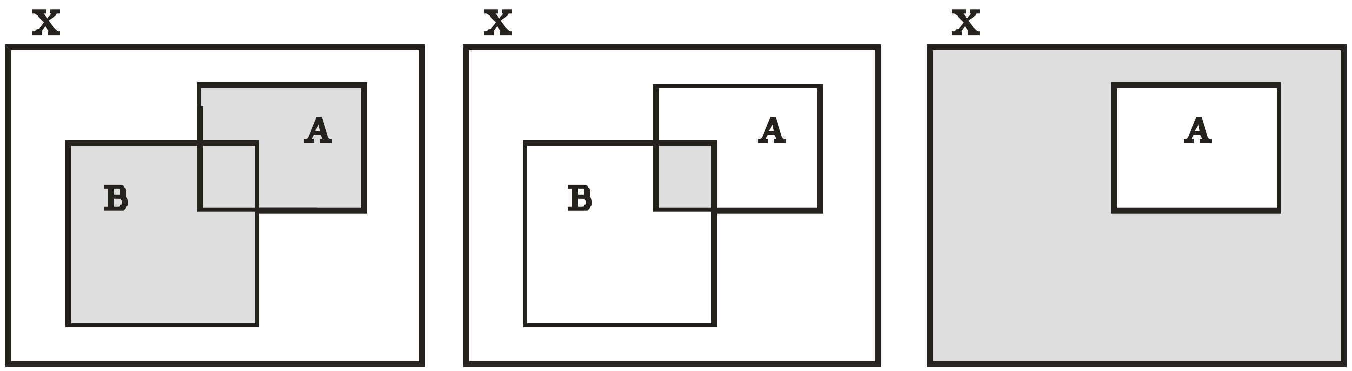
Fuzzy Sets*

L. A. Zadeh

Department of Electrical Engineering and Electronics Research Laboratory, University of California, Berkeley, California

A fuzzy set is a class of objects with a continuum of grades of membership. Such a set is characterized by a membership (characteristic) function which assigns to each object a grade of membership ranging between zero and one. The notions of inclusion, union, intersection, complement, relation, convexity, etc., are extended to such sets, and various properties of these notions in the context of fuzzy sets are established. In particular, a separation theorem for convex fuzzy sets is proved without requiring that the fuzzy sets be disjoint.

Characteristic function $f_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$



The Law of Non-Contradiction $A \cap \overline{A} = \varnothing$,

The Law of Excluded Middle $A \cup \overline{A} = X$.

Fuzzy Sets

$$A = \{(x, \mu_A(x)) \mid x \in X\}$$

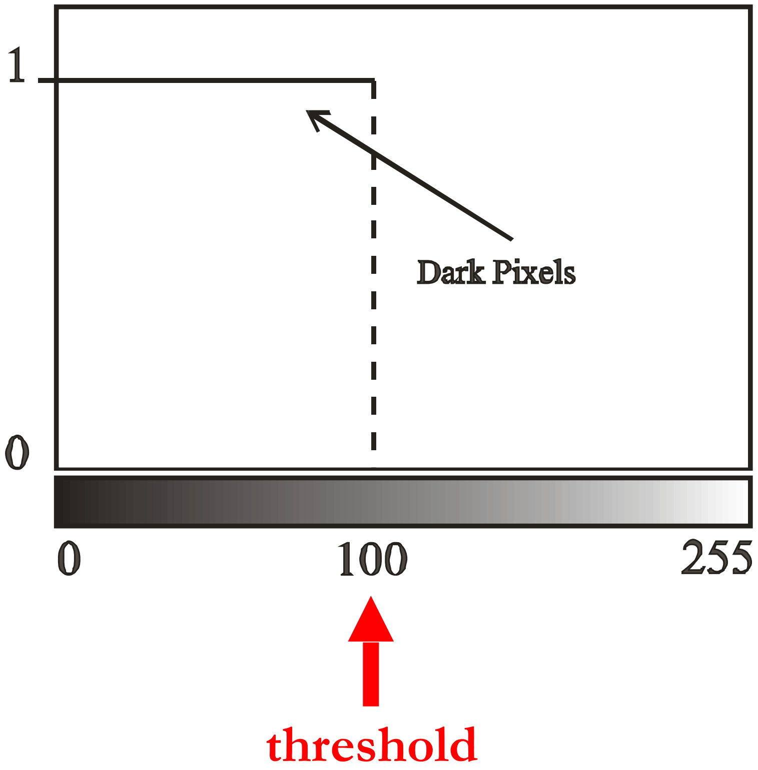
$$A = \int_X \mu_A(x) / x$$

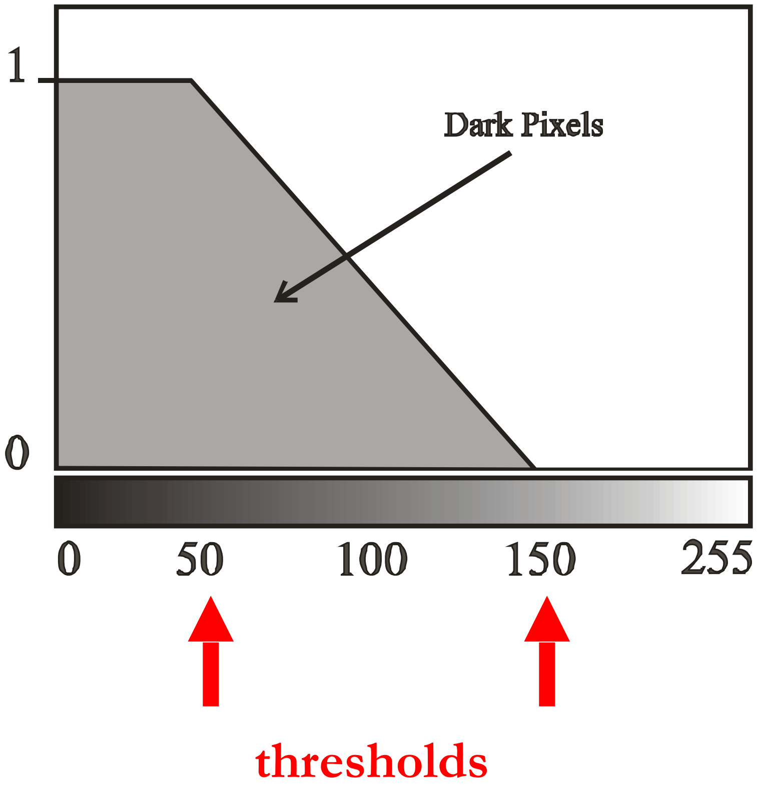
Example: Consider the set *Neighbors of 4*

$$A_{classic} = \{3, 4, 5\}$$

$$A_{Fuzzy} = \left\{ \frac{0.6}{1}, \frac{0.9}{2}, \frac{1.0}{3}, \frac{1.0}{4}, \frac{1.0}{5}, \frac{0.9}{6}, \frac{0.6}{7} \right\}$$

Fuzzy Sets

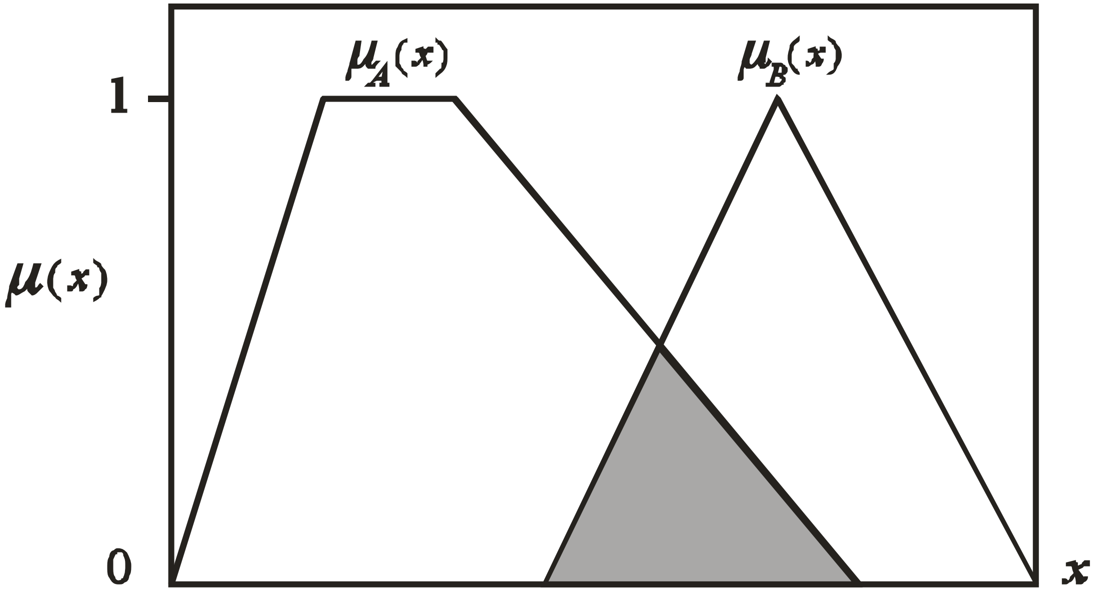

1
0
Dark Pixels
0
100
255
threshold


1
0
Dark Pixels
0
50
100
150
255
thresholds

Fuzzy Sets

Intersection

$$\mu_{A \cap B}(x) = \min\left(\mu_A(x), \mu_B(x)\right) \forall x \in X$$



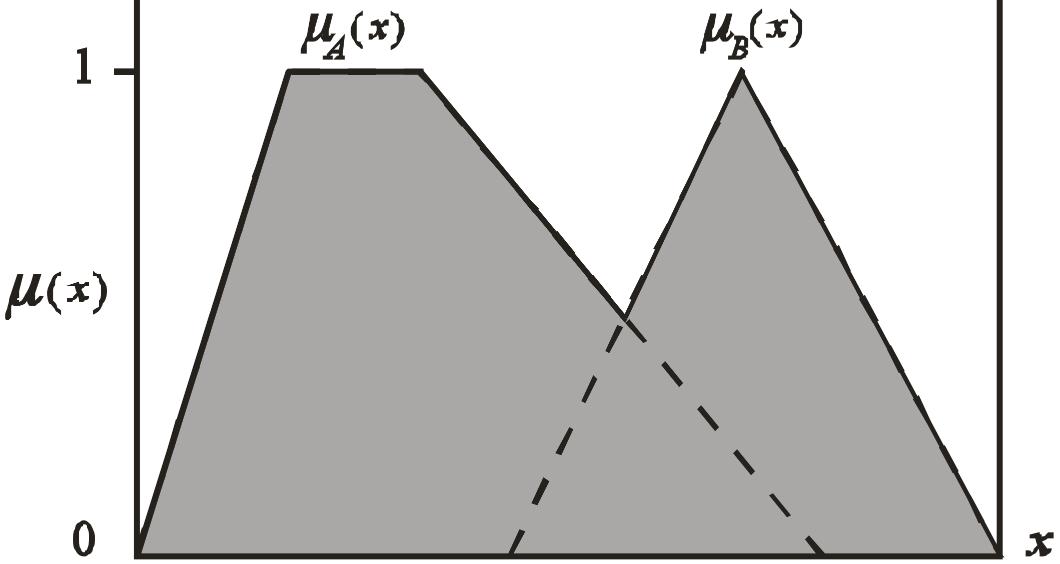
Source: H.R.Tizhoosh, Fuzzy Bildverarbeitung, Springer, 1998

$$A \cap \bar{A} \neq \varnothing$$

Fuzzy Sets

Union

$$\mu_{A\cup B}(x) = \max\big(\mu_A(x), \mu_B(x)\big) \forall x \in X$$



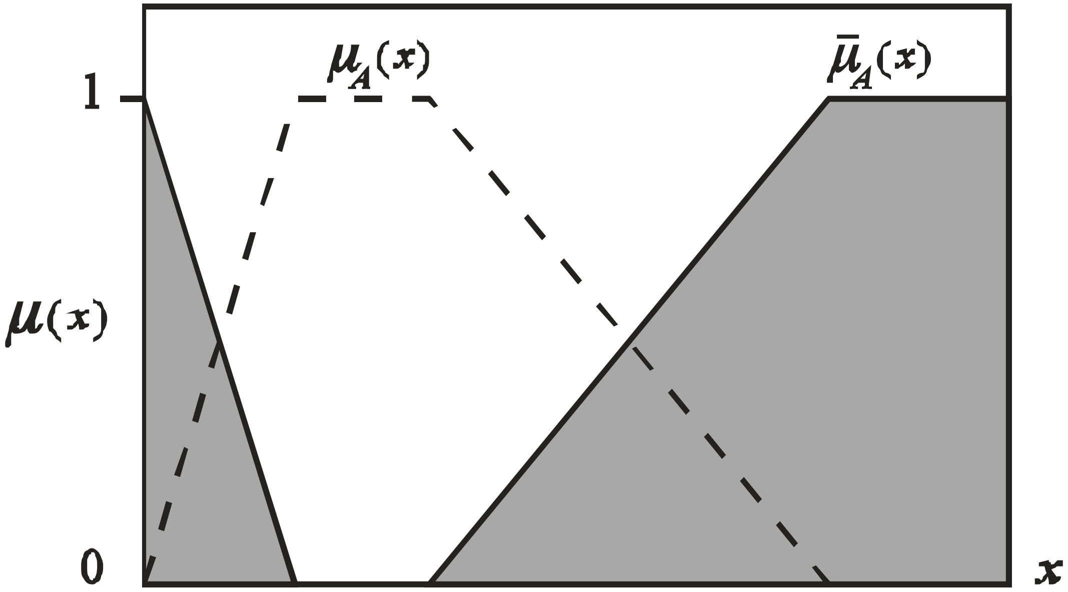
Source: H.R.Tizhoosh, Fuzzy Bildverarbeitung, Springer, 1998

$$A \cup \bar{A} \neq X$$

Fuzzy Sets

Complement

$$\bar{\mu}_A(x) = 1 - \mu_A(x)\ \forall x \in X$$



Source: H.R.Tizhoosh, Fuzzy Bildverarbeitung, Springer, 1998

Fuzzy Sets

T-Norm

Boundary	T(0, 0) =0 , T(a, 1) = T(1, a) = a
Monotonicity	a<=c & b<= d -> T(a, b)<=T(c, d)
Commutativity	T(a, b) = T(b, a)
Associativity	T(a, T(b, c)) = T(T(a, b), c)

Example:

Yager

$$1 - \min\left[1, \left((1-a)^w + (1-b)^w\right)^{\frac{1}{w}}\right] \quad w \in (0, \infty)$$

Dubois & Prade

$$\frac{ab}{\max(a, b, \alpha)} \quad \alpha \in (0,1)$$

Fuzzy Sets

T-CoNorm (S-Norm)

Boundary S(1, 1) =1 , S(a, 0) = S(0, a) = a

Monotonicity a<=c & b<=d -> S(a,b)<= S(c, d)

Commutativity S(a, b) = S(b, a)

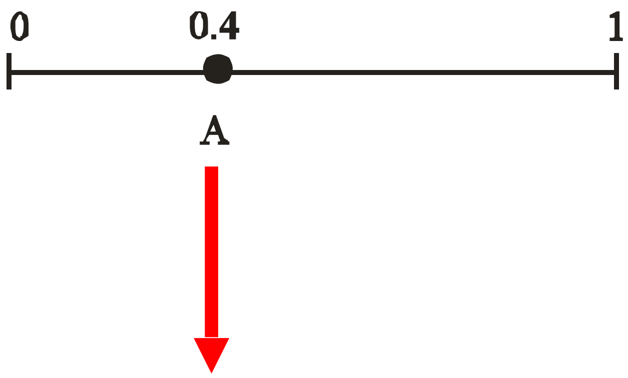
Associativity S(a, S(b,c)) = S(S(a,b), c)

Example:

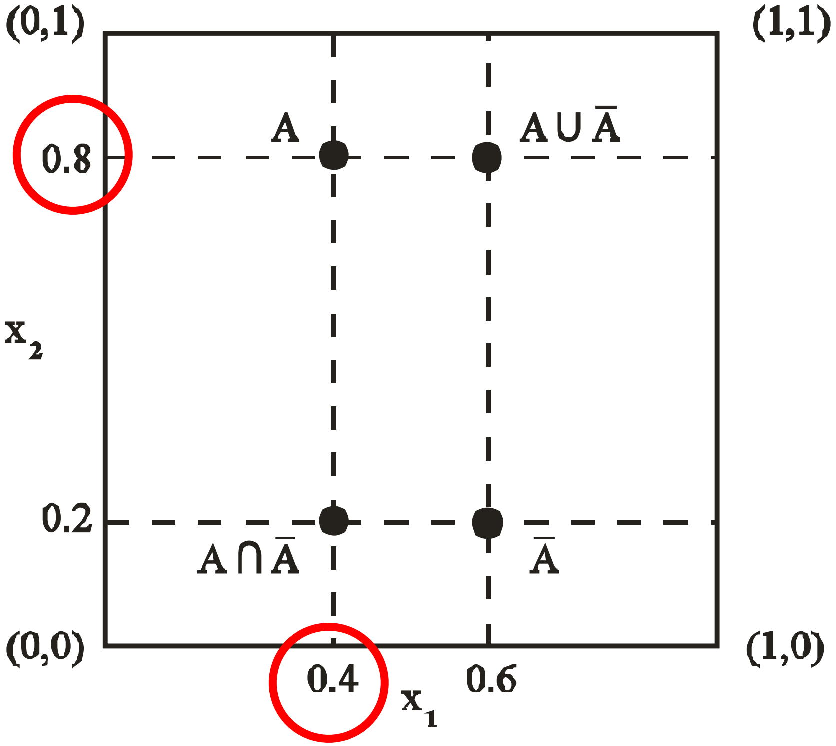
Yager $$\min\left[1, \left(a^w + b^w\right)^{\frac{1}{w}}\right] \quad w \in (0, \infty)$$

Dubois & Prade $$\frac{a + b - ab - \min(a, b, 1-\alpha)}{\max(1-a, 1-b, \alpha)} \quad \alpha \in (0,1)$$

Hypercube Presentation



fuzzy set with one member



fuzzy set with two members

Fuzzy Sets

Memberships

Membership as similarity

Membership as probability

Membership as intensity

Membership as approximation

Membership as compatibility

Membership as possibility

Generating MFs

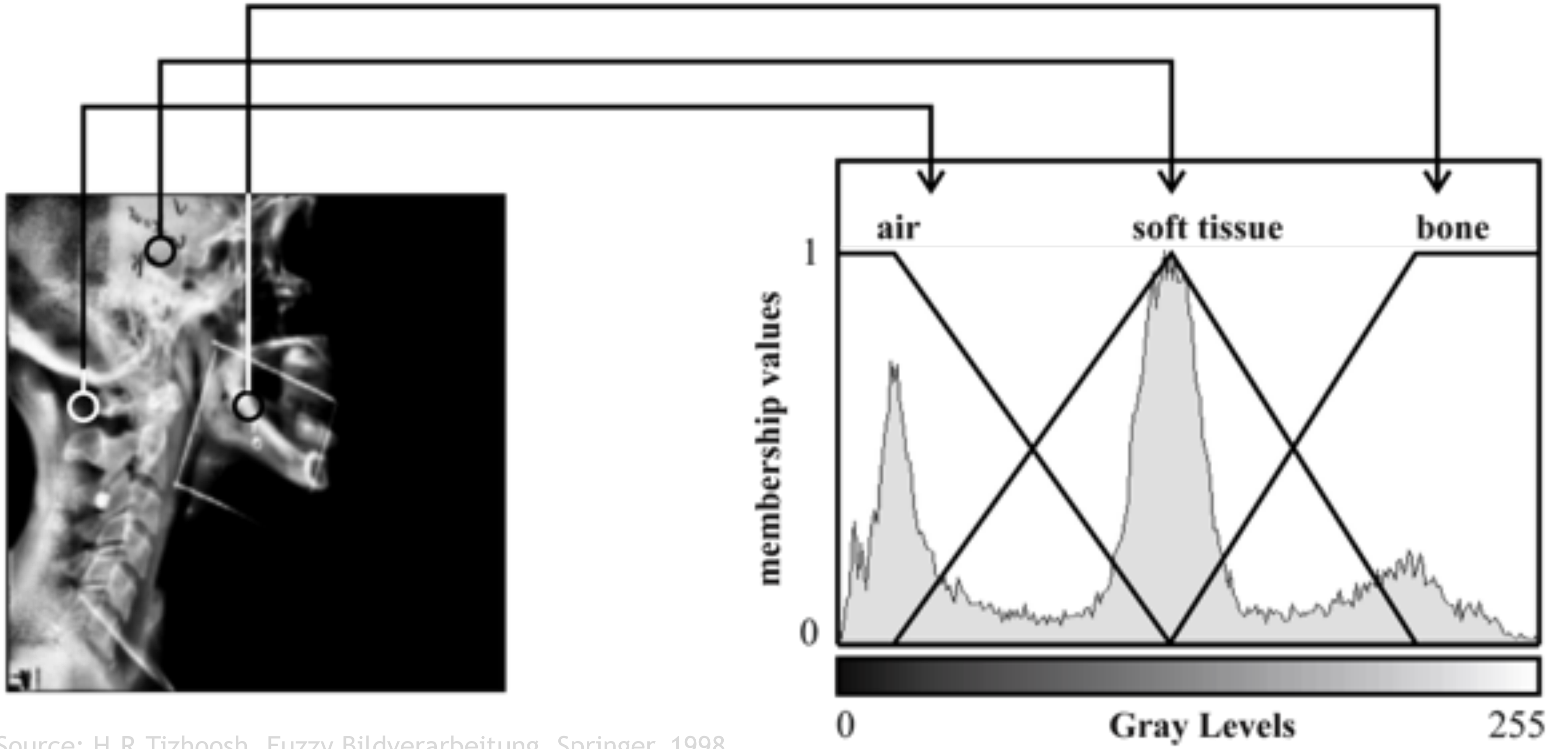
1. Subjective:

   intuition, expertise, knowledge

2. Automatic:

   Clustering, Neural nets, Genetic Algorithms

Memberships



Source: H.R.Tizhoosh, Fuzzy Bildverarbeitung, Springer, 1998

Fuzzy Sets

- 1972: Fuzzy controller for steam engine (Assilian and Mamdani)

The Plant to be Controlled

The plant for which the controller was implemented comprises a steam engine and boiler combination. The model of the plant used has two inputs: heat input to the boiler and throttle opening at the input of the engine cylinder, and two outputs: the steam pressure in the boiler and the speed of the engine.

Int. J. Man–Machine Studies (1975) 7, 1–13

An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller

E. H. MAMDANI AND S. ASSILIAN
Queen Mary College, London University, U.K.

(*Received 2 November 1973*)

This paper describes an experiment on the "linguistic" synthesis of a controller for a model industrial plant (a steam engine). Fuzzy logic is used to convert heuristic control rules stated by a human operator into an automatic control strategy. The experiment was initiated to investigate the possibility of human interaction with a learning controller. However, the control strategy set up linguistically proved to be far better than expected in its own right, and the basic experiment of linguistic control synthesis in a non-learning controller is reported here.

Fuzzy Controller

- 1972: Fuzzy controller for steam engine (Aslani and Mamdani)

(1) *PE*—Pressure Error, defined as the difference between the present value of the variable and the set point.

(2) *SE*—Speed Error, defined as in (1).

(3) *CPE*—Change in pressure error, defined as the difference between present *PE* and last (corresponding to last sampling instant).

(4) *CSE*—Change in speed error, defined as in (3).

(5) *HC*—Heat Change (action variable).

(6) *TC*—Throttle Change (action variable).

Int. J. Man-Machine Studies (1975) 7, 1–13

An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller

E. H. MAMDANI AND S. ASSILIAN
Queen Mary College, London University, U.K.

(*Received 2 November 1973*)

This paper describes an experiment on the "linguistic" synthesis of a controller for a model industrial plant (a steam engine). Fuzzy logic is used to convert heuristic control rules stated by a human operator into an automatic control strategy. The experiment was initiated to investigate the possibility of human interaction with a learning controller. However, the control strategy set up linguistically proved to be far better than expected in its own right, and the basic experiment of linguistic control synthesis in a non-learning controller is reported here.

Fuzzy Controller

- 1972: Fuzzy controller for steam engine (Aslani and Mamdani)

HEATER ALGORITHM

If $PE = NB$
and $CPE =$ not $(NB$ or $NM)$
and $SE = ANY$
and $CSE = ANY$
then $HC = PB$

THROTTLE ALGORITHM

If $PE = ANY$
and $CPE = ANY$
and $SE = NB$
and $CSE =$ not $(NB$ or $NM)$
then $TC = PB$

Int. J. Man–Machine Studies (1975) 7, 1–13

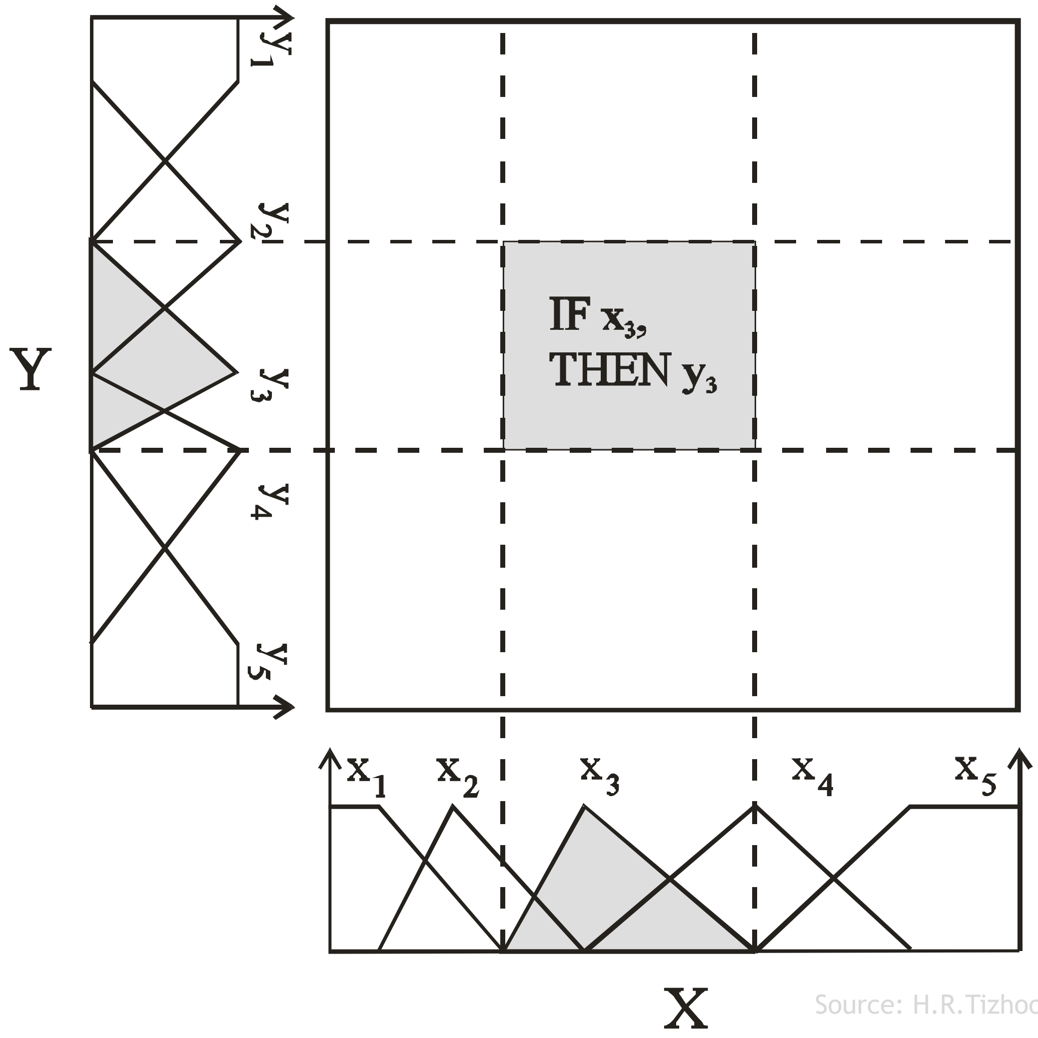
An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller

E. H. Mamdani and S. Assilian
Queen Mary College, London University, U.K.

(*Received 2 November 1973*)

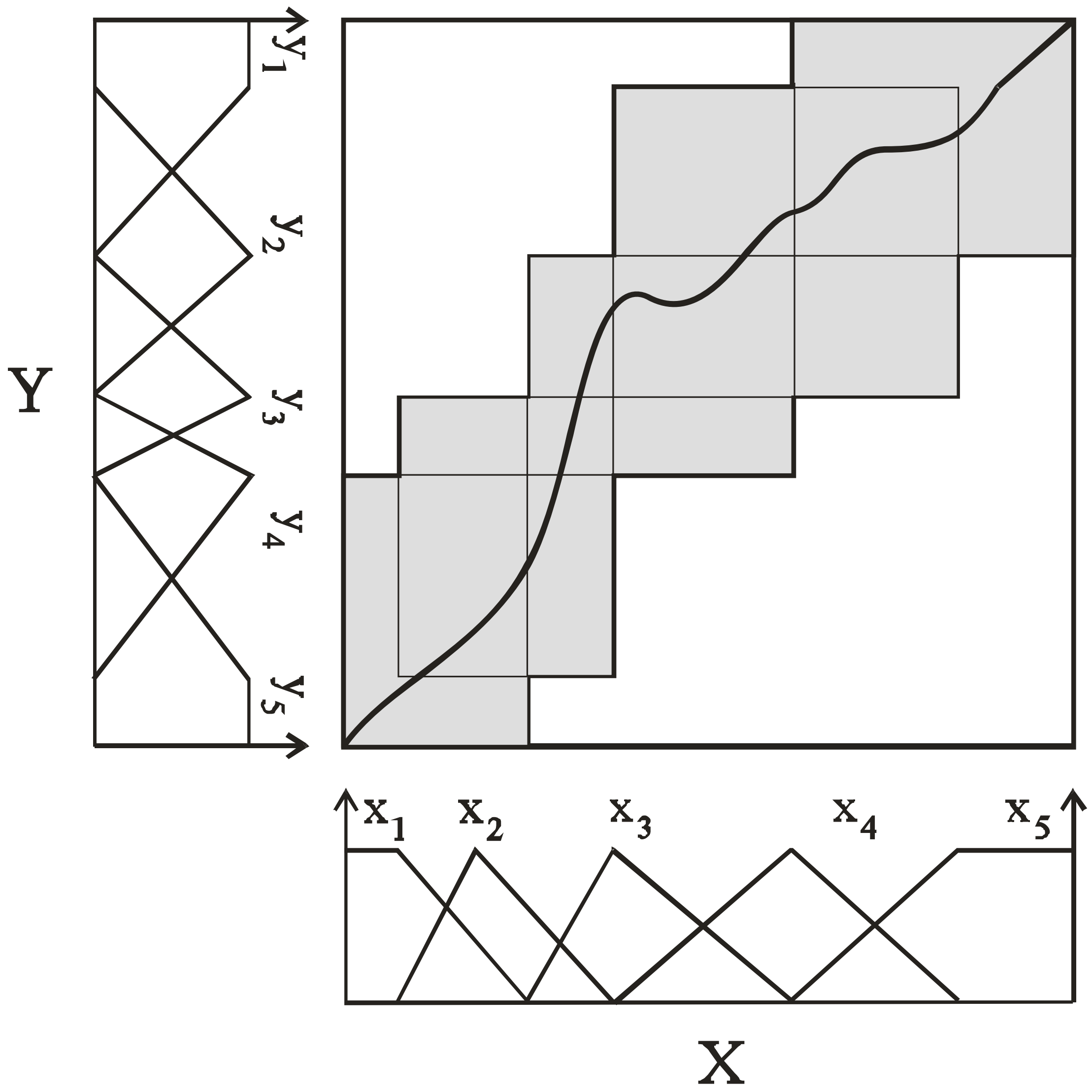
This paper describes an experiment on the "linguistic" synthesis of a controller for a model industrial plant (a steam engine). Fuzzy logic is used to convert heuristic control rules stated by a human operator into an automatic control strategy. The experiment was initiated to investigate the possibility of human interaction with a learning controller. However, the control strategy set up linguistically proved to be far better than expected in its own right, and the basic experiment of linguistic control synthesis in a non-learning controller is reported here.

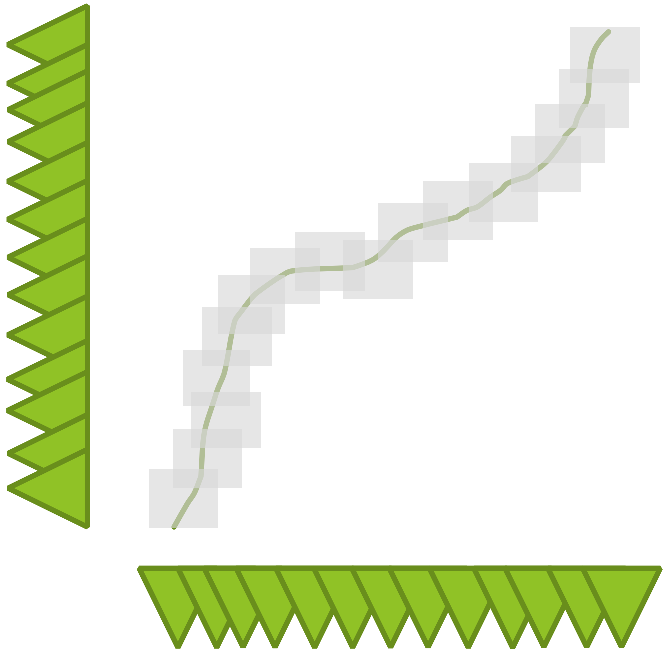
Fuzzy Controller



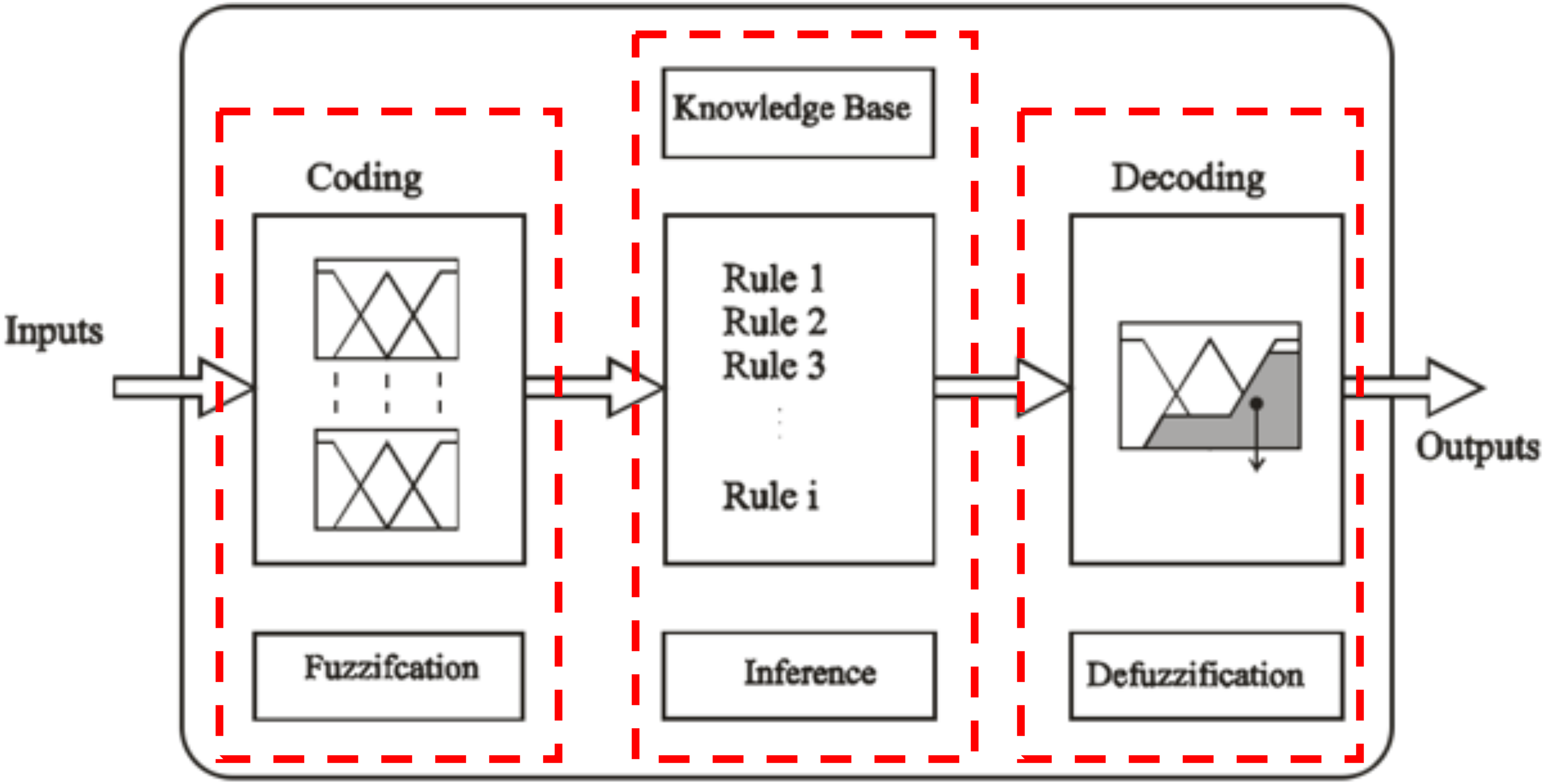
Source: H.R.Tizhoosh, Fuzzy Bildverarbeitung, Springer, 1998

Fuzzy Controller



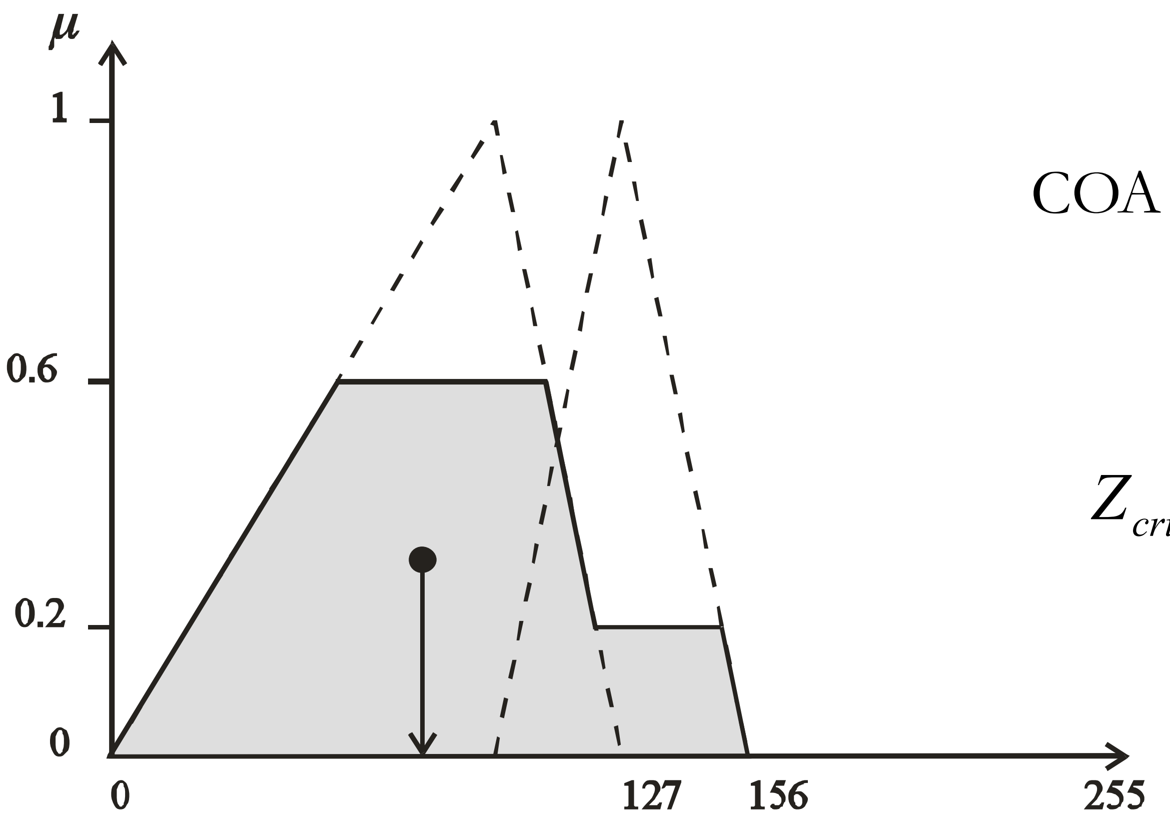


Fuzzy Controller



Source: H.R.Tizhoosh, Fuzzy Bildverarbeitung, Springer, 1998

Fuzzy Controller



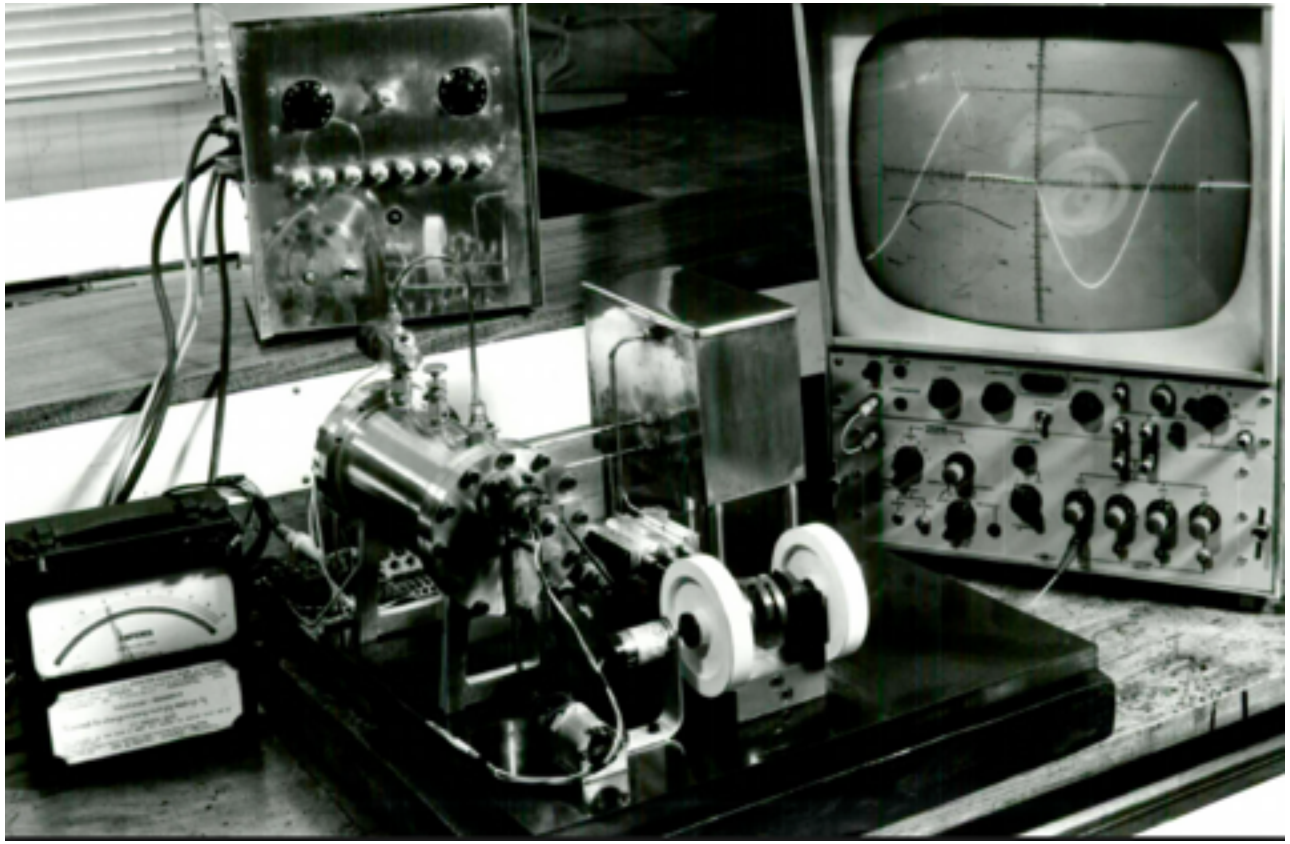
$Z_{Crisp} = 79$

Source: H.R.Tizhoosh, Fuzzy Bildverarbeitung, Springer, 1998

COA Center of Area

$$Z_{crisp} = \frac{\sum_{z=\alpha}^{z=\beta} z \cdot \mu(z)}{\sum_{z=\alpha}^{z=\beta} \mu(z)}$$

Fuzzy Controller




Turning Around the Ideas of 'Meaning' and 'Complement'

Article · December 2015

DOI: 10.1007/978-3-319-26986-3_1

1st Enric Trillas
35.52 · European Centre for Soft Computing

2nd Rudolf Seising
23.94 · Friedrich-Schiller-University Jena

Photograph of the "Fuzzy steam engine", Queen Mary College, 1974, reprint courtesy of Brian Gaines

Fuzzy Controller

- 1969: Concept of Fuzzy Partitioning (Enrique Ruspini)

INFORMATION AND CONTROL 15, 22–32 (1969)

A New Approach to Clustering

ENRIQUE H. RUSPINI
Space Biology Laboratory, University of California, Los Angeles

A general formulation of data reduction and clustering processes is proposed. These procedures are regarded as mappings or transformations of the original space onto a "representation" or "code" space subjected to some constraints. Current clustering methods, as well as three other data reduction techniques, are specified within the framework of this formulation. A new method of representation of the reduced data, based on the idea of "fuzzy sets," is proposed to avoid some of the problems of current clustering procedures and to provide better insight into the structure of the original data.

FIG. 1. Nagy (1968) has illustrated the major difficulties found in cluster analysis: (a) and (c) bridges between clusters; (b) nonspherical clusters; (d) linearly nonseparable clusters; (e) unequal cluster populations.

Fuzzy Clustering

- 1973: First Fuzzy Clustering algorithm (FCM by Dunn and Bezdek)

$$\hat{u}_{jk} = \frac{1}{\sum_{i=1}^{c} \left(\frac{\|\mathbf{x}_k - \hat{\mathbf{v}}_j\|^2}{\|\mathbf{x}_k - \hat{\mathbf{v}}_i\|^2} \right)^{\frac{1}{m-1}}} \quad \forall \quad \begin{array}{l} 1 \leqslant j \leqslant c \\ 1 \leqslant k \leqslant n \end{array}$$

U is nondegenerate, i.e., has no empty subsets

$$\hat{\mathbf{v}}_j = \frac{\sum_{k=1}^{n} (\hat{u}_{jk})^m \mathbf{x}_k}{\sum_{k=1}^{n} (\hat{u}_{jk})^m} \quad \forall \; 1 \leqslant j \leqslant c$$

Journal of Cybernetics, 1974, 3, 3, pp. 58-73

Cluster Validity with Fuzzy Sets

James C. Bezdek†
Department of Mathematics
State University College
Oneonta, New York

Fuzzy Clustering

Fuzzy C-Means $M_{fc} = \left\{ U \,\middle|\, \mu_{ik} \in [0,1];\ \sum_{i=1}^{c} \mu_{ik} = 1;\ 0 < \sum_{k=1}^{N} \mu_{ik} < N \right\}$

1. Initialization

2. Class Centers $$v_i = \frac{\sum_{k=1}^{N} (\mu_{ik})^m x_k}{\sum_{k=1}^{N} (\mu_{ik})^m}$$

3. Update Membership $$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}}}$$

4. Condition $$\left\| U^t - U^{t-1} \right\| \leq \varepsilon$$

Fuzzy Clustering

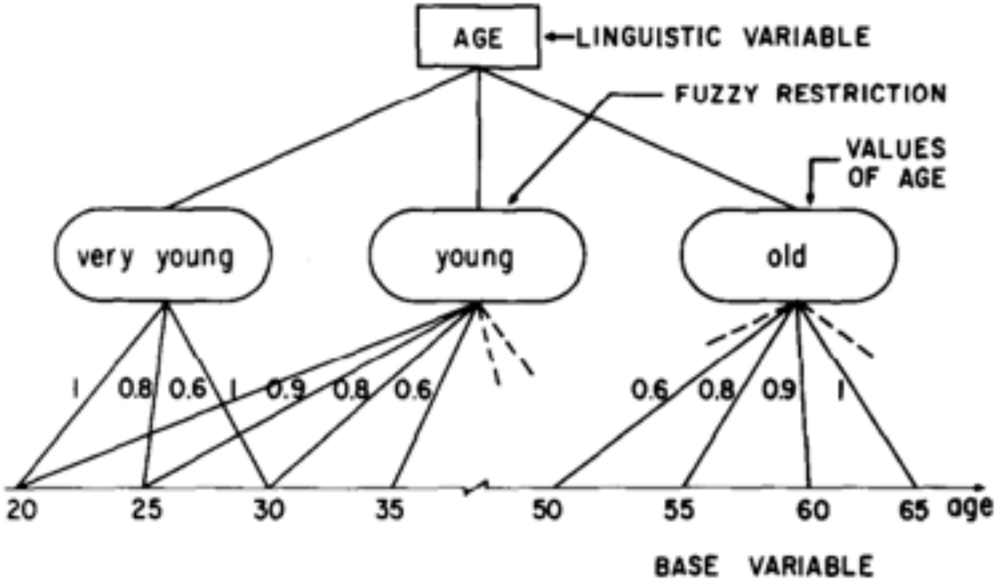
FCM

With 2 features

Source: H.R.Tizhoosh, Fuzzy Bildverarbeitung, Springer, 1998

Fuzzy Clustering

- 1975: The Concept of Linguistic Variables (Lotfi Zadeh)



INFORMATION SCIENCES 8, 199–249 (1975)

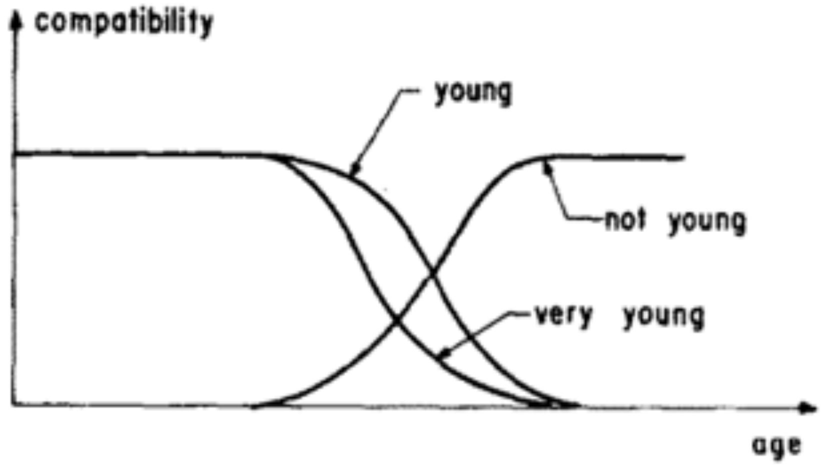
The Concept of a Linguistic Variable and its Application to Approximate Reasoning—I

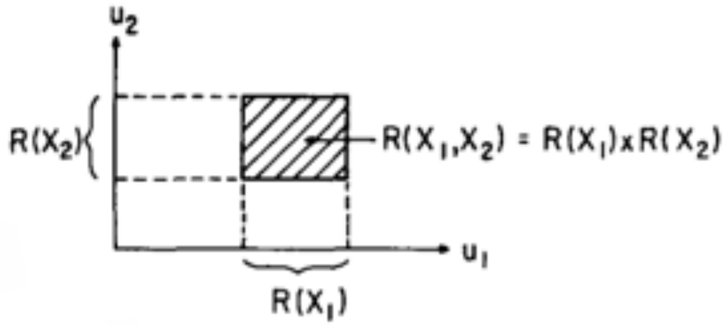
L. A. ZADEH

Computer Sciences Division, Department of Electrical Engineering and Computer Sciences, and the Electronics Research Laboratory, University of California, Berkely, California 94720

Linguistic Variables

- 1975: The Concept of Linguistic Variables (Lotfi Zadeh)





INFORMATION SCIENCES 8, 199-249 (1975)

The Concept of a Linguistic Variable and its Application to Approximate Reasoning–I

L. A. ZADEH

Computer Sciences Division, Department of Electrical Engineering and Computer Sciences, and the Electronics Research Laboratory, University of California, Berkely, California 94720

Linguistic Variables

„In retreating from precision in the face of overpowering complexity, it is natural to explore the use of what might be called **linguistic variables**, that is, variables whose values are not numbers but **words** or **sentences** in a natural or artificial language.“

Zadeh

Linguistic Variables

$$\left(x,\ T(x),\ U,\ G,\ M\right)$$

Variable name

Set of its terms

Universe of discourse

Syntax rules

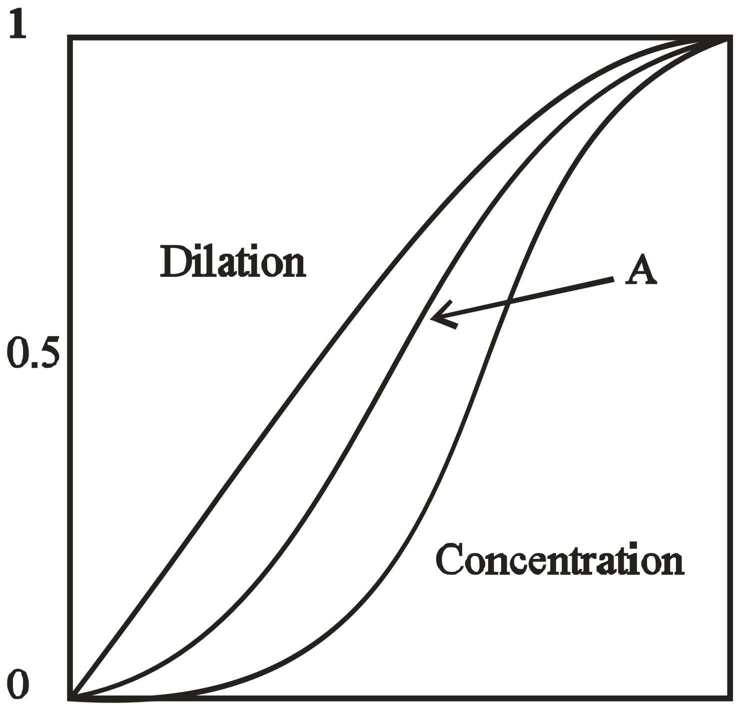
Semantic rules

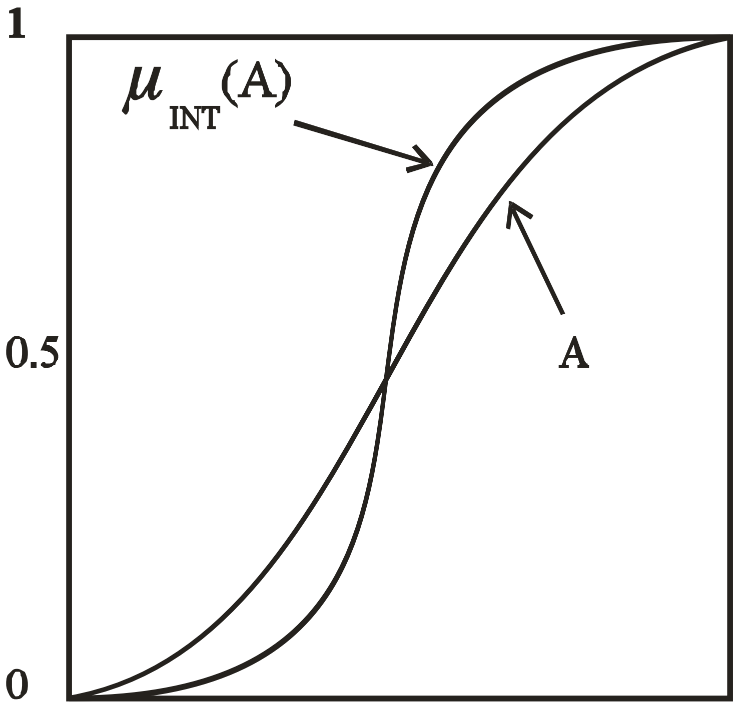
Linguistic Variables

Word	Average
always	99
very often	88
ususally	85
often	78
relatively often	65
balanced	50
from time to time	20
sometimes	20
not usually	10
seldom	10
very seldom	6
almost never	3
never	0

(Simpson, 1944)

Linguistic Hedges





$$\mu_{CON}(x) = [\mu(x)]^2$$

$$\mu_{DIL}(x) = [\mu(x)]^{0.5}$$

$$\mu_{INT}(x) = \begin{cases} 2[\mu(x)]^2 & \textit{for } 0 \le \mu(x) \le 0.5 \\ 1 - 2[1 - \mu(x)]^2 & \textit{else} \end{cases}$$

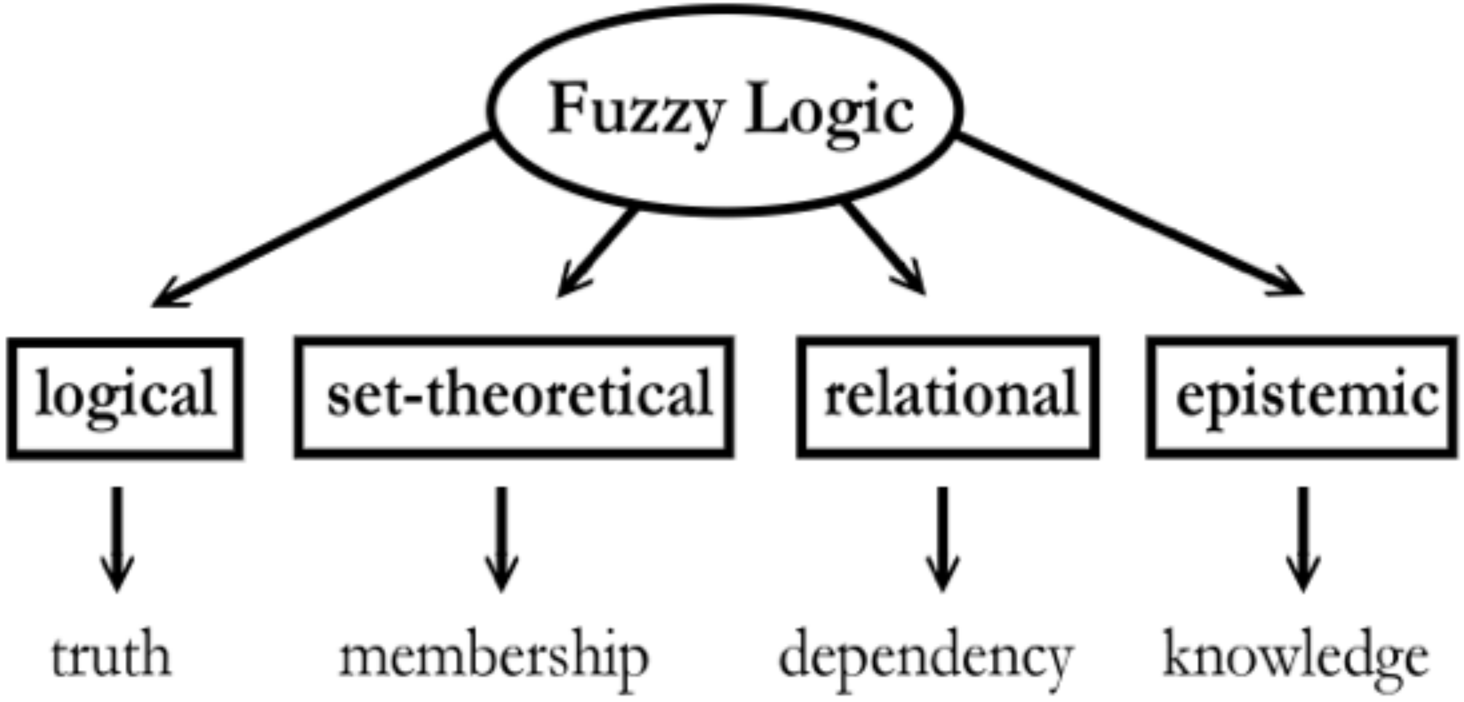
Source: H.R.Tizhoosh, Fuzzy Bildverarbeitung, Springer, 1998

Linguistic Hedges

μ_A is known (*bright, cold, tall, old,...*)

$\mu_{\text{very A}}$	= CON(A)
$\mu_{\text{more or less A}}$	= DIL(A)
$\mu_{\text{very very A}}$	= CON(CON((A))
$\mu_{\text{not very A}}$	= 1– (CON(A))
$\mu_{\text{ more A}}$	$= (A)^{1.25}$
$\mu_{\text{ less A}}$	$= (A)^{0.75}$

Linguistic Hedges


Fuzzy Logic
logical
set-theoretical
relational
epistemic
truth
membership
dependency
knowledge

Fuzzy Logic

If a pixel is bright, then it is with high probability noise

Relaxation

Logic	Truth	Sets	Identity
Classical Reasoning	{0,1}	Crisp	Yes
Fuzzy Reasoning	[0,1]	Crisp	Yes
Approximate Reasoning	[0,1]	Fuzzy	Yes
Plausible Reasoning	[0,1]	Fuzzy	NO

Fuzzy Logic

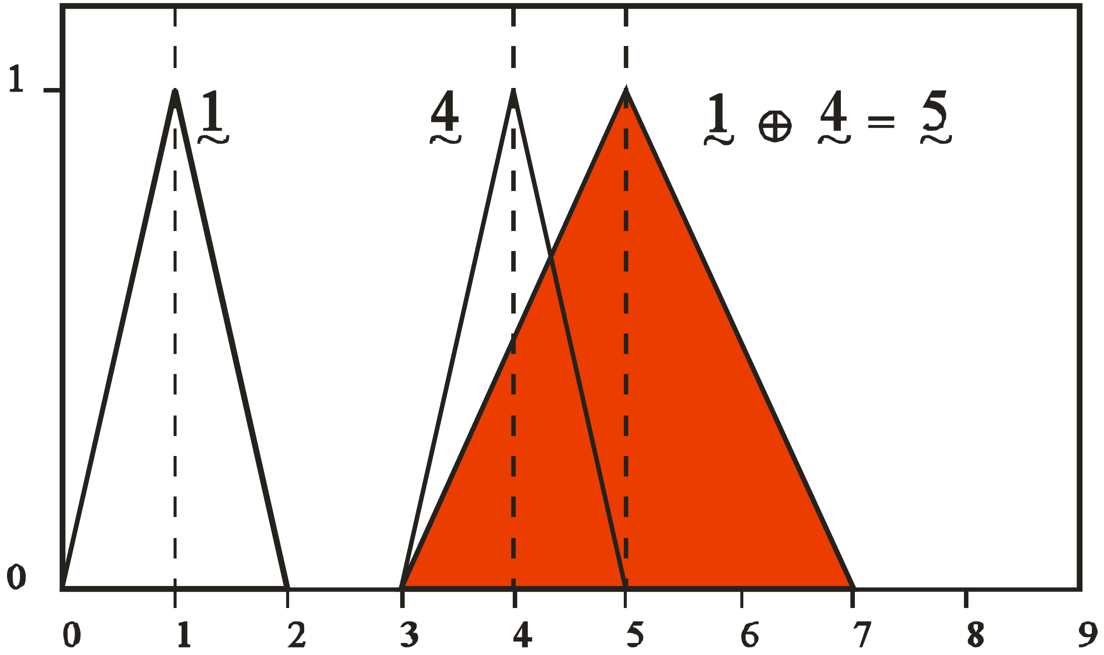
$$1 + 4 = 5$$

'about 1' + 'approximately 4' = ?

Fuzzy Arithmetic

Extension Principle

$$\mu_{A\oplus B}(z) = \sup_{x+y=z}\left[\min\left[\mu_A(x), \mu_B(y)\right]\right]$$



Source: H.R.Tizhoosh, Fuzzy Bildverarbeitung, Springer, 1998

Fuzzy Arithmetic

Dual Set

$$A = \{x \mid x \in X, x \text{ has a certain property}\}$$

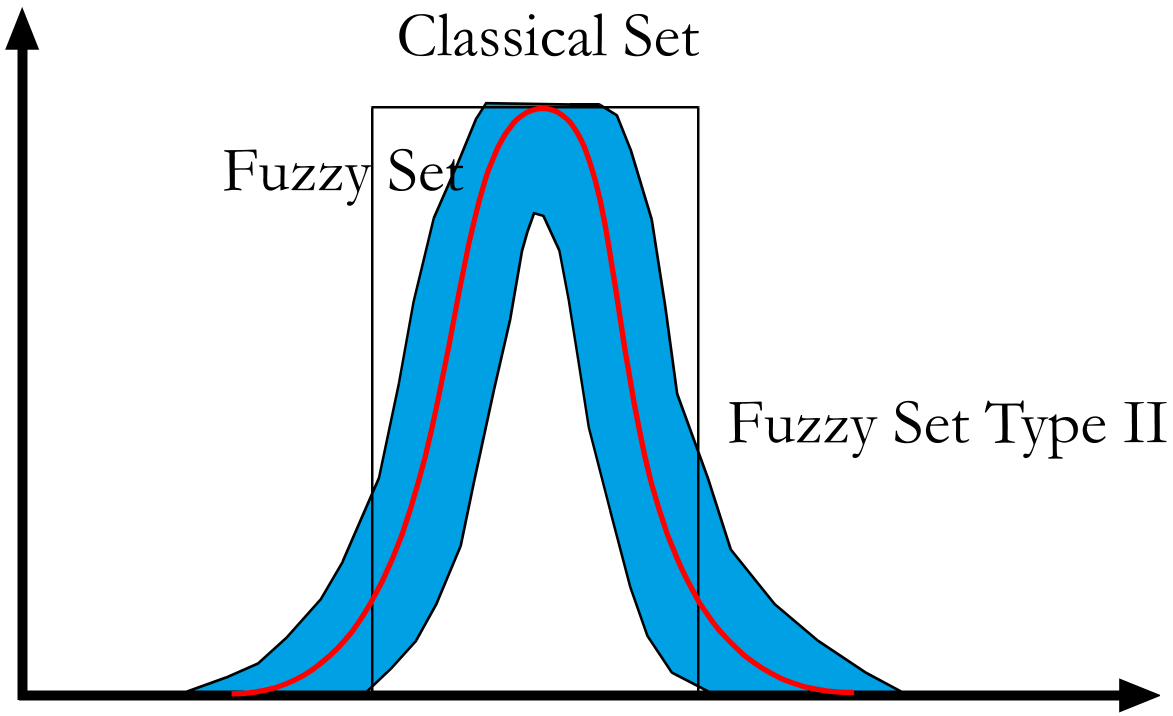
Fuzzy Set

$$\tilde{A} = \left\{\left(x, \mu_{\tilde{A}}(x)\right) \mid x \in X, \mu_{\tilde{A}}(x) \in [0,1]\right\}$$

Fuzzy Set Type II

$$\underset{\sim}{\tilde{A}} = \left\{\left((x,u), \mu_{\underset{\sim}{\tilde{A}}}(x,u)\right) \mid x \in X, u \in J_x \subseteq [0,1]\right\}$$

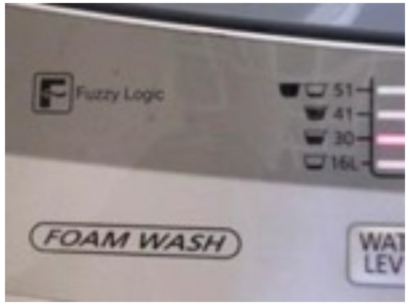
Type II Fuzzy Sets



Type II Fuzzy Sets

- Mid 1970s – late 1980s: Booming of fuzzy applications in Japan and Europe

Rice Cooker



Washing Machines

Sendai Subway

Fuzzy Boom

- 1993: ANFIS (J.S.R. Jang)

  - ANFIS (adaptive-network-based fuzzy inference system)
  - A fuzzy inference system implemented in the framework of adaptive networks
  - hybrid learning procedure
  - ANFIS can construct an input-output mapping based on both human knowledge (in the form of fuzzy if-then rules) and stipulated input-output data pairs

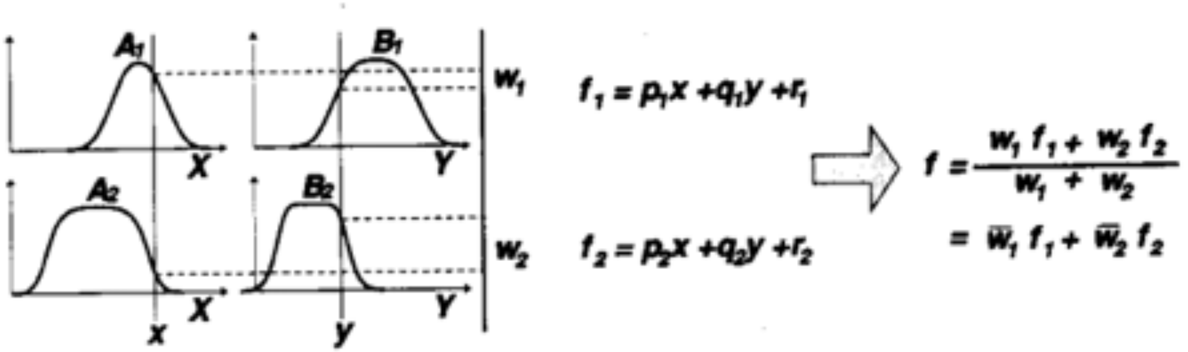
Fuzzy Learning

IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, VOL. 23, NO. 3, MAY/JUNE 1993

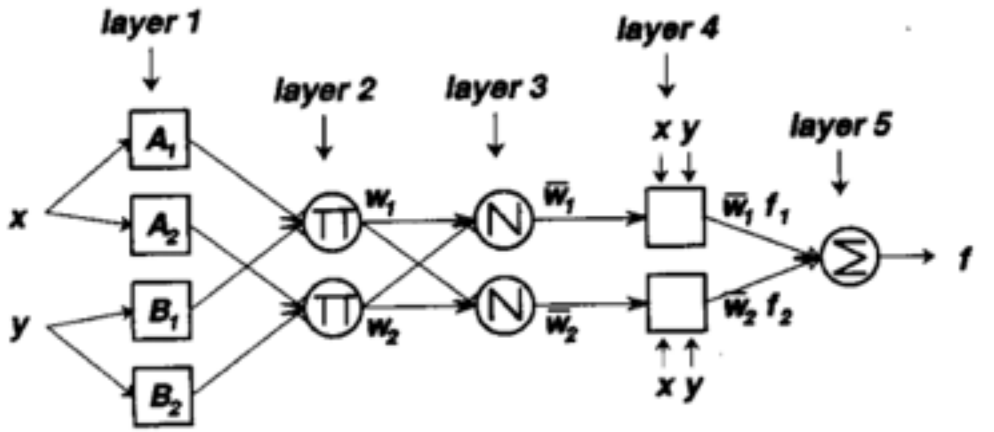
ANFIS: Adaptive-Network-Based Fuzzy Inference System

Jyh-Shing Roger Jang

- 1993: ANFIS (J.S.R. Jang)



Fuzzy Reasoning



Equivalent ANFIS

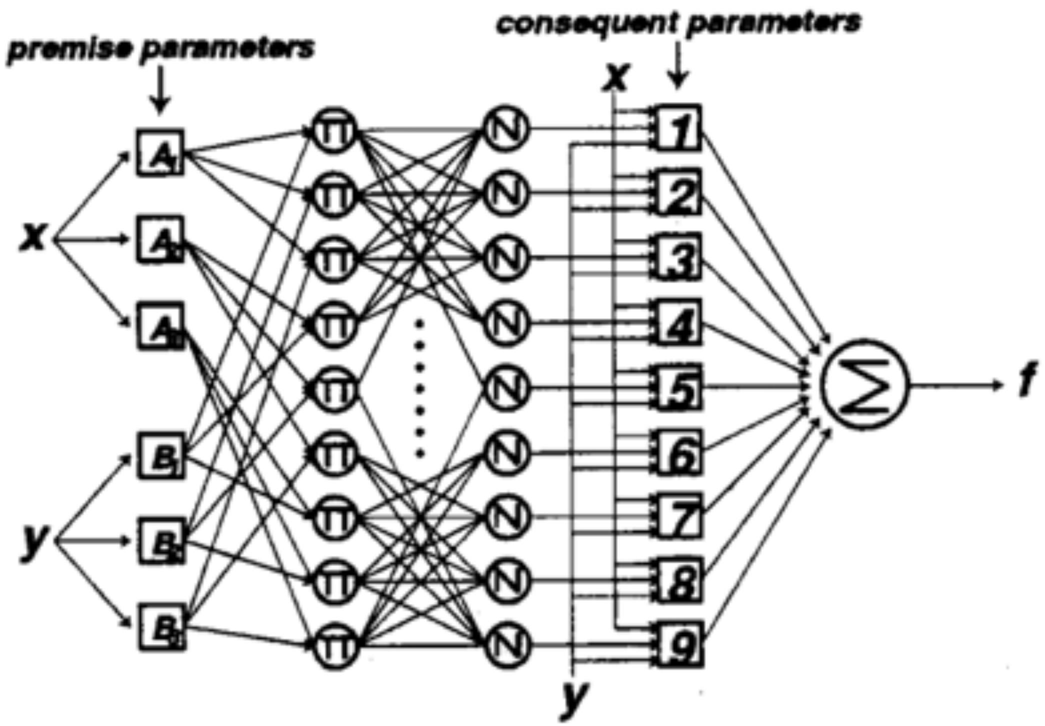
Fuzzy Learning

IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, VOL. 23, NO. 3, MAY/JUNE 1993

ANFIS: Adaptive-Network-Based Fuzzy Inference System

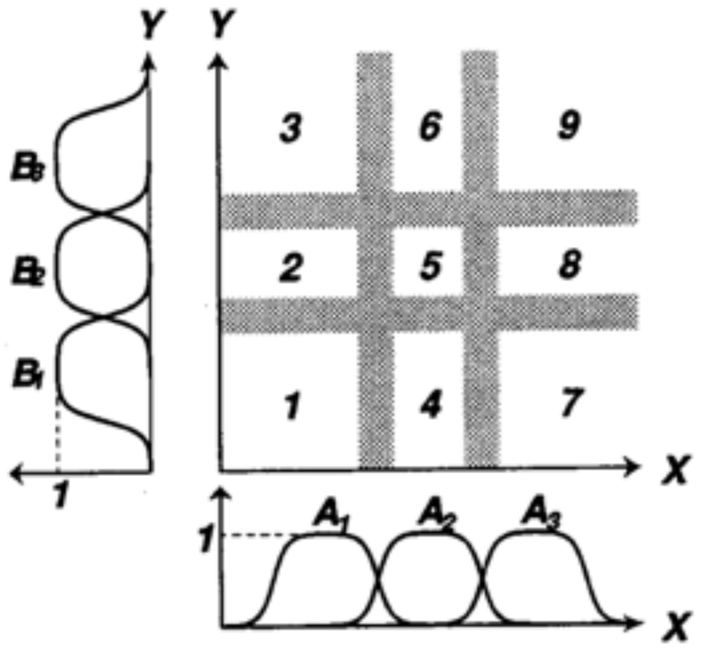
Jyh-Shing Roger Jang

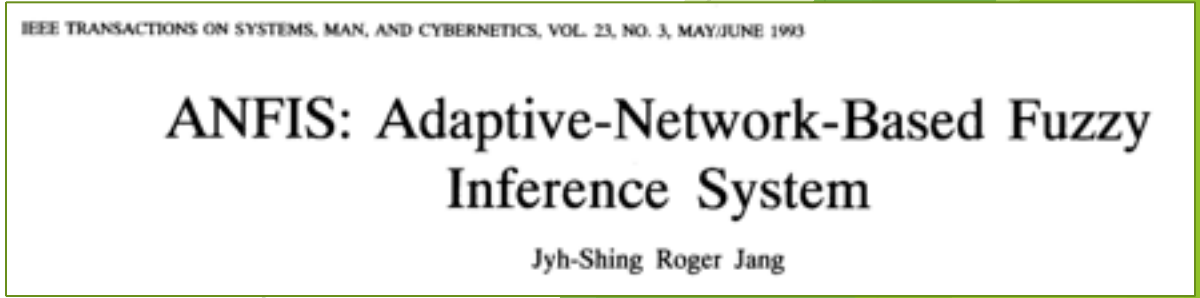
- 1993: ANFIS (J.S.R. Jang)



ANFIS with 9 rules

Fuzzy Subspace




IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, VOL. 23, NO. 3, MAY/JUNE 1993

ANFIS: Adaptive-Network-Based Fuzzy Inference System

Jyh-Shing Roger Jang

Fuzzy Learning

IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 34, NO. 1, FEBRUARY 2004

An Approach to Online Identification of Takagi-Sugeno Fuzzy Models

Plamen P. Angelov, *Member, IEEE*, and Dimitar P. Filev, *Senior Member, IEEE*

- 2004: Evolving Fuzzy Systems (P.P. Angelov et al.)

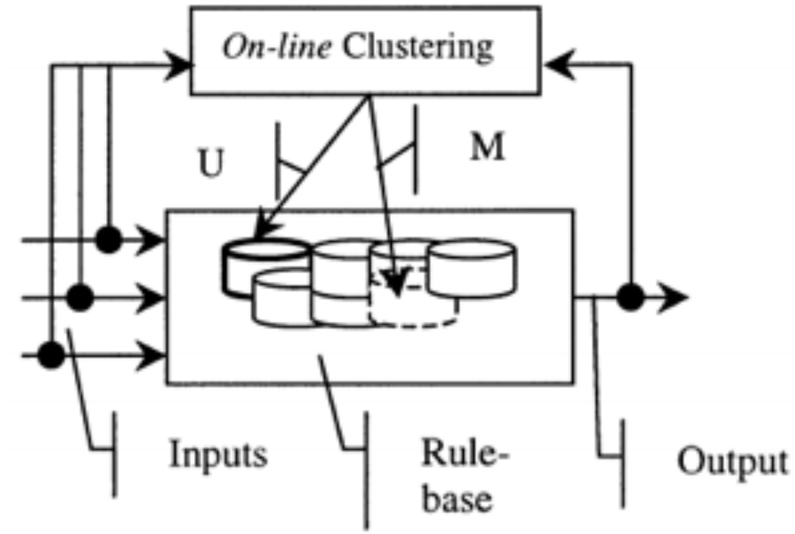


Fig. 1. Schematic representation of the rule-base evolution based on the data samples potential (M—modification/replacement); U—up-grade of a rule).

Fuzzy Evolution

IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 34, NO. 1, FEBRUARY 2004

An Approach to Online Identification of Takagi-Sugeno Fuzzy Models

Plamen P. Angelov, *Member, IEEE,* and Dimitar P. Filev, *Senior Member, IEEE*

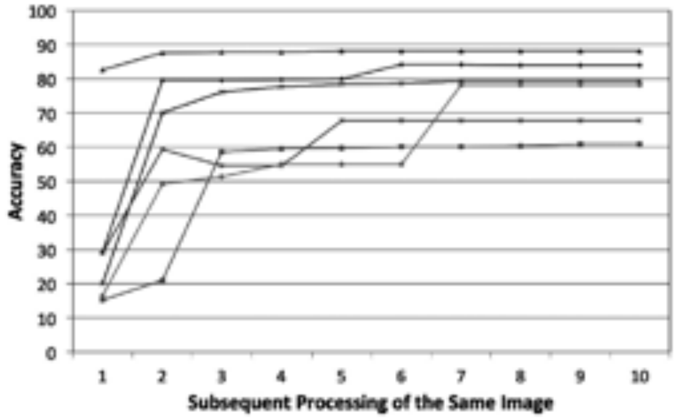
1) Stage 1: Initialization of the rule-base structure (antecedent part of the rules).
2) Stage 2: At the next time step reading the *next* data sample.
3) Stage 3: *Recursive* calculation of the potential of each new data sample to influence the *structure* of the rule-base.
4) Stage 4: *Recursive* up-date of the potentials of old centers taking into account the influence of the *new* data sample.
5) Stage 5: Possible ***modification*** or up-grade *of* the rule-base *structure* based on the potential of the new data sample in comparison to the potential of the *existing* rules' centers (focal points).
6) Stage 6: *Recursive* calculation of the consequent parameters.
7) Stage 7: Prediction of the output for the next time step by the ETS model.

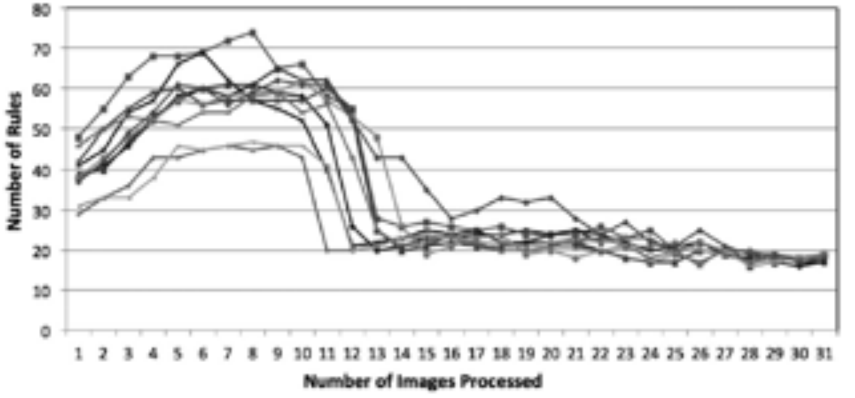
Fuzzy Evolution

Evolving Fuzzy Image Segmentation

Algorithm 2 EFIS Online for Image Segmentation

1: *LFR* – Load the fuzzy rules $R_1, R_2, \cdots$ and the rule matrix M
2: *RNI* – Read a new image I
3: *DRO* – Detect ROI
4: *DSP* – Determine seed points (x_s^j, y_s^j) inside ROI
5: *EXF* – Extract features $f_1, f_2, \cdots, f_m$ from the seed point's neighbourhood
6: *PFI* – Perform fuzzy inference to generate output(s): $p_1, p_2, \cdots, p_k$ = FUZZY-INFERENCE$(R_1, R_2, \cdots)$
7: *APS* – Apply the parameters to segment I
8: *FED* – Display the segment S and wait for the user feedback (user generates a gold standard image G by editing S)
9: ——— ***Rule Evolution - Invisible to User*** ———
10: *DPA* – Determine the best output(s) $p_1^*, p_2^*, \cdots, p_k^*$ (via comparison of S with G)
11: *PRU* – Pruning step: AddRow = FALSE if M contains a row similar to $\{f_1, f_2, \cdots, f_m, p_1^*, p_2^*, \cdots, p_k^*\}$, otherwise AddRow = TRUE
12: **if** AddRow=TRUE **then**
13: Add a new row to the rule matrix: $M(N_R+1, 1:m+k) = \{f_1, f_2, \cdots, f_m, p_1^*, p_2^*, \cdots, p_k^*\}$
14: *GFR* – Generate fuzzy rules $R_1, R_2, \cdots$ from the rule matrix M (e.g., using clustering)
15: *SAV* – Save the rule matrix M and the generated rules
16: **end if**





Fuzzy Evolution

IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 22, NO. 1, FEBRUARY 2014

EFIS—Evolving Fuzzy Image Segmentation

Ahmed A. Othman, Hamid R. Tizhoosh, *Member, IEEE*, and Farzad Khalvati

Machine Learning

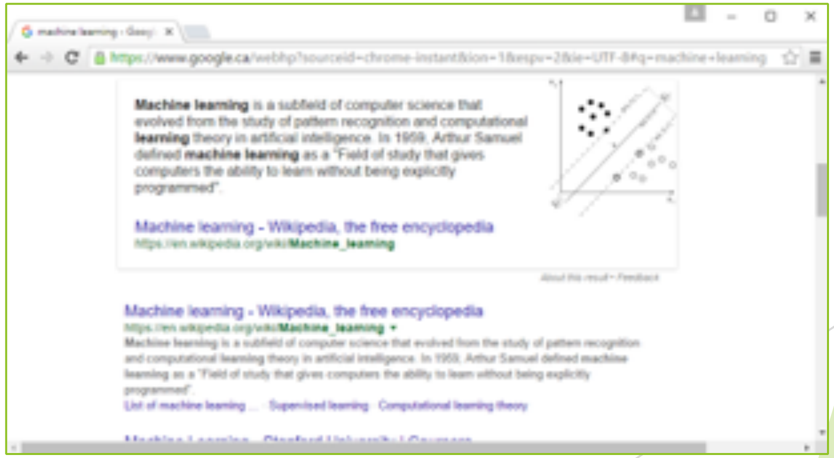

ASIMO
HONDA


IBM
POWER7

Heating Pressure Rice Cooker & Warmer

Bar Code Reader

OCR Pen



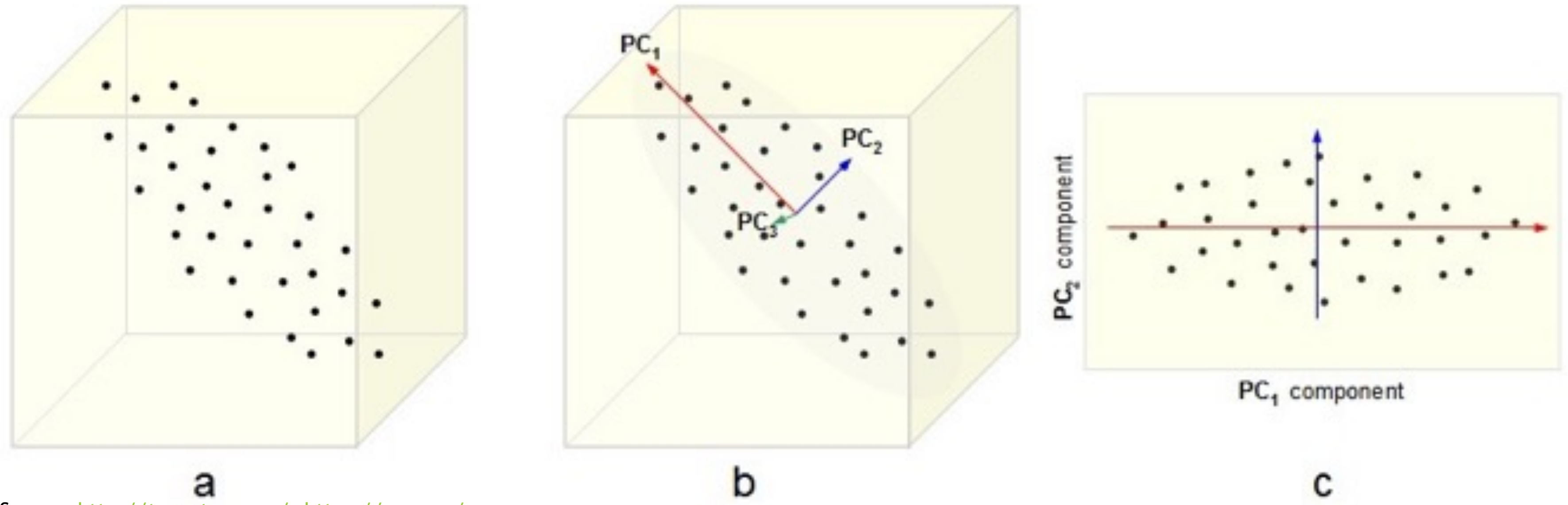
Smart Search

Source: http://earthxxii.deviantart.com/art/Etnogenez-Birth-of-Artificial-Intelligence-402715744

https://wall.alphacoders.com/by_sub_category.php?id=205999

A Brief History of Machine Learning

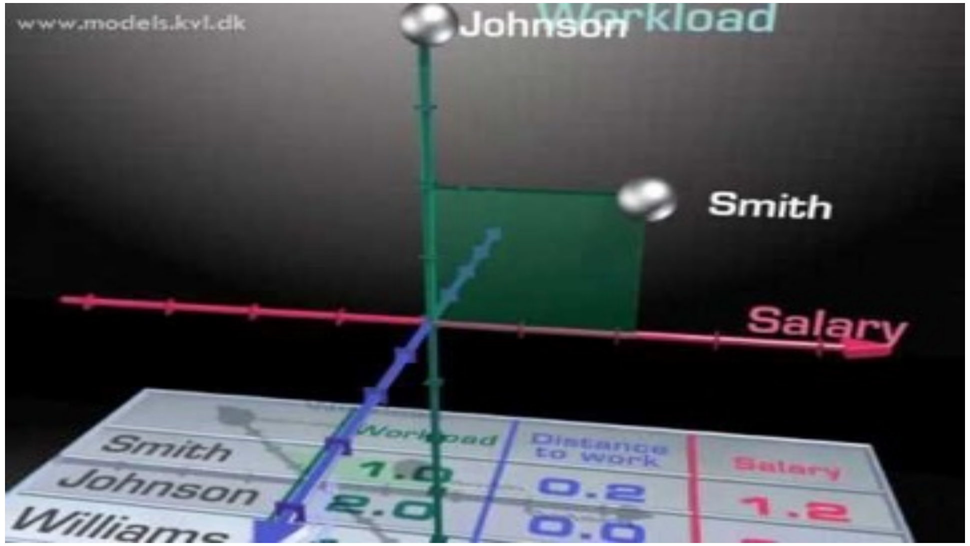
- 1901: First works on PCA (K.Person)
- 1933: PCA development (H. Hotelling)
- 2002: Principal Component Analysis (book by I. Jolliffe)



Source: http://taygetea.com/, https://cnx.org/

PCA

A Brief History of Machine Learning



Source: https://www.youtube.com/watch?v=4pnQd6jnCWk

PCA

A Brief History of Machine Learning

```
>>> import numpy as np
>>> from sklearn.decomposition import PCA
>>> X = np.array([[-1, -1], [-2, -1], [-3, -2], [1, 1], [2, 1], [3, 2]])
>>> pca = PCA(n_components=2)
>>> pca.fit(X)
PCA(copy=True, n_components=2, whiten=False)
>>> print(pca.explained_variance_ratio_)
[ 0.99244...  0.00755...]
```

Source: http://scikit-learn.org/

PCA

- 1950: Alan Turing’s seminal paper

Source: Copeland, B. J. *Artificial Intelligence* (Oxford: Blackwell, 1993)

A. M. Turing (1950) Computing Machinery and Intelligence. *Mind 49:* 433-460.

COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

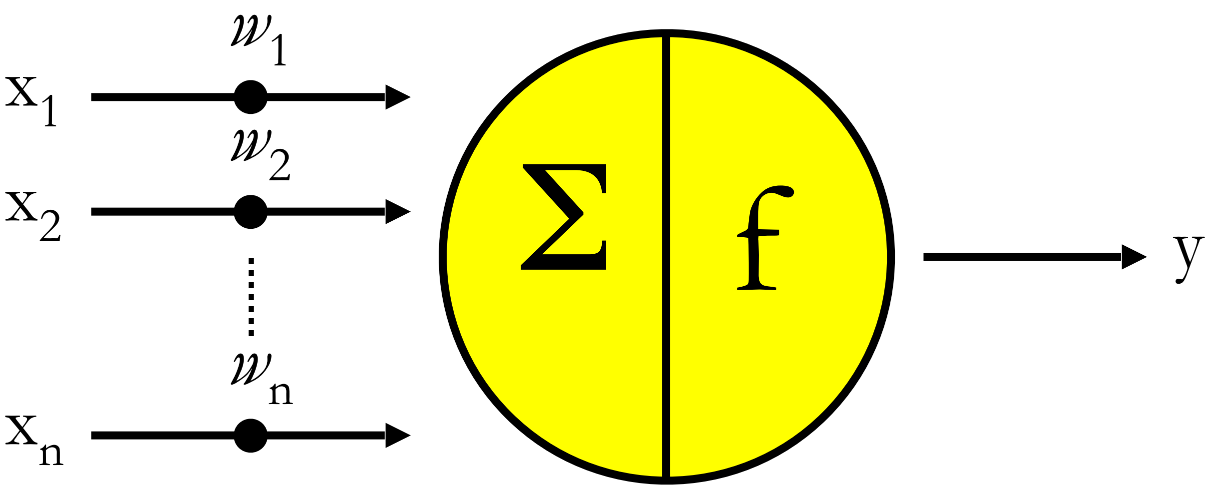
Turing Test

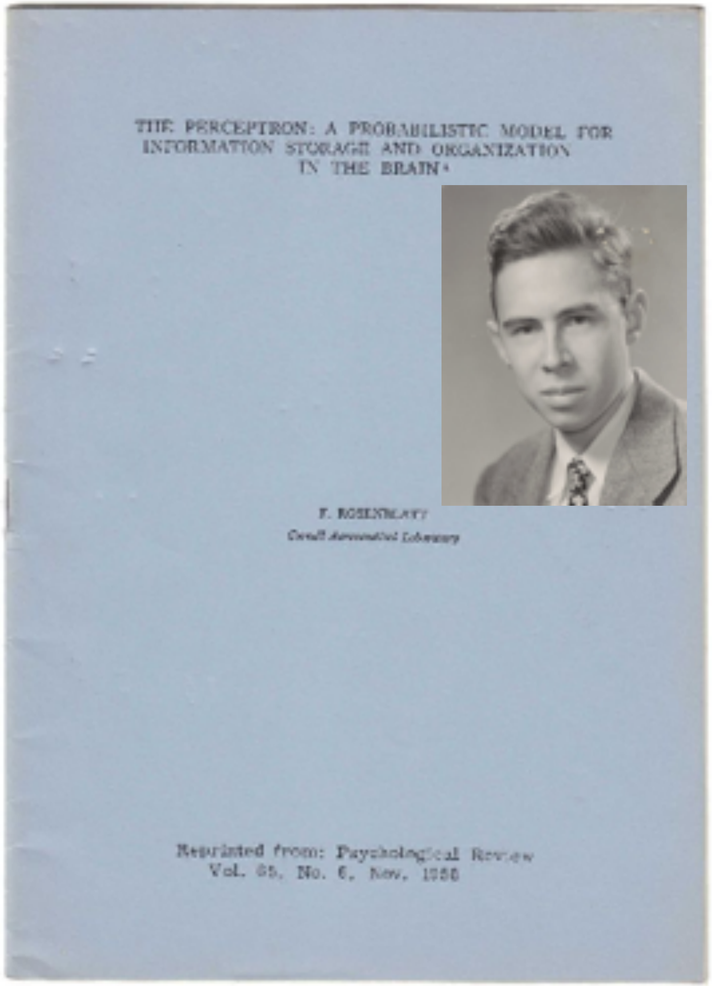
- 1980: Searle, John R., Minds, Brains, and programs

Source: macrovu.com

The Chinese Room

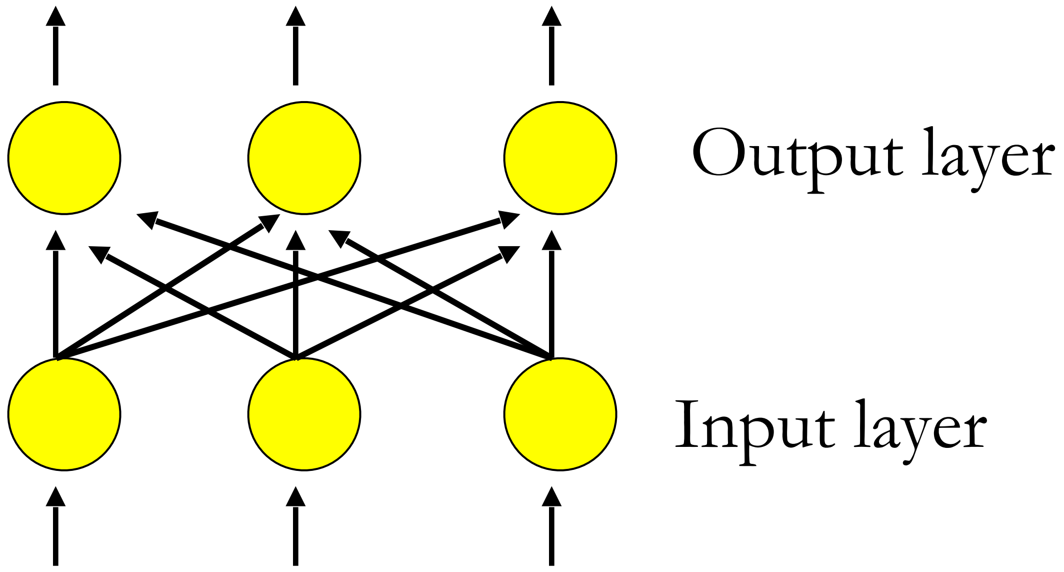
- 1958: Perceptrons (F. Rosenblatt)
- 1969: Limitations of Perceptrons (Minsky and Papert)





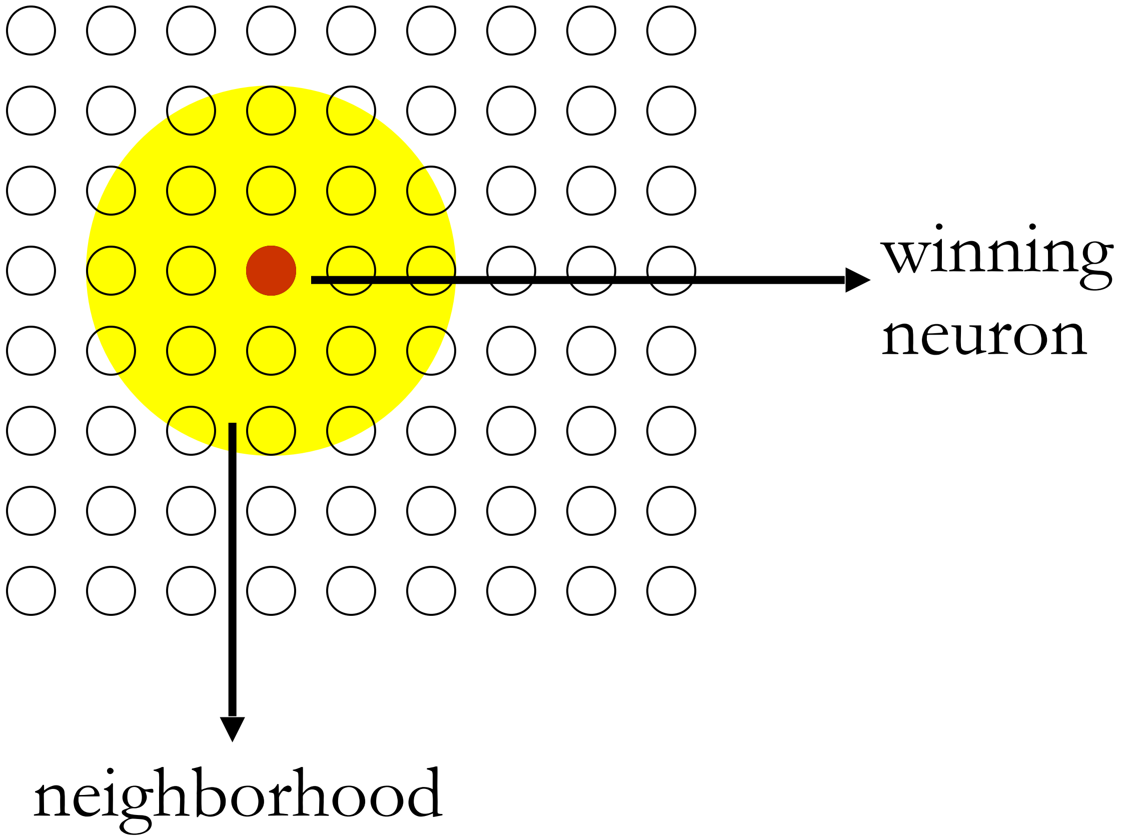
Perceptron

- 1982: Self-Organizing Maps (T. Kohonen)



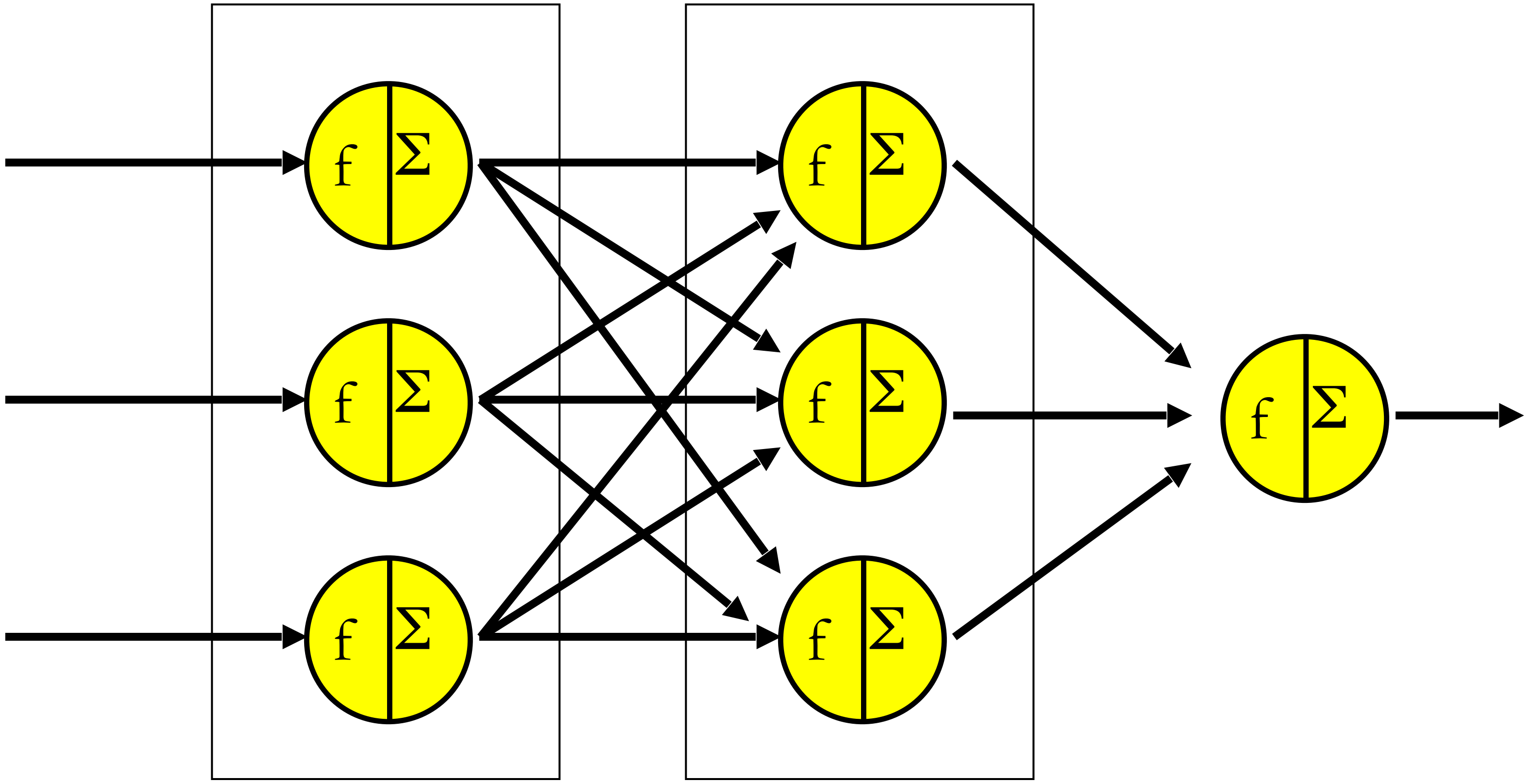
SOM

- 1982: Self-Organizing Maps (T. Kohonen)



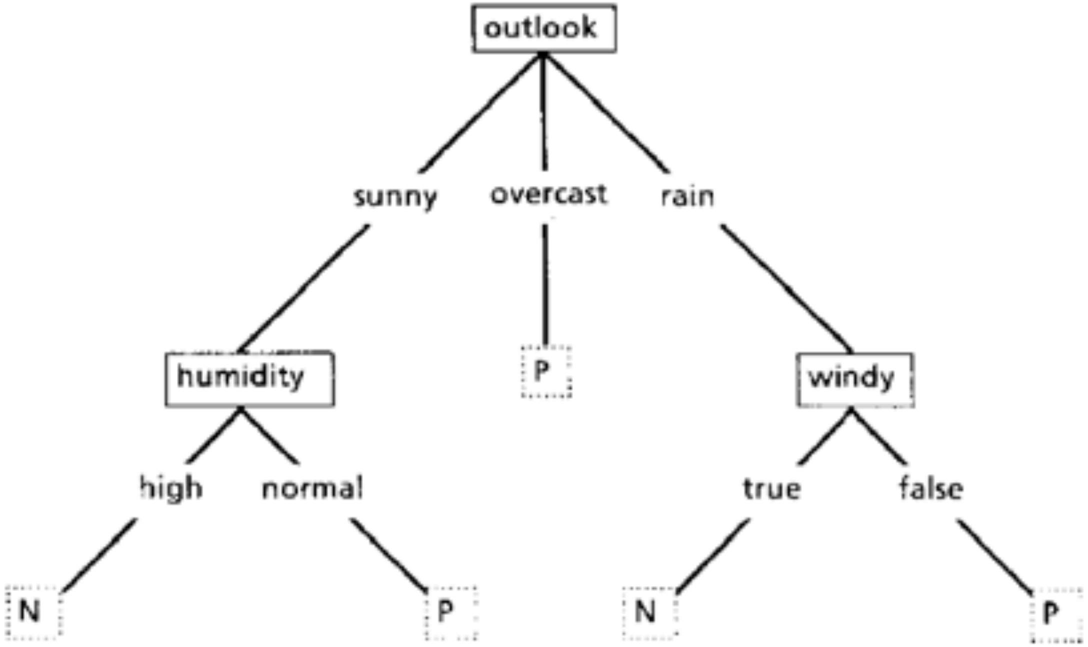
SOM

- 1986: Backpropagation (Rumelhart et al.)



BackProp

- 1986: ID3 algorithm (J.R.Quinlan)
- 1993: C 4.5 algorithm (J.R.Quinlan)

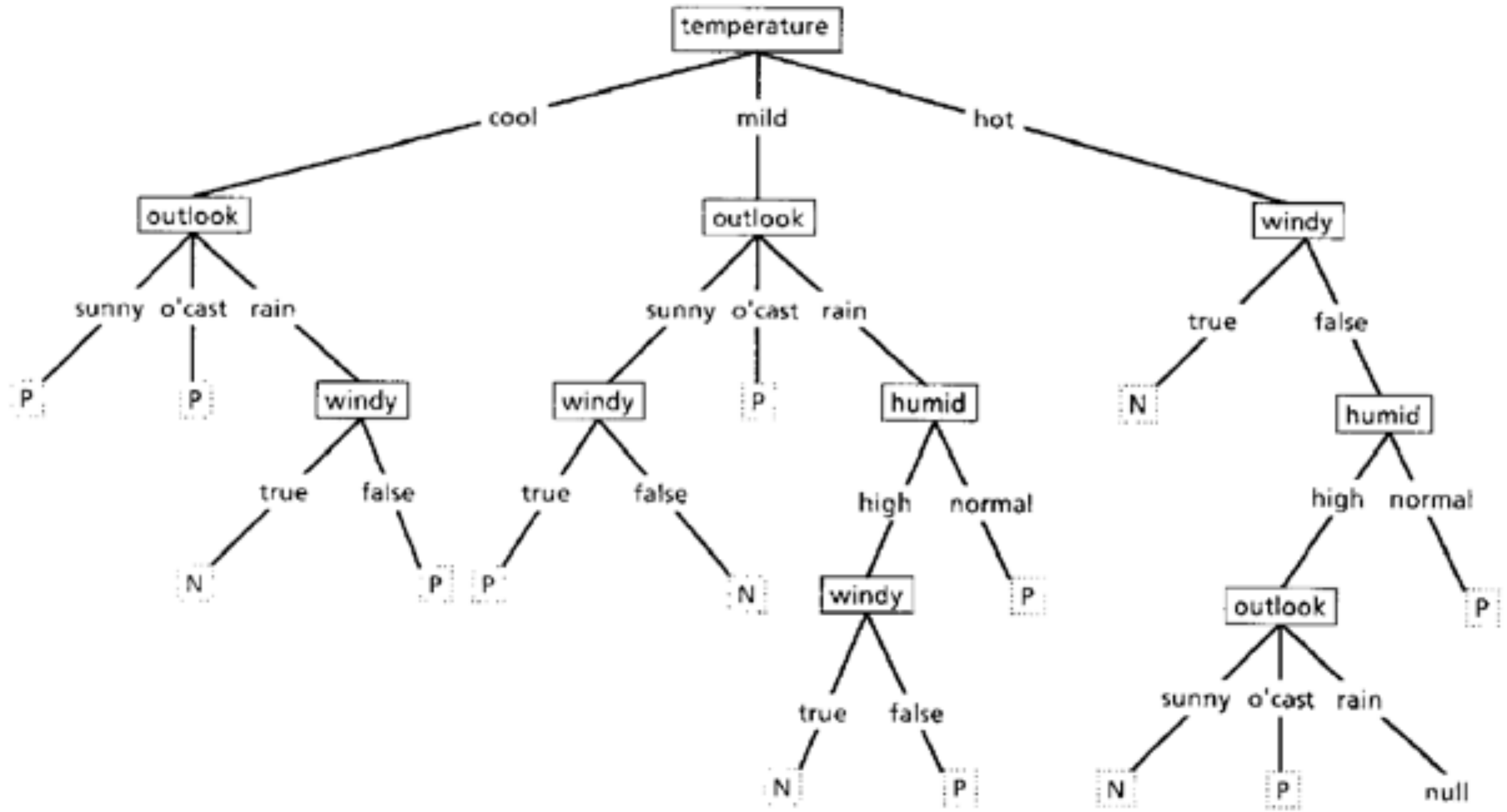


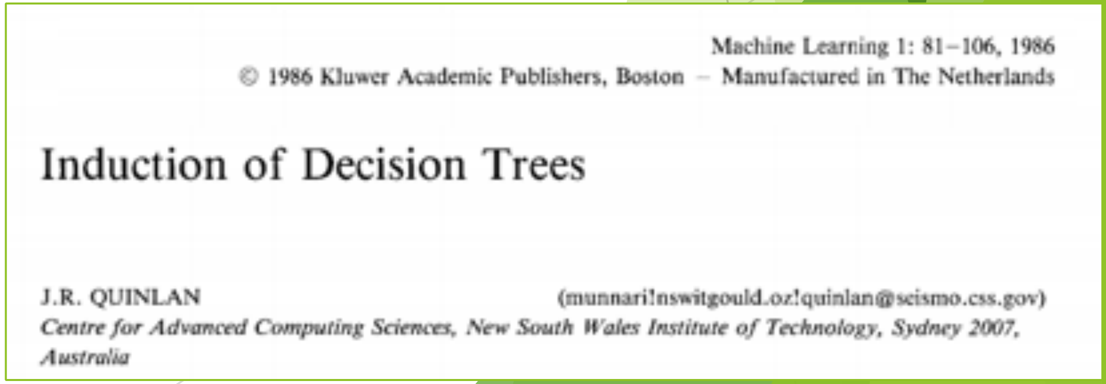
Machine Learning 1: 81–106, 1986

Induction of Decision Trees

J.R. QUINLAN (munnari!nswitgould.oz!quinlan@seismo.css.gov)
Centre for Advanced Computing Sciences, New South Wales Institute of Technology, Sydney 2007, Australia

Decision Trees




Machine Learning 1: 81–106, 1986

Induction of Decision Trees

J.R. QUINLAN (munnari!nswitgould.oz!quinlan@seismo.css.gov)
Centre for Advanced Computing Sciences, New South Wales Institute of Technology, Sydney 2007, Australia

Decision Trees

Positive and negative Samples:

$$I(p, n) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{n}{p+n}$$

Expected information of the tree with A as root:

$$E(A) = \sum_{i=1}^{v}\frac{p_i+n_i}{p+n}I(p_i, n_i)$$

Information gain for taking A:

$$\text{gain}(A) = I(p, n) - E(A)$$

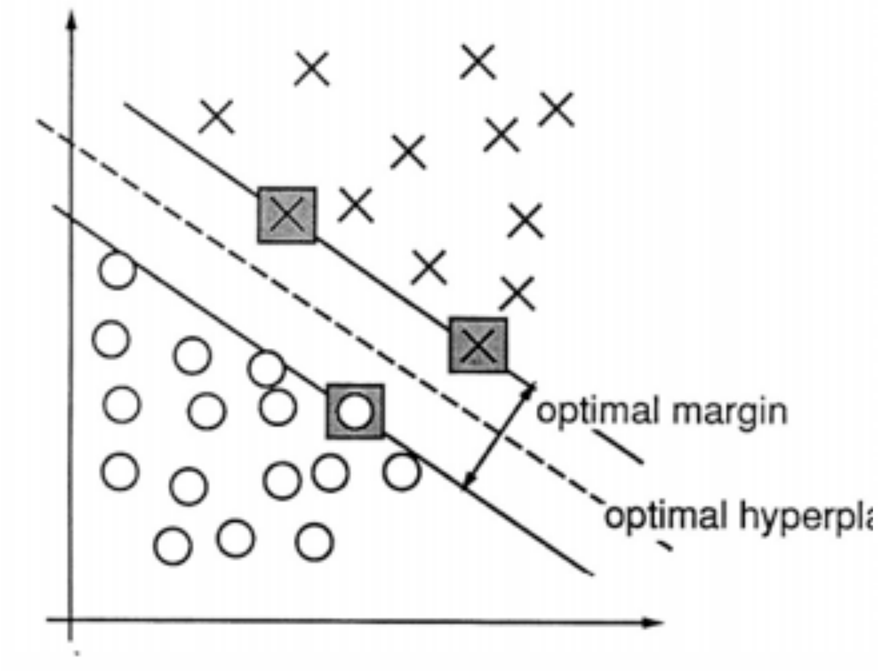
Machine Learning 1: 81–106, 1986

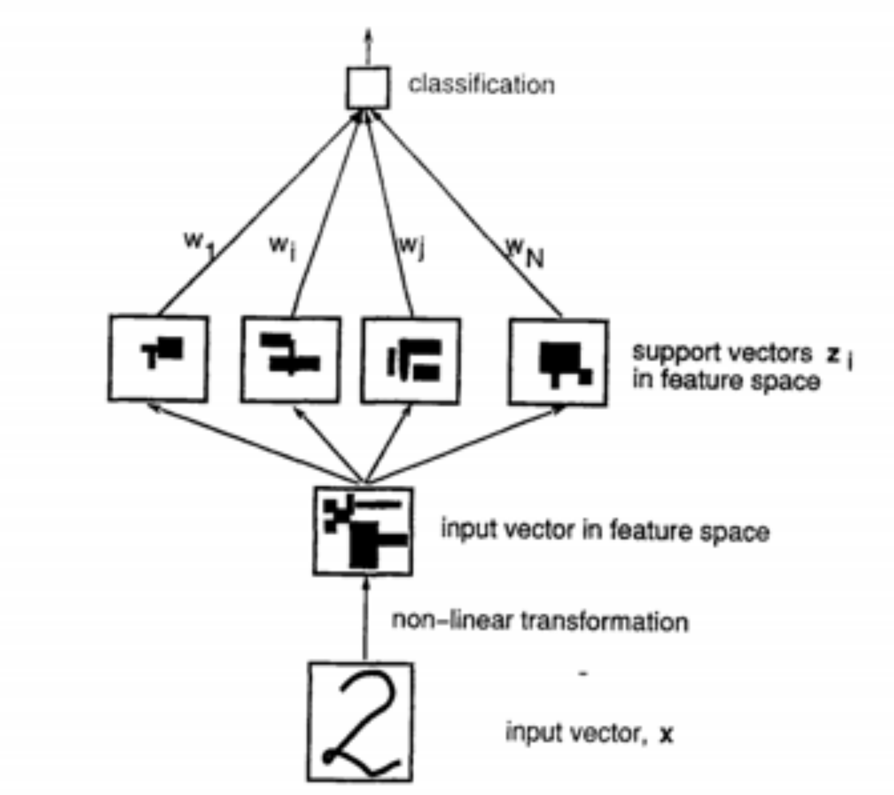
Induction of Decision Trees

J.R. QUINLAN (munnari!nswitgould.oz!quinlan@seismo.css.gov)
Centre for Advanced Computing Sciences, New South Wales Institute of Technology, Sydney 2007, Australia

Decision Trees

- 1995: Support Vector Machines (Cortes and Vapnik)





Machine Learning, 20, 273–297 (1995)

Support-Vector Networks

CORINNA CORTES corinna@neural.att.com
VLADIMIR VAPNIK vlad@neural.att.com
AT&T Bell Labs., Holmdel, NJ 07733, USA

SVM

The optimal hyperplane

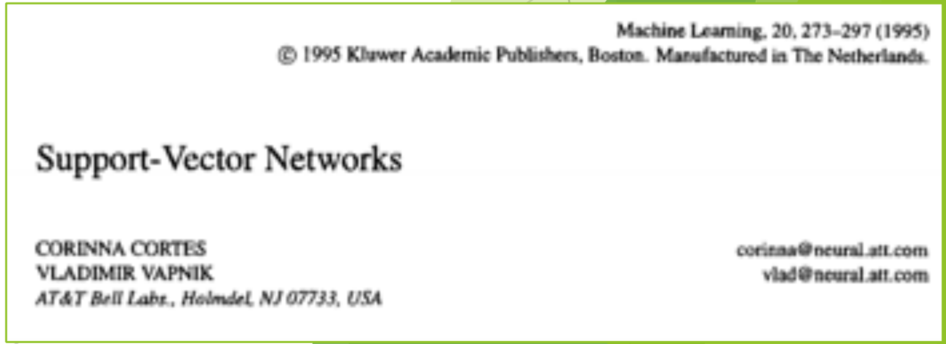
$$\mathbf{w}_0 \cdot \mathbf{z} + b_0 = 0$$

Optimal hyperplane as linear combination of support vectors

$$\mathbf{w}_0 = \sum_{\text{support vectors}} \alpha_i \mathbf{z}_i .$$

Linear decision in the feature space

$$I(\mathbf{z}) = \operatorname{sign}\left(\sum_{\text{support vectors}} \alpha_i \mathbf{z}_i \cdot \mathbf{z} + b_0 \right)$$



Machine Learning, 20, 273–297 (1995)
© 1995 Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.

Support-Vector Networks

CORINNA CORTES corinna@neural.att.com
VLADIMIR VAPNIK vlad@neural.att.com
AT&T Bell Labs., Holmdel, NJ 07733, USA

SVM

SVM

Labelled training data

$$(y_1, \mathbf{x}_1), \ldots, (y_\ell, \mathbf{x}_\ell), \qquad y_i \in \{-1, 1\}$$

Linearly separable

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x}_i + b \geq 1 \quad & \text{if} \quad y_i = 1, \\ \mathbf{w} \cdot \mathbf{x}_i + b \leq -1 \quad & \text{if} \quad y_i = -1, \end{aligned}$$

Or rewritten

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \qquad i = 1, \ldots, \ell.$$

The optimal hyperplane

$$\mathbf{w}_0 \cdot \mathbf{z} + b_0 = 0$$

Distance between two classes

$$\rho(\mathbf{w}, b) = \min_{\{\mathbf{x}: y=1\}} \frac{\mathbf{x} \cdot \mathbf{w}}{|\mathbf{w}|} - \max_{\{\mathbf{x}: y=-1\}} \frac{\mathbf{x} \cdot \mathbf{w}}{|\mathbf{w}|}.$$

..and for the optimal hyperplane

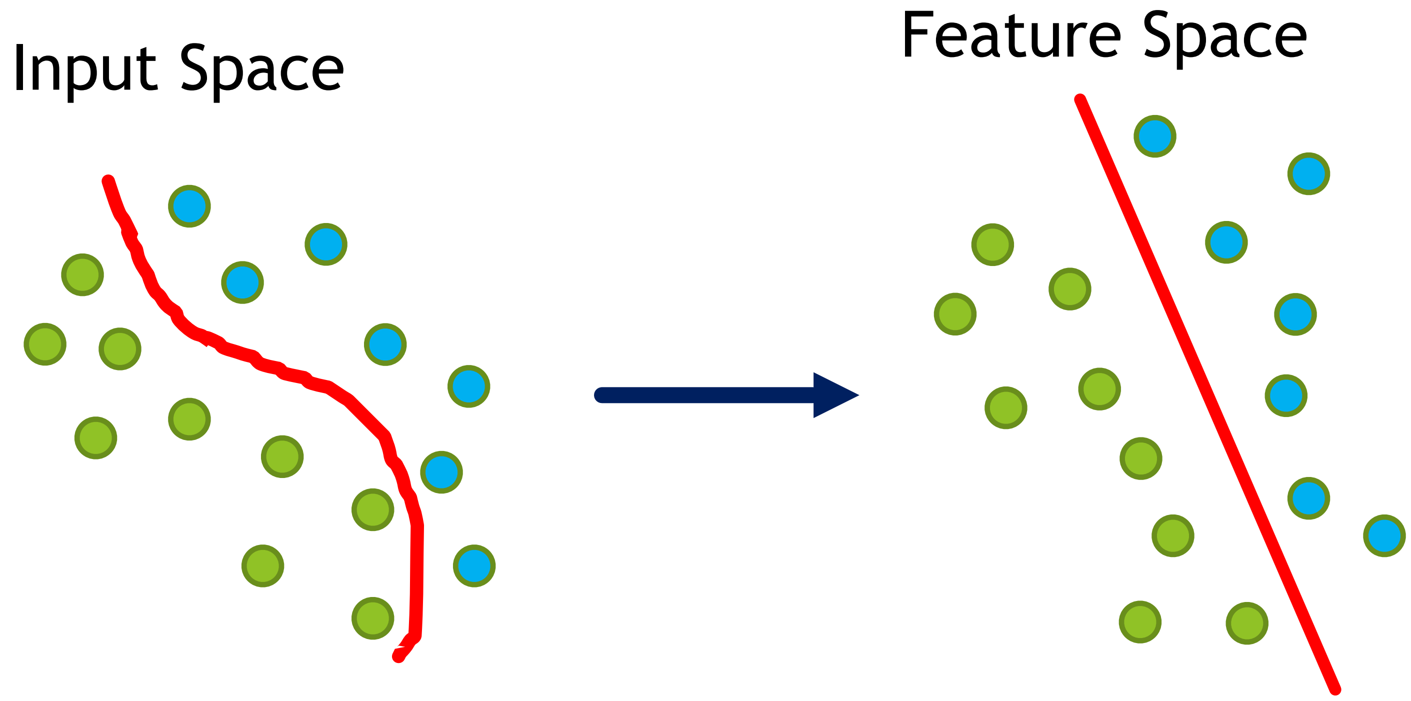
$$\rho(\mathbf{w}_0, b_0) = \frac{2}{|\mathbf{w}_0|} = \frac{2}{\sqrt{\mathbf{w}_0 \cdot \mathbf{w}_0}}$$

Machine Learning, 20, 273–297 (1995)

Support-Vector Networks

CORINNA CORTES corinna@neural.att.com
VLADIMIR VAPNIK vlad@neural.att.com
AT&T Bell Labs., Holmdel, NJ 07733, USA

Non-Linearly Separable Cases



SVM

- 1995: Boosting (Freund and Shapire)

Main idea:

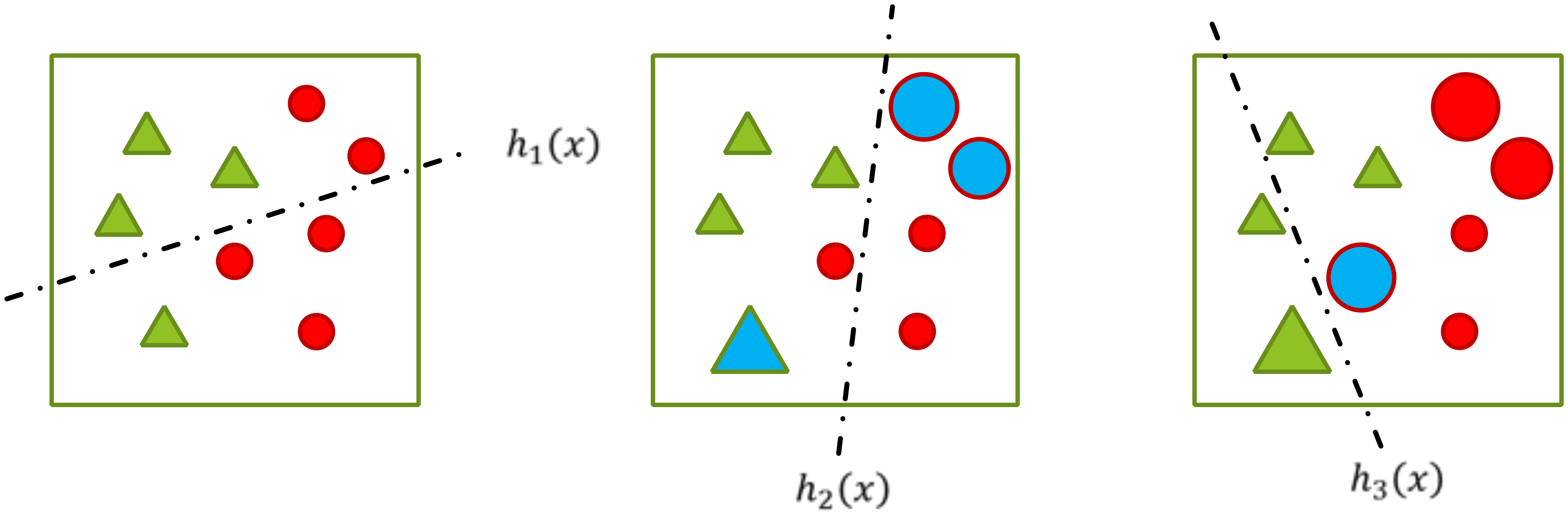
Combine multiple ***weak*** classifier (instead of trying to find the best classifier) in order to increase classification accuracy

Boosting

A decision-theoretic generalization of on-line learning and an application to boosting*

Yoav Freund   Robert E. Schapire

AT&T Labs



$$H(x) = sign(w_1 h_1(x) + w_2 h_2(x) + w_3 h_3(x))$$

Boosting

- 1995: Boosting (Freund and Shapire)

Given: $(x_1, y_1), \ldots, (x_m, y_m)$ where $x_i \in X, y_i \in Y = \{-1, +1\}$

Initialize $D_1(i) = 1/m$.

For $t = 1, \ldots, T$

Train weak learner with Dt

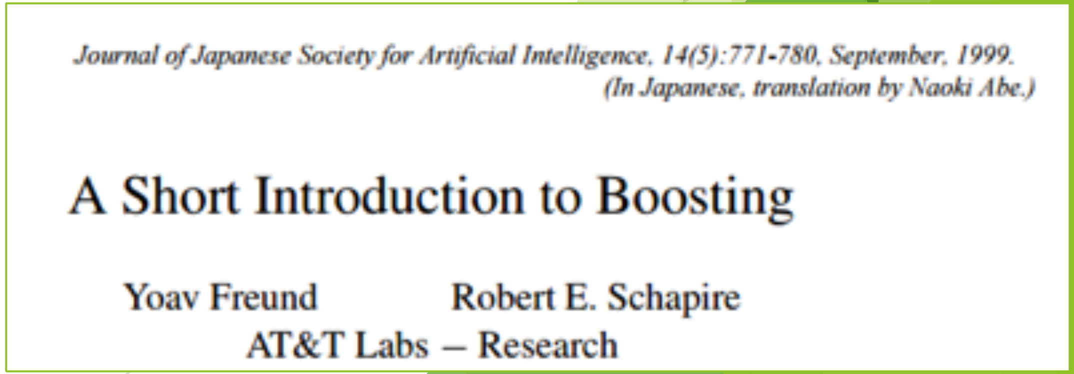
Get weak hypothesis and its error: $h_t : X \to \{-1, +1\}$ $\epsilon_t = \Pr_{i \sim D_t}[h_t(x_i) \neq y_i]$

Set $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right)$

Update
$$\begin{aligned} D_{t+1}(i) &= \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{if } h_t(x_i) \neq y_i \end{cases} \\ &= \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t} \end{aligned}$$

Output final solution:

$$H(x) = \operatorname{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x) \right)$$


Journal of Japanese Society for Artificial Intelligence, 14(5):771-780, September, 1999.
(In Japanese, translation by Naoki Abe.)

A Short Introduction to Boosting

Yoav Freund Robert E. Schapire
AT&T Labs — Research

Boosting

- 1995: Random Decision Forests (T.K. Ho)
- 2001: Random Forests (L.Breiman)

- A forest is a collection of several hundred to several thousand trees.
- The forest error depends on:
  1. Correlation between trees: more correlation more error
  2. The quality of each individual tree (lower error)

Random Forests

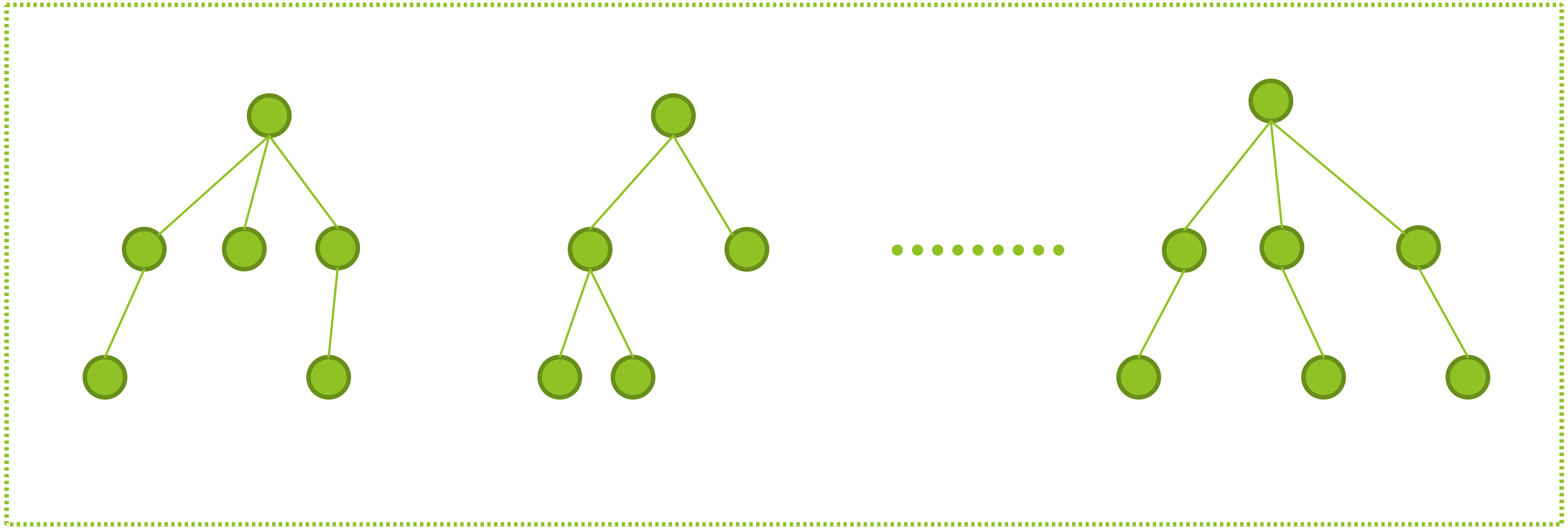
RANDOM FORESTS

Leo Breiman
Statistics Department
University of California
Berkeley, CA 94720

Random Decision Forests

Tin Kam Ho
AT&T Bell Laboratories
600 Mountain Avenue, 2C–548C
Murray Hill, NJ 07974, USA

Ensemble learning with a multitude of trees



Forest

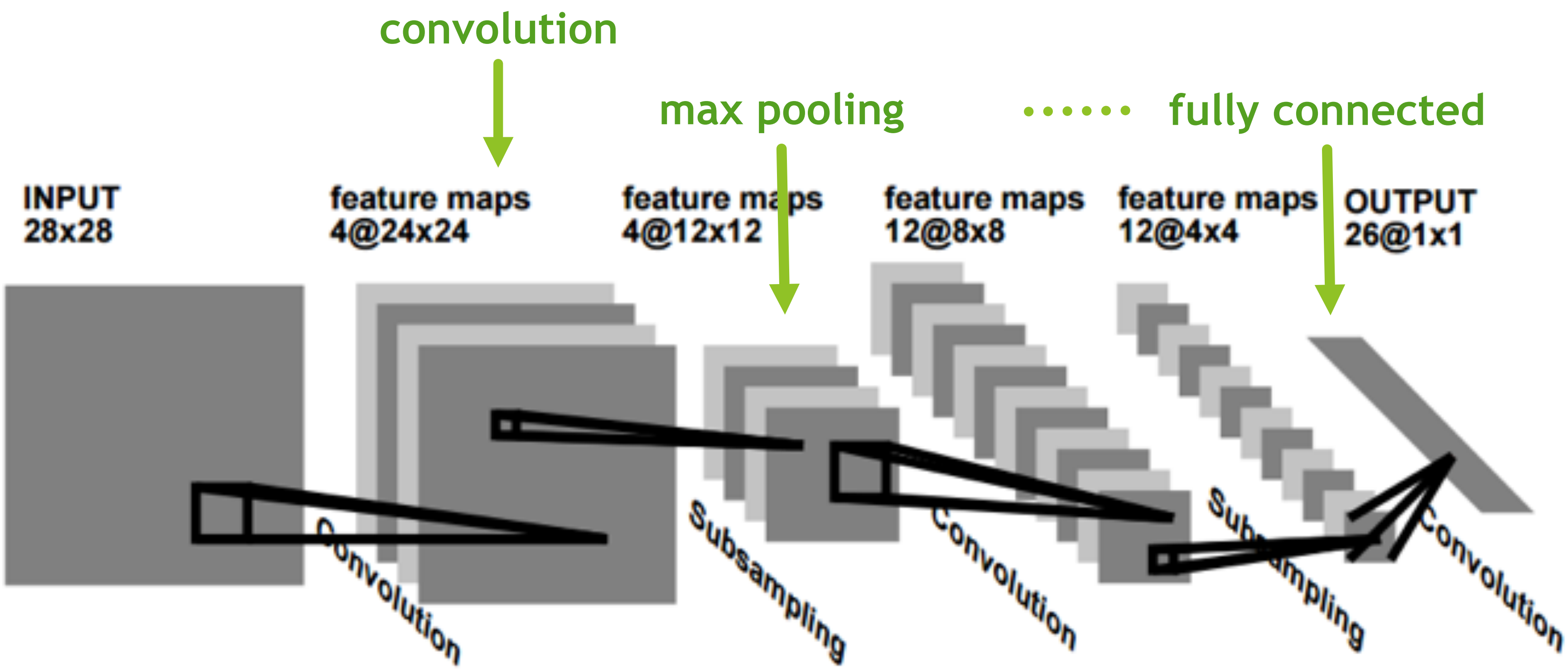
Random Forests

- 1995: Convolutional Neural Networks (LeCun and Bengio)
- 2006: Fast learning for deep belief nets (Hinton et al.)
- 2007: Greedy layer-wise training for deep nets (Bengio et al.)

Idea:
Neural networks can learn difficult recognition tasks if designed with more hidden layers
(i.e., more than 5 hidden layers)

Challenge:
Training of deep networks was until recently practically impossible

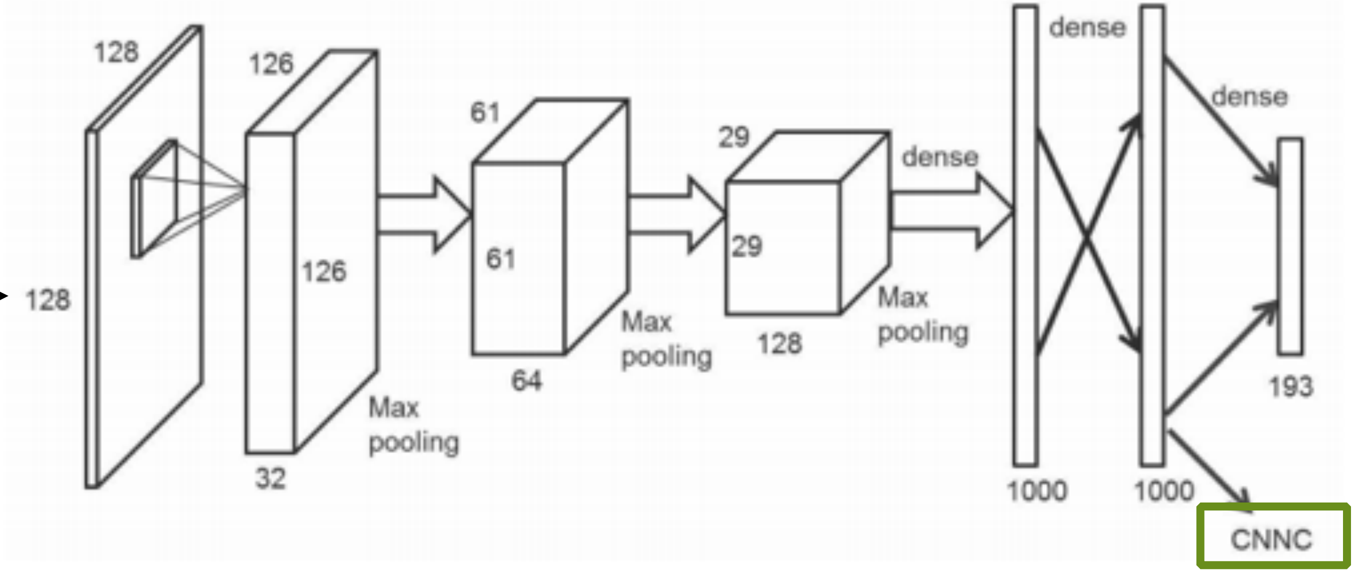
Deep Nets



Convolutional Networks for Images, Speech, and Time-Series

Yann LeCun   Yoshua Bengio

Deep Nets



[-1.9 3.7 12.9 -23.8 … 21.6]

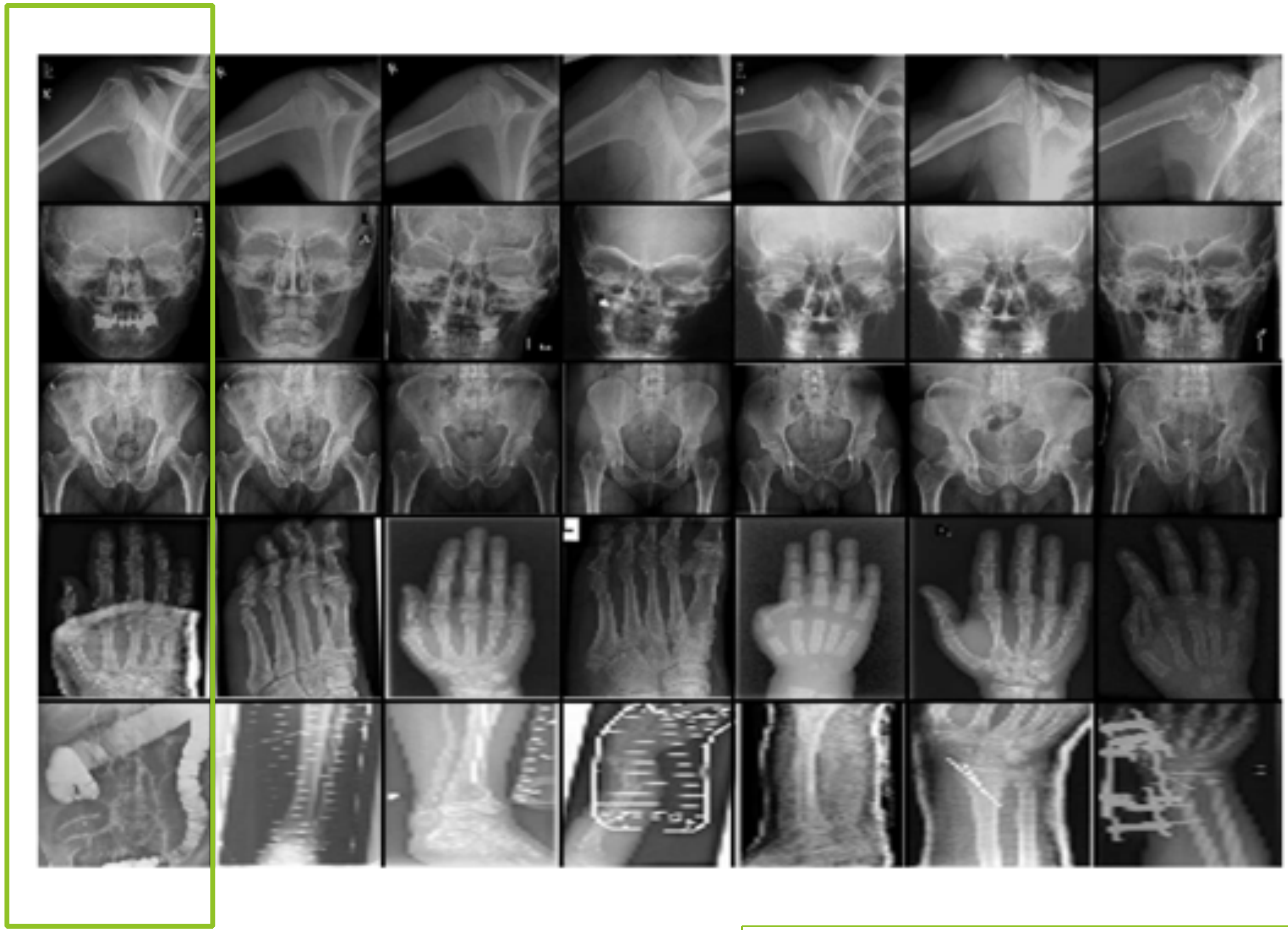
Deep Nets


To appear in proceedings of The 2016 IEEE International Joint Conference on Neural Networks (IJCNN 2016), July 24-29, 2016, Vancouver, Canada

GENERATING BINARY TAGS FOR FAST MEDICAL IMAGE RETRIEVAL BASED ON CONVOLUTIONAL NETS AND RADON TRANSFORM

Xinran Liu†, H.R.Tizhoosh‡, J.Kofman†‡

Query
Images



Deep Nets

To appear in proceedings of The 2016 IEEE International Joint Conference on Neural Networks (IJCNN 2016), July 24-29, 2016, Vancouver, Canada

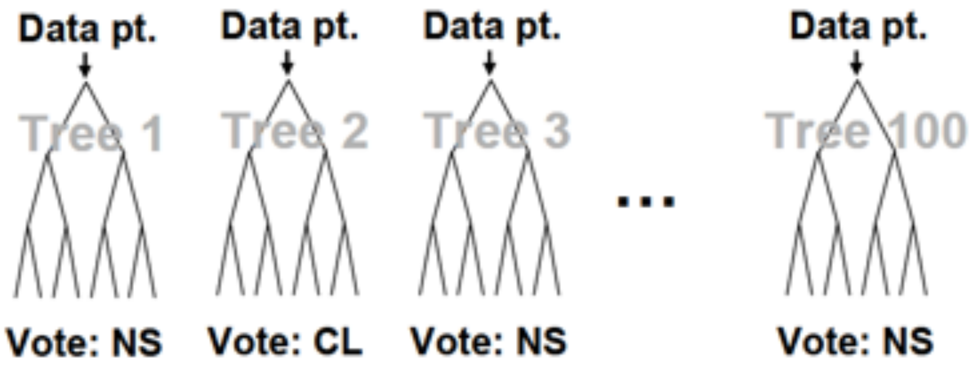
GENERATING BINARY TAGS FOR FAST MEDICAL IMAGE RETRIEVAL BASED ON CONVOLUTIONAL NETS AND RADON TRANSFORM

Xinran Liu[†], *H.R.Tizhoosh*[‡], *J.Kofman*[†‡]

Fuzzy Logic & Machine Learning

Fuzzy Logic & Random Forest

- Random forests, ensembles of weakly-correlated decision trees, can be used in concert with fuzzy logic concepts to both classify storm types based on a number of radar-derived storm characteristics and provide a measure of "confidence" in the resulting classifications.
- The random forest technique provides measures of variable importance and interactions, as well as methods for addressing missing data for transforming the input data and structuring the final classification algorithm.




USING RANDOM FORESTS AND FUZZY LOGIC FOR AUTOMATED STORM TYPE IDENTIFICATION

John K. Williams* and Jennifer Abernethy
National Center for Atmospheric Research, Boulder, Colorado

Fuzzy Logic & Random Forest

- Fuzzy Random Forest: A Multi-classifier based on a “forest” of randomly generated fuzzy decision trees
- Combining
  - the robustness of multi-classifiers
  - the construction efficiency of decision trees
  - the power of the randomness o increase the diversity, and
  - the flexibility of fuzzy logic for data managing

A Fuzzy Random Forest: Fundamental for Design and Construction

P.P. Bonissone
GE Global Research
One Research Circle
Niskayuna, NY 12309. USA
bonissone@crd.ge.com

J.M. Cadenas M.C. Garrido
Dept. Ingeniería de la Información
y las Comunicaciones
Universidad de Murcia. Spain
jcadenas@um.es carmengarrido@um.es

R.A. Díaz-Valladares
Dept. Ciencias Computacionales
Universidad de Montemorelos
Mexico
rdiaz@um.edu.mx

Fuzzy Logic & Random Forest

1. Start with examples set of entry
2. At any node N still to be expanded, compute the number of examples of each class.
3. Compute the standard information content.
4. At each node search the set of remaining attributes to split the node
    - Select with any criteria, the candidate attributes set to split the node.
    - Compute the information content to each child node obtained from each candidate attribute.
    - Select the candidate attribute such that information gain is maximal.
5. Divide N in sub-nodes according to possible outputs of the attribute selected in the previous step.
6. Repeat steps 2-5 to stop criteria is satisfied in all nodes.

A Fuzzy Random Forest: Fundamental for Design and Construction

P.P. Bonissone
GE Global Research
One Research Circle
Niskayuna, NY 12309. USA
bonissone@crd.ge.com

J.M. Cadenas **M.C. Garrido**
Dept. Ingeniería de la Información y las Comunicaciones
Universidad de Murcia. Spain
jcadenas@um.es carmengarrido@um.es

R.A. Díaz-Valladares
Dept. Ciencias Computacionales
Universidad de Montemorelos
Mexico
rdiaz@um.edu.mx

Fuzzy Logic & Random Forest

Begin
  TreeClasification.
  ForestClasification.
End
TreeClasification $(Random\ Forest, A[t][i])$
Begin
  For each Tree t
    For each Class i
      $A[t][i] = Faggre(K(1, \chi_{t,1}(e), \omega_{t,1}), \ldots,$
        $K(1, \chi_{t,N_t}(e), \omega_{t,N_t}), \ldots, K(I, \chi_{t,1}(e), \omega_{t,1}),$
        $, \ldots, K(I, \chi_{t,N_t}(e), \omega_{t,N_t}))$
    End For each Class
  End For each Tree
End
ForestClasification $(A[t][i], F[i])$
Begin
  For each Class i
    $F[i] = Faggre(A[1][i], \ldots, A[T][i])$
  End For each Class
End

Algorithm 4. FRF Inference (Strategy 1)

A Fuzzy Random Forest: Fundamental for Design and Construction

P.P. Bonissone
GE Global Research
One Research Circle
Niskayuna, NY 12309. USA
bonissone@crd.ge.com

J.M. Cadenas **M.C. Garrido**
Dept. Ingeniería de la Información
y las Comunicaciones
Universidad de Murcia. Spain
jcadenas@um.es carmengarrido@um.es

R.A. Díaz-Valladares
Dept. Ciencias Computacionales
Universidad de Montemorelos
Mexico
rdiaz@um.edu.mx

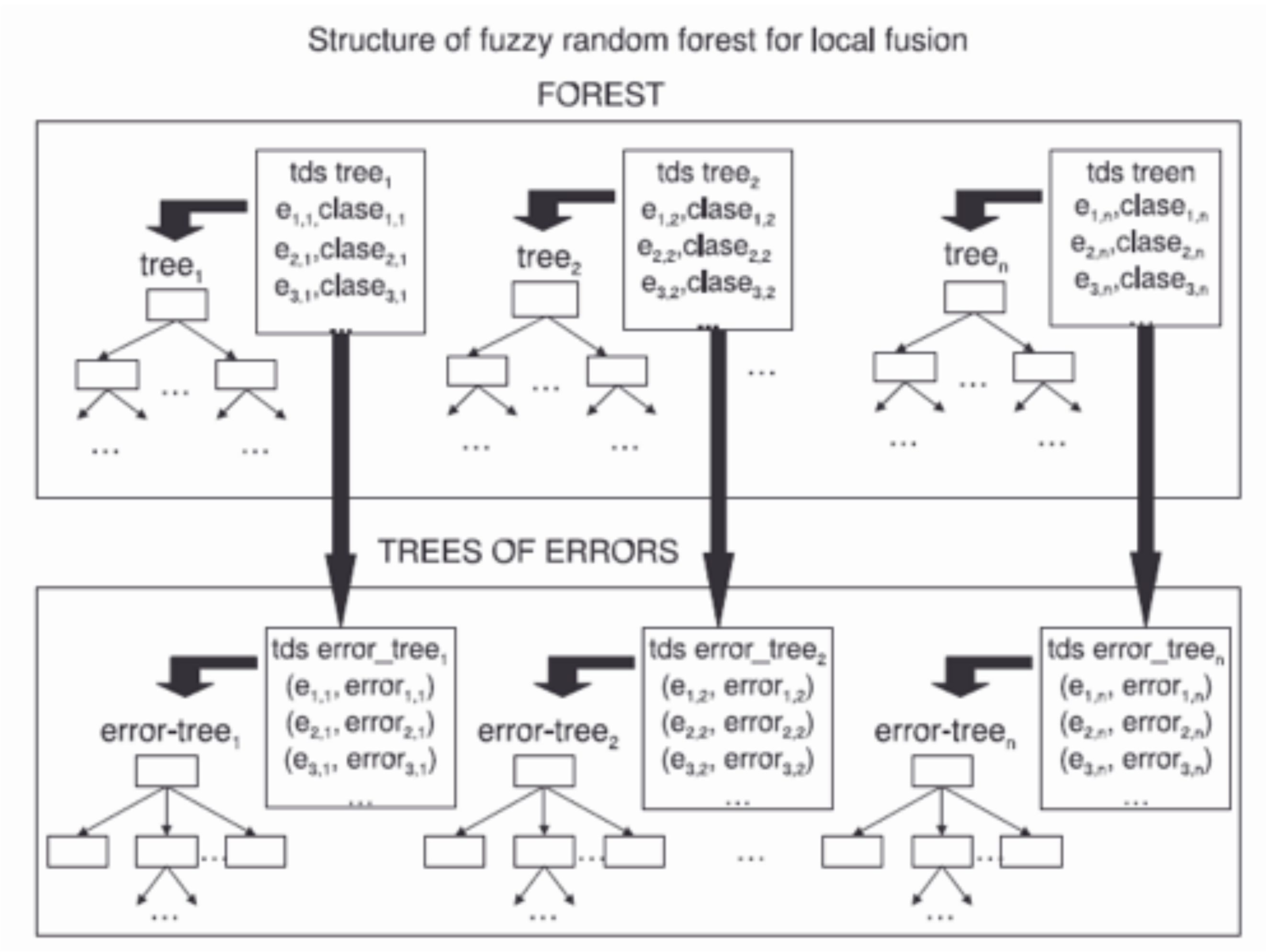
Fuzzy Logic & Random Forest

- *A multi-classifier system based on a forest of randomly generated fuzzy decision trees (Fuzzy Random Forest),*
- *New method to combine their decisions to obtain the final decision of the forest.*
- *The proposed combination is a weighted method based on the concept of local fusion and on the data set Out Of Bag (OOB) error.*

IFSA-EUSFLAT 2009

Weighted decisions in a Fuzzy Random Forest

P.P. Bonissone[1] J.M. Cadenas[2] M.C. Garrido[2] R.A. Díaz-Valladares[3] R. Martínez[2]



IFSA-EUSFLAT 2009

Weighted decisions in a Fuzzy Random Forest

P.P. Bonissone[1] J.M. Cadenas[2] M.C. Garrido[2] R.A. Díaz-Valladares[3] R. Martínez[2]

Fuzzy Logic & PCA

- Principal Components Analysis is sensitive to outliers, missing data, and poor linear correlation between variables.
- Data transformations have a large impact upon PCA.
- Robust fuzzy PCA algorithm (FPCA): The matrix data is fuzzified, thus diminishing the influence of the outliers.

STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume XLVI, Number 2, 2001

PRINCIPAL COMPONENTS ANALYSIS BASED ON A FUZZY SETS APPROACH

HORIA F. POP

Fuzzy Logic & PCA

- Nonlinear fuzzy robust principal component analysis (NFRPCA) algorithm
- After this preprocessing step the similarity classifier is then used for the actual classification.
- The procedure was tested for dermatology, hepatitis and liver-disorder data.
- Compared to results with classical PCA and the similarity classifier, higher accuracies were achieved with the approach using nonlinear fuzzy robust principal component analysis and the similarity classifier.

International Journal of Fuzzy Systems, Vol. 13, No. 3, September 2011 153

A New Nonlinear Fuzzy Robust PCA Algorithm and Similarity Classifier in Classification of Medical Data Sets

Pasi Luukka

Fuzzy Logic & PCA

- A new method called PCA-TF is proposed that allows performing PCA on data sets of trapezoidal (or triangular) fuzzy numbers, that may contain also real numbers and intervals.
- A group of orthogonal axes is found that permits the projection of the maximum variance of a real numbers' matrix, where each number represents a trapezoidal fuzzy number.
- The initial matrix of fuzzy numbers is projected to these axes by means of fuzzy numbers arithmetic, which yields Principal Components and they are also fuzzy numbers.
- Based on these components it is possible to produce graphs of the individuals in two-dimensional plane.
- It is also possible to evaluate the shape of the ordered pai visualize the membership function for each point on the z xy plane.

Principal Components Analysis for Trapezoidal Fuzzy Numbers

Alexia Pacheco[1] and Oldemar Rodríguez[2]

[1] Costarican Institute of Electricity, San José, Costa Rica apacheco@ice.go.cr
[2] School of Mathematics, University of Costa Rica, San José, Costa Rica oldemar.rodriguez@ucr.ac.cr

Fuzzy Logic & PCA

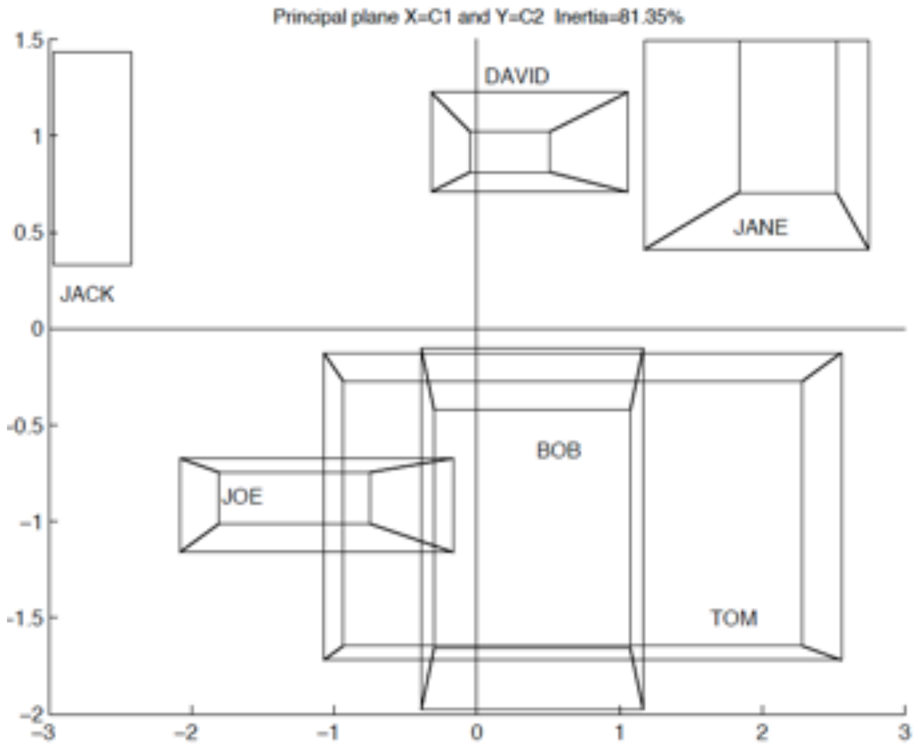


Fig. 1. Principal plane using the first two principal components.

Table 1. Student dataset [4]

	M1	M2	P1	P2
TOM	15	fairly good	unknown	[14,16]
DAVID	9	good	fairly good	10
BOB	6	[10,11]	[13,20]	bien
JANE	very bad	very good	19	[10,12])
JOE	(0,0,2,6)	fairly good	[10,14]	14
JACK	(1,1,1,1)	[4,6]	9	[6,9]

[4] T. Denoeux, M. Masson Principal Component Analysis of Fuzzy Data using Autossociative Neural Networks. IEEE Transactions on fuzzy systems, 12:336-349, 2004.

Principal Components Analysis for Trapezoidal Fuzzy Numbers

Alexia Pacheco[1] and Oldemar Rodríguez[2]

[1] Costarican Institute of Electricity, San José, Costa Rica apacheco@ice.go.cr
[2] School of Mathematics, University of Costa Rica, San José, Costa Rica oldemar.rodriguez@ucr.ac.cr

Fuzzy Logic & PCA

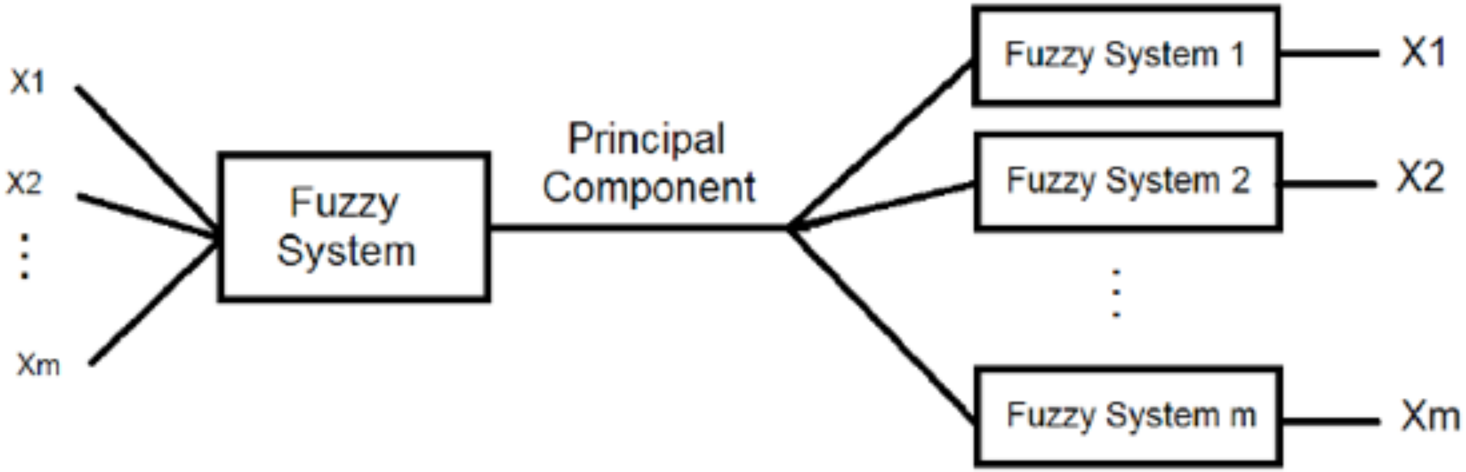
- *Hybrid approach: Fuzzy + PCA*
- *A new PCA based monitoring that uses fuzzy logic capability*
- *The reason to use fuzzy logic: its good ability to approximate nonlinear function with arbitrary accuracy*
- *Tested on Tennessee Eastman Process*

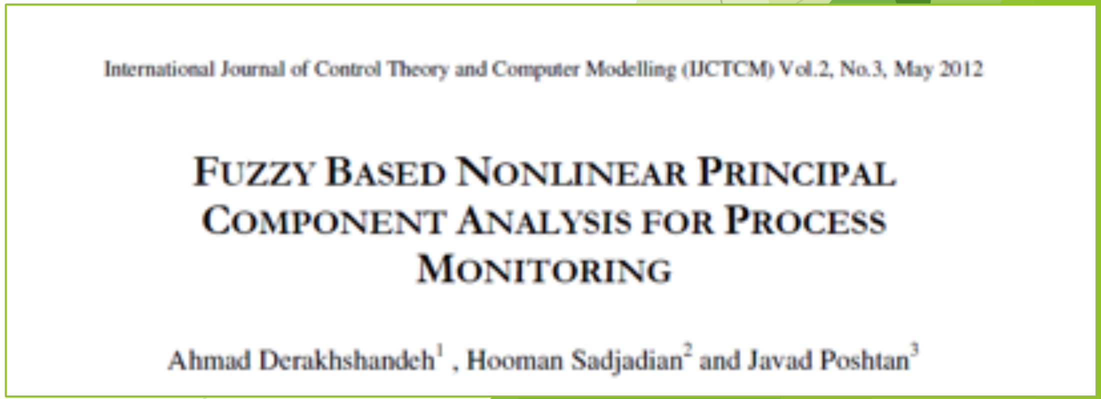
International Journal of Control Theory and Computer Modelling (IJCTCM) Vol.2, No.3, May 2012

FUZZY BASED NONLINEAR PRINCIPAL COMPONENT ANALYSIS FOR PROCESS MONITORING

Ahmad Derakhshandeh[1] , Hooman Sadjadian[2] and Javad Poshtan[3]

Fuzzy Logic & PCA

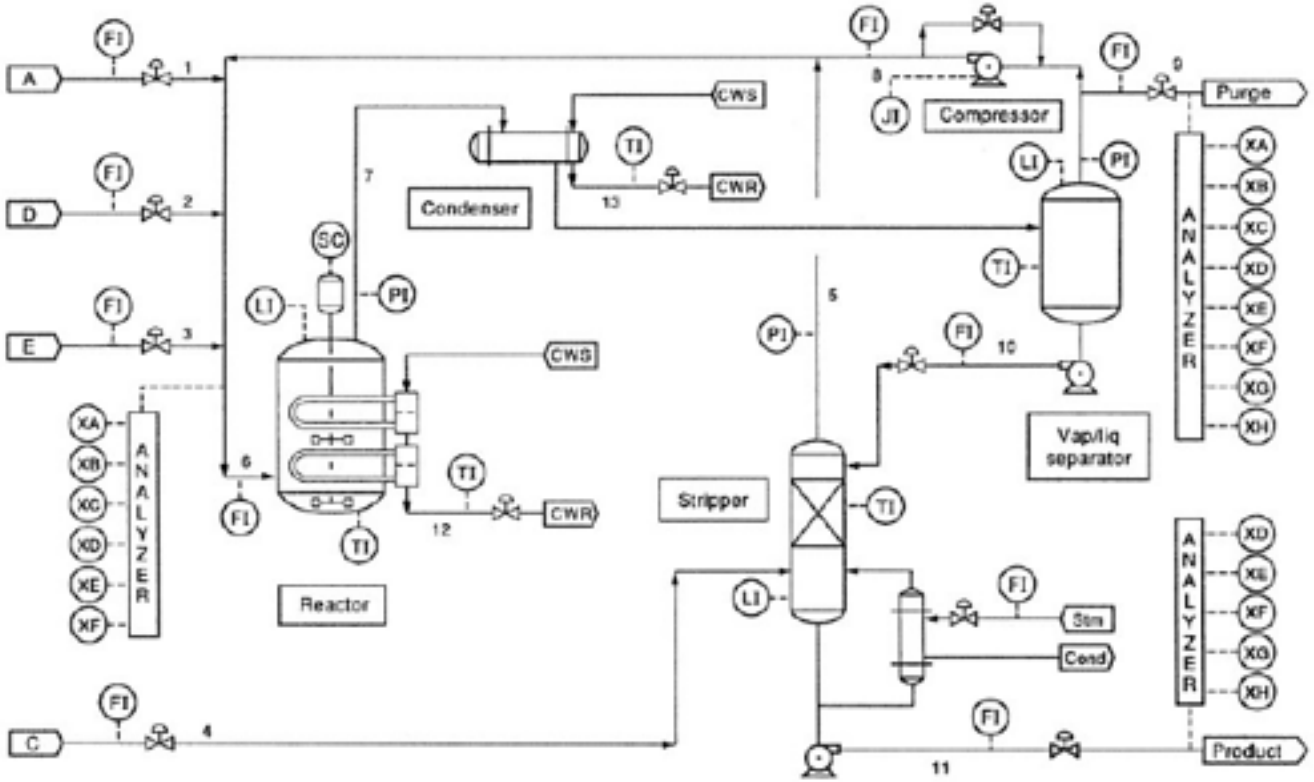



International Journal of Control Theory and Computer Modelling (IJCTCM) Vol.2, No.3, May 2012

FUZZY BASED NONLINEAR PRINCIPAL COMPONENT ANALYSIS FOR PROCESS MONITORING

Ahmad Derakhshandeh[1], Hooman Sadjadian[2] and Javad Poshtan[3]

Fuzzy Logic & PCA



The Tennessee Eastman (TE) Challenge Process

International Journal of Control Theory and Computer Modelling (IJCTCM) Vol.2, No.3, May 2012

Fuzzy Based Nonlinear Principal Component Analysis for Process Monitoring

Ahmad Derakhshandeh[1] , Hooman Sadjadian[2] and Javad Poshtan[3]

Fuzzy Logic & SVM

- A support vector machine (SVM) learns the decision surface from two distinct classes of the input points.
- A fuzzy membership is applied to each input point
- The SVMs are reformulated such that different input points can make different contributions to the learning of decision surface.

464 IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 13, NO. 2, MARCH 2002

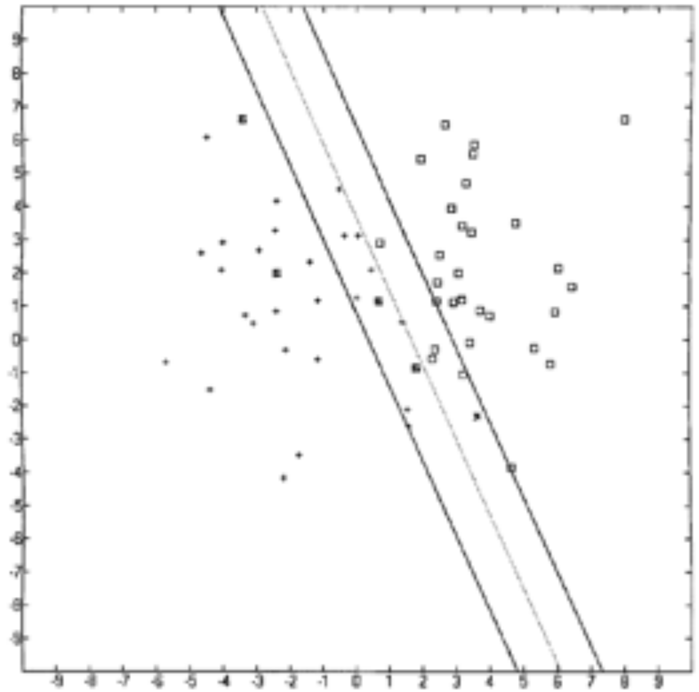
Fuzzy Support Vector Machines

Chun-Fu Lin and Sheng-De Wang

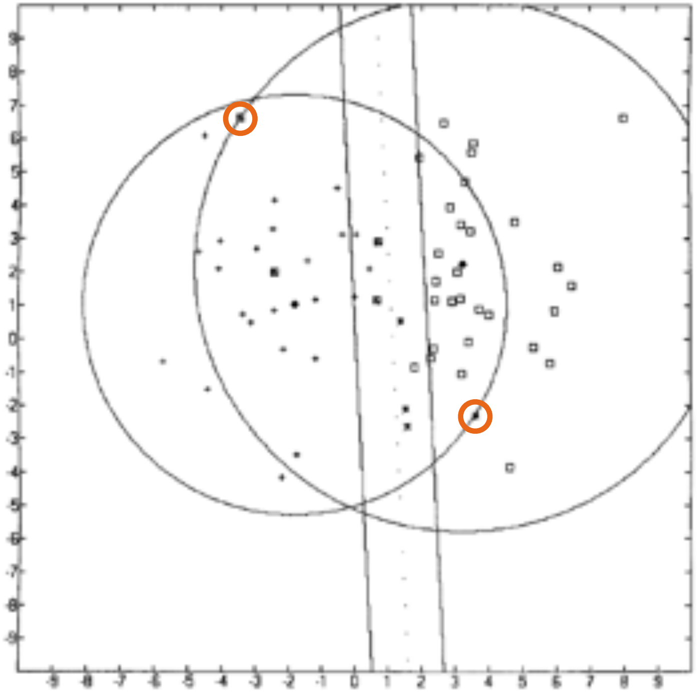
Abstract—A support vector machine (SVM) learns the decision surface from two distinct classes of the input points. In many applications, each input point may not be fully assigned to one of these two classes. In this paper, we apply a fuzzy membership to each

will be derived in Section III. Three experiments are presented in Section IV. Some concluding remarks are given in Section V.

Fuzzy Logic & SVM



The result of SVM learning for data sets with outliers



The result of FSVM learning for data sets with outliers.

Since the fuzzy membership is a function of the mean and radius of each class, these two outliers are regarded as less important in FSVM training

464 IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 13, NO. 2, MARCH 2002

Fuzzy Support Vector Machines

Chun-Fu Lin and Sheng-De Wang

Abstract—A support vector machine (SVM) learns the decision surface from two distinct classes of the input points. In many applications, each input point may not be fully assigned to one of these two classes. In this paper, we apply a fuzzy membership to each

will be derived in Section III. Three experiments are presented in Section IV. Some concluding remarks are given in Section V.

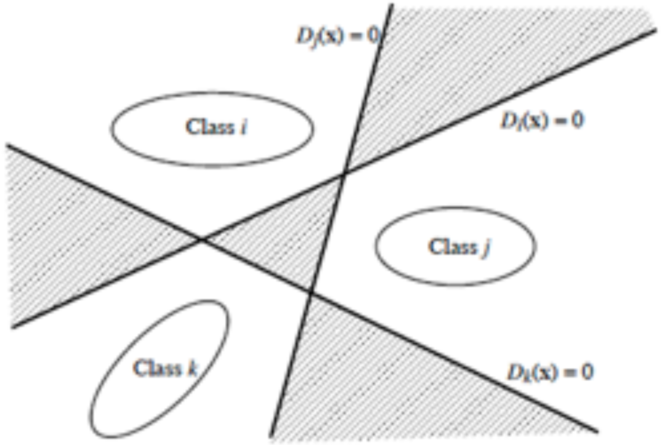
Fuzzy Logic & SVM

- In SVM, an n-class problem is converted into n two-class problems.
- For the i-th two-class problem ones determines the optimal decision function which separates class i from the remaining classes.
- Using the decision functions obtained by training the SVM, for each class, one can define a truncated polyhedral pyramidal membership function.
- Since, for the data in the classifiable regions, the classification results are the same for the two methods, the generalization ability of the FSVM is the same with or better than that of the SVM.

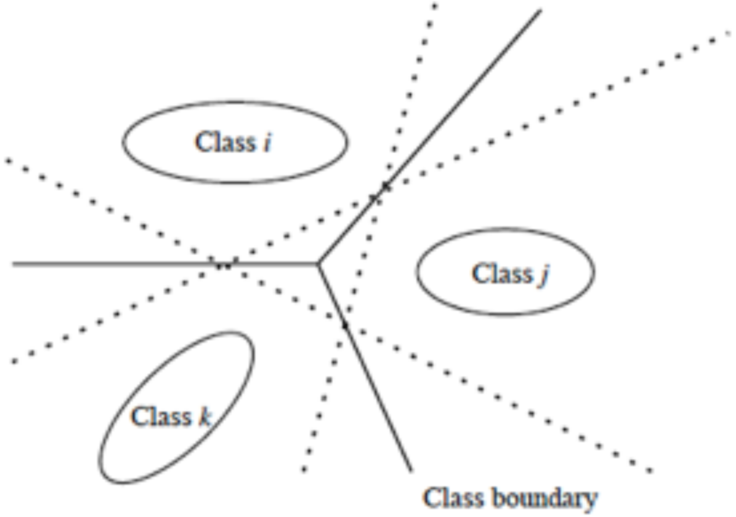
Fuzzy Support Vector Machines for Pattern Classification

Takuya Inoue and Shigeo Abe
Graduate School of Science and Technology, Kobe University, Kobe, Japan
E-mail: abe@eedept.kobe-u.ac.jp

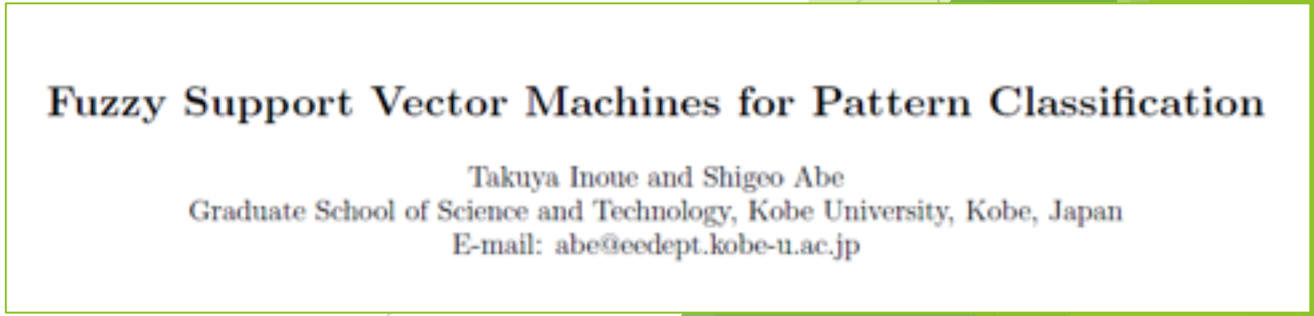
Fuzzy Logic & SVM



Unclassifiable region by the two-class formulation



Class boundary with membership functions



Fuzzy Support Vector Machines for Pattern Classification

Takuya Inoue and Shigeo Abe
Graduate School of Science and Technology, Kobe University, Kobe, Japan
E-mail: abe@eedept.kobe-u.ac.jp

Fuzzy Logic & SVM

- In the absence of additional information, fuzzy membership values are usually selected based on the distribution of training vectors
- A number of assumptions are made about the underlying shape of this distribution.
- An alternative method of generating membership values: generate membership values iteratively based on the positions of training vectors relative to the SVM decision surface itself.
- The algorithm is capable of generating results equivalent to an SVM with a modified (non distance based) penalty (risk) function.

Iterative Fuzzy Support Vector Machine Classification

Alistair Shilton and Daniel T. H. Lai

Abstract—Fuzzy support vector machine (FSVM) classifiers are a class of nonlinear binary classifiers which extend Vapnik's support vector machine (SVM) formulation. In the absence of additional information, fuzzy membership values are usually

In this paper we introduce a new algorithm for the calculation of the membership values using an iterative FSVM (I-FSVM) that makes no a-priori assumptions about the shape

Fuzzy Logic & SVM

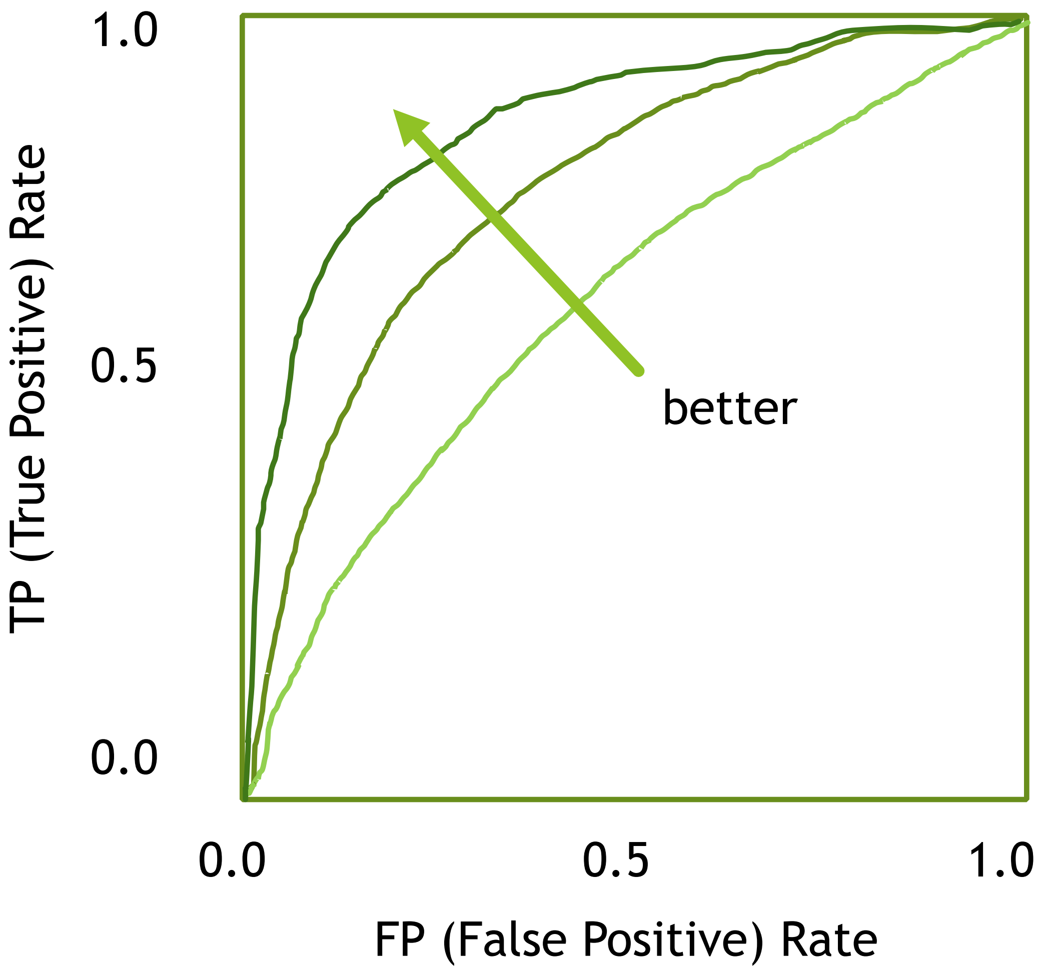
- The use of Receiver Operating Characteristic (ROC) Curve and the area under the ROC Curve (AUC) has been used as a measure of the performance of machine learning algorithms.
- A SVM classifier fusion model using genetic fuzzy system.
- Genetic algorithms are applied to tune the optimal fuzzy membership functions.
- The performance of SVM classifiers are evaluated by their AUCs.
- AUC-based genetic fuzzy SVM fusion model produces not only better AUC but also better accuracy than individual SVM classifiers.

Combining SVM Classifiers Using Genetic Fuzzy Systems Based on AUC for Gene Expression Data Analysis

Xiujuan Chen[1], Yichuan Zhao[2], Yan-Qing Zhang[1], and Robert Harrison[1]

[1] Department of Computer Science, Georgia State University, Atlanta, GA 30302, USA
[2] Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30302, USA
{xchen8@, matyiz@langate, zhang@taichi.cs, cscrwh@asterix.cs}gsu.edu

Fuzzy Logic & SVM



Fuzzy Logic & SVM

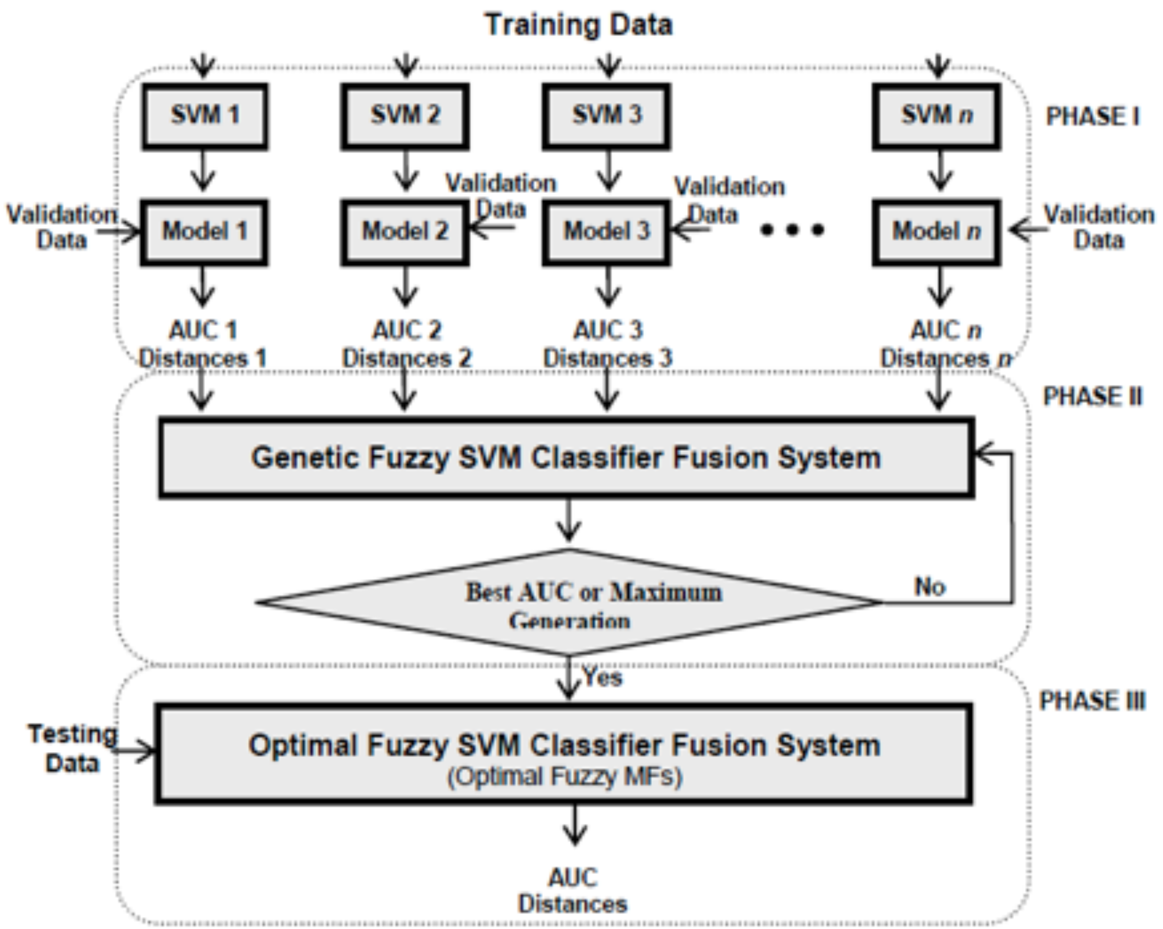
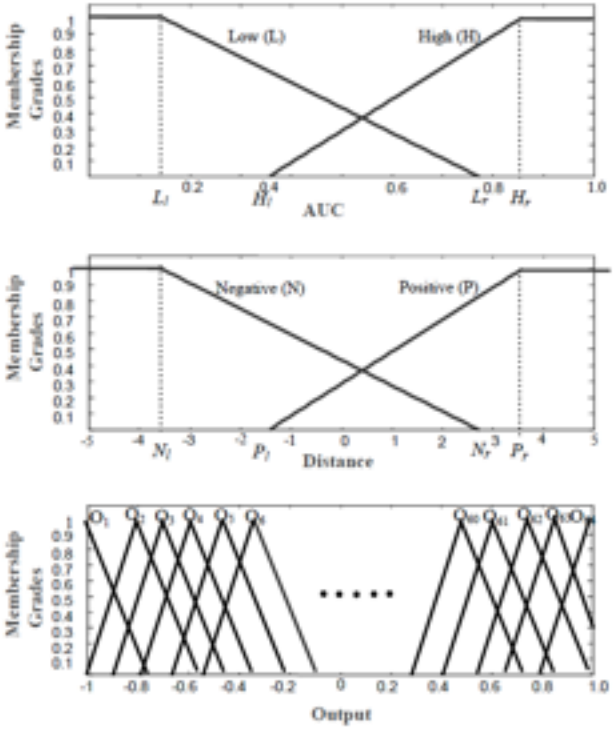


Fig. 2. Genetic fuzzy fusion system for SVM classifiers



Combining SVM Classifiers Using Genetic Fuzzy Systems Based on AUC for Gene Expression Data Analysis

Xiujuan Chen[1], Yichuan Zhao[2], Yan-Qing Zhang[1], and Robert Harrison[1]

[1] Department of Computer Science, Georgia State University, Atlanta, GA 30302, USA
[2] Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30302, USA
{xchen8@,matyiz@langate,zhang@taichi.cs, cscrwh@asterix.cs}gsu.edu

Fuzzy Logic & SVM

- How to solve the sensitivity of SVM to noise and outliers
- Characterizations of fuzzy support vector machine (FSVM) can be analyzed.
- But the determination of fuzzy membership is a difficulty.
- New fuzzy membership function is proposed.
- Each sample points is given the tightness arranged forecasts by this method

A New Fuzzy SVM based on the Posterior Probability Weighting Membership

Yan Wei*, Xiao Wu
College of Computer and Information Science, Chongqing Normal University, Chongqing, China
Email: weiyan@cqnu.edu.cn , wuxiao1985@sina.com

Abstract—To solve the sensitivity to the noises and outliers in support vector machine (SVM), the characterizations of fuzzy support vector machine (FSVM) are analyzed. But the

bayesian decision theory of inspiration, in [9], by using a posterior probability to express the uncertainty of samples. Gaowei Wu etc. established a posterior

Fuzzy Logic & SVM

- Fuzzy support vector machines (FSVMs) and the extraction of fuzzy rules from SVMs.
- An FSVM is identical to a special type of SVM.
- Categorization and analysis of existing approaches to obtain fuzzy rules from SVMs.
- Questioning the sense of extracting fuzzy rules from SVs:
  - Simpler methods that output prototypical points (e.g., clustering approaches) can used.

EUSFLAT-LFA 2011 July 2011 Aix-les-Bains, France

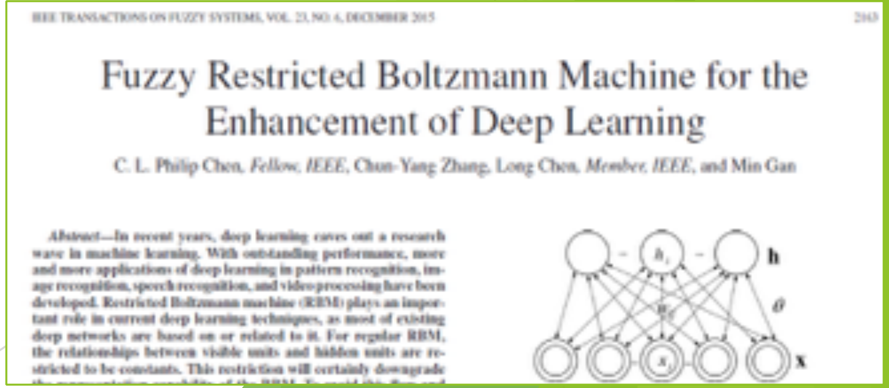
On the usefulness of fuzzy SVMs and the extraction of fuzzy rules from SVMs

Christian Moewes*, Rudolf Kruse†

Faculty of Computer Science, University of Magdeburg, Germany

Fuzzy Logic & Deep Learning

- Fuzzy Restricted Boltzmann Machine (FRBM): the parameters governing the model are replaced by fuzzy numbers.
- The original RBM becomes a special case in the FRBM, when there is no fuzziness in the FRBM model.
- In the process of learning FRBM, the fuzzy free energy function is defuzzified before the probability is defined.

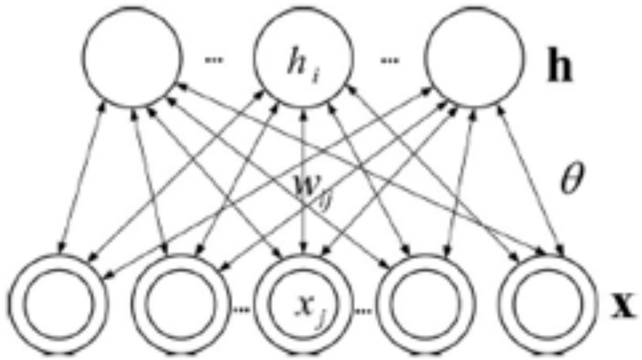

IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 23, NO. 6, DECEMBER 2015 2163

Fuzzy Restricted Boltzmann Machine for the Enhancement of Deep Learning

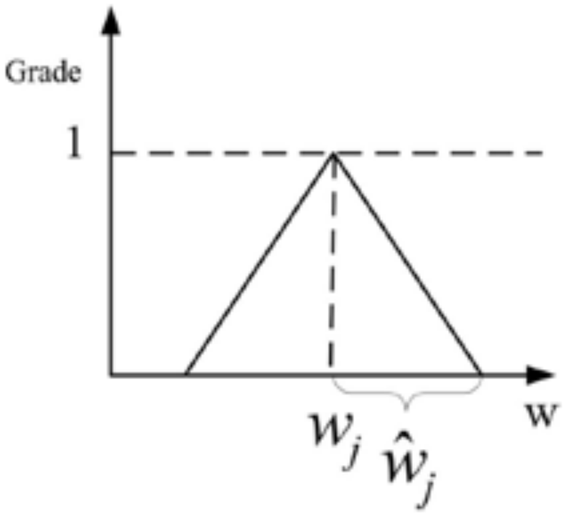
C. L. Philip Chen, *Fellow, IEEE*, Chun-Yang Zhang, Long Chen, *Member, IEEE*, and Min Gan

Abstract—In recent years, deep learning caves out a research wave in machine learning. With outstanding performance, more and more applications of deep learning in pattern recognition, image recognition, speech recognition, and video processing have been developed. Restricted Boltzmann machine (RBM) plays an important role in current deep learning techniques, as most of existing deep networks are based on or related to it. For regular RBM, the relationships between visible units and hidden units are restricted to be constants. This restriction will certainly downgrade

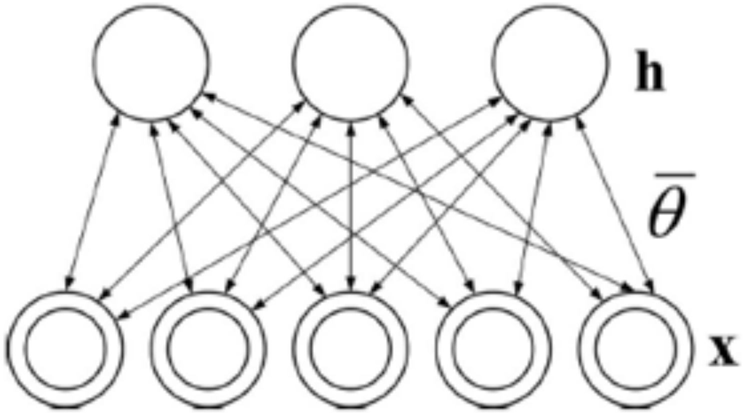
Fuzzy Logic & Deep Learning



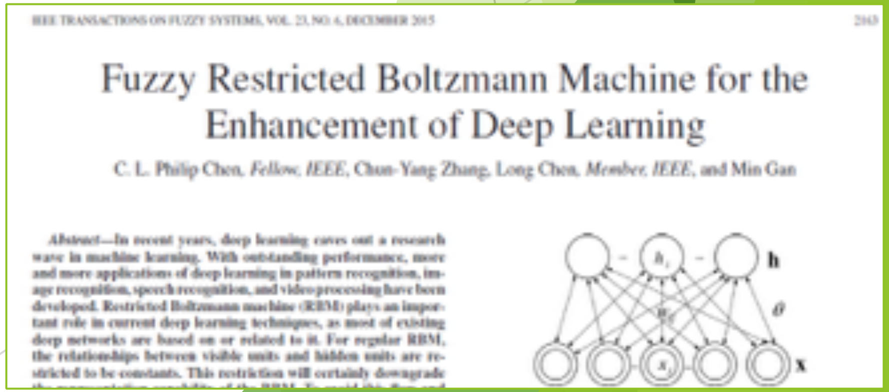
RBM



parameters as fuzzy numbers



FRBM



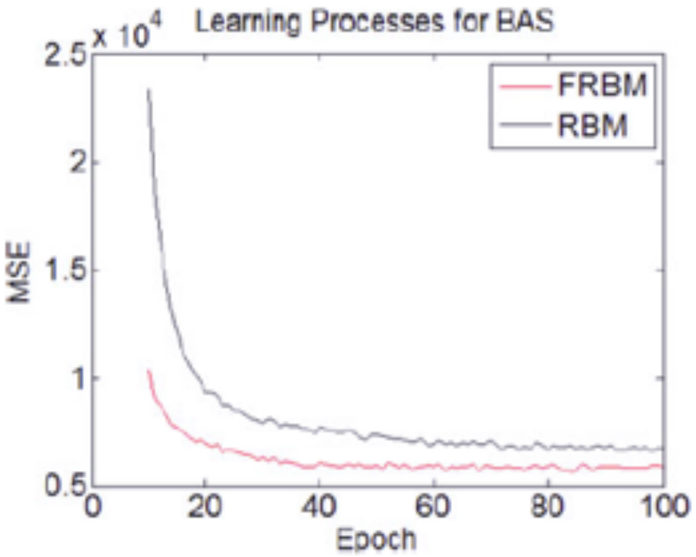
IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 23, NO. 6, DECEMBER 2015

Fuzzy Restricted Boltzmann Machine for the Enhancement of Deep Learning

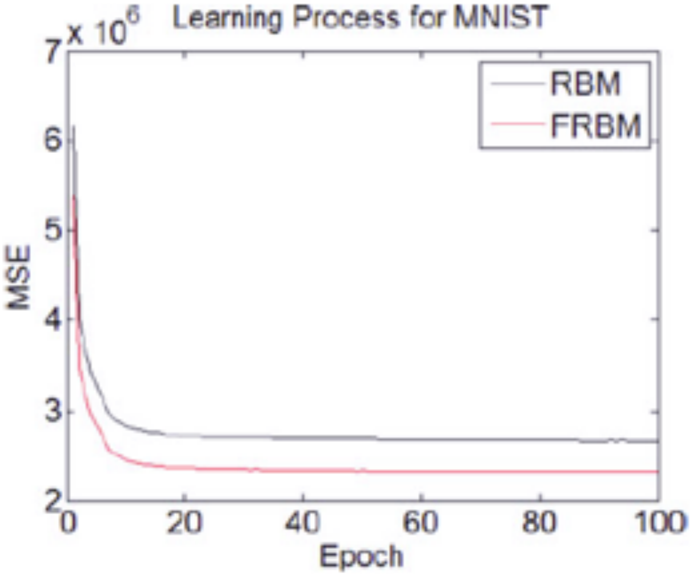
C. L. Philip Chen, *Fellow, IEEE*, Chun-Yang Zhang, Long Chen, *Member, IEEE*, and Min Gan

Abstract—In recent years, deep learning caves out a research wave in machine learning. With outstanding performance, more and more applications of deep learning in pattern recognition, image recognition, speech recognition, and video processing have been developed. Restricted Boltzmann machine (RBM) plays an important role in current deep learning techniques, as most of existing deep networks are based on or related to it. For regular RBM, the relationships between visible units and hidden units are restricted to be constants. This restriction will certainly downgrade

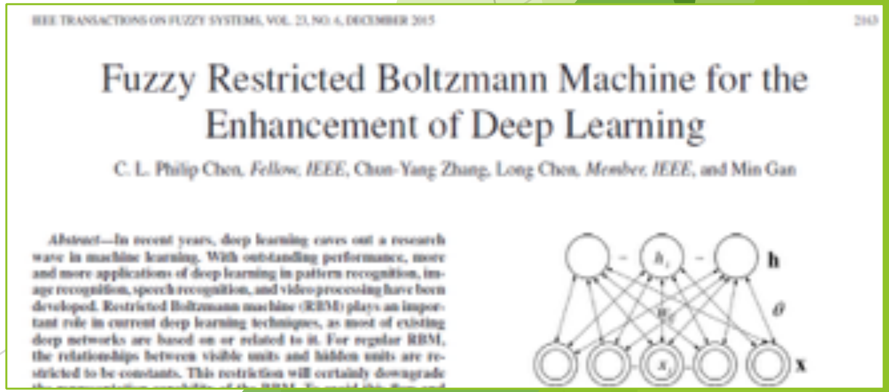
Fuzzy Logic & Deep Learning



Learning processes of RBM and FRBM based on BAS dataset.



Learning processes of RBM and FRBM based on MNIST dataset.



IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 23, NO. 6, DECEMBER 2015

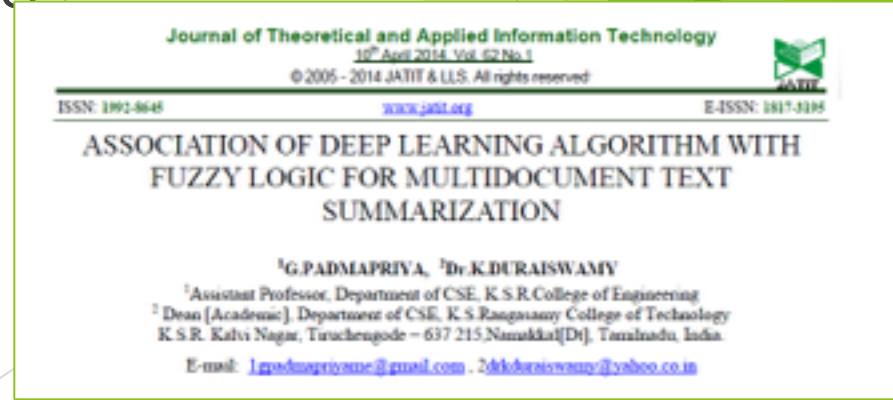
Fuzzy Restricted Boltzmann Machine for the Enhancement of Deep Learning

C. L. Philip Chen, *Fellow, IEEE*, Chun-Yang Zhang, Long Chen, *Member, IEEE*, and Min Gan

Abstract—In recent years, deep learning caves out a research wave in machine learning. With outstanding performance, more and more applications of deep learning in pattern recognition, image recognition, speech recognition, and video processing have been developed. Restricted Boltzmann machine (RBM) plays an important role in current deep learning techniques, as most of existing deep networks are based on or related to it. For regular RBM, the relationships between visible units and hidden units are restricted to be constants. This restriction will certainly downgrade

Fuzzy Logic & Deep Learning

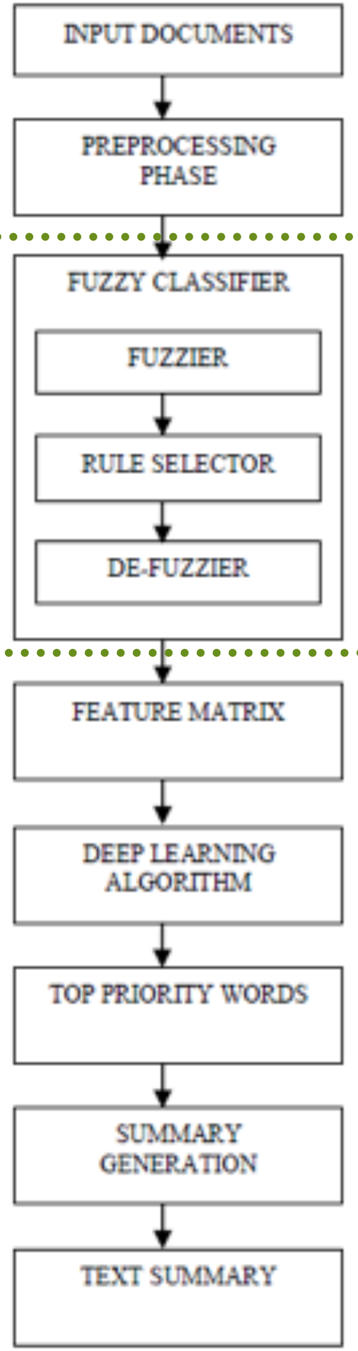
- Text summarization: Measuring the worth of sentences for a summary
- Associating the Deep learning algorithm with fuzzy logic
- The fuzzifier is a process of translating the inputs into feature values. Based on fuzzy values, rules are generated for each sentence by the weight given to the features.
- A rule can be defined for the proposed approach as, a set of features value is considered for judging the importance of sentences. The rules are composed based on the importance of the each sentence.


Journal of Theoretical and Applied Information Technology
10th April 2014. Vol. 62 No.1

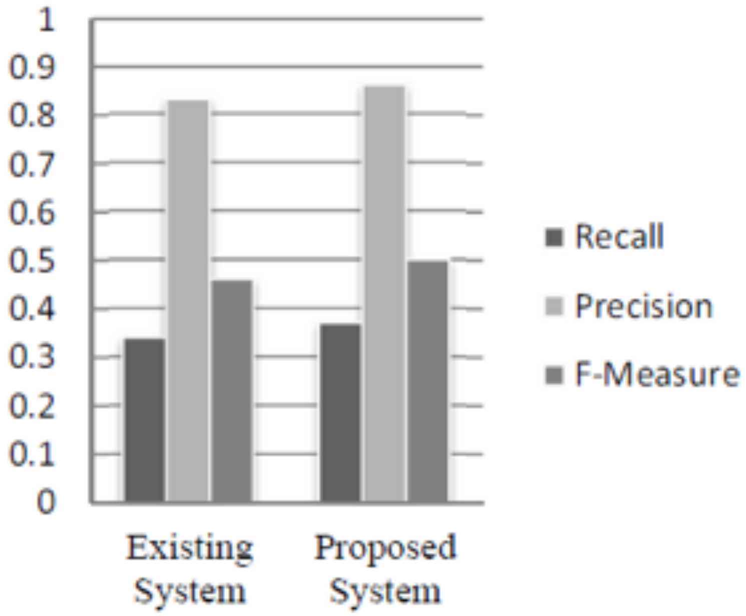
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

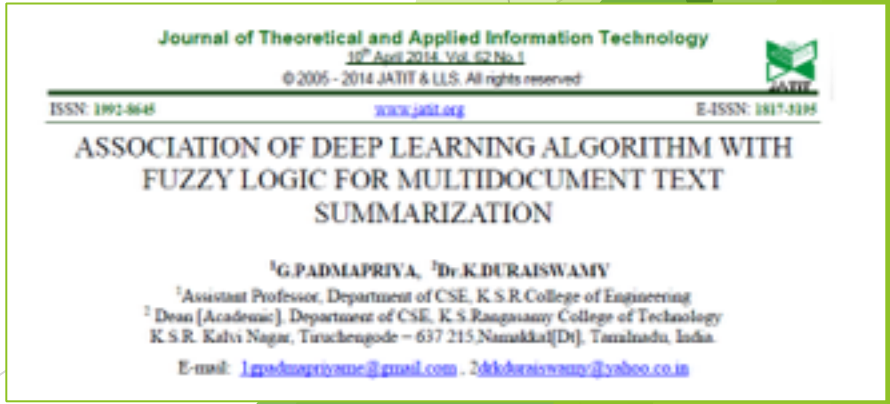
ASSOCIATION OF DEEP LEARNING ALGORITHM WITH FUZZY LOGIC FOR MULTIDOCUMENT TEXT SUMMARIZATION

[1]G.PADMAPRIYA, [2]Dr.K.DURAISWAMY
[1]Assistant Professor, Department of CSE, K.S.R.College of Engineering
[2] Dean [Academic], Department of CSE, K.S.Rangasamy College of Technology
K.S.R. Kalvi Nagar, Tiruchengode – 637 215,Namakkal[Dt], Tamilnadu, India
E-mail: 1gpadmapriyame@gmail.com , 2drkduraiswamy@yahoo.co.in



"The proposed text summarization algorithm uses the fuzzy logic system has to assign class labels for the sentences, in order to compute the importance of each sentence. The fuzzy logic system accepts the pre summarized set of documents as input"




Journal of Theoretical and Applied Information Technology
10th April 2014. Vol. 62 No.1
© 2005 - 2014 JATIT & LLS. All rights reserved

ISSN: 1992-8645 | www.jatit.org | E-ISSN: 1817-3195

ASSOCIATION OF DEEP LEARNING ALGORITHM WITH FUZZY LOGIC FOR MULTIDOCUMENT TEXT SUMMARIZATION

[1]G.PADMAPRIYA, [2]Dr.K.DURAISWAMY
[1]Assistant Professor, Department of CSE, K.S.R.College of Engineering
[2] Dean [Academic], Department of CSE, K.S.Rangasamy College of Technology
K.S.R. Kalvi Nagar, Tiruchengode – 637 215,Namakkal[Dt], Tamilnadu, India.
E-mail:

Fuzzy Logic & Deep Learning

- Embedding prior knowledge into the learning structure
- Two-step semi-supervised learning method called fuzzy deep belief networks (FDBN) for sentiment classification
- A general deep belief networks (DBN) is trained by the semi-supervised learning taken on training dataset.
- Then, a fuzzy membership function for each class of reviews is designed based on the learned deep architecture.
- Since the training of DBN maps each review into the DBN output space, the distribution of all training samples in the space is treated as prior knowledge and is encoded by a series of fuzzy membership functions.
- A new FDBN architecture is constructed and the supervised learning s applied to improve the classification performance of the FDBN.


Contents lists available at ScienceDirect

Neurocomputing

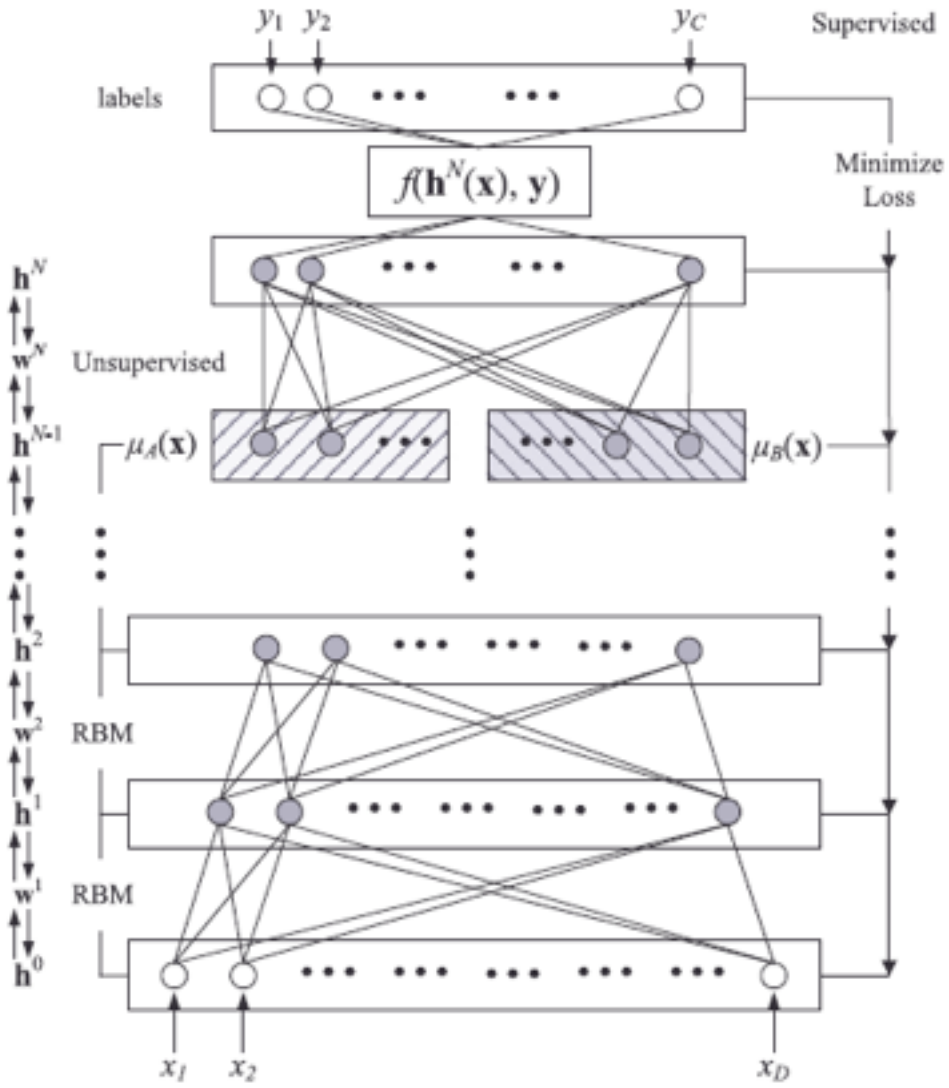
journal homepage: www.elsevier.com/locate/neucom

ELSEVIER

Fuzzy deep belief networks for semi-supervised sentiment classification

Shusen Zhou[a,*], Qingcai Chen[b], Xiaolong Wang[b]

Fuzzy Logic & Deep Learning



Training of FDBN is two stages:

1. The parameters of $\mu_A(\mathbf{x})$ and $\mu_B(\mathbf{x})$ are estimated based on the mapping result of trained DBN architecture.
2. The result of membership function $\mu_A(\mathbf{x})$ and $\mu_B(\mathbf{x})$ is used with $\mathbf{X}$ and $\mathbf{Y}^L$ at the same time to refine the FDBN.


Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

ELSEVIER

Fuzzy deep belief networks for semi-supervised sentiment classification

Shusen Zhou, Qingcai Chen, Xiaolong Wang

Fuzzy Logic & Deep Learning

- Neural networks increasingly adopted in the prediction of exchange rate
- However, most of them predict a specific number
- Small gap between the predicted values and the actual values may lead to disastrous consequences.
- Forecast the fluctuation range of the exchange rate by combining Fuzzy Granulation with Continuous-valued Deep Belief Networks (CDBN)


2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

A Model with Fuzzy Granulation and Deep Belief Networks for Exchange Rate Forecasting

Ren Zhang, Furao Shen*, and Jinxi Zhao

Fuzzy Logic & Deep Learning

- Detecting abnormal behaviors in surveillance videos
- Using fuzzy clustering and multiple Auto-Encoders (FMAE).
- Many types of normal behaviors in the daily life: fuzzy clustering
- Multiple Auto-Encoders to estimate different types of normal behaviors
- Auto-Encoder is a good tool to capture common structures of normal video due to large redundancies

DETECTING ABNORMAL BEHAVIORS IN SURVEILLANCE VIDEOS BASED ON FUZZY CLUSTERING AND MULTIPLE AUTO-ENCODERS

Zhengying Chen, Yonghong Tian, Wei Zeng, Tiejun Huang*

School of Electronics Engineering and Computer Science, Peking University, Beijing, P.R.China 100871
{cathychenchn, yhtian, weizeng, tjhuang}@pku.edu.cn

A Critical Review

A Critical Review

- Machine learning has been changing rapidly
- “Hot topics” are changing quickly
- ML researchers are working on deep learning, manifold learning, structured output pre-diction, sparsity and compressed sensing, constructive induction, etc.,
- Most fuzzy researchers are still occupied with rule induction (“a topic that matured and essentially stopped in ML research in the 1990s”).

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 281 (2015) 292–299

FUZZY sets and systems

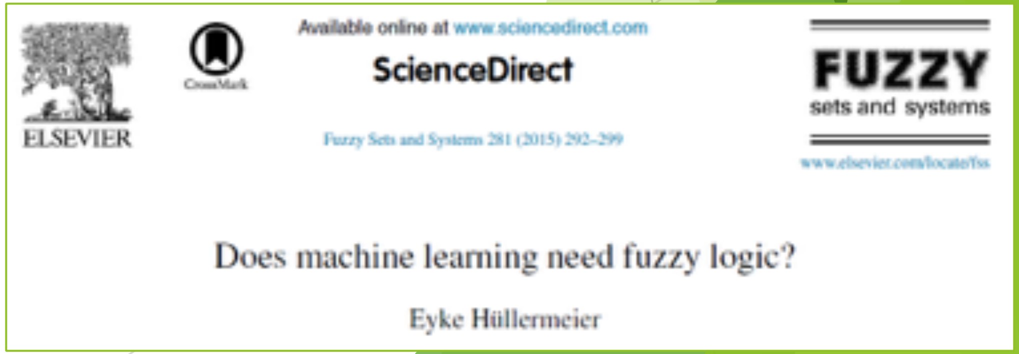
www.elsevier.com/locate/fss

Does machine learning need fuzzy logic?

Eyke Hüllermeier

A Critical Review

- Most works extend ML methods by delivery a fuzzy extension, e.g.,
  - from rule induction to fuzzy rule induction,
  - from decision trees to fuzzy decisions trees,
  - from nearest neighbor estimation to fuzzy nearest neighbor estimation,
  - from support vector machines to fuzzy support vector machines

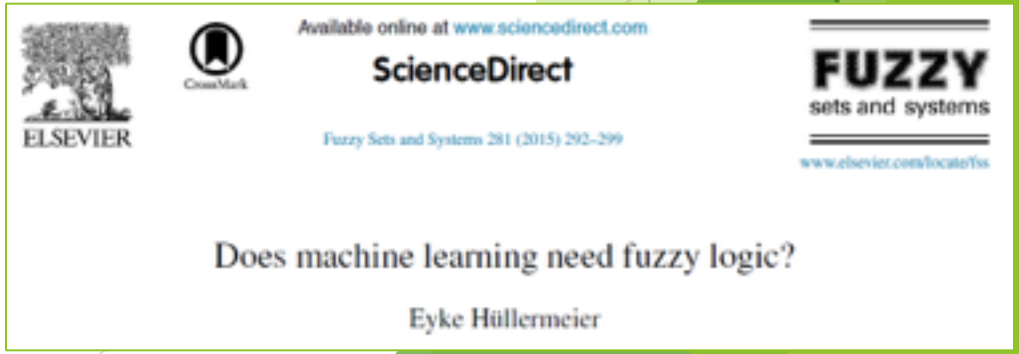

ELSEVIER
Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems 281 (2015) 292–299
FUZZY sets and systems
www.elsevier.com/locate/fss

Does machine learning need fuzzy logic?

Eyke Hüllermeier

A Critical Review

- "Fuzzification" of ML methods can be questioned:
- The intellectual challenge is typically not very high (scientific contribution most likely not very deep)
- Increased flexibility through fuzzification could also be achieved by passing to a more flexible non-fuzzy model class (e.g., using SVMs with Gaussian instead of linear kernels)
- More flexibility may be a disadvantage (i.e., risk of overfitting)
- Increased computational complexity
- In some cases, the link to fuzzy sets and fuzzy logic appears to be somewhat artificial (member-ship functions are used as a weighting function)


ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 281 (2015) 292–299

FUZZY sets and systems

www.elsevier.com/locate/fss

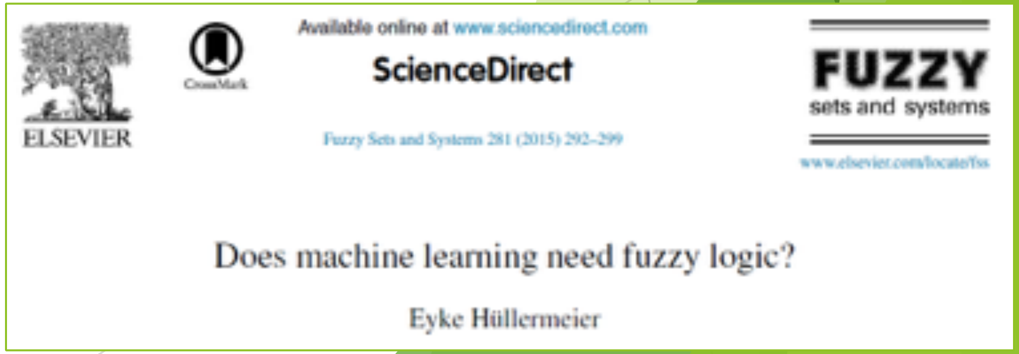
Does machine learning need fuzzy logic?

Eyke Hüllermeier

…however, potentials are There

Modeling:

- We need a suitable formalization of the problem
- Often overlooked in machine learning
- Fuzzy logic has much to offer in this regard


ELSEVIER

CrossMark

Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 281 (2015) 292–299

FUZZY sets and systems

www.elsevier.com/locate/fss

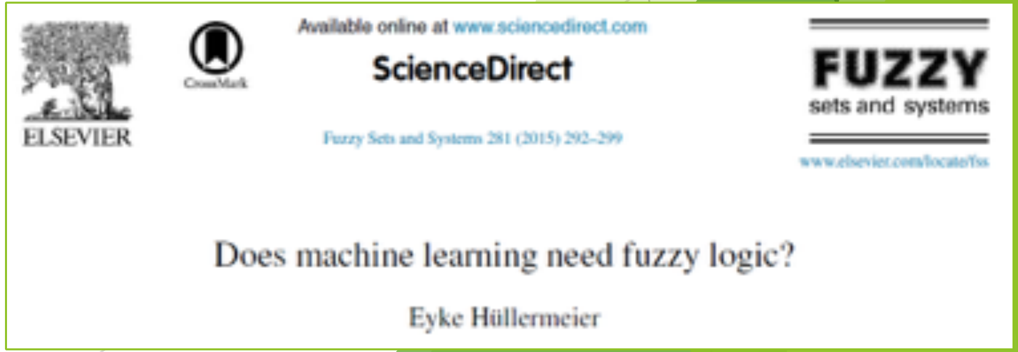
Does machine learning need fuzzy logic?

Eyke Hüllermeier

…however, potentials are There

Non-Inductive Inference:

- Transfer learning: Taking advantage of what has been learned in one domain while learning in another domain
- *Knowledge transfer:* largely *similarity-based* or *analogical*
- Fuzzy inference can support that kind of formal reasoning



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

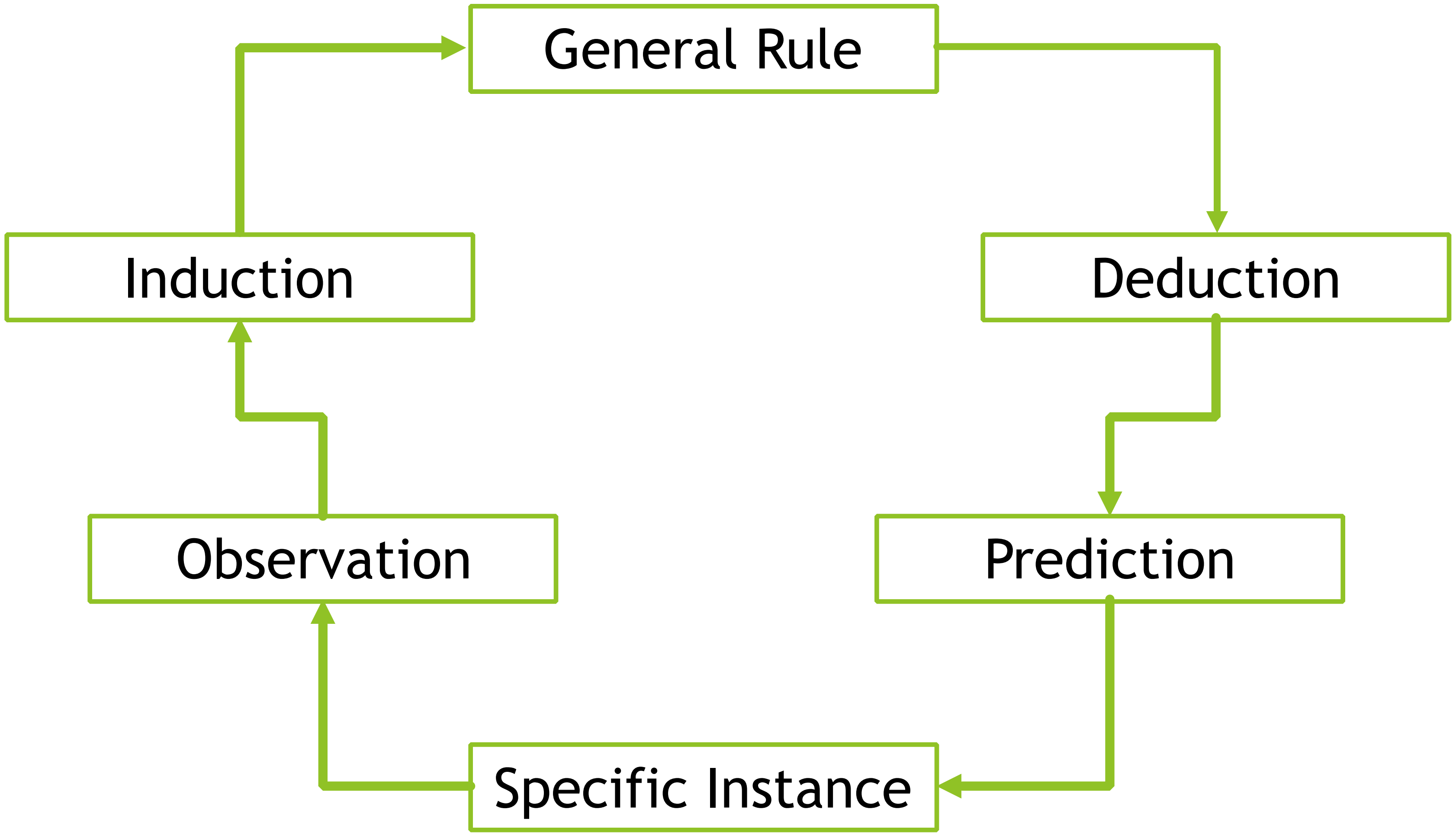
Fuzzy Sets and Systems 281 (2015) 292–299

FUZZY sets and systems

www.elsevier.com/locate/fss

Does machine learning need fuzzy logic?

Eyke Hüllermeier


General Rule
Induction
Deduction
Observation
Prediction
Specific Instance

Induction

- Going from specific to general
- Reasoning from evidence (observations/data) to draw a conclusion (establish a hypothesis)
- Always less certain than evidence
- Used because examining all observations may be impossible/infeasible
- Example: Neural Networks, all types of classifiers

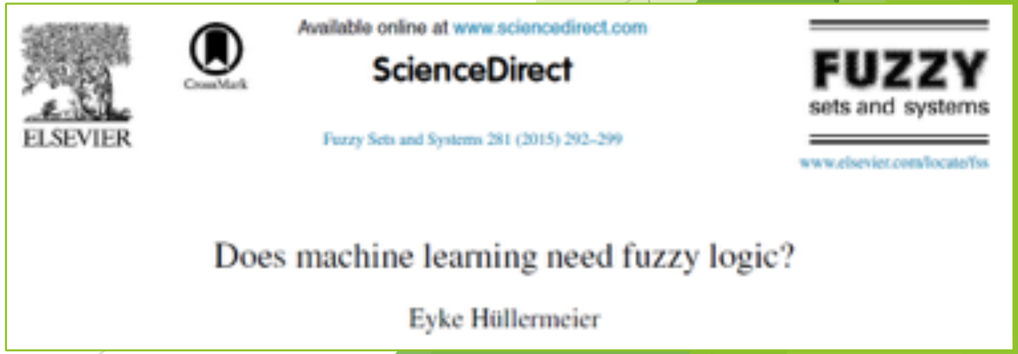
Deduction

- Going from general to specific
- Draw a conclusion follows logically
- Conclusion true when premise true
- Used to exploit linguistically formulated knowledge
- Example: Fuzzy rules established by experts

…however, potentials are There

Uncertainty:

- Uncertainty is everywhere
- Fuzzy framework can contribute to representation and handling of uncertainty
- *Possibility theory* for uncertainty formalisms, such as belief functions, and imprecise probabilities, can be used for this purpose


ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 281 (2015) 292–299

FUZZY sets and systems

www.elsevier.com/locate/fss

Does machine learning need fuzzy logic?

Eyke Hüllermeier

Possibility Theory

Question: John is **about 18**, is he allowed to vote?

Answer: It is *quite possible*, but *not certain*

The *uncertainty* of an event is described in possibility theory at the same time :

- by the degree of possibility of this event and
- by the degree of possibility of the opposite event

Possibility Theory

Ω	Reference Set
$A \subseteq \Omega$	Event
$g(A)$	Real number measuring the confidence in the occurrence of A

$$g(\varnothing) = 0, \quad g(\Omega) = 1$$

Impossible event   Sure event

Possibility Theory

Monotonicity with respect to inclusion:

$$A \subseteq B \Rightarrow g(A) \leq g(B)$$

Such *set functions* are called:

Fuzzy measures, valuations or confidence measures

Possibility Theory

Consider the confidence measure concerning disjunctions:

$$\forall A, B \subseteq \Omega, \quad g(A \cup B) \geq \max(g(A), g(B))$$

The **possibility measure**:

$$\forall A, B \subseteq \Omega, \quad \Pi(A \cup B) = \max(\Pi(A), \Pi(B))$$

Finite reference set:

$$\forall A, \quad \Pi(A) = \sup\{\pi(\omega) | \omega \in A\}$$

$$\pi(\omega) = \Pi(\{\omega\})$$

Possibility Theory

Consider the confidence measure concerning conjunctions:

$$\forall A, B \subseteq \Omega, \quad g(A \cap B) \leq \min(g(A), g(B))$$

The **necessity measure**:

$$\forall A, B \subseteq \Omega, \quad N(A \cup B) = \min(N(A), N(B))$$

Finite reference set:

$$\forall A, \quad N(A) = \inf\{1 - \pi(\omega) | \omega \notin A\}$$

$$\pi(\omega) = \Pi(\{\omega\})$$

Possibility Theory

$$\forall A, \quad \Pi(A) = 1 - N(\overline{A})$$

Contradictory events:

$$\max\left(\Pi(A), \Pi(\overline{A})\right) = 1$$

$$\min\left(N(A), N(\overline{A})\right) = 0$$

Possibility Theory

Probability

$$P(A \cup B) = P(A) + P(B)$$

$$A \cap B = \varnothing$$

Possibility

$$\Pi(A \cup B) = \max(\Pi(A) + \Pi(B))$$

Necessity

$$N(A \cap B) = \min(N(A) + N(B))$$

Possibility Theory

$$\forall A, \quad P(A) = \sum_{\omega \in A} p(\omega)$$

$$\forall A, \quad \Pi(A) = \sup\{\pi(\omega) \mid \omega \in A\}$$

$$N(A) = \inf\{1 - \pi(\omega) \mid \omega \notin A\}$$

$$\forall \omega, \quad \sum_{\omega \in \Omega} p(\omega) = 1$$

$$\exists \omega, \quad \pi(\omega) = 1$$

Possibility Theory

$$P(A) + P(\bar{A}) = 1$$

$$\max\left(\Pi(A), \Pi(\bar{A})\right) = 1$$

$$\Pi(A) + \Pi(\bar{A}) \geq 1$$

$$N(A) + N(\bar{A}) \leq 1$$

$$\forall A, \quad N(A) \leq P(A) \leq \Pi(A)$$

Possibility Theory

Total Ignorance

$$\forall \omega \in \Omega, \quad p(\omega) = \frac{1}{|\Omega|}$$

100 events: p=0.01

$$\forall \omega \in \Omega, \quad \pi(\omega) = 1$$

100 events: π=1