Fuzzy Formal Concept Analysis

Marcos Evandro Cintra – *mecintra@gmail.com*
São Paulo University – Brazil
University of Bristol – UK

Outline

- FCA
  - Definitions
  - History
  - Software available
  - Basic concepts
  - Scaling
- Fuzzy FCA
- Related work

Formal Concept Analysis (FCA)

- A method for:
  - Data analysis
  - Knowledge representation
  - Information management
- Non supervised machine learning method
- Clustering based

History

- Proposed by Rudolf Wille in the early 80s
- First 10 years restricted to a small group of researchers and Wille's students in Germany
- Because of the mathematical nature of most of the publications of that time, knowledge of FCA remained restricted to a group of "insiders"

History

- Through funded research projects, FCA was implemented in several larger-scale applications:
- An implementation of a knowledge exploration system for civil engineering in cooperation with the Ministry for Civil Engineering of North-Rhine Westfalia

Last 10 years

- Applications in many disciplines, such as:
  - Linguistics
  - Software engineering
  - Psychology
  - Information retrieval
- Reasons: Influential papers in several fields

FCA – introductions, bibliographies and software

- English-language introductory material has been scarce
- Some material in German
- Almost no material in Portuguese ☹

FCA introductions, bibliographies and software

- Main conferences:
  - Int. Conf. on Conceptual Structures (ICCS), since 1995
  - Int. Conference on FCA (ICFCA), since2003
- Software available (probably another contributing factor for the recent growth of interest in FCA):
  - ConExp
  - ToscanaJ

Software availability – Issues

- Due to the complexity of the underlying lattice data structures and of the visualizations, FCA software can usually NOT be developed in the kinds of short-term projects which are normally funded by national research agencies!

Software companies

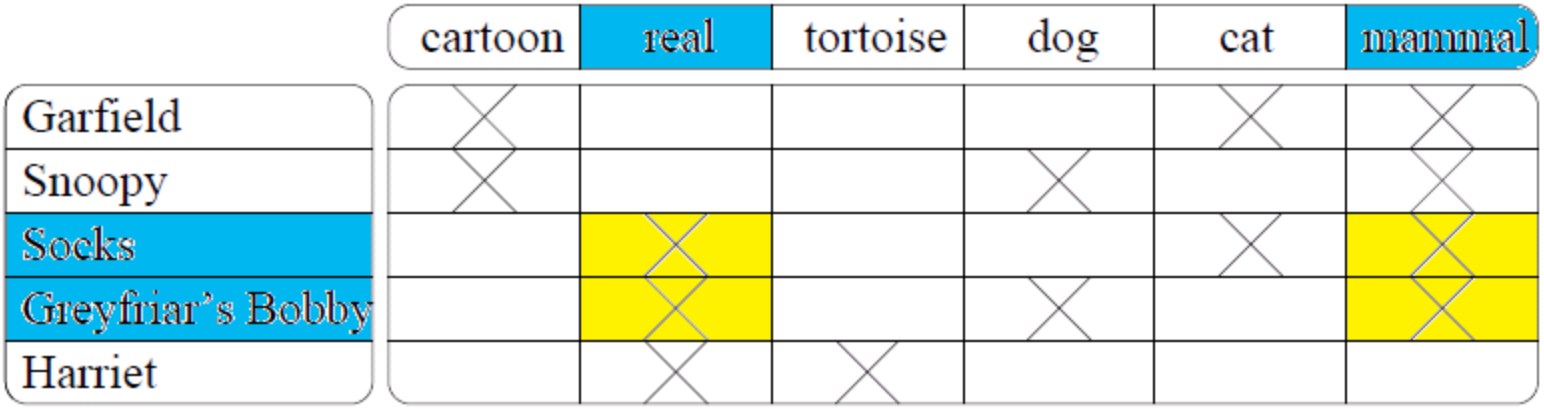
- German Navicon (www.navicon.de):
  - Founded 13 years ago by some of Wille's former students
  - Employ FCA software for information management tasks
  - The company is doing well but is nowadays more focused on database technologies than on FCA
- Australian company (www.mailsleuth.com) that is marketing an email analysis tool based on FCA

FCA – Basic Concepts

Toy Example

	cartoon	real	tortoise	dog	cat	mammal
Garfield	X				X	X
Snoopy	X			X		X
Socks		X			X	X
Greyfriar's Bobby		X		X		X
Harriet		X	X			

Definitions



	cartoon	real	tortoise	dog	cat	mammal
Garfield	X				X	X
Snoopy	X			X		X
Socks		X			X	X
Greyfriar's Bobby		X		X		X
Harriet		X	X			

- **Formal concept**: a pair of objects and attributes that is closed
- Example:
  - Set of all objects that are real and mammal: {Socks, Bobby}

Important

	cartoon	real	tortoise	dog	cat	mammal
Garfield	X				X	X
Snoopy	X			X		X
Socks		X			X	X
Greyfriar's Bobby		X		X		X
Harriet		X	X			

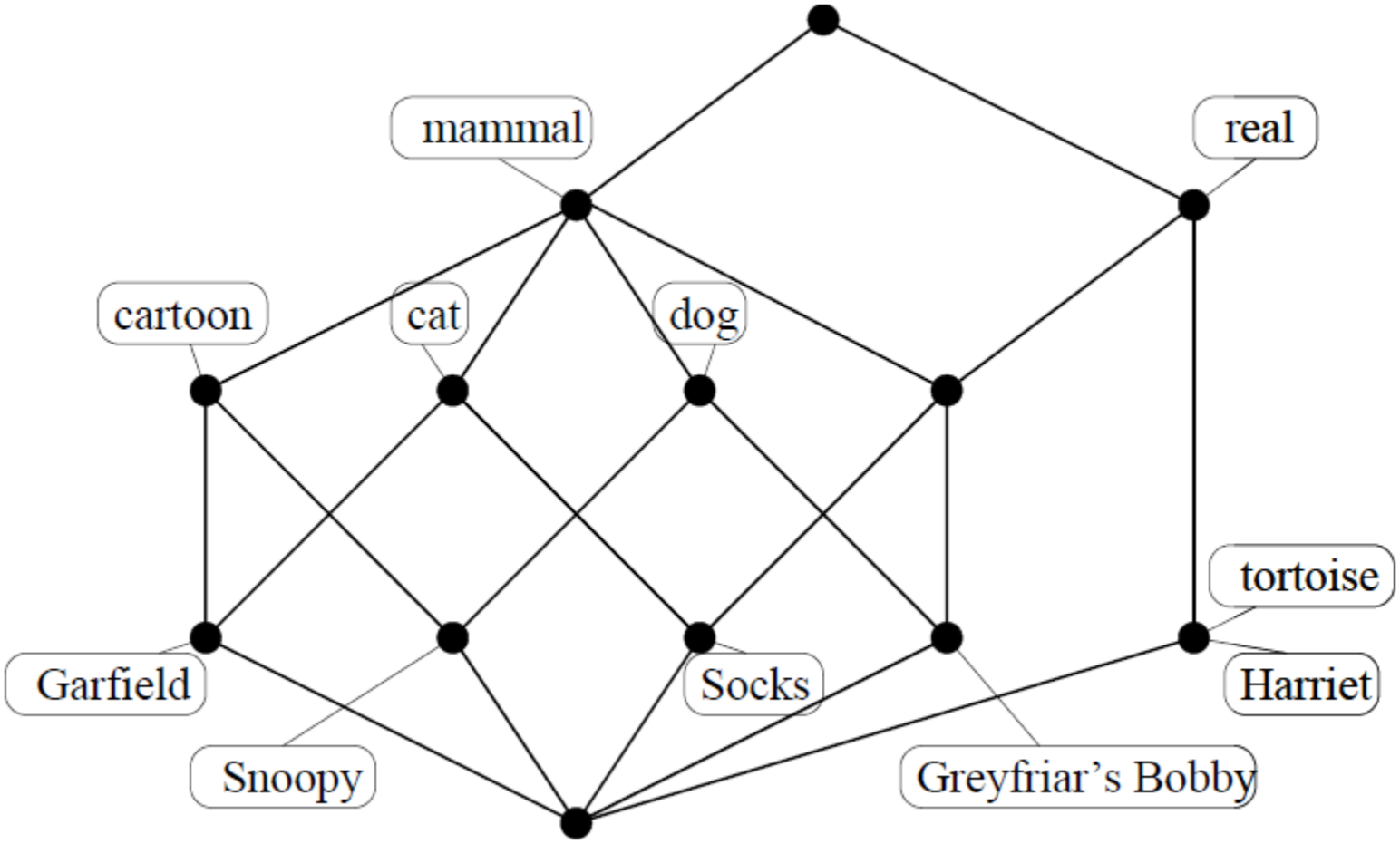
1. ***Novice users often need to be reminded that although many formal concepts in an application may correspond to intuitive notions they might have, not all formal concepts need to do so.***
2. The existence of a class is not necessary for a formal context
3. Attributes have only two options

Definitions

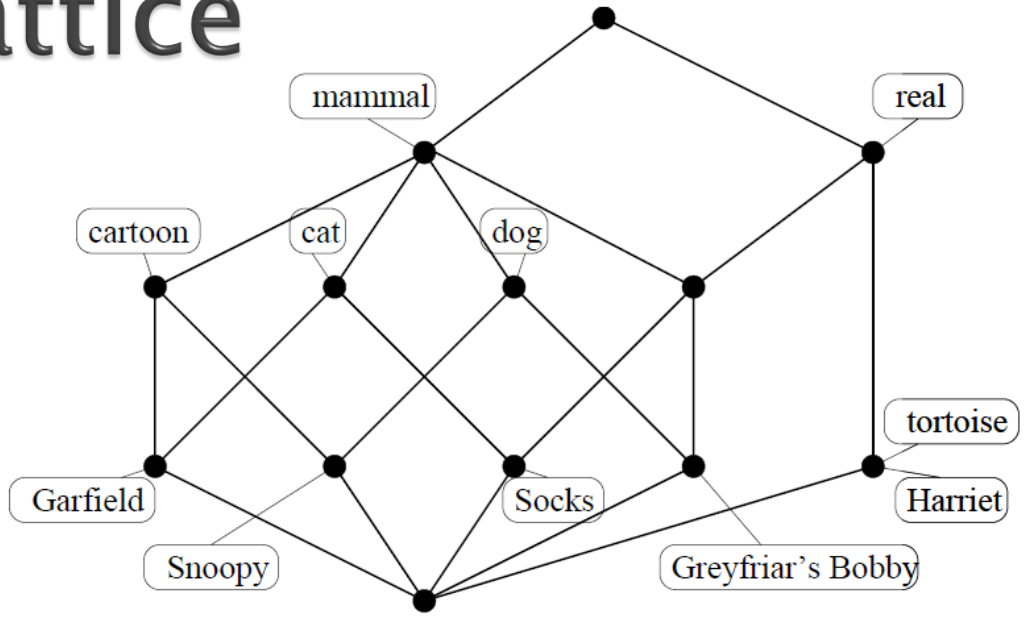
	cartoon	real	tortoise	dog	cat	mammal
Garfield	X				X	X
Snoopy	X			X		X
Socks		X			X	X
Greyfriar's Bobby		X		X		X
Harriet		X	X			

- **Extension**: set of objects of a formal concept
- **Intension**: set of attributes of a formal concept

Concept Lattice

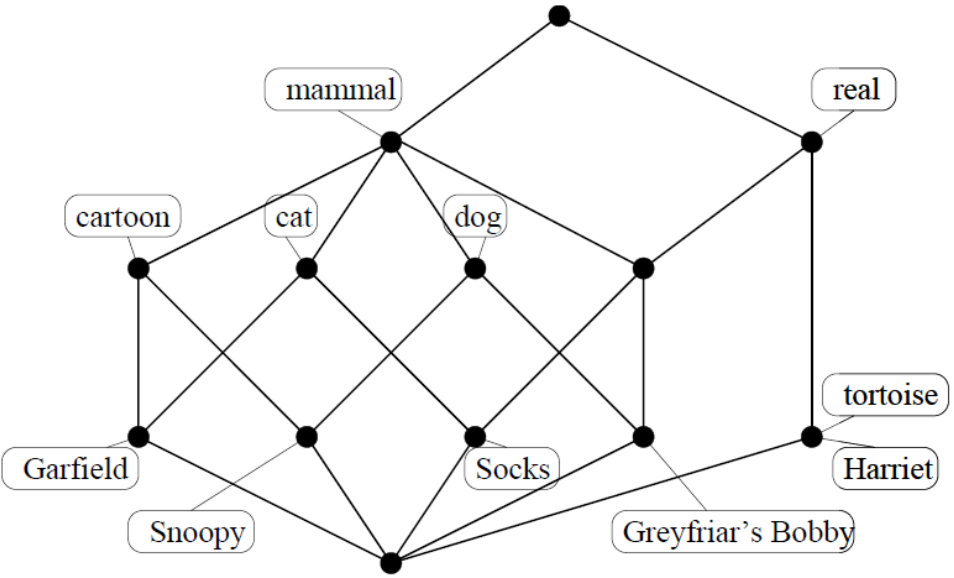


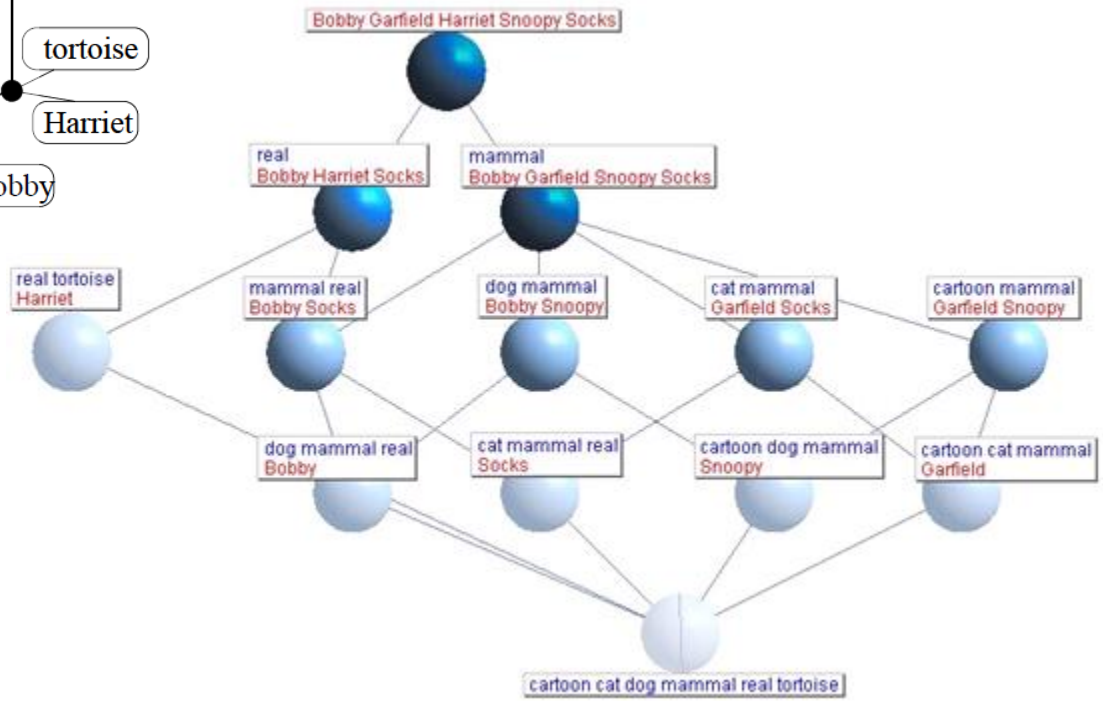
A concept Lattice



- **Formal concepts** are represented by the nodes
- Attributes are noted slightly above nodes
- Objects are noted slightly under nodes
- Retrieving **Extensions**: trace all paths leading down from the node
- Retrieving **Intension**: trace all paths leading up from the node

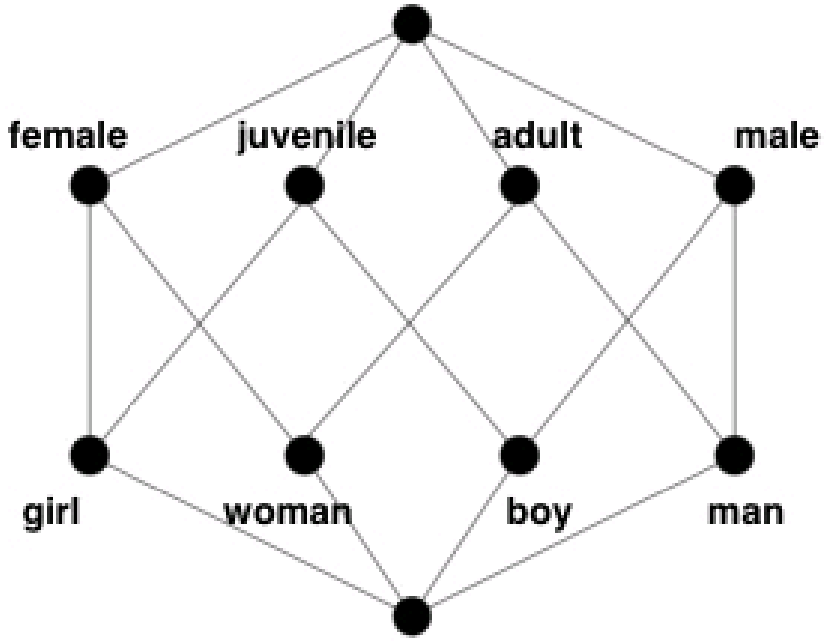
A concept Lattice





Another toy example

	female	juvenile	adult	male
girl	x	x		
woman	x		x	
boy		x		x
man			x	x



Dimensionality Issues

- Large sets of attributes and objects produce messy lattices
- A messy visualization only tells us that the lattice is complex
- Concept lattices should be human-readable or the information should be displayed in another format

Alternatives

- Make concept lattice visualizations expandable (such as files/folders)
- Dividing them into different components based on groupings of related attributes

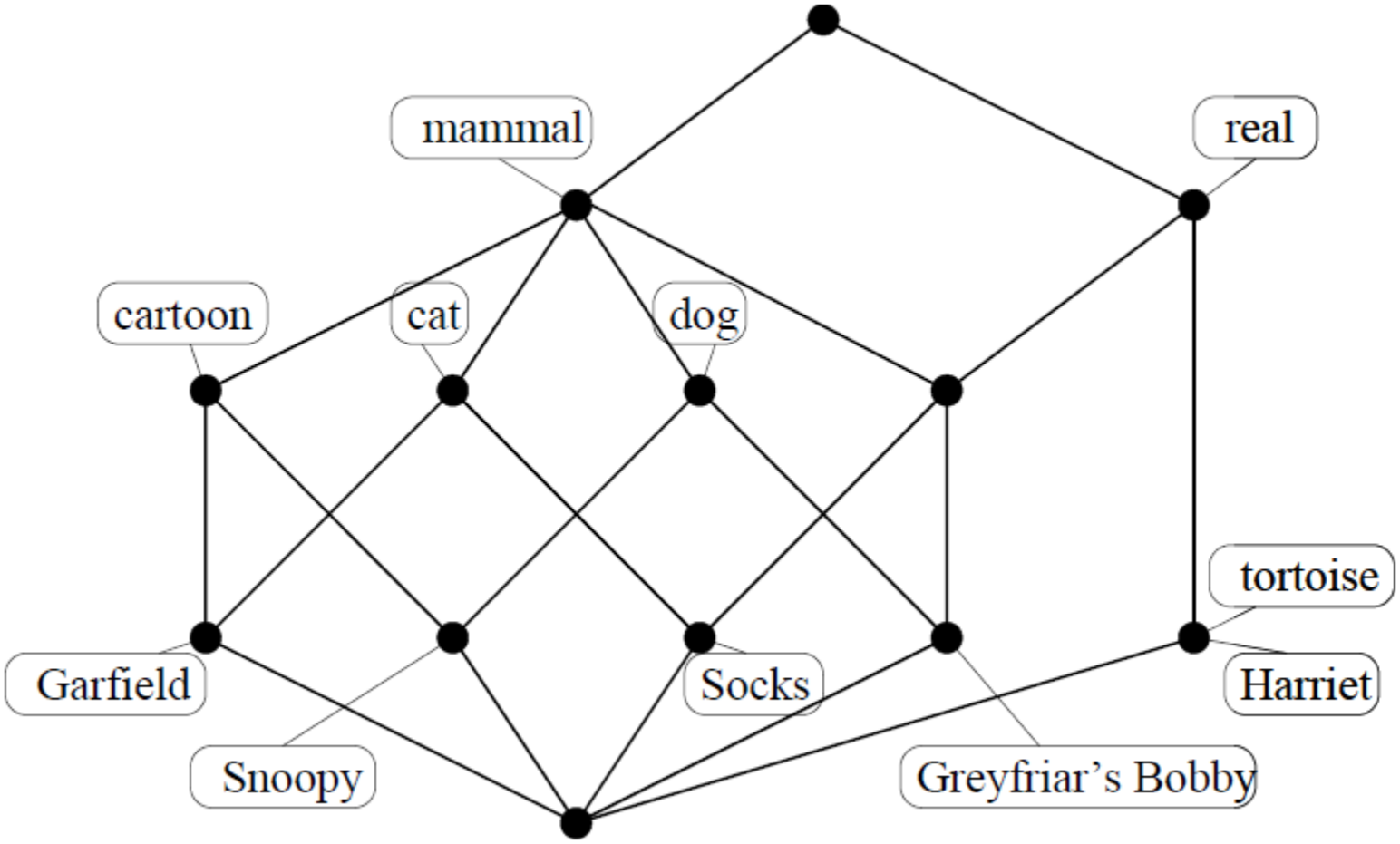
Definition: Conceptual Scales

- Consider a survey with a structure of data representing rank orders (“strongly agree”, “agree”, “neutral”, “disagree”, “strongly disagree”):
  - The lattice can be drawn without considering the actual results from the survey

Implications

- All implications are correct for the given formal context;
- Some might be incorrect in real life:
  - Dog and Cat implies Mammal
  - Cartoon implies Mammal ?!
- Support for attribute exploration (Burmeister, 1999a):
  - Questions are made to check all implications
  - Counter examples are requested to update the formal context

A concept Lattice



Applications

- Visualization
- Decision trees
- Rule extraction for:
  - Classification
  - Association
- Generation of ontologies
- Attribute selection

SCALING

- It is the transformation of data into contexts
- It is also an information granulation task

K_0	sex	age
ADAM	m	21
BETTY	f	50
CHRIS	/	66
DORA	f	88
EVA	f	17
FRED	m	/
GEORGE	m	90
HARRY	m	50

SCALING

	sex		age				
K	m	f	<18	<40	≤65	>65	≥80
ADAM	×			×	×		
BETTY		×			×		
CHRIS						×	
DORA		×				×	×
EVA		×	×	×	×		
FRED	×						
GEORGE	×					×	×
HARRY	×				×		

Scaling

	age	height	symptom
Alice	23	165	1
Boris	30	180	0
Cyril	31	167	1
David	43	159	0
Ellen	24	155	1
Fred	64	170	0
George	30	190	0

Scaling

- Age: young[0,30], medium[31,50], old[51,∞]
- Height: short[0,160], medium[161,180],tall[181,∞]

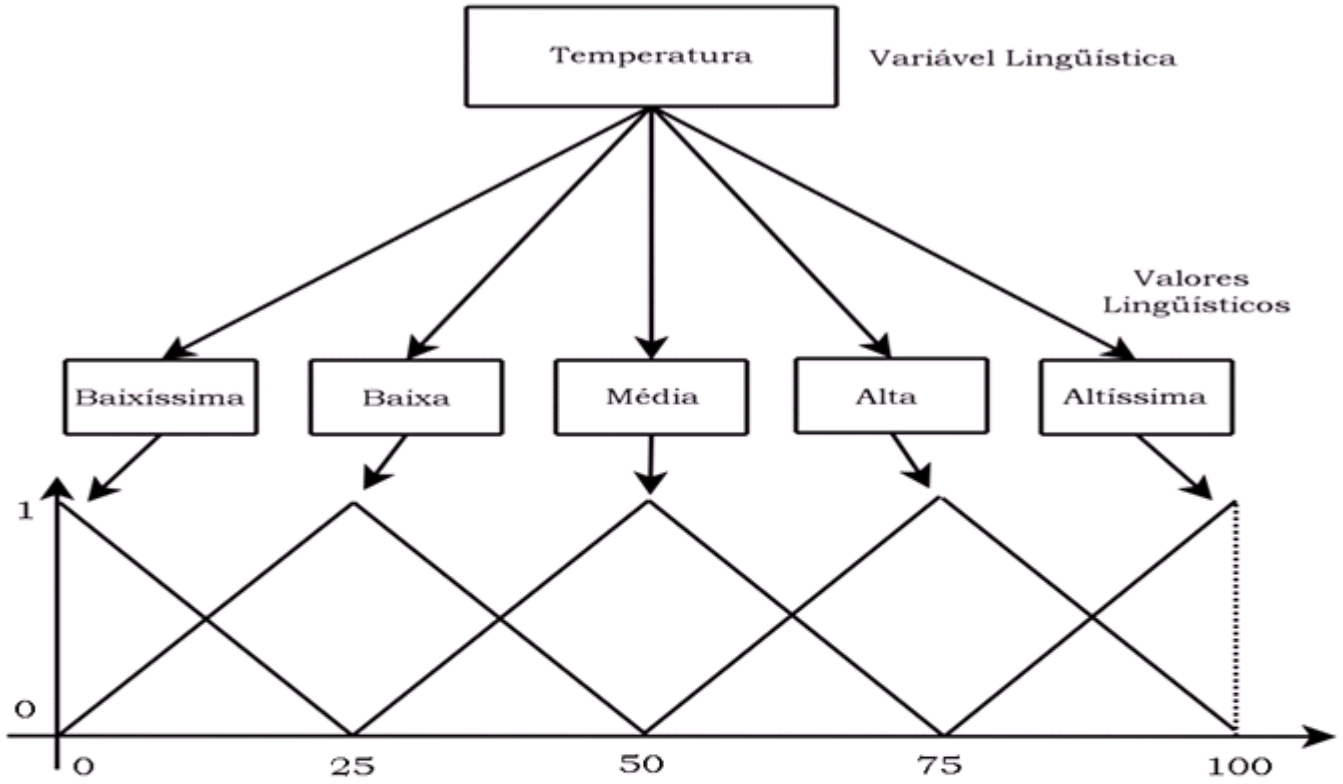
	a_y	a_m	a_o	h_s	h_m	h_t	symptom
Alice	1	0	0	0	1	0	1
Boris	1	0	0	0	1	0	0
Cyril	0	1	0	0	1	0	1
David	0	1	0	1	0	0	0
Ellen	1	0	0	1	0	0	1
Fred	0	0	1	0	1	0	0
George	1	0	0	0	0	1	0

Fuzzy – Advantages

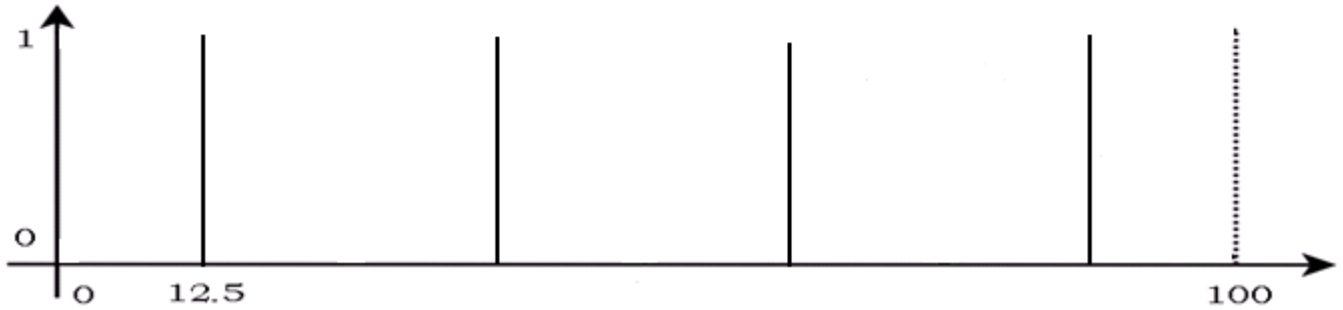
- Avoids the creation of unnatural separations when defining variables in terms of fuzzy sets (instead of the traditional alternatives available)
- Most people think in a fuzzy way, rather than crisply

Fuzzy definition of variables

Fuzzy



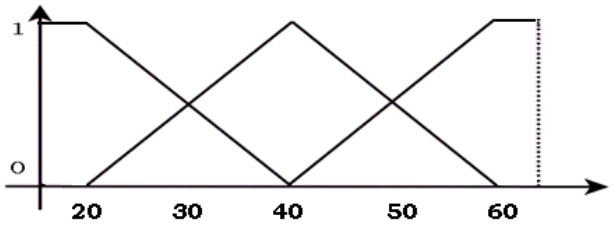
Crisp

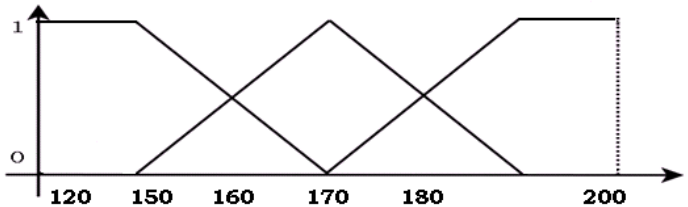


Fuzzy Scalling

- Age: young[0,30], medium[31,50], old[51,∞]
- Height: short[0,160], medium[161,180],tall[181,∞]

	age	height	symptom
Alice	23	165	1
Boris	30	180	0
Cyril	31	167	1
David	43	159	0
Ellen	24	155	1
Fred	64	170	0
George	30	190	0





	a_y	a_m	a_o	h_s	h_m	h_t	symptom
Alice	1	0	0	0	1	0	1
Boris	1	0	0	0	1	0	0
Cyril	0	1	0	0	1	0	1
David	0	1	0	1	0	0	0
Ellen	1	0	0	1	0	0	1
Fred	0	0	1	0	1	0	0
George	1	0	0	0	0	1	0

	a_y	a_m	a_o	h_s	h_m	h_t	symptom
Alice	1	0.5	0	0.5	1	0	1
Boris	1	0.5	0	0	0.5	0.5	0
Cyril	0.5	1	0	0.5	1	0	1
David	0	1	0.5	1	0.5	0	0
Ellen	1	0.5	0	1	0.5	0	1
Fred	0	0	1	0.5	1	0	0
George	1	0.5	0	0	0	1	0

Neural Networks and FCA

- Select a representative dataset
- Define neural network
- Build synthetic dataset
- Present synthetic dataset to neural network
- Build a formal context
- Obtain formal concepts
- Extract implication rules using nextClosure algorithm

Solar Energy

	Tamb(°C)	*Tin(°C)*	*G(W/m2)*	*Tout(°C)*
maximum	27,18	60,81	1084,95	64,73
minimum	22,48	20,29	731,71	26,42

Tamb(°C)	*Tin(°C)*	*G(W/m2)*	*Tout(°C)(Net)*
22.48	20.29	731.71	26.38
22.82	28.97	933.56	35.33
23.15	20.29	958.79	28.42
24.16	40.55	933.56	46.79
24.159	40.55	958.79	46.96
25.17	31.87	857.87	37.70
26.17	40.55	984.02	47.26
27.18	46.34	731.71	51.29
27.18	46.34	756.94	51.47
27.18	60.81	1084.95	64.31

Discretization of variables (2 sets)

	Ranges
Tamb(ºC)	[22.48, 24.83 > [24.83, 27.18]
Tin(ºC)	[20.29, 40.55 > [40.55, 60.81]
G(W/m2)	[731.71, 908.33 > [908.33, 1084.95]
Tout(ºC)	[26.42, 45.575 > [45.575, 64.73]

Tamb(ºC)	*Tin(ºC)*	*G(W/m2)*	*Tout(ºC)(Net)*
22.48	20.29	731.71	26.38
22.82	28.97	933.56	35.33
23.15	20.29	958.79	28.42
24.16	40.55	933.56	46.79
24.159	40.55	958.79	46.96
25.17	31.87	857.87	37.70
26.17	40.55	984.02	47.26
27.18	46.34	731.71	51.29
27.18	46.34	756.94	51.47
27.18	60.81	1084.95	64.31

index	*Tamb(ºC)*	*Tin(ºC)*	*G(W/m2)*	*Tout(ºC)*
1	1	1	1	1
2	1	1	2	1
3	1	1	2	1
4	1	2	2	2
5	1	2	2	2
6	2	1	1	1
7	2	2	2	2
8	2	2	1	2
9	2	2	1	2
10	2	2	2	2

Rule Extraction: NextClosure

1	***If***	Tout = 2	***Then***	Tin = 2		
2	***If***	Tin = 1	***Then***	Tout = 1		
3	***If***	Tin =2	***And***	G = 2	***Then***	Tout =2
4	***If***	G =2	***And***	Tout =1	***Then***	Tin = 1
5	***If***	Tin = 2	***And***	Tout =1	***Then***	G = 1

1	{Tin_[4]} ===> {Tout_[4]}
2	{Tout_[1]} ===> {Tin_[1]}
3	{G_[1] Tout_[4]} ===> {Tin_[4]}
4	{Tin_[1] G_[1]} ===> {Tout_[1]}
5	{G_[1] Tout_[2]} ===> {Tin_[2]}
6	{G_[2] Tout_[4]} ===> {Tin_[4]}
7	{Tin_[1] G_[2]} ===> {Tout_[1]}
8	{Tin_[3] G_[1]} ===> {Tout_[3]}
9	{G_[2] Tout_[2]} ===> {Tin_[2]}
10	{Tin_[3] G_[2]} ===> {Tout_[3]}
11	{Tamb_[1] Tin_[1] G_[3]} ===> {Tout_[1]}
12	{Tamb_[1] G_[3] Tout_[2]} ===> {Tin_[2]}
13	{Tamb_[1] Tin_[1] Tout_[2]} ===> {G_[4]}

References

- Websites with information on FCA:
  - http://www.upriss.org.uk/fca/fca.html
  - http://fca.radvansky.net/news.php

References

- Belohlávek, R., Sklenar, V., and Zacpal, J. (2005). Crisply generated fuzzy Concepts. In ICFCA, pages 269–284.
- Ganter, B. (1984). Two basic algorithms in concept analysis. Technical report, Technische Hochschule Darmstadt. 28 p.
- Ganter, B. (2002). Formal concept analysis: Algorithmic aspects. Technical report, Technische Universitat Dresden.
- Wolff, K. E. (2002). Concepts in fuzzy scaling theory: order and granularity. Fuzzy Sets and Systems, 132(1):63–75. Zheng, S., Zhou, Y., and Martin, T. (2009). A new method for fuzzy formal concept analysis. In ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies, pages 405–408.