

***Projet de Recherche – DIC9410***  
***Doctorat Informatique Cognitive***

***Un Modèle Intelligent d'Estimation des  
Coûts de Développement de Logiciels***

Préparé par:

**Ali Idri**

Directeurs de recherche:

**Professeur Alain Abran, ETS**

**Professeur Serge Robert, UQÀM**

**22 Août 2002, UQÀM, Montréal**

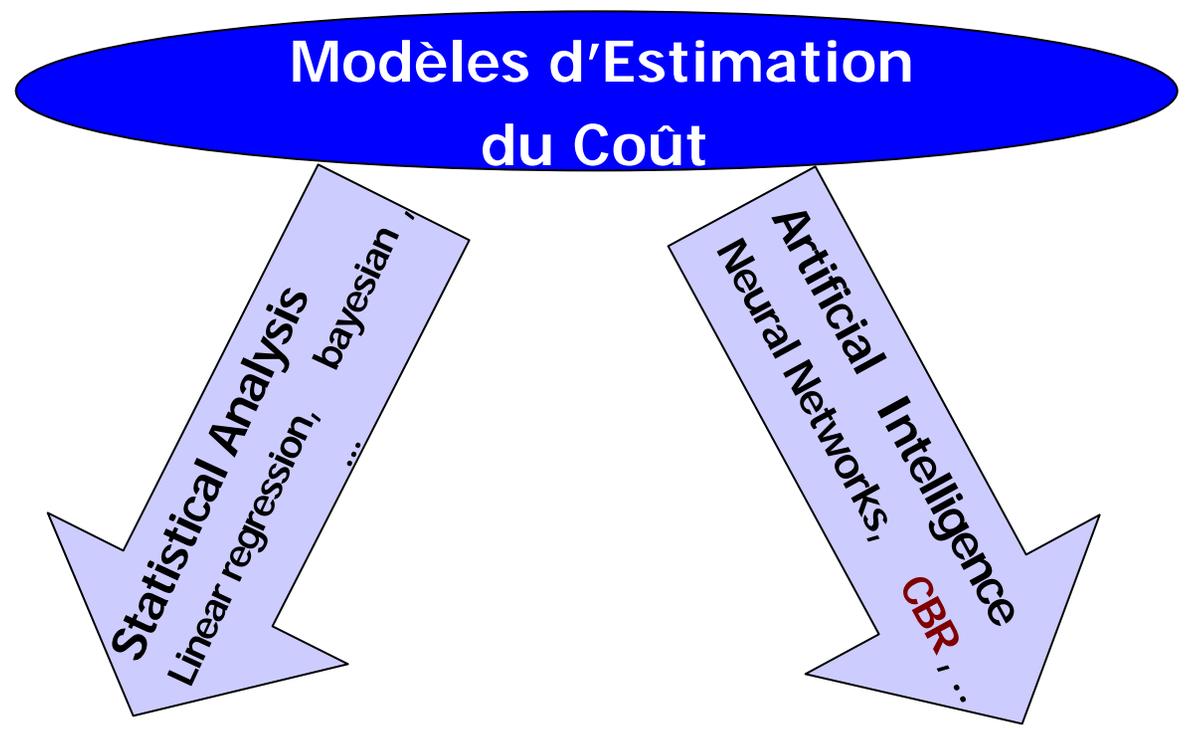
# Plan

---

- ✍ **Problématiques de recherche**
- ✍ **Approches adoptées pour le développement du modèle d'estimation du coût**
- ✍ **Contributions du projet de recherche**
- ✍ **État d'avancement des travaux**
- ✍ **Dernière étape à entreprendre avant d'entamer la rédaction**
- ✍ **Plan de la thèse**

# Problématiques de recherche

- ✍ Estimation du coût de développement d'un logiciel est une tâche très critique en gestion de logiciels (Halstead 75, Walston-Felix 77, Putnam 78, Bohem 81, ...)



Modèles Algorithmiques

Modèles Non-Algorithmiques

- ✍ **Aucune technique ne performe mieux que les autres dans toutes les situations**
  
- ✍ **La performance est définie en terme de l'exactitude des estimations fournies par le modèle (MRE, Pred (0.25))**
  
- ✍ **La performance dépend :**
  - ✍ Du modèle
  - ✍ Taille de la base de données des projets logiciels
  - ✍ Le nombre de variables décrivant les projets logiciels
  - ✍ La présence des points extrêmes
  
- ✍ **L'estimation du coût par Analogie est une technique prometteuse qui attire de plus en plus les chercheurs**  
**Vacinanza 1990, Shepperd 1996, Angelis 2000, ...**

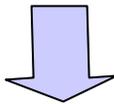
## Processus d'estimation par analogie

-  Caractérisation des projets logiciels par un ensemble de variables (complexité logicielle, compétence des analystes, méthodes de développement utilisées, ...)
-  Evaluation de la similarité entre le nouveau projet et les projets logiciels historiques
-  Utilisation des valeurs des coûts réels des projets similaires au nouveau projet pour en déduire une estimation à son coût

## Limites de l'estimation par Analogie

-  Elle ne traite pas convenablement le cas des projets logiciels décrits par des valeurs linguistiques telles que '**bas**', '**élevé**', et '**complexe**'
-  Elle ne permet pas la gestion de l'incertitude au niveau des estimations fournies
-  Elle ne permet pas l'apprentissage

## ✍ Problématique 1 : Tolérance des imprécisions

- ✍ En génie logiciel, les projets logiciels sont souvent décrits par des valeurs linguistiques plutôt que par des valeurs numériques:
    - ✍ Le génie logiciel est une science encore jeune !!
    - ✍ Les humains sont directement impliqués dans le processus de mesures de presque tous les attributs d'un logiciel
- 
- ✍ La plupart des attributs sont mesurés sur une échelle au plus ordinale
    - ✍ Cas du COCOMO'81, 15 parmi 17 attributs sur une échelle ordinale de six valeurs linguistiques 'very low', 'low', 'nominal', 'high', 'very high', 'extra high'
    - ✍ Cas du COCOMO II, 22 parmi 24
  - ✍ Tous les modèles d'estimation du coût, y compris l'estimation par Analogie, utilise une représentation classique pour traiter ces valeurs

DATA	Low	Nominal	High	Very High
D/P	D/P < 10	10 ? D/P < 100	100 ? D/P < 1000	D/P ?? 1000

## ✍ **Problématique 2 : Gestion de l'incertitude**

- ✍ Une estimation est une évaluation d'une situation future d'une entité particulière
- ✍ Les modèles d'estimation existants, y compris l'estimation par analogie, ne génère qu'une seule valeur estimée => possibilités d'erreurs fatales au niveau de la gestion du projet
- ✍ Solution: Générer l'estimation sous forme d'un ensemble de valeurs avec une possibilité de distribution
- ✍ En estimation du coût par analogie, la nécessité de tenir compte des incertitude découle de:

### **Similar software projects have similar costs**

- ✍ La conséquence de cette hypothèse est imprécise
- ✍ L'hypothèse peut être non-deterministique

## ✍ Problématique 3 : Apprentissage

- ✍ Un modèle d'estimation doit être capable d'apprendre afin d'adapter ses estimations dépendamment des changements survenus dans l'organisme:
  - ✍ On ne développe plus le même type d'application
  - ✍ La complexité est de plus en plus élevée
  - ✍ Le personnel est de plus en plus expérimenté
  - ✍ Etc.
  
- ✍ Deux avantages de l'apprentissage:
  - ✍ L'apprentissage est un moyen efficace pour améliorer la précision du modèle au niveau des estimations du coût qu'il fournit
  - ✍ L'apprentissage peut être utilisé pour calibrer et adapter un modèle d'estimation du coût, développé dans un environnement, à un autre environnement différent.
  
- ✍ L'estimation par analogie n'intègre pas pour le moment des mécanismes d'apprentissage dans son processus

## Objectifs

**Développer une nouvelle technique  
d'estimation du coût par analogie**

- (0) Exactitude des estimations**
- (1) Tolérance des imprécisions au niveau des descriptions des projets logiciels**
- (2) Gestion de l'incertitude au niveau des estimations fournies**
- (3) Apprentissage**

# Approches Adoptées

---

## ✍ C'est quoi un modèle intelligent d'estimation du coût?

- ✍ Le concept de l'intelligence!!
- ✍ L'intelligence regroupe plusieurs caractéristiques
  - ✍ Apprentissage
  - ✍ Créativité
  - ✍ Facilité de comprendre
  - ✍ Etc

*« Someones's intelligence is their ability to understand and learn things »*

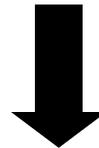
*« Intelligence is the ability to think and undestand instead of doing things by instinct or automatically »*

- ✍ Les caractéristiques de l'intelligence ne sont pas présentes avec les mêmes degrés chez les humains
- ✍ Les quatre caractéristiques d'un modèle intelligent d'estimation du coût
  - ✍ Exactitude
  - ✍ Tolérance des imprécisions
  - ✍ Gestion de l'incertitude
  - ✍ Apprentissage

- ✍ **Quelle est l'approche de modélisation à utiliser pour modéliser la relation entre le coût et les les facteurs l'affectant?**
  
- ✍ **Importance du raisonnement dans notre intelligence**
- ✍ **Exemples de types de raisonnement dans l'IA**
  - ✍ **Raisonnement logique et formel (Dédution)**
  - ✍ **Raisonnement par généralisation (Induction)**
  - ✍ **Raisonnement par analogie**
  
- ✍ **Le raisonnement par analogie consiste en la mise en correspondance entre deux situations (**ancienne** et **nouvelle**) afin de fournir un comportement devant une situation nouvelle en fonction d'informations mémorisées sur l'ancienne situation**
  
- ✍ **Il a été utilisé pour l'explication de certains concepts qui ne sont pas facilement explicables directement (cas de religions:  $R(\text{God, People}) = R(\text{Father, Children})$ )**
  
- ✍ **Il a été utilisé aussi pour le développement de nouvelles théorie scientifiques (symbolisme et connexionnisme en sciences cognitives)**

- ✍ Le raisonnement par analogie en IA n'a pas eu le même intérêt que le raisonnement formel ou le raisonnement par généralisation
- ✍ Une version simplifiée du raisonnement par analogie: le raisonnement à base de cas (**Case-Based Reasoning** – CBR)
  - ✍ Origines : **Schank (1982)** et **Gentner(1983)**
  - ✍ **Aomodt, Plaza, Aha, kolodner, Leake,...**
- ✍ **Processus de la technique CBR:**
  - ✍ Retrouver les cas similaires au cas étudié
  - ✍ Utiliser les cas similaires retrouvés pour résoudre le problème
  - ✍ Réviser la solution proposée
  - ✍ Retenir les éléments de cette expérience qui serviront à la résolution des problèmes futurs
- ✍ **Avantages de la technique CBR**
  - ✍ Elle est facile à expliquer contrairement au cas des réseaux de neurones.
  - ✍ Elle peut modéliser les relations complexes existantes entre le coût et les facteurs affectant le coût contrairement au cas des méthodes statistiques.
  - ✍ Elle peut être utilisée dans le cas où on a peu de connaissances sur le problème étudié contrairement au cas des systèmes à base de règles.

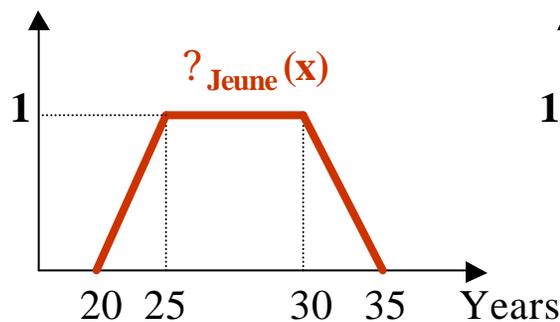
- ✍ Elle utilise la connaissance dans sa forme 'brute' contrairement au cas des systèmes à base de règles où la connaissance doit être encodée dans des règles (ce processus d'encodage n'est pas souvent facile)
- ✍ Elle peut être utilisée dans le cas où une solution optimale du problème n'est pas évidente
- ✍ Elle généralise plusieurs types de modélisation en IA :
  - ✍ Cas d'un arbre de régression
  - ✍ Cas de certains réseaux de neurones (Perceptron simple, Réseau de Kohonen, Radial Basis Function Network, Réseau de Hopfield)
- ✍ **Cependant, le raisonnement par analogie chez les humains est souvent approximatif plutôt qu'exact**
  - ✍ Le cerveau humain manipule des informations imprécises, incertaines et partielles dans le processus de l'analogie
  - ✍ Le cerveau humain gère l'incertitude au niveau des conclusions déduites de l'analogie
  - ✍ Le cerveau humain apprend et ajuste ses connaissances après un processus d'analogie



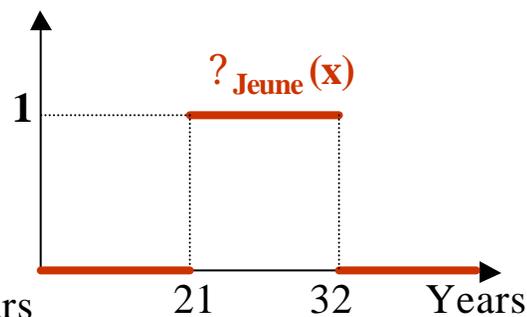
**Nécessité d'intégration de techniques dans le processus d'estimation par analogie afin qu'il soit au niveau de celui adopté par les humains**

## ✍ Quelles sont les techniques que l'estimation par analogie doit utiliser pour satisfaire les quatre critères d'intelligence?

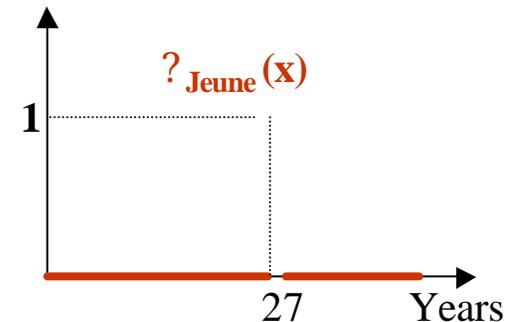
- ✍ La logique floue pour le traitement des imprécisions et des incertitudes (Zadeh 1965)
- ✍ Les réseaux de neurones pour l'apprentissage
- ✍ La logique floue = la théorie des ensembles flous
- ✍ Ensemble flou est un ensemble dont la fonction d'appartenance prend des valeurs dans l'intervalle  $[0,1]$  plutôt que dans la paire  $\{0,1\}$
- ✍ La logique floue offre des outils pour la représentation et le traitement des valeurs linguistiques (situations floues)



Ensemble flou



Intervalle classique



Nombre

## ✍ Pourquoi la logique floue est-elle intéressante pour les systèmes intelligents?

- ✍ Le grand défi pour l'IA se trouve dans les domaines où la connaissance est imprécise et incertaine (Robotique, Industries de produits de consommation)
- ✍ La logique floue avantage la métaphore du raisonnement approximatif utilisé par les humains (exemple du stationnement d'une voiture)
- ✍ La logique floue offre des outils puissants pour l'imitation de l'esprit humain
  - ✍ Computing with Words (CW)
  - ✍ Computational Theory of Perceptions (CTP)

## ✍ Logique floue et Intelligence Artificielle

- ✍ L'IA ne manipulait que deux valeurs de vérités (vraie et fausse)
- ✍ L'intelligence artificielle classique ne peut traiter les situations où l'information est imprécise:

**Le problème de stationnement d'une voiture**

# Contributions de notre projet

---

## ✍ Estimation du coût de logiciels

- ✍ Une nouvelle technique d'estimation par analogie qui permet:
  - ✍ Tolérance des imprécisions dans la description des projets logiciels
  - ✍ Gestion de l'incertitude au niveau des estimations fournies
  - ✍ Apprentissage
- ✍ Une technique facilement interprétable, flexible et facilement adaptable aux besoins des utilisateurs

## ✍ Raisonnement à base de cas

- ✍ Montrer l'utilité du raisonnement à base de cas dans la résolution des problèmes complexes
- ✍ Développer de nouvelles mesures de similarités
- ✍ Développer une nouvelle technique d'adaptation

## **Logique floue et Intelligence artificielle**

-  Illustrer le rôle de la logique floue dans le développement des systèmes intelligents
-  Symbolisme et connexionnisme sont complémentaires plutôt que exclusives

## **Mesure de logiciels**

-  Montrer l'utilité d'utilisation des valeurs linguistiques plutôt que des valeurs numériques dans le processus de mesurage des attributs d'un logiciel
-  Software perceptions or Software measures?

## **Computation intelligente en génie logiciel**

-  Logiciels intelligents ?
-  Estimation du coût par une approche intelligente

# État d'avancement des travaux

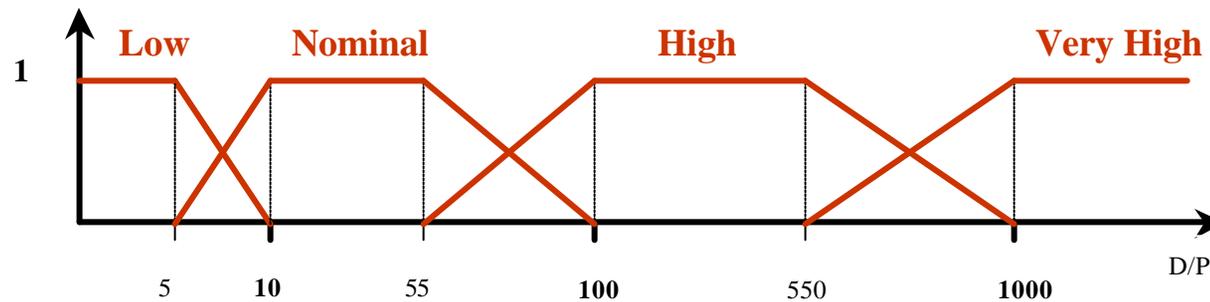
- ✍ Incorporation de la logique floue dans le modèle COCOMO'81 (7<sup>th</sup> FT&T, Atlantic City, 2000)

$$MM_{est} = A \cdot Size^B \cdot \prod_{i=1}^{15} C_{ij}$$

- ✍ Size est mesurée en terme de KDSI (nombre)
- ✍ 15 facteurs sont mesurés sur une échelle ordinaire de six valeurs linguistique **very low, low, nominal, high, very high, extra high**

Low	Nominal	High	Very high
D/P < 10	10 ? D/P < 100	100 ? D/P < 1000	D/P ? 1000

- ✍ Si on considère deux projets P<sub>1</sub> et P<sub>2</sub> ayant des valeurs au voisinage des limites pour tous les 15 attributs, on aura un effort pour P<sub>1</sub> égal à **15H-M** alors que pour P<sub>2</sub>, l'effort est à **52H-M**



'fuzzy' intermediate COCOMO'81				
	Database #1	Database #2	Database #3	COCOMO'81
Pred(20) (%)	62.14	46.86	41.27	68
Min RE <sub>i</sub> (%)	0.11	0.40	0.06	0.02
Max RE <sub>i</sub> (%)	88.60	3233.03	88.03	83.58
Mean RE <sub>i</sub> (%)	22.50	78.45	30.80	18.52
Standard deviation RE <sub>i</sub>	19.69	404.40	22.95	16.97

classical intermediate COCOMO'81				
	Database #1	Database #2	Database #3	COCOMO'81
Pred(20) (%)	68	68	68	68
Min RE <sub>i</sub> (%)	0.02	0.02	0.02	0.02
Max RE <sub>i</sub> (%)	83.58	83.58	83.58	83.58
Mean RE <sub>i</sub> (%)	18.52	18.52	18.52	18.52
Standard deviation RE <sub>i</sub>	16.97	16.97	16.97	16.97

- ✍ **Un ensemble de mesures de similarité entre les projets logiciels** (7<sup>th</sup> IEEE International Symposium Software Metrics, London, 2001; 9<sup>th</sup> IFSA/20<sup>th</sup> NAFIPS, Vancouver, 2001)

- ✍ Shepperd et al. (1997)

$$d(P_1, P_2, V) = \frac{1}{\sum_{v_j} d_{v_j}(P_1, P_2)}$$

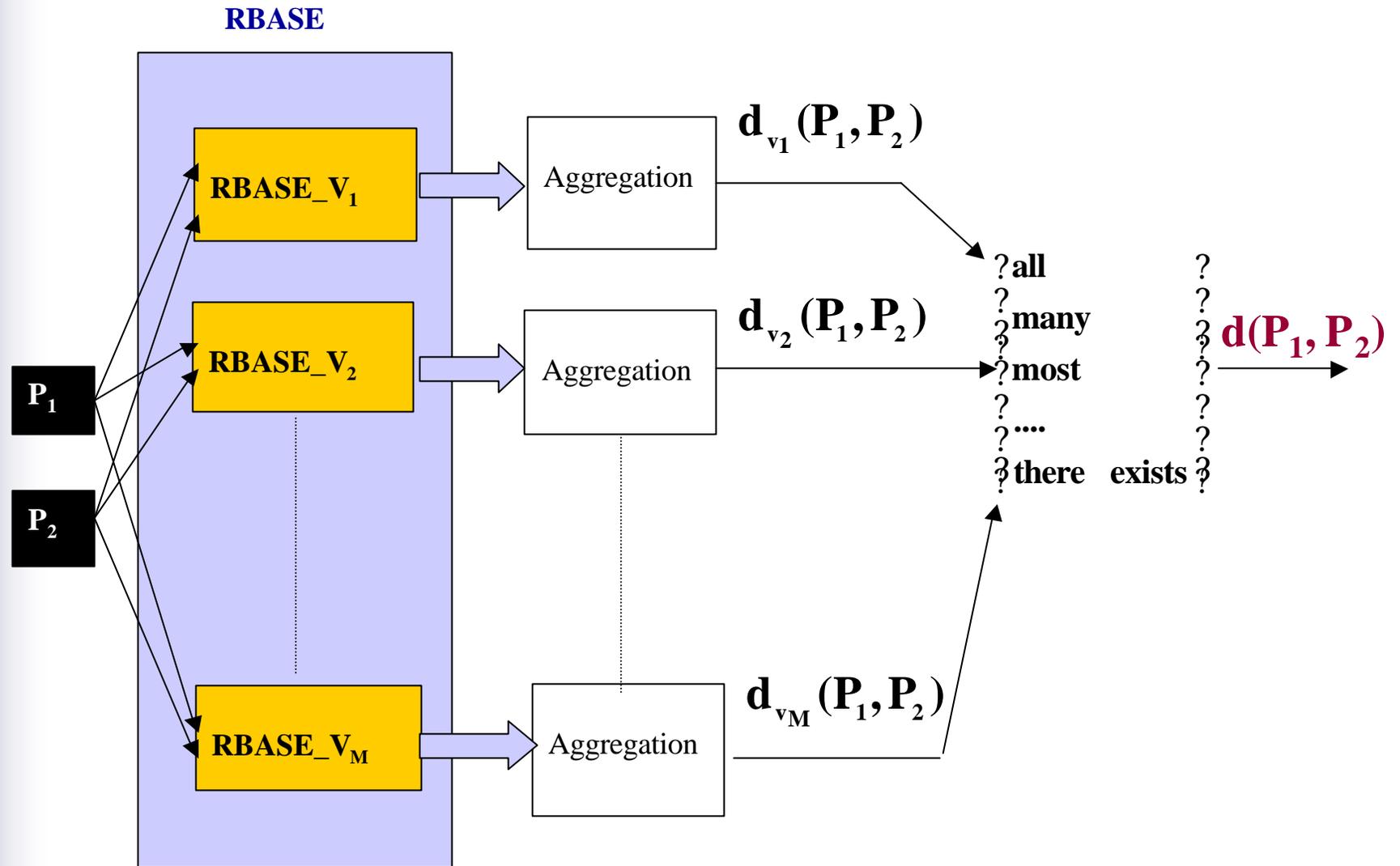
$$d_{v_j}(P_1, P_2) = \frac{1}{\sum_{v_j} (V_j(P_1) \oplus V_j(P_2))^2}$$

$\oplus = 0$  si  $V_j(P_1) = V_j(P_2)$   
 $\oplus = 1$  si  $V_j(P_1) \neq V_j(P_2)$

Classical Logic

Fuzzy Logic

Valeurs linguistiques  
faible, élevé, Excellent ???



**Individual similarities**

$$d_{v_j}(P_1, P_2) = \frac{1}{k} \sum_{k=1}^k \min(\max(A_k^j(P_1), A_k^j(P_2)), \max(A_k^j(P_1), A_k^j(P_2)))$$

$\frac{1}{k} \sum_{k=1}^k \min(\max(A_k^j(P_1), A_k^j(P_2)), \max(A_k^j(P_1), A_k^j(P_2)))$   
 $\frac{1}{k} \sum_{k=1}^k \min(\max(A_k^j(P_1), A_k^j(P_2)), \max(A_k^j(P_1), A_k^j(P_2)))$

**Overall Similarity**

The implementation of the chosen linguistic quantifier is realized by an Ordered Weight Averaging operator

- $d(P_1, P_2)$  all of  $d_{v_j}(P_1, P_2)$
- $d(P_1, P_2)$  most of  $d_{v_j}(P_1, P_2)$
- $d(P_1, P_2)$  many of  $d_{v_j}(P_1, P_2)$
- $d(P_1, P_2)$  at least four of  $d_{v_j}(P_1, P_2)$
- $d(P_1, P_2)$  ...
- $d(P_1, P_2)$  there exists of  $d_{v_j}(P_1, P_2)$

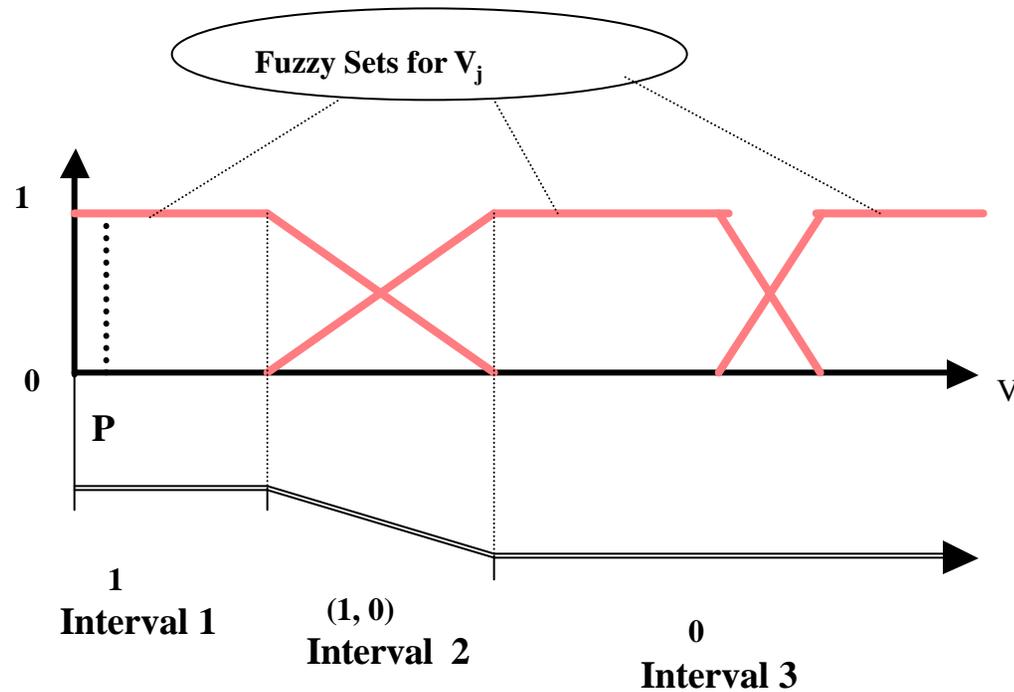


$$d(P_1, P_2) = \sum_{j=1}^M w_j d_{v_j}(P_1, P_2)$$

$$w_j(P_1, P_2) = Q\left(\frac{k-1}{T}\right) = Q\left(\frac{j-1}{T}\right)$$

## Axiome 0

$$d_{v_j}(P_1, P_2) = 0 \text{ iff } A_k^j / A_k^j(P_1) = 0 \text{ and } A_k^j(P_2) = 0$$



## Axiome 1

$$d(P_1, P_2) = 0; d(P, P) > 0$$

 **Axiom 2**

$$d(P, P_i) \neq d(P, P)$$

 **Axiom 3**

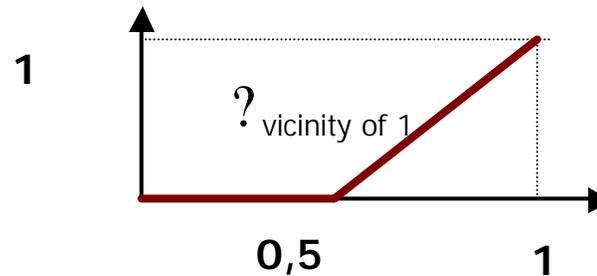
$$d(P_1, P_2) = d(P_2, P_1)$$

 **Résultats de la validation axiomatique**

	$d_{v_j}(P_1, P_2) / d(P_1, P_2)$		
	Max-min	Sum-product	Kleene-Dienes
<b>Axiom 0</b>	Yes/	Yes/	No/
<b>Axiom 1</b>	Yes/Yes	Yes/Yes	Yes/Yes
<b>Axiom 2</b>	Yes/Yes	No/No	Yes/Yes if NC
<b>Axiom 3</b>	Yes/Yes	Yes/Yes	No/No

## ✍ Estimation par analogie utilisant la logique floue (8<sup>th</sup> IEEE International Symposium on Software Metrics, Ottawa, 2002)

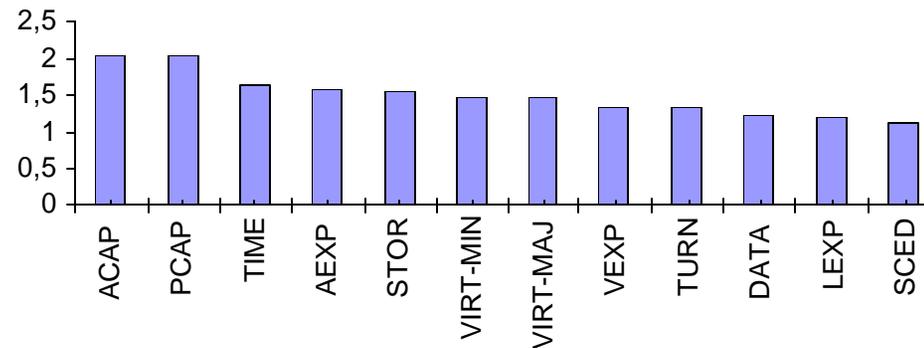
- ✍ Identification des projets logiciels
- ✍ Évaluation de la similarité entre les projets logiciels
- ✍ Adaptation
  - ✍ Combien de projets similaires seront utilisés? **En général k=2**
  - ✍ Comment calculer le coût du nouveau projet à partir des coûts réels des projets similaires? **(la moyenne, la médiane)**



$$Effort(P) = \frac{\sum_{i=1}^N \text{? ?}_{\text{vicinity of 1}}(d(P, P_i)) \cdot Effort(P_i)}{\sum_{i=1}^N \text{? ?}_{\text{vicinity of 1}}(d(P, P_i))}$$

si  $\text{?}_{\text{vicinity of 1}}(x) = x$ , on retrouve la formule de la moyenne pondérée

- ✍ La base de données COCOMO'81 contient 63 projets logiciels
- ✍ Nos mesures de similarités engage beaucoup de calculs, nous avons donc développé un prototype pour automatiser ces calculs : **F\_ANGEL**
- ✍ **F\_ANGEL** permet l'utilisation de plusieurs quantificateurs linguistiques de la forme  $Q(x) = x^? , ? > 0$
- ✍ Les poids  $U_k$  sont calculés en utilisant le ratio de productivité d'un projet:



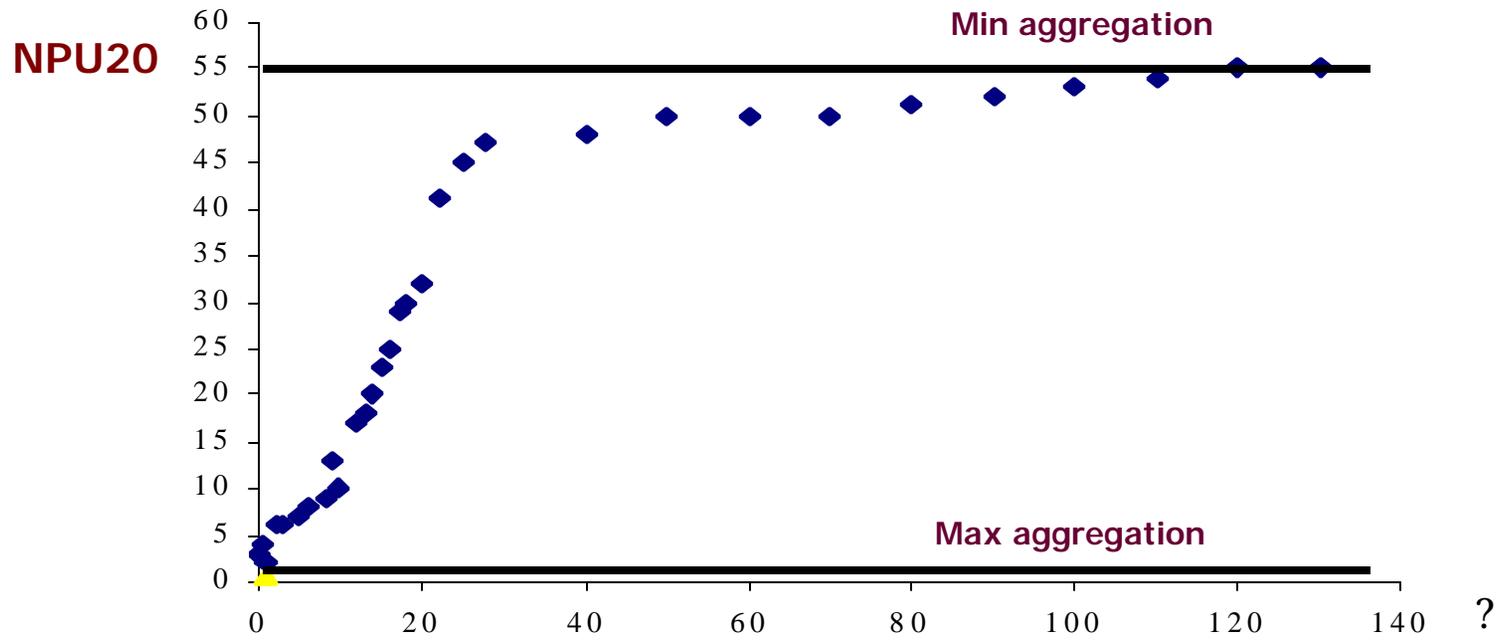
✂ Comparaison avec trois autres modèles

- ✂ Classical Analogy
- ✂ Intermediate COCOMO'81
- ✂ Fuzzy Intermediate COCOMO

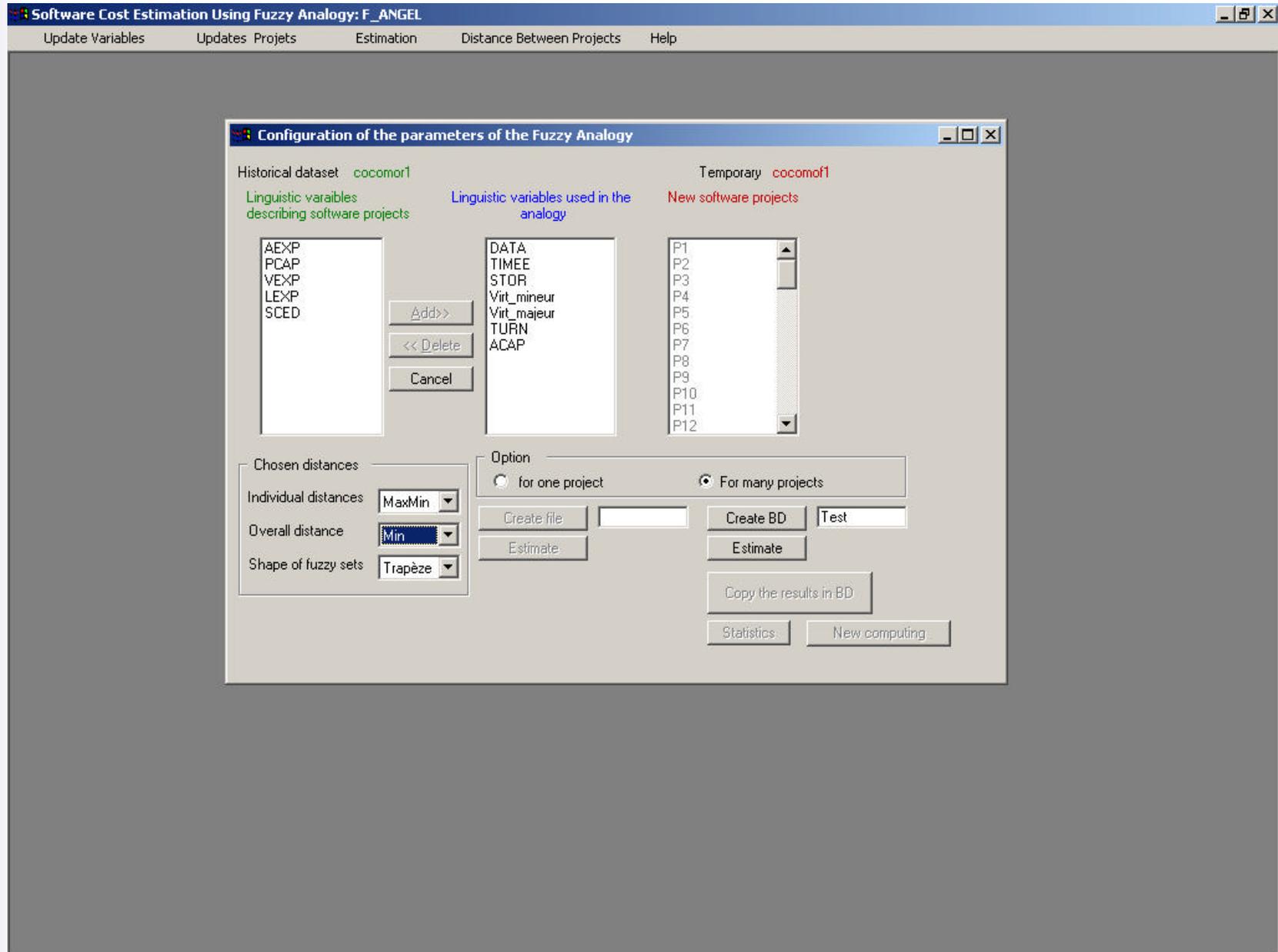
✂ L'exactitude des estimations est évaluée par:

- ✂ MRE, Pred (0.20)
- ✂ Quatre autres quantités, MMRE, min of MRE, max of MRE, and SDMRE

		<i>Dataset #1</i>	<i>Dataset #2</i>	<i>Dataset #3</i>
		<i>Pred(0.20) (%)</i>	<i>Pred(0.20) (%)</i>	<i>Pred(0.20) (%)</i>
?-RIM	<b>Max (?)</b>	<b>4,76</b>	<b>4,76</b>	<b>4,76</b>
	<b>1/100</b>	<b>4,76</b>	<b>4,76</b>	<b>4,76</b>
	<b>1/30</b>	<b>4,76</b>	<b>4,76</b>	<b>4,76</b>
	<b>1/15</b>	<b>4,76</b>	<b>4,76</b>	<b>4,76</b>
	<b>1/10</b>	<b>4,76</b>	<b>4,76</b>	<b>4,76</b>
	<b>1/7</b>	<b>4,76</b>	<b>4,76</b>	<b>4,76</b>
	<b>1/3</b>	<b>6,34</b>	<b>6,34</b>	<b>6,34</b>
	<b>1</b>	<b>6,34</b>	<b>3,17</b>	<b>7,93</b>
	<b>3</b>	<b>6,34</b>	<b>9,52</b>	<b>9,52</b>
	<b>7</b>	<b>9,52</b>	<b>14,28</b>	<b>9,52</b>
	<b>10</b>	<b>15,87</b>	<b>20,63</b>	<b>23,80</b>
	<b>15</b>	<b>38,09</b>	<b>36,50</b>	<b>44,44</b>
	<b>30</b>	<b>74,60</b>	<b>77,77</b>	<b>66,66</b>
	<b>100</b>	<b>92,06</b>	<b>84,12</b>	<b>87,63</b>
	<b>Min (All)</b>	<b>92,06</b>	<b>87,30</b>	<b>88,88</b>



	<b>'fuzzy'/classical intermediate COCOMO'81</b>						<b>Classical Analogy (three datasets)</b>	
	<b>Dataset #1</b>		<b>Dataset #2</b>		<b>Dataset #3</b>		<b>K</b>	<b>Pred(0.20) %</b>
Pred(20) (%)	62.14	68	46.86	68	41.27	68	<b>2</b>	31,75
Min MRE (%)	0.11	0.02	0.40	0.02	0.06	0.02	<b>3</b>	25,40
Max MRE (%)	88.60	83.58	3233.03	83.58	88.03	83.58	<b>4</b>	19,05
Mean MRE <sub>i</sub> (%)	22.50	18.52	78.45	18.52	30.80	18.52	<b>5</b>	12,70
Standard deviation MRE	19.69	16.97	404.40	16.97	22.95	16.97	<b>6</b>	12,70



Software Cost Estimation Using Fuzzy Analogy: F\_ANGEL

Update Variables    Update Projets    Estimation    Distance between Projects    Help

Updating Historical and Temporary datasets

Add    Modify    Delete    Close

BDHIST     BDTMP

Name of variable:

Projet:	P29
Cout:	7
DATA:	23.65806007
TIMEE:	29.64956093
STOR:	47.81184387
Virt_mineur:	13.88129997
Virt_majeur:	89.42086792
TURN:	1.72944355
ACAP:	34.39971161
AEXP:	0.457331181
PCAP:	74.68506622
VEXP:	9.063806534
LEXP:	8.598918915
SCED:	67.69272614

29/63

Software Cost Estimation Using Fuzzy Analogy: F\_ANGEL

Update Variables    Update Projets    Estimation    Distance between Projects    Help

	Très faible	Faible	Moyen	Elevé	Très élevé	Extra élevé	dvj(P1,P3)
Data		0	0	0	1		
		0	0	0	1		1
Time			1	0	0	0	
			0,9331519316	0683999999E-02	0	0	0,9331519316
Stor			0,8384077452	0,1615922548	0	0	
			0	0,840962601	0,159037399	0	0,1615922548
Virt_mineur		1	0	0	0		
		0	0,7252386092	0,2747613908	0		0
Virt_majeur		1	0	0	0		
		0	0	1	0		0
Turn		0,643229514	0,356770486	0	0		
		0	0	1	0		0
Acap	0	0	0	1	0		
	0	1	0	0	0		0
Aexp	0	0	0	0	1		
	0	0,4505198	0,5494802	0	0		0
Pcap	0	0	0	2116088899999	78839111000001		
	0	1	0	0	0		0
Vexp	0	0	0	1			
	0	0,478552818	0,521447182	0			0
Lexp	0	0	0	1			
	0	0	1	0			0
Sced	0	0	1	0	0		
	0	0	0	0,58192444	0,41807556		0
d(P1,P3)							1

Individual similarities: MaxMin

Overall Similarity P1 and P3: Max

Historical project: P1, P2, P3, P4, P5, P6

New project: P1, P2, P3, P4, P5, P6

## ✍ La gestion de l'incertitude dans l'estimation par analogie (soumis à International Journal of Engineering Intelligent Systems)

### ✍ Deux sources d'incertitudes dans l'estimation par analogie:

- ✍ Mesures des facteurs affectant le coût
- ✍ Performance de l'estimation par analogie

### ✍ Pourquoi l'hypothèse 'similar projects have similar costs' peut être non-déterministique?

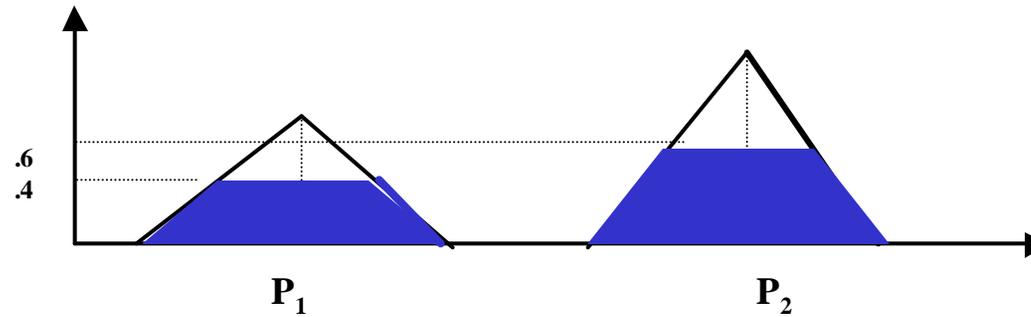
- ✍ On ne peut inclure tous les attributs dans la description des projets logiciels
- ✍ Les mesures de similarités sont-elles valides?

### ✍ Évaluation de l'hypothèse 'similar projects have similar costs' en utilisant la base de données COCOMO'81

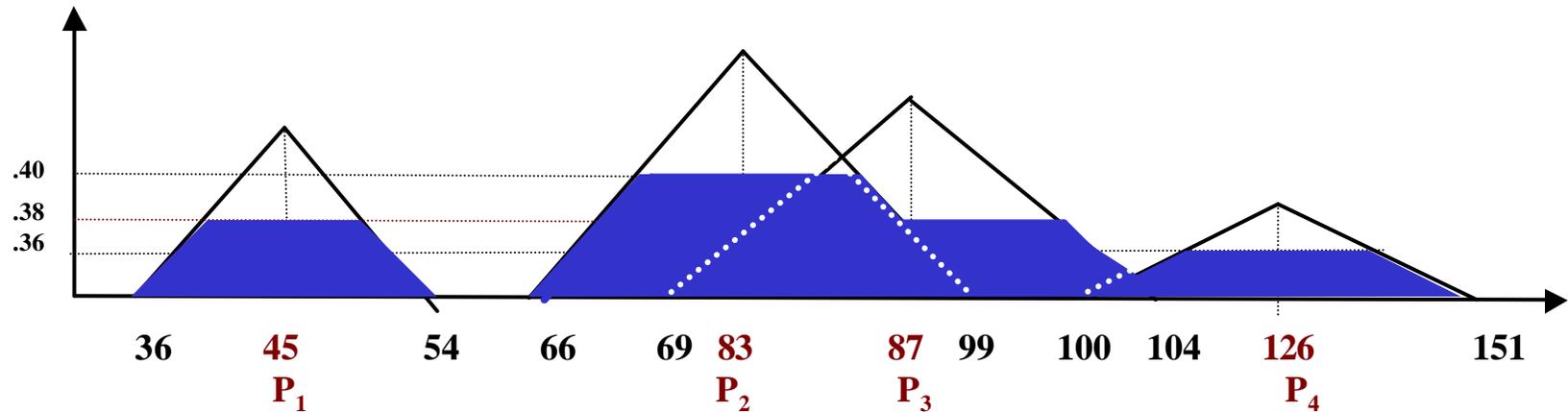
- ✍  $S(P_1, P_2) ? C(C_1, C_2) \Rightarrow$  déterministique
- ✍  $S(P_1, P_2) > C(C_1, C_2) \Rightarrow$  non-déterministique

## ✂ Cas de l'hypothèse déterministique

### ✂ Problème d'existence d'une solution



## ✂ Cas de l'hypothèse non-déterministique



### ✂ Gestion des risques

$$\text{Risque} = (E_2 - E_1) \cdot \text{Poss}(E_2)$$

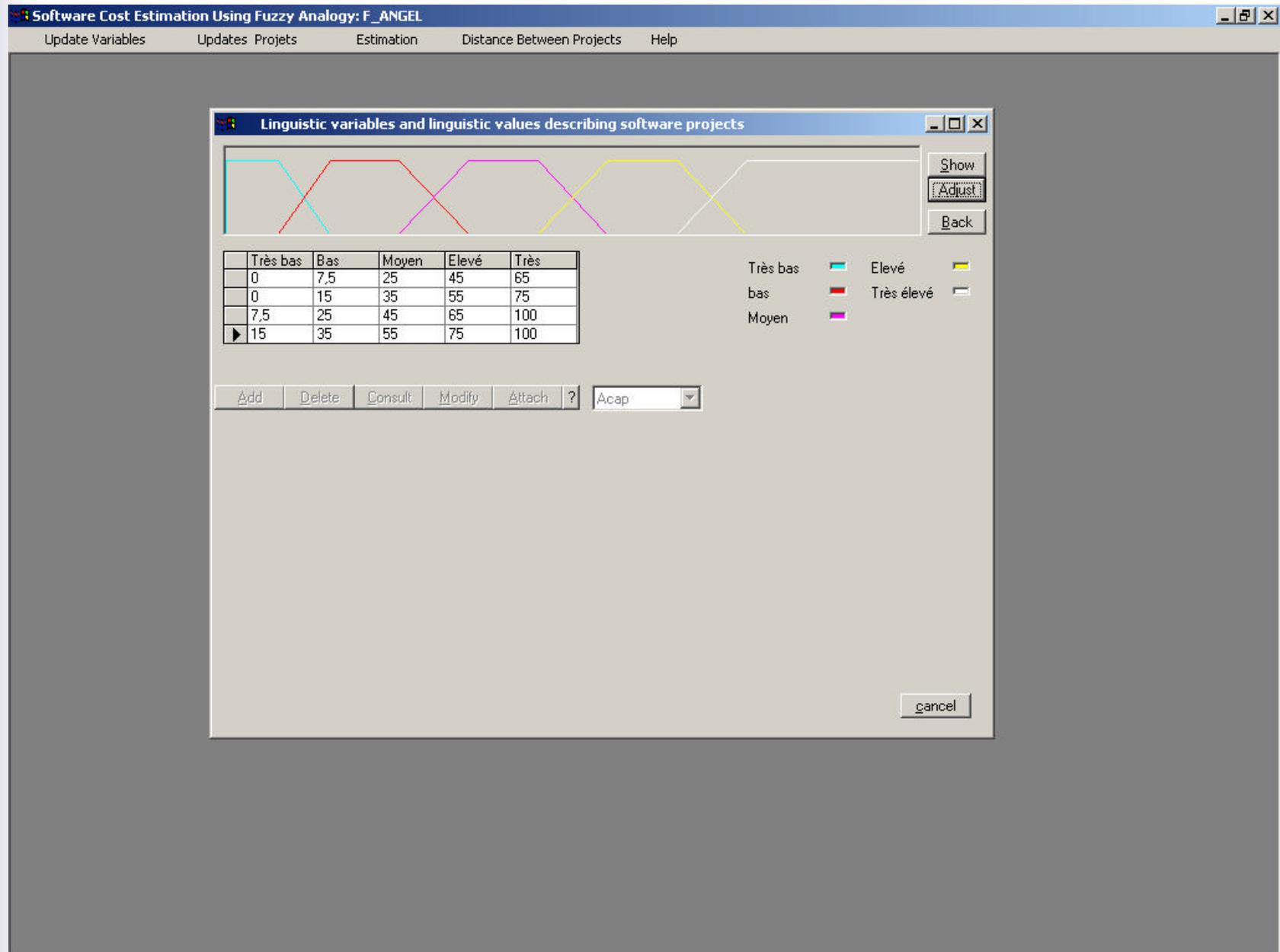
$E_1 = 45\text{K}\$, E_2 = 126\text{K}\$ \text{ alors Risque} = 29\text{K}\$$

# Dernière étape à entreprendre

---

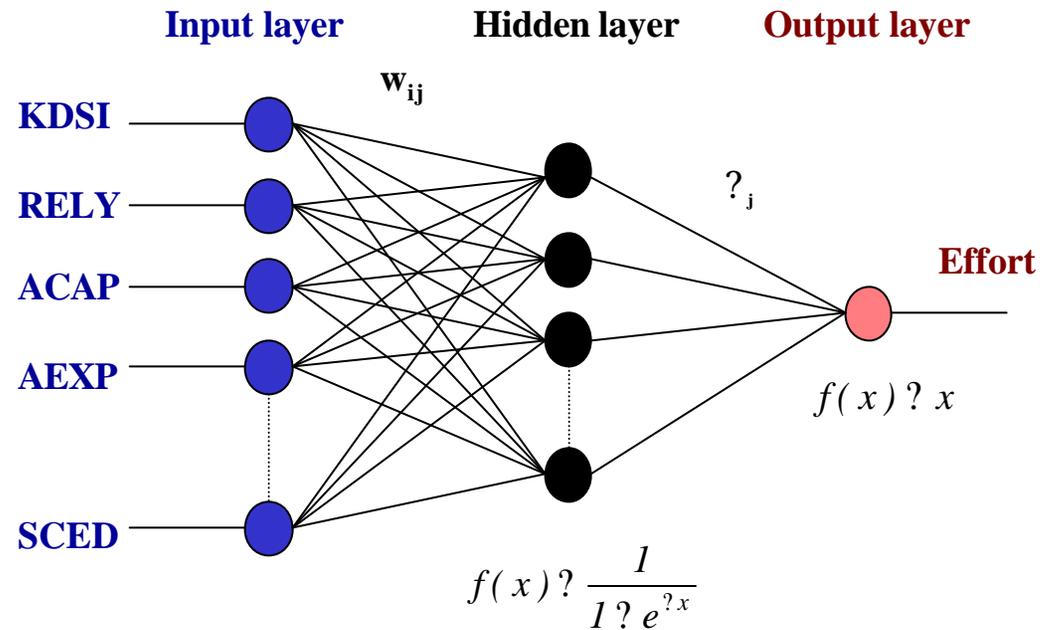
## ✍ Introduire l'apprentissage

- ✍ Certaines fonctionnalités d'apprentissage sont déjà introduites dans notre approche:
  - ✍ Suppression des attributs qui ne sont plus significatifs (capacité de stockage)
  - ✍ Ajout de nouveaux attributs significatifs (sécurité)
  - ✍ Modification des informations concernant les valeurs d'un attribut
    - ✍ Ajout de nouvelles valeurs linguistiques
    - ✍ Suppression des anciennes valeurs linguistiques
    - ✍ Changement de la représentation des valeurs linguistiques
    - ✍ Changement du poids associé à l'attribut
  - ✍ Modification de la définition du quantificateur linguistique utilisé dans l'évaluation de la similarité entre les projets logiciels
  - ✍ Modification de la la définition de la valeur linguistique 'Closely similar' utilisée dans l'étape d'adaptation



## Apprentissage au niveau des coûts estimés

- Les réseaux de neurones sont-ils 'black Boxes'?
- Interprétation d'un réseau de neurone en estimation du coût de développement de logiciels (FUZZ-IEEE, Hawaii, 2002)



✂ Validation du réseau de neurones sur la base de données COCOMO'81

✂ Même données d'apprentissage et de test

MRE %	NN avec 13 unités cachées et 300 000 itérations d'apprentissage
<b>Max</b>	16,67
<b>Moyen</b>	1,50
<b>Min</b>	0,00

✂ Différentes données d'apprentissage et de test

Indicateurs	NN avec 40 projets pour l'apprentissage	NN avec 62 projets pour l'apprentissage
<b>Pred(25)</b>	13,04	34,78
<b>MMRE%</b>	203,66	84,35

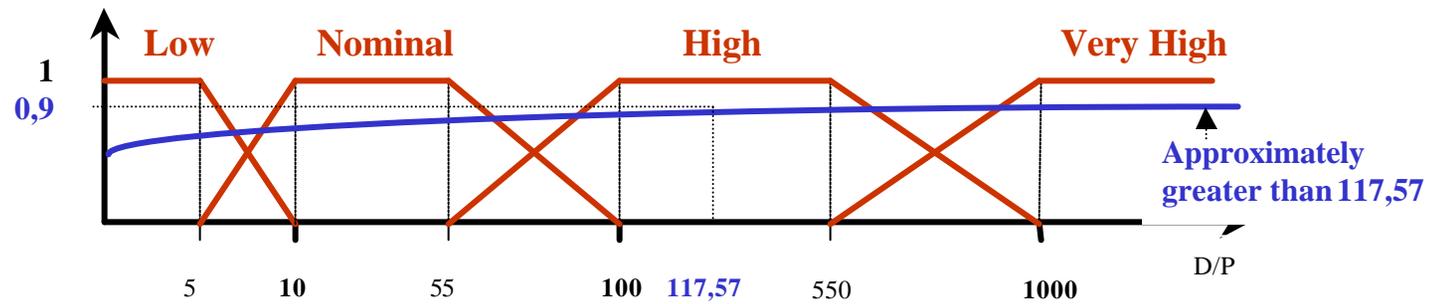
## ✍ Transformation d'un NN en un système à base de règles floues

### ✍ Règles floues

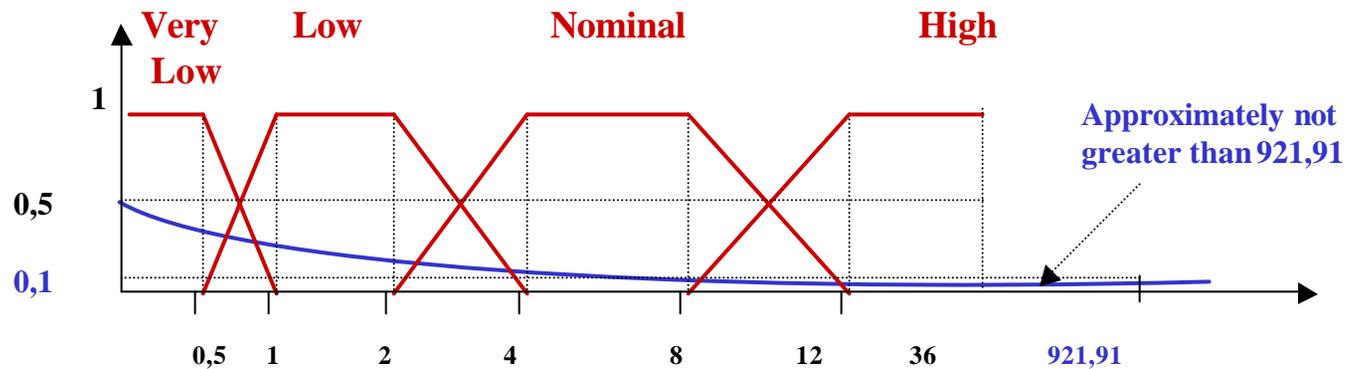
**Mamdani Rule :**  $r_k$  : if  $x_1$  is  $A_1^k$  and  $x_2$  is  $A_2^k$  and...and  $x_{N_x}$  is  $A_n^k$   
 then  $y_1$  is  $B_1^k$ , ...,  $y_{N_y}$  is  $B_{N_y}^k$

**Sugeno Rule :**  $r_k$  : if  $x_1$  is  $A_1^k$  and  $x_2$  is  $A_2^k$  and...and  $x_{N_x}$  is  $A_n^k$   
 then  $y_1$  ?  $f_{1,k}(x_1, \dots, x_{N_x})$ , ...,  $y_{N_y}$  ?  $f_{N_y,k}(x_1, \dots, x_{N_x})$

$R_1$ :		
<b>If :</b>		
<b>DATA</b>	is approximately greater than	117,57
<b>VIRTmi</b>	is approximately not greater than	30,14
<b>TIME</b>	is approximately not greater than	156,28
<b>STORE</b>	is approximately not greater than	155,72
<b>VIRTma</b>	is approximately not greater than	448,88
<b>TURN</b>	is approximately not greater than	25,76
<b>ACAP</b>	is approximately not greater than	203,72
<b>AEXP</b>	is approximately not greater than	400,37
<b>PCAP</b>	is approximately not greater than	331,69
<b>VEXP</b>	is approximately not greater than	1958,86
<b>LEXP</b>	is approximately not greater than	921,91
<b>SCED</b>	is approximately not greater than	794,68
<b>KDSI</b>	is approximately not greater than	132,27
<b>Then</b>	<b>Effort</b> = 6049,77	



DATA is approximately greather than 117,58 ? DATA is high or very high



LEXP is approximately not greather than 921,91 ? LEXP is Q(very low)

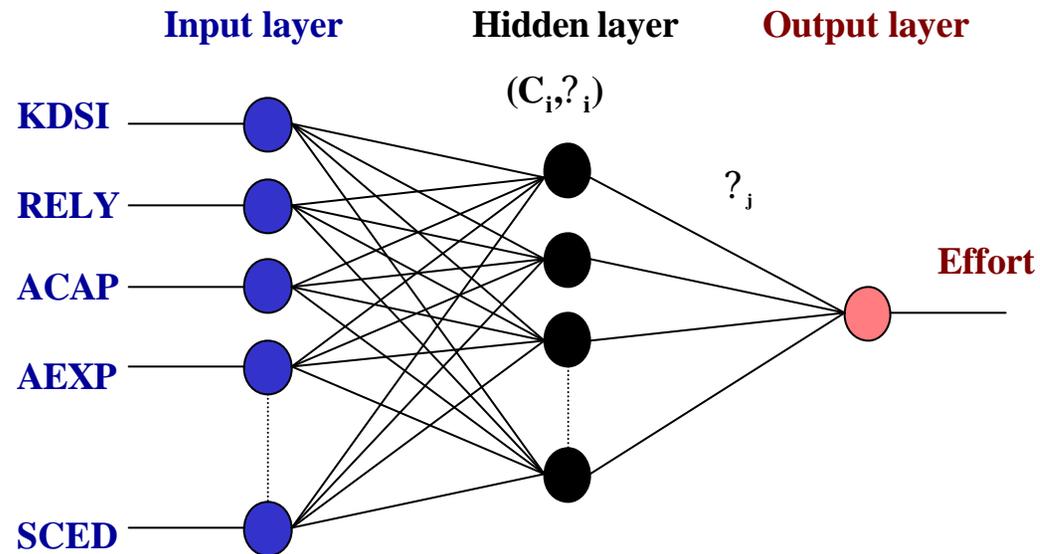
## ✍ Interprétation des règles floues obtenues

$r_k$  : if  $x_1$  is  $A_1^k$   $i$  ? or  $x_2$  is  $A_2^k$   $i$  ? or... $i$  ? or  $x_{N_x}$  is  $A_n^k$  then Effort ? CST

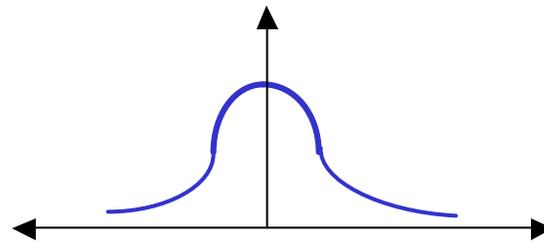
$$i \text{ ? or}(a_1, a_2, \dots, a_n) \text{ ? } \frac{a_1 a_2 \dots a_n}{(1 \text{ ? } a_1)(1 \text{ ? } a_2) \dots (1 \text{ ? } a_n) \text{ ? } a_1 a_2 \dots a_n}$$

- ✍ **i-or** n'est pas adéquat pour représenter la relation coût-facteurs
- ✍ **i-or** n'est pas facilement interprétable (comportement entre le **And** et le **Or**)
- ✍ Si on considère que le degré de vérité d'une seule règle floue est égal à 1 et que toutes les autres règles floues ont des degrés au voisinage de 0!!

✂ Radial Basis Function Network (Moody et Darken, 89)



$$f(x) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right)$$



# *Plan de la thèse*

---

- ✍ **Introduction**
- ✍ **Intelligence Artificielle**
- ✍ **Logique floue**
- ✍ **Estimation du coût de logiciels et Logique floue**
- ✍ **Estimation du coût par analogie et logique floue**
- ✍ **Gestion de l'incertitude dans l'estimation par l'analogie floue**
- ✍ **Apprentissage dans l'estimation par analogie floue**
- ✍ **Conclusions et Perspectives**

# *Questions?*