An Intelligent Approach to Software Cost Prediction

Xishi Huang, Danny Ho[1], Luiz F. Capretz, Jing Ren

Dept. of ECE, University of Western Ontario, London, Ontario, N6G 1H1, Canada
1Toronto Design Center, Motorola Canada Ltd., Markham, Ontario, L6G 1B3, Canada

{xhuang22, lcapretz, jren2}@uwo.ca, Danny.Ho@motorola.com

October 22, 2003

Agenda

- **Problem**
- The Novel Neuro-Fuzzy Model
- Validation by Industry Project Data
- Conclusion and Future Work

Software Cost Estimation

Motivation:

- Many software projects go over time and budget
- Software development has become an essential investment for many organizations

Problem Statement:

- We use project information such as software size and other attributes to predict software cost or effort

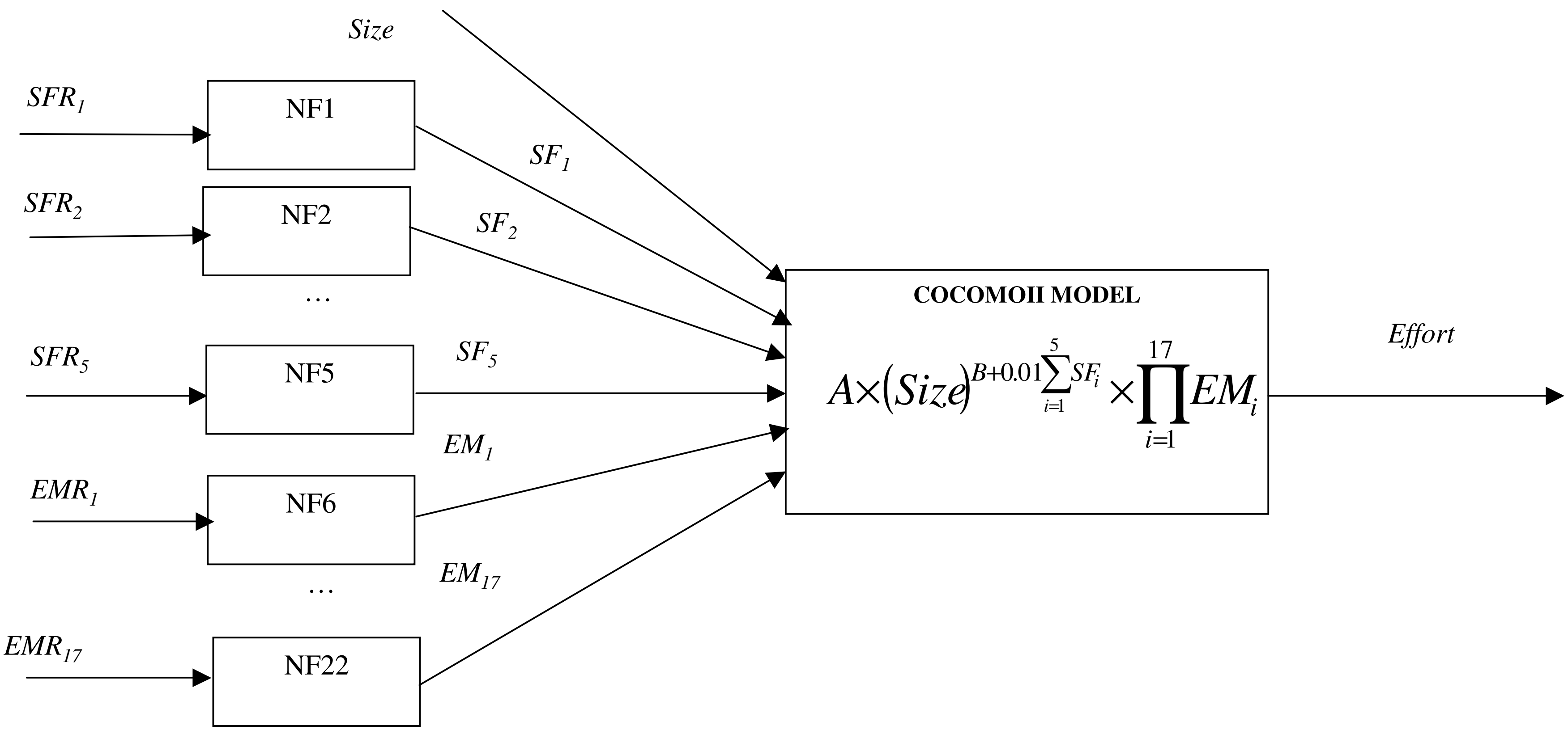
Characteristics of Software Cost Estimation:

- Complex nonlinear relationships between software development cost and cost drivers
- Imprecise and uncertain measurement
- Rapid change of software technology and processes

Agenda

- Problem
- **A Novel Neuro-Fuzzy Model**
- Validation by Industry Project Data
- Conclusion and Future Work

A Novel Neuro-Fuzzy COCOMO Model



COCOMO Model

- COCOMO II Post Architecture Model

$$Effort = A \times (Size)^{B+0.01\times\sum_{i=1}^{5} SF_i} \times \prod_{i=1}^{17} EM_i$$

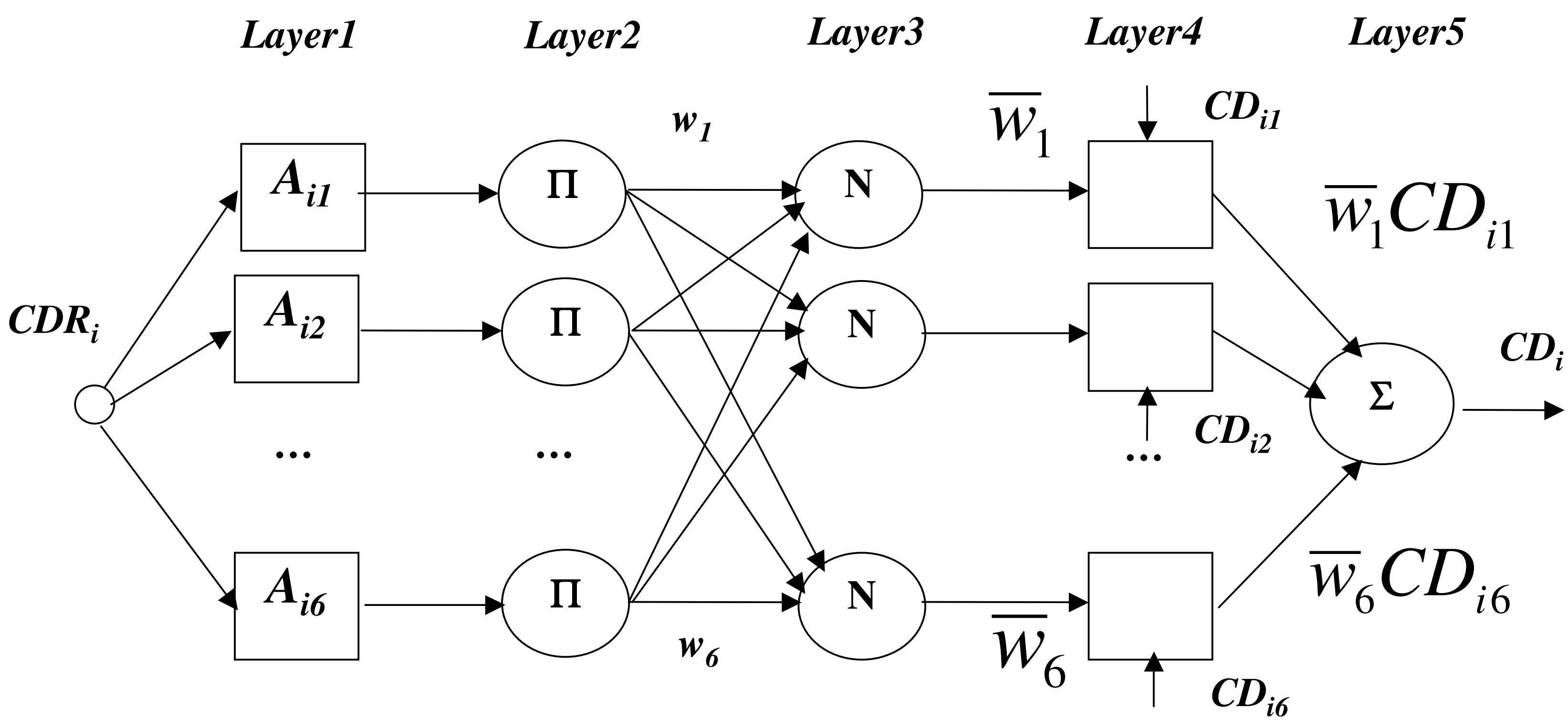
where:

- effort is in staff-month (SM)
- A,B are constants
- Size is in KSLOC
- SF_i's are scale factors
- EM_i's are effort multipliers

Some Notations

Fuzzy Sets		Rating Levels	Parameter Values
A_{i1}	1	Very Low (VL)	CD_{i1}
A_{i2}	2	Low (L)	CD_{i2}
A_{i3}	3	Nominal (N)	CD_{i3}
A_{i4}	4	High (H)	CD_{i4}
A_{i5}	5	Very High (VH)	CD_{i5}
A_{i6}	6	Extra High (XH)	CD_{i6}

Sub-Model NF_i



CDR_i: continuous rating value

CD_i: corresponding numerical value

Reasoning Phase: Fuzzy rules

- If CDR_i is A_{i1}(VL), then $CD_i = CD_{i1}$,
- If CDR_i is A_{i2}(L), then $CD_i = CD_{i2}$,
- If CDR_i is A_{i3}(N), then $CD_i = CD_{i3}$,
- If CDR_i is A_{i4}(H), then $CD_i = CD_{i4}$,
- If CDR_i is A_{i5}(VH), then $CD_i = CD_{i5}$,
- If CDR_i is A_{i6}(XH), then $CD_i = CD_{i6}$.

  (i,k), $i=1,2,\ldots,22$, $k=1,2,\ldots,6$

Sub-Model NF_i (Cont'd)

- Layer 1: $$O_k^1 = \mu_{ik}(x)$$
- Layer 2: $$w_k = \prod \mu_{ikj}(x)$$
- Layer 3: $$\overline{w}_k = \frac{w_k}{\sum w_j}$$
- Layer 4: $$O_k^4 = \overline{w}_k CD_{ik}$$
- Layer 5: $$CD_i = \sum_k \overline{w}_k CD_{ik}$$

Interpretation of sub-model NF_i

Linear interpolation:

$$\mu_{ik}(x)=\begin{cases} x-(k-1), & k-1\le x\le k \\ (k+1)-x, & k\le x\le k+1 \\ 0, & otherwise \end{cases}$$

$$CD_i = CD_{ik} + (CD_{ik+1} - CD_{ik})(CDR_i - k)$$
$$k \le CDR_i \le k+1, k = 1,2,\cdots,6$$

Learning Algorithms

Optimization problem:

$$E = \sum_{n=1}^{NN} \frac{1}{2} w_n \left(\frac{E_n - E_{dn}}{E_{dn}} \right)^2$$

subject to monotonic constraints:

$$CD_{i1} \le CD_{i2} \le CD_{i3} \le CD_{i4} \le CD_{i5} \le CD_{i6}, i \in I_{INC}(CD)$$

$$CD_{i1} \ge CD_{i2} \ge CD_{i3} \ge CD_{i4} \ge CD_{i5} \ge CD_{i6}, i \in I_{DEC}(CD)$$

Learning Algorithm:

$$CD_{ik}^{l+1} = CD_{ik}^{l} - \alpha \frac{\partial E}{\partial CD_{ik}}$$

Advantage of Continuous Rating Values

Two similar projects P1 and P2 have the same nominal effort, say 100 staff-months.

The COCOMO II.2000 model,

- For P1, 203 staff-months
- For P2, 2886 staff-months!

The difference is over 14 times.

Our Model: 809 staff-months for both projects

Agenda

- Problem
- A Novel Neuro-Fuzzy Model
- **Validation by Industry Project Data**
- Conclusion and Future Work

Validation by Industry Project Data

Sources of Project Data

- Industrial project data: 6 projects
- COCOMO81 database: 63 projects

Validation Results

- Case I. Learning with all project data
- Case II. Learning with part of project data
- Case III. Use larger weights for local data
- Case IV. Learning without monotonic constraints

Effort estimation for all 69 project data points

ARE	COCOMO81 Model	Neuro-Fuzzy Model							
		Case I		Case II		Case III		Case IV	
	PERC	PERC	IMPRV	PERC	IMPRV	PERC	IMPRV	PERC	IMPRV
20%	71%	86%	15%	88%	17%	88%	17%	89%	18%
30%	81%	92%	11%	92%	11%	92%	11%	94%	13%

Effort estimation for industrial project data

Project No	Actual Effort	COCOMO81 Model		Neuro-Fuzzy Model					
				Case I		Case II		Case III	
		Estimate	Error	Estimate	Error	Estimate	Error	Estimate	Error
P1	638.0	827.0	29%	745.0	16%	739.7	15%	728.6	14%
P2	185.0	152.2	-17%	167.3	-9%	166.1	-10%	163.6	-11%
P3	332.0	279.8	-15%	322.0	-3%	306.8	-7%	325.5	-1%
P4	619.9	701.4	13%	651.5	5%	651.7	5%	642.0	3%
P5	64.8	71.1	9%	63.4	-2%	60.8	-6%	64.1	-1%
P6	76.6	83.1	8%	72.2	-5%	72.3	-5%	73.8	-3%

Agenda

- Problem
- A Novel Neuro-Fuzzy Model
- Validation by Industry Project Data
- **Conclusion and Future Work**

Conclusion

- **Propose a novel neuro-fuzzy COCOMO model**
  - Neuro-fuzzy COCOMO structure
  - Monotonic constraints
  - Learning algorithm
  - Fuzzy rules

Conclusion (Cont’d)

Distinguishing Features of the Proposed Model:

- Learning ability
- Robust to imprecise and uncertain inputs
- Good Interpretability
- Knowledge integration
- Reduced number of learning parameters
- Good generalization
- Local learning

Future Work

Extending the neuro-fuzzy approach to

- Other cost estimation models, e.g. SLIM
- Quality estimation models, e.g. COQUALMO

References

1. B. Boehm et al., *Software Cost Estimation with COCOMO II*, Prentice Hall PTR, Upper Saddle River, New Jersey 07458, 2000.
2. B. Boehm, *Software Engineering Economics*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632, 1981.
3. S. MacDonell and A. Gray, "A comparison of modeling techniques for software development effort prediction," in *Proceedings of the 1997 International Conference on Neural Information Processing and Intelligent Information Systems*, Springer-Verlag, 1997, pp. 869–872.
4. S. Chulani, *Bayesian Analysis of Software Cost and Quality Models*, Ph.D. Dissertation, University of Southern California, 1999.
5. M. Shepperd and G. Kadoda, "Comparing software prediction techniques using simulation", *IEEE Transactions on Software Engineering*, Vol. 27, No. 11, pp. 1014-1022, November 1999.
6. S. Mitra, "Neuro-fuzzy rule generation: survey in soft computing framework," *IEEE Trans. on Neural Networks*, Vol. 11, pp. 748–768, May 2000.
7. R. Fuller. *Introduction to Neuro-Fuzzy Systems*, Physica-Verlag, Heidelberg, 2000.
8. R. J. S. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Trans. Systems, Man, and Cybernetics*, Vol. 23, pp. 665-685, 1993.
9. D. Ho, "Experience report on COCOMO and the Costar tool from Nortel's Toronto Laboratory," in *Eleventh International Forum on COCOMO and Software Cost Modeling*, University of Southern California, Los Angeles, October 1996.
10. N. Panlilio-Yap and D. Ho, "Deploying software estimation technology and tools: the IBM SWS Toronto Lab experience," in *Ninth International Forum on COCOMO and Software Cost Modeling*, University of Southern California, Los Angeles, October 1994.
11. T. Takaki and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control", *IEEE Trans. Syst., Man, Cybern.*, Vol. 15, pp.116-132, 1985.
12. X. Huang, *A Neuro-Fuzzy Model for Software Cost Estimation*, Master Thesis, Department of Electrical and Computer Engineering, The University of Western Ontario, July 2003.

THANKS !

Any Questions?