

Information Reuse and Integration
November 1-3, 2000

A FRAMEWORK FOR INTEGRATING ENSEMBLE CLASSIFIERS AND SCIENTIFIC DATABASE TECHNOLOGIES

by R.S. Renner

Dept. of Computer Science
California State University, Chico
Chico, CA 95929 USA



Copyright © 2000 R.S. Renner

ABSTRACT

A general framework is presented for the design and development of an integrated system of ensemble network generation, storage, analysis and retrieval.

The system is comprised of two independent components:

- 1) a simulation tool for generating ensembles of neural network classifiers, and,
- 2) a scientific database for storing classifiers, predictions, trial conditions, performance-related data, and meta-data.

A model implementation is discussed.

Copyright © 2000 R.S. Renner

NNES SciDB

A Database to support *NNES* scientific data

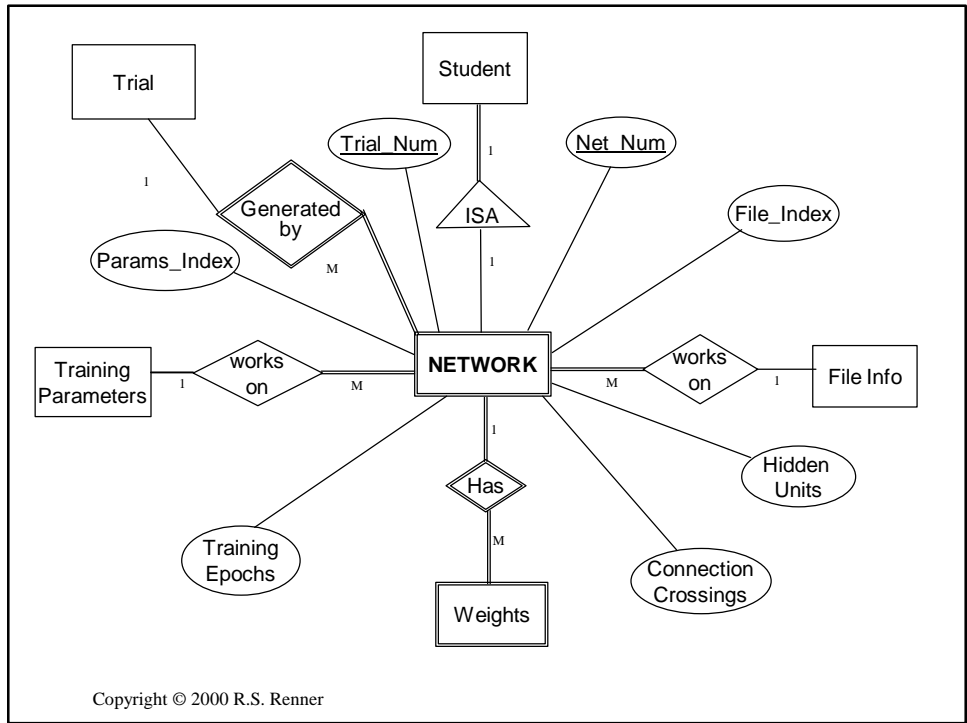
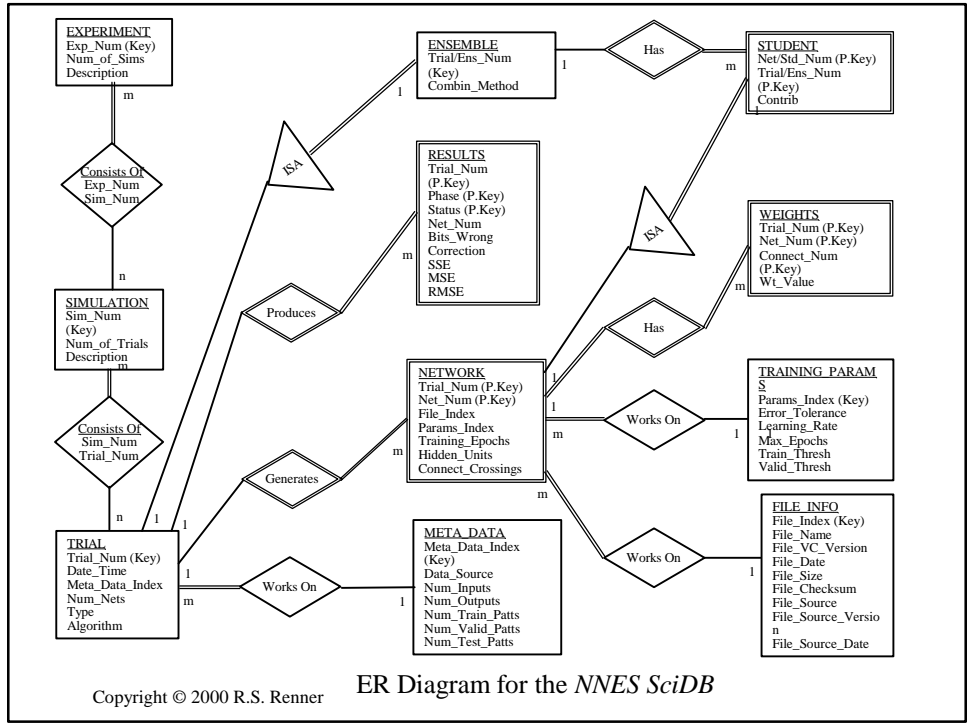
- storage
 - trials, ensembles, networks, students
 - results, file-info, meta-data, training parameters, weights
- analysis
 - experiments, simulations, trials, ensembles,
- retrieval
 - best, worst, average results; comparisons
 - error, training time, parameters, algorithms

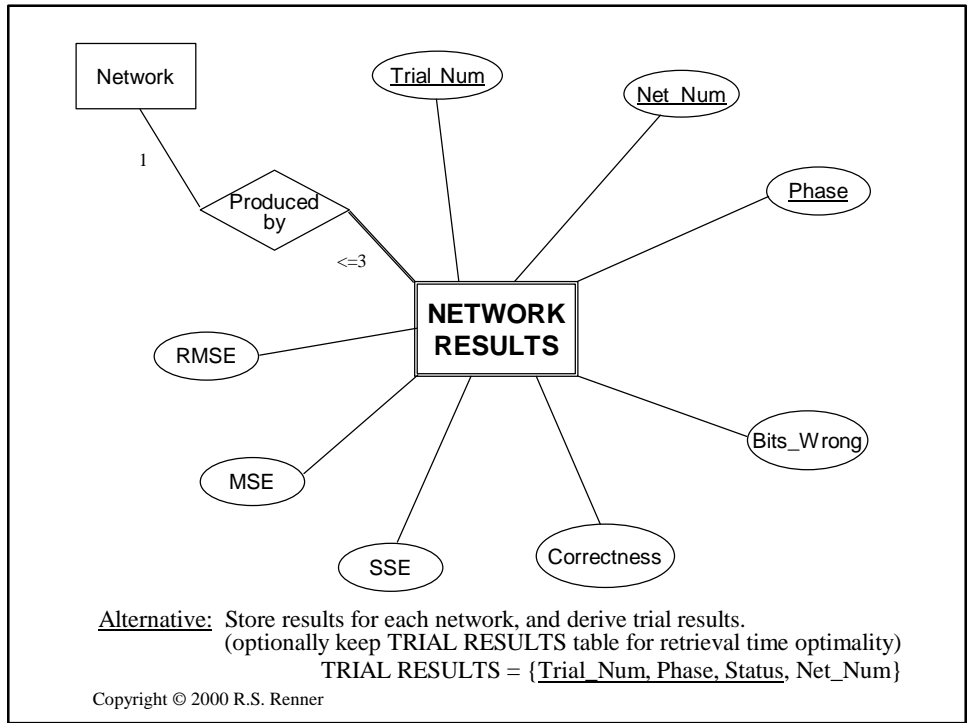
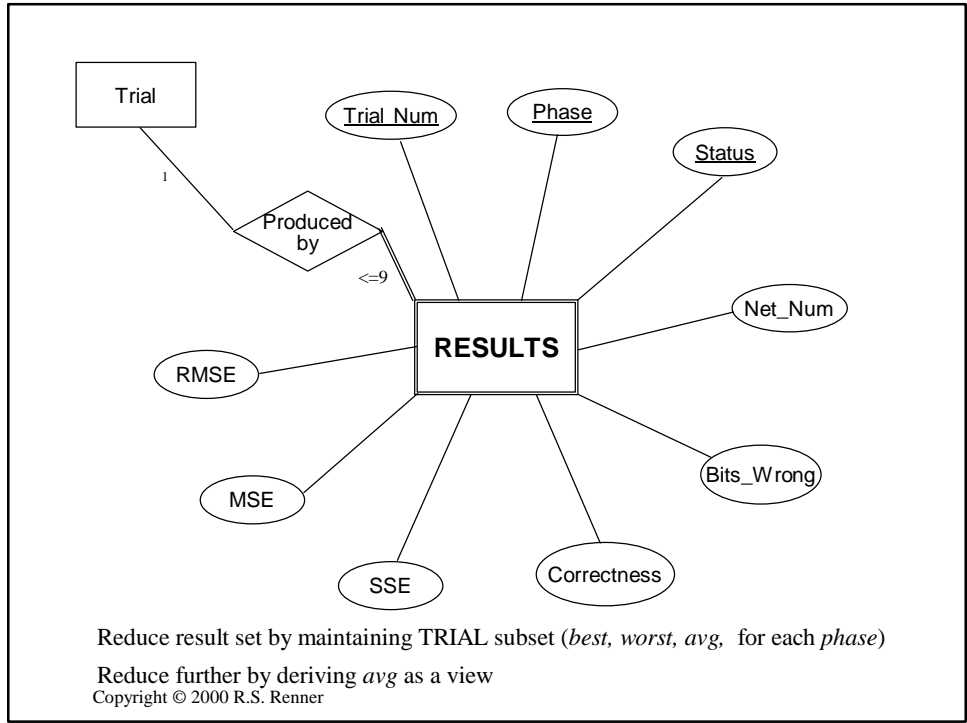
Copyright © 2000 R.S. Renner

User Requirements

- compare networks, students, and ensembles
- view parameter settings for any particular network
- query the networks based on training algorithm
- query the networks based on combination algorithm
- determine network rankings & criteria
 - (best, worst or average)
- associate training data with particular net or ensemble
- specify criteria for compiling experiments and simulations, from existing trials

Copyright © 2000 R.S. Renner





NNES *SciDB* Implementation

- Storage and retrieval - Oracle 8 / SQL
- Data acquisition from NNES - Perl scripts
- ‘Experiments’ & ‘Simulations’
 - defined dynamically in *NNES SciDB* as needed
 - implemented as views
 - based on keyword(s) search of existing trials/ensembles
- Prototype stage

Copyright © 2000 R.S. Renner

```
-- generate the TRIAL table --
CREATE TABLE TRIAL
(
  Trial_Num          INTEGER
  Date_Time         DATE
  Meta_Data_Index   INTEGER
  Num_Nets          INTEGER
  Type              VARCHAR2(10)
  Algorithm         VARCHAR2(10)
  PRIMARY KEY (Trial_Num),
  FOREIGN KEY (Meta_Data_Index)
  REFERENCES META_DATA );

-- after script processing simulator outfiles
-- loop for each trial & increment trialnum
INSERT INTO TRIAL VALUES
  (trialnum, sysdate, meta_index, numnets,
   type, algo);

-- sample query to get CCEns training results
CREATE VIEW CCEns_Results AS
SELECT
  Sim_Num, Trial_Num, Status, Training_Epochs,
  Hidden_Units, Bits_Wrong, Correction, MSE
FROM
  SIM_TRIAL, TRIAL, RESULTS, NETWORK
WHERE
  SIM_TRIAL.Trial_Num=TRIAL.Trial_Num AND
  TRIAL.Trial_Num=Result.Trial_Num AND
  TRIAL.Trial_Num=NETWORK.Trial_Num AND
  NETWORK.Net_Num=RESULTS.Net_Num AND
  TRIAL.Algorithm='Cascor' AND
  TRIAL.Type='Ensemble' AND
  RESULTS.Phase='Training';
```

Copyright © 2000 R.S. Renner

Upgrades to *NNES SciDB*

- Porting of *SciDB* stored networks back into *NNES*
- Autonomous integration
- Global GUI interface / Web-based interface
- enhanced query set (detailed analysis)

General Framework

- integration of heterogeneous systems
- integration of distributed systems

Copyright © 2000 R.S. Renner

Abbreviated References

- P. Sollich and A. Krogh, "Learning with ensembles: How over-fitting can be useful," *NIPS* 8, 1996.
- L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE TPAMI*, 1990.
- C. Ji and S. Ma, "Combinations of weak classifiers," *IEEE TNN*, vol. 8, 1997.
- T. Heskes, "Balancing between bagging and bumping," *NIPS* 9, 1997.
- A. Krogh & J. Vedelsby, "Neural network ensembles, cross validation, ...," *NIPS* 7, 1995.
- S. Hashem & B. Schmeiser, "Improving model accuracy ... linear combinations ...," *IEEE TNN*, 1995.
- H. Drucker, R. Schapire, & P. Simard, "Boosting performance in neural networks," *IJPRAI*, 1993.
- D. H. Wolpert, "Stacked generalization," *Neural Networks*, 1992.
- M. P. Perrone & L. N. Cooper, "When networks disagree: ...," in *NNNSVNCs*, 1993.
- R. S. Renner, "Improving Generalization of Constructive Neural Networks ...," FSU 1999.
- R. S. Renner, et.al., "Combining Constructive Neural Networks for Ensemble ...," *JCIS*, 2000.
- R. S. Renner, "Ensemble Systems of Cascor Networks Demonstrate Superiority over ...," *ICAI*, 2000.
- L. Prechelt, "PROBEN1 Benchmark Set," Universitat Karlsruhe, 1998.
- C. E. Rasmussen, et.al., "Delve," University of Toronto, 1997.
- R. S. Renner, et.al., "A Simulation Tool for Managing Intelligent Ensembles," *ICAI*, 1999.

Copyright © 2000 R.S. Renner