Combining Multiple Expert Systems using Combinatorial Fusion Analysis

D. Frank Hsu, Ph.D.
Clavius Distinguished Professor
hsu (at) cis (dot) fordham (dot) edu
Fordham University, New York, NY 10023
(Joint work with Christina Schweikert and Roger Tsai)

Oct 24-25, 2011
DIMACS Workshop on Science of Expert Opinions
Rutgers University
New Brunswick, NJ, USA

Outline

(A) Information Fusion

(B) Combinatorial Fusion Analysis

(C) Multiple Expert Systems Applications

(D) Remarks and Acknowledgement.

(A) Information Fusion

(A) Information Fusion

Pointers for IF:

* Complexity: Multiple sensors, multiple sources, multiple systems.

* Levels: Data fusion, Feature fusion, Decision fusion.

* Computing, Informatics, and Analytics:

  Data-Information-Knowledge-Wisdom-Enlightenment.

FAQ's for IF:

* What: Combination of data or information from multiple sensors, sources, features, systems, cues, classifiers, or decisions.

* Why: To improve the quality (better accuracy and higher effectiveness) of data, feature characteristics, decisions and actions.

* When: To Fuse or Not To Fuse.

* How: A diverse array of combination methods.

Crossing the Street

Figure Skating Judgment

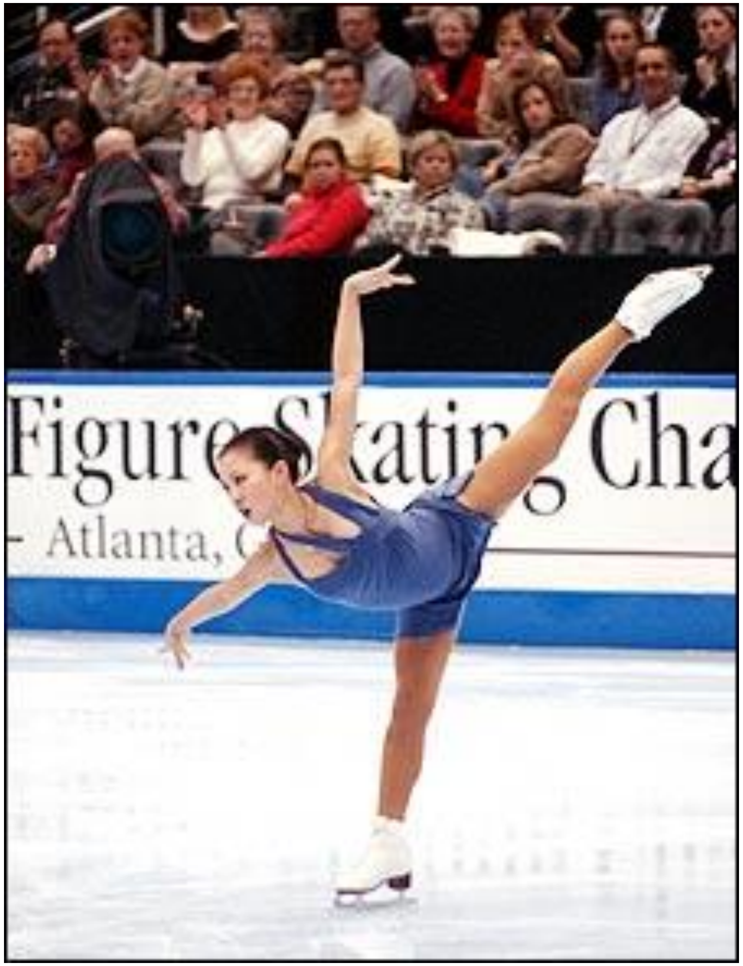


Figure Skating Judgment

	J1	J2	J3	SC	D		J1	J2	J3	RC	C
d_1	9.6	9.7	9.8	29.1	2		5	3	3	11	3
d_2	9.8	9.2	9.9	28.9	3		3	8	2	13	4
d_3	9.7	9.9	10	29.6	1		4	2	1	7	1
d_4	9.5	9.3	9.7	28.5	6		6	7	4	17	7
d_5	9.9	9.4	9.5	28.8	4		2	6	6	14	5
d_6	9.4	9.6	9.6	28.6	5		7	4	5	16	6
d_7	9.3	9.5	9.4	28.2	7		8	5	7	20	8
d_8	10	10	7	27	8		1	1	8	10	2

Internet Search Strategy

Internet Search Strategy

	A		B		C Rank Comb		D Score Comb	
d_1	1.00	1	0.80	2	1.5	1	0.90	1
d_2	0.40	7	1.00	1	4.0	4	0.70	3
d_3	0.70	4	0.35	5	4.5	5	0.525	5
d_4	0.90	2	0.60	3	2.5	2	0.75	2
d_5	0.80	3	0.40	4	3.5	3	0.60	4
d_6	0.60	5	0.25	7	6.0	6	0.425	6
d_7	0.20	9	0.30	6	7.5	8	0.25	8
d_8	0.50	6	0.20	8	7.0	7	0.35	7
d_9	0.30	8	0.10	10	9.0	9	0.20	9
d_{10}	0.10	10	0.15	9	9.5	10	0.125	10

(B) Combinatorial Fusion Analysis (CFA)

(1) Multiple Scoring Systems and RSC Functions

(2) Applications

(a) Science and Technology:

Target Tracking and Computer Vision

(b) Biomedical Informatics and Pharmacogenomics:

Virtual Screening and Drug Discovery

(c) Information Retrieval:

Biomedical Literature Collections

(d) Information Retrieval:

Search Engine Optimization

(e) On-line Learning

(f) Classifier Ensemble

(1) Multiple Scoring Systems (MSS) and RSC Functions

- Score function, rank function, and rank/score function of system A.

  s(A)

  s(A) → r(A), sorting

  s(A), r(A) → f(A) ?

- Score combination and rank combination

  Scoring Systems **A**, **B**: $\mathbf{Com_s(A,B) = C}$, $\mathbf{Com_r(A,B) = D}$

- Performance evaluation (criteria)

- Diversity measure: Diversity between **A** and **B**, **d(A, B)**, is equal to **d(s(A)**, **s(B))** or **d(r(A)**, **r(B))**, or **d(f(A), f(B))**?

- Two main questions:

  (1) When are **P(C)** or **P(D)** greater than or equal to **P(A)** and **P(B)**?

  (2) When is **P(D)** greater or equal to **P(C)**?

Ref: Hsu, D.F., Chung, Y.S., and Kristal, B.S. Combinatorial fusion analysis: methods and practice of combining multiple scoring systems, in: H.H. Hsu (Ed.), Advanced Data Mining Technologies in Bioinformatics, Idea Group Inc., (2006), pp. 32-62.

The Rank Score Characteristic Function

D= set of classes, documents, forecasts, price ranges with $|D| = n$.

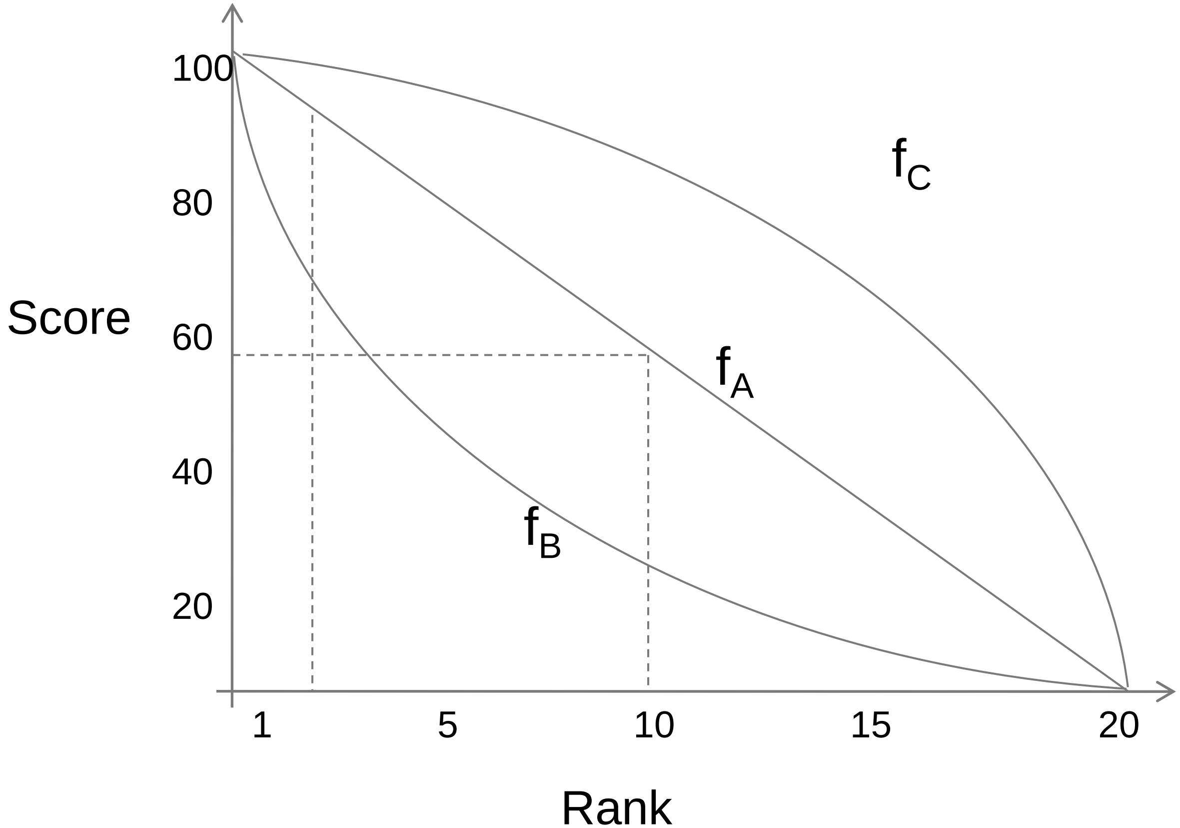
N= the set $\{1,2,\ldots,n\}$

R= a set of real numbers

$$f(i)=(s \circ r^{-1})(i)$$

$$=s(r^{-1}(i))$$

The RSC Function



Three RSC functions: f_A, f_B and f_C

Cognitive Diversity between A and B = $d(f_A, f_B)$

The RSC Function

D	Score function s:D→R	Rank function r:D→N		RSC function f:N→R	
d_1	3	10		1	10
d_2	8.2	3		2	9.8
d_3	7	4		3	8.2
d_4	4.6	7		4	7
d_5	4	8		5	5.4
d_6	10	1		6	5
d_7	9.8	2		7	4.6
d_8	3.3	9		8	4
d_9	1	12		9	3.3
d_{10}	2.5	11		10	3
d_{11}	5	6		11	2.5
d_{12}	5.4	5		12	1

How do we compute the RSC function?

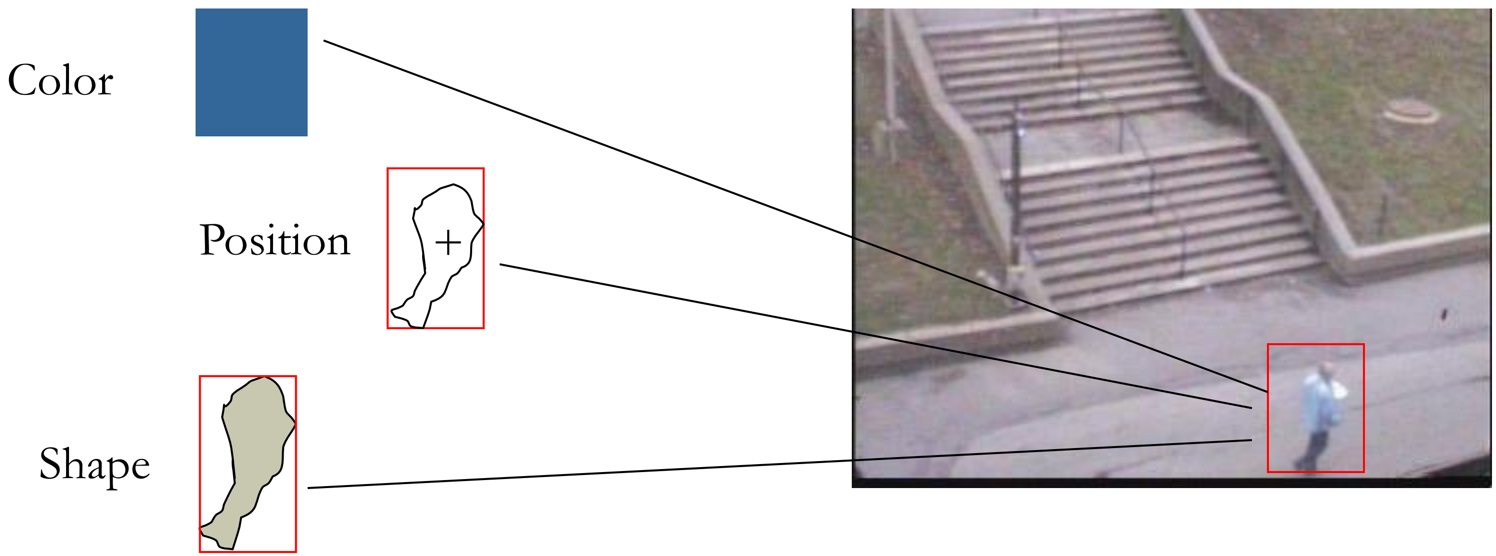
Sorting the score value by using its rank value as the key.

(2) Applications

(a) Science and Technology: Target Tracking and Computer Vision

We use three features:

- Color – average normalized RGB color.
- Position – location of the target region centroid
- Shape – area of the target region.



Ref: Lyons, D.M., Hsu, D.F. Information Fusion 10(2): 124-136 (2009).

(a) Science and Technology: Target Tracking and Computer Vision

Experimental Results

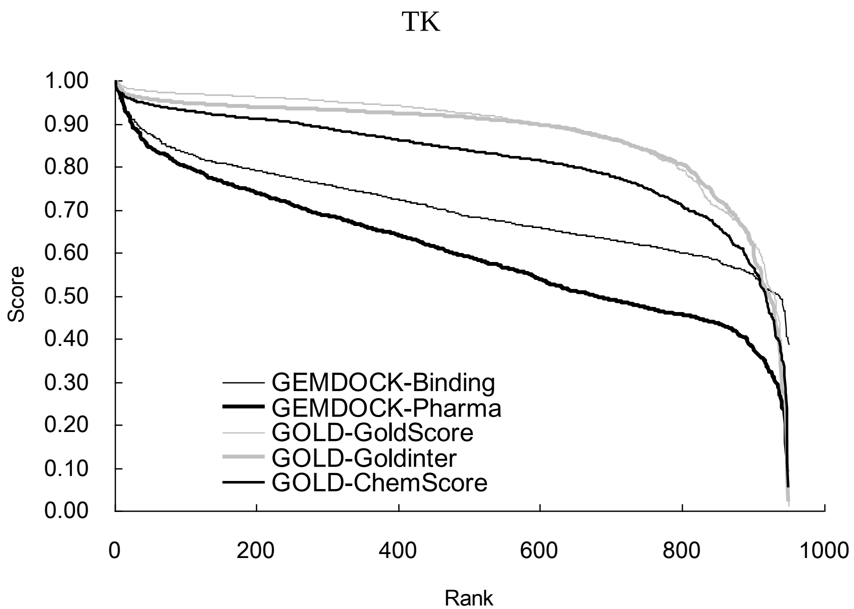
Seq.	RUN2 Score fusion **MSSD Avg.**	**MSSD Var.**	RUN3 Score and rank fusion using ground truth to select **MSSD Avg.**	**MSSD Var.**	RUN4 Score and rank fusion using rank-score function to select **MSSD Avg.**	**MSSD Var.**
1	1537.22	694.47	1536.65	695.49	1536.9	694.24
2	816.53	8732.13	723.13	3512.19	723.09	3511.41
3	108.89	61.61	108.34	60.58	108.89	61.61
4	23.14	2.39	23.04	2.30	23.14	2.39
5	334.13	120.11	332.89	119.39	334.138	120.11
6	96.40	119.22	66.9	12.91	67.28	13.38
7	577.78	201.29	548.6	127.78	577.78	201.29
8	538.35	605.84	500.9	57.91	534.3	602.85
9	143.04	339.73	140.18	297.07	142.33	294.94
10	260.24	86.65	252.17	84.99	258.64	85.94
11	520.13	2991.17	440.98	2544.69	470.27	2791.62
12	1188.81	745.01	1188.81	745.01	1188.81	745.01

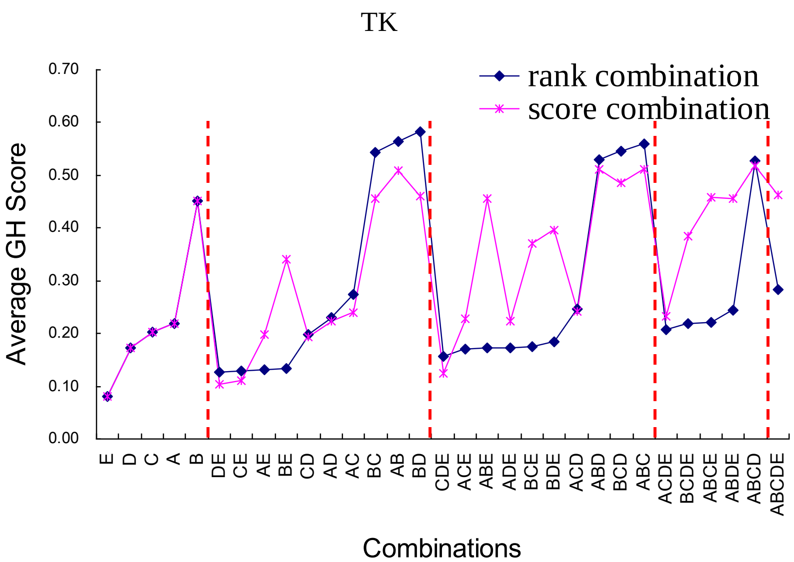
- RUN4 is as good or better (highlighted in gray) than RUN2 in all cases
- RUN4 is, predictably, not always as good as RUN3 ('best case').

Note: Lower MSSD implies better tracking performance.

(b) Biomedical Informatics and Pharmacogenomics: Virtual Screening

The Performance of Thymidine Kinase (TK)





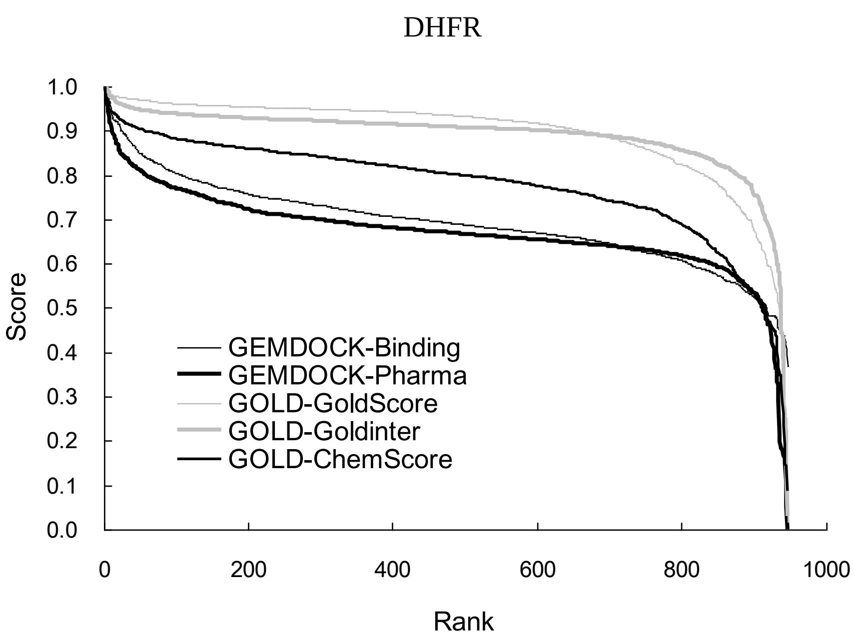
•Combinations of different methods improve the performances

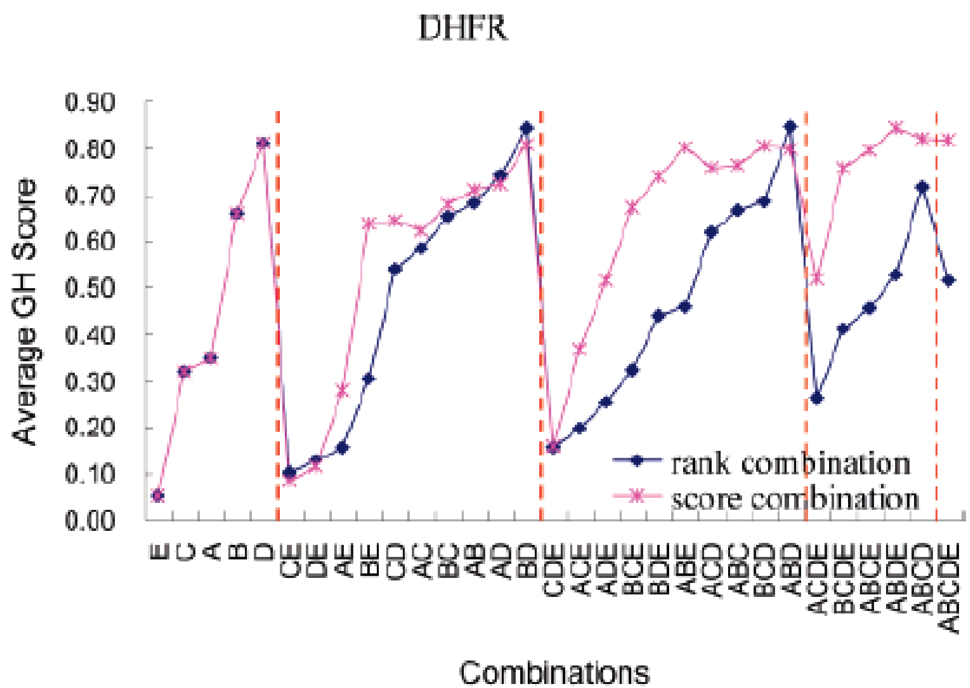
•The combination of B and D works best on thymidine kinase (TK)

Ref: Yang et al. Journal of Chemical Information and Modeling. 45 (2005) 1134-1146.

(b) Biomedical Informatics and Pharmacogenomics: Virtual Screening

The Performance of Dihydrofolate Reductase (DHFR)

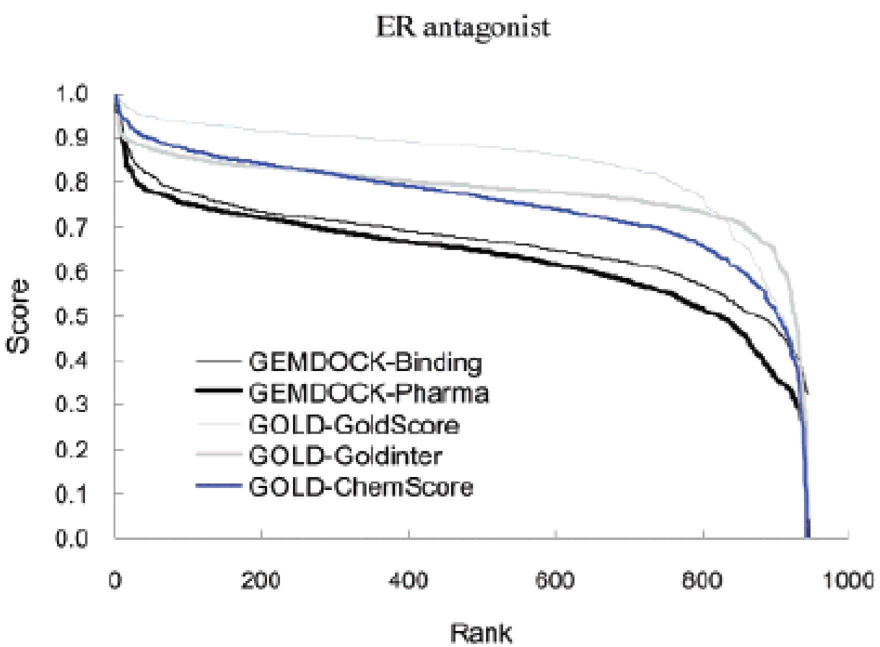


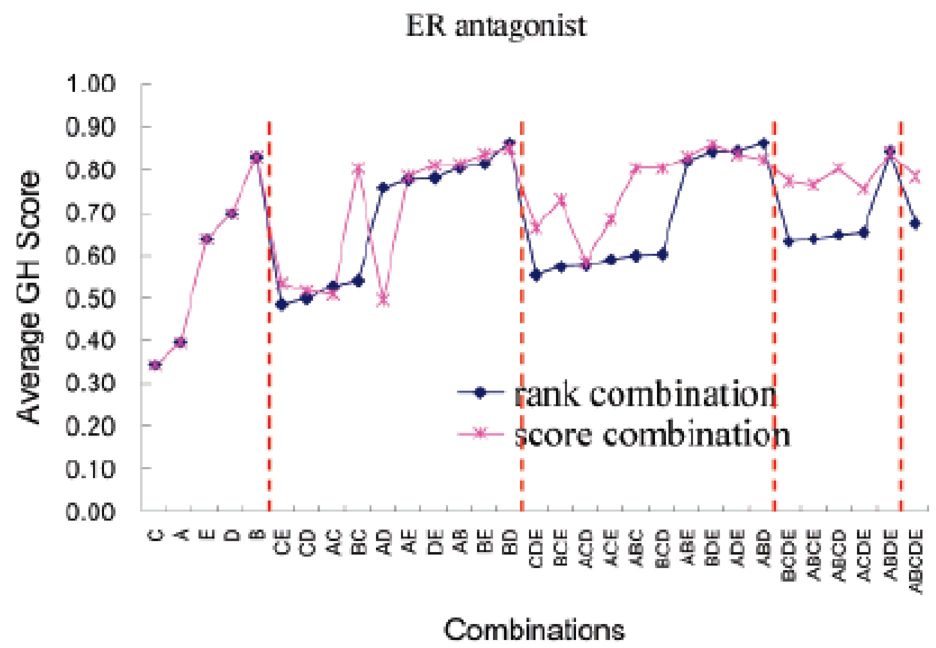


- Combinations of different methods improve the performances
- The combination of B and D works best on dihydrofolate reductase (DHFR)

(b) Biomedical Informatics and Pharmacogenomics: Virtual Screening

The Performance of ER-Antagonist Receptor (ER)

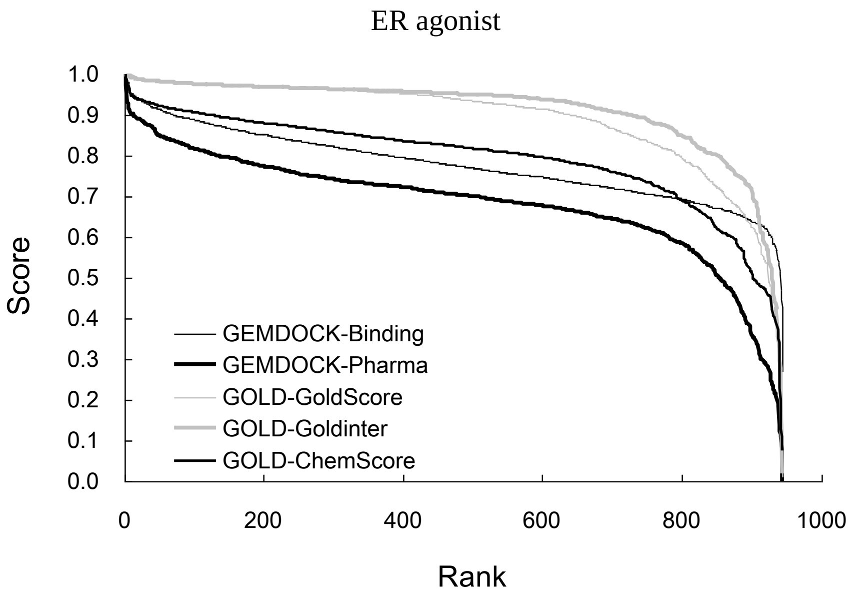


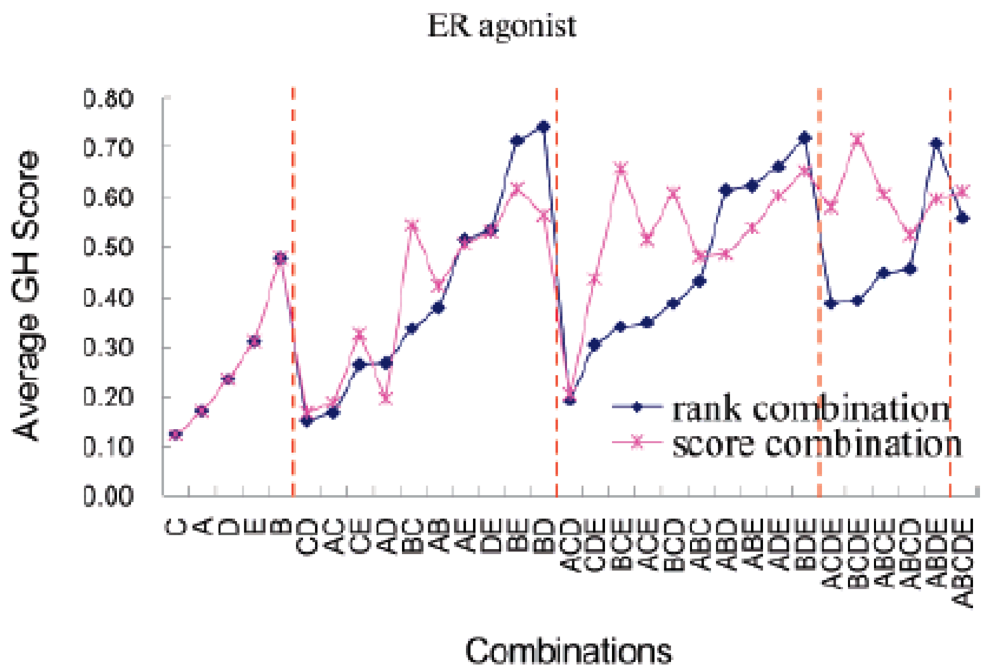


- Combinations of different methods improve the performances
- The combination of B and D works best on ER-antagonist receptor (ER)

(b) Biomedical Informatics and Pharmacogenomics: Virtual Screening

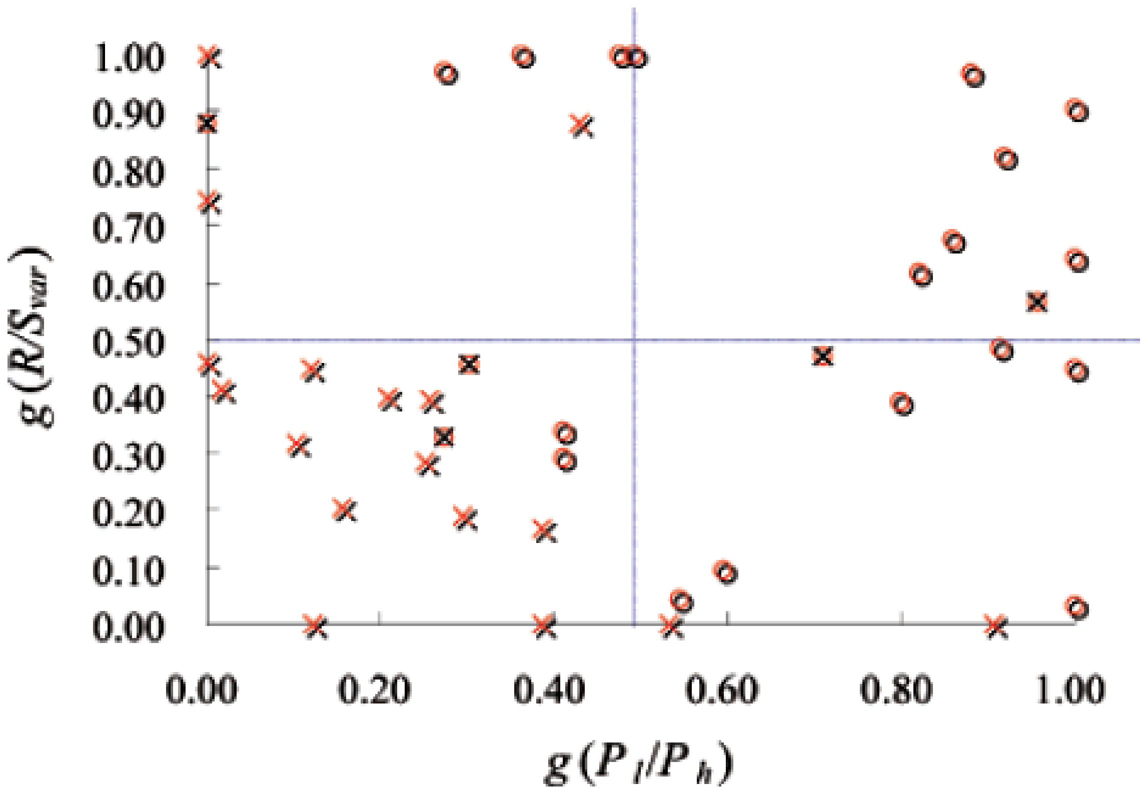
The Performance of ER-Agonist Receptor (ERA)



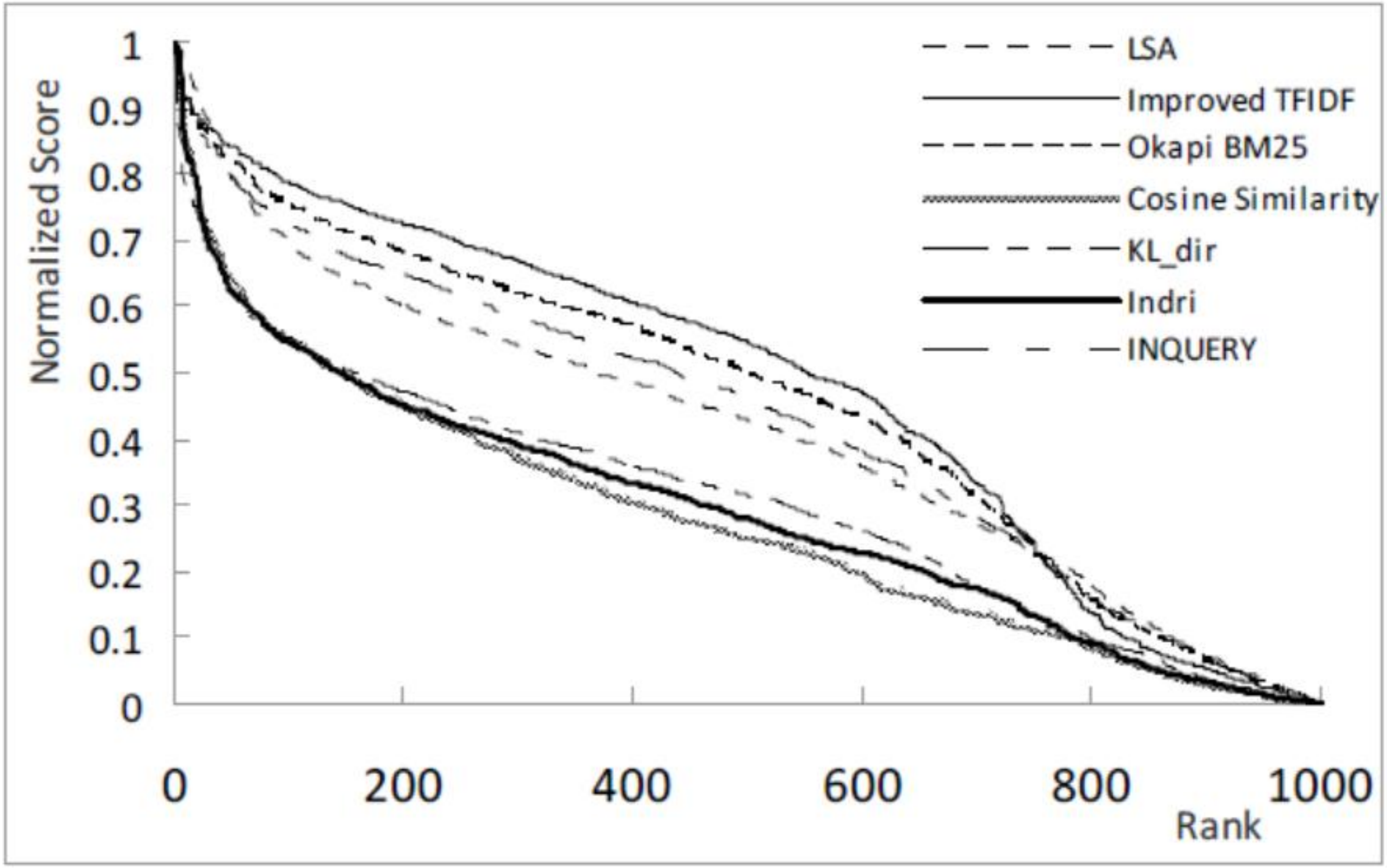


- Combinations of different methods improve the performances
- The combination of B and D works best on ER-agonist receptor (ERA)

(b) Biomedical Informatics and Pharmacogenomics: Virtual Screening

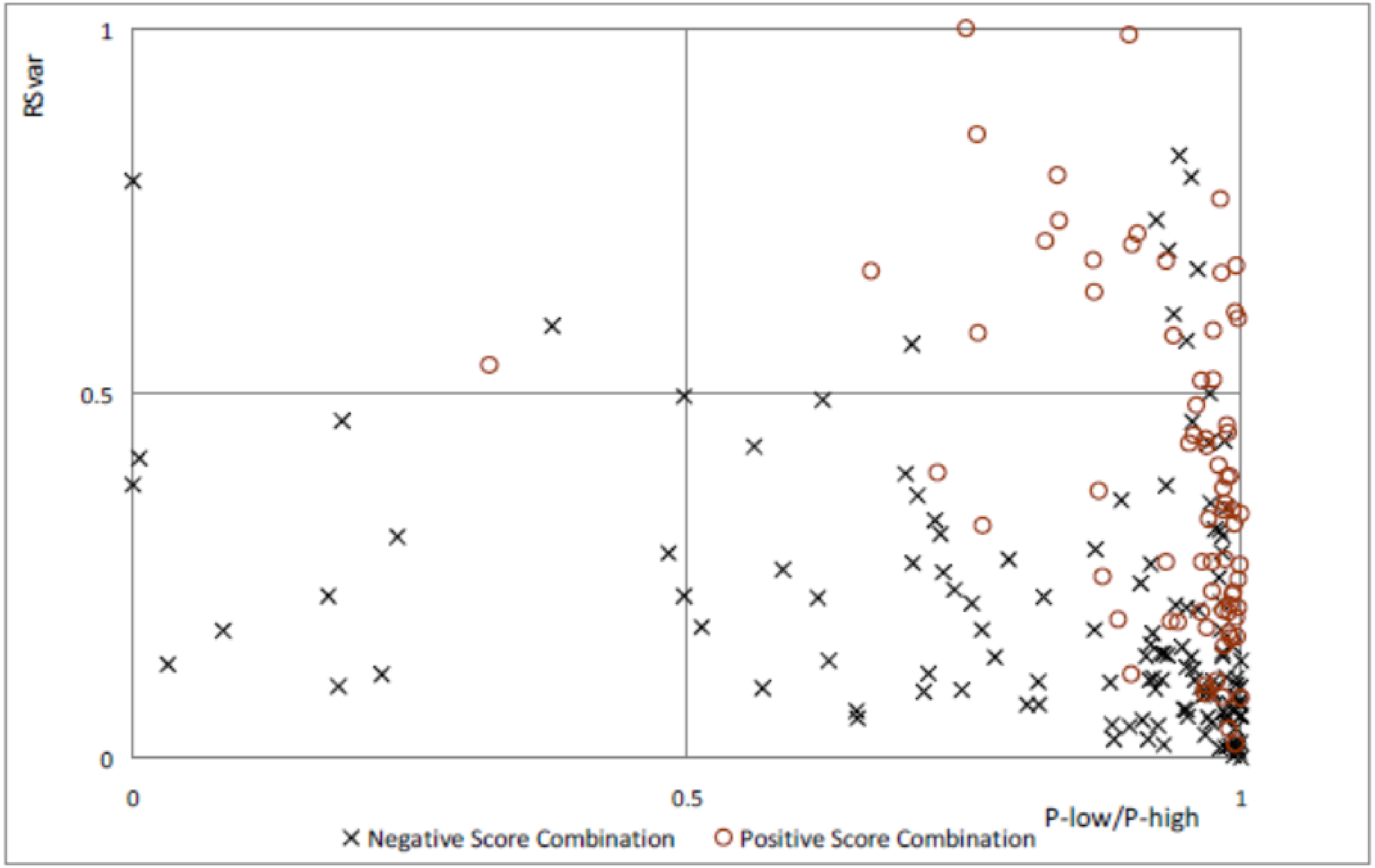


(c) Information Retrieval: Biomedical Literature Collections



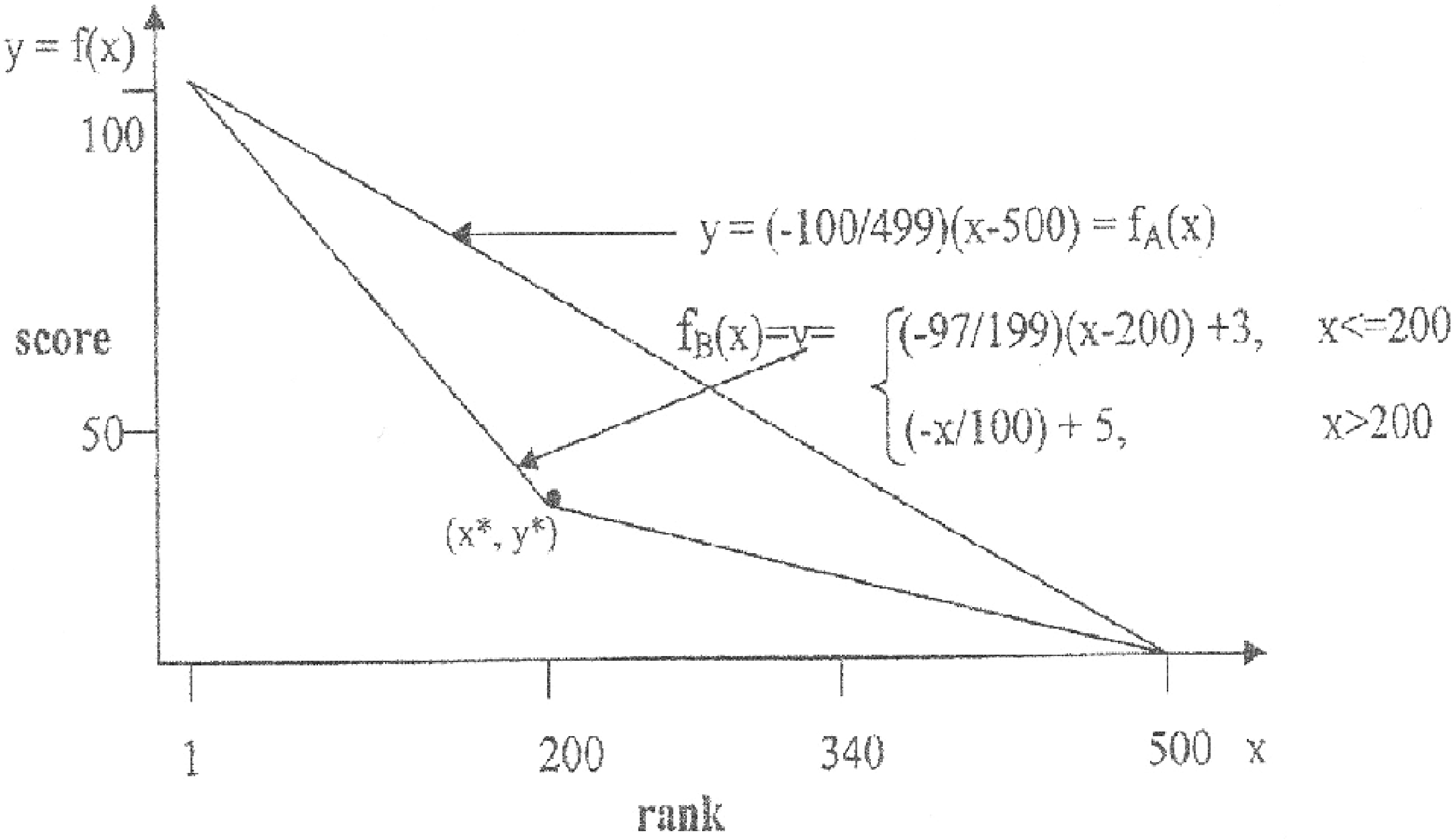
Rank-Score Characteristic Graphs of Seven IR Models

(c) Information Retrieval: Biomedical Literature Collections



RSvar vs. Performance Ratio

(d) Information Retrieval: Search Engine Optimization



Ref: Hsu, D.F., Taksa, I. Information Retrieval 8(3) (2005) 449–480

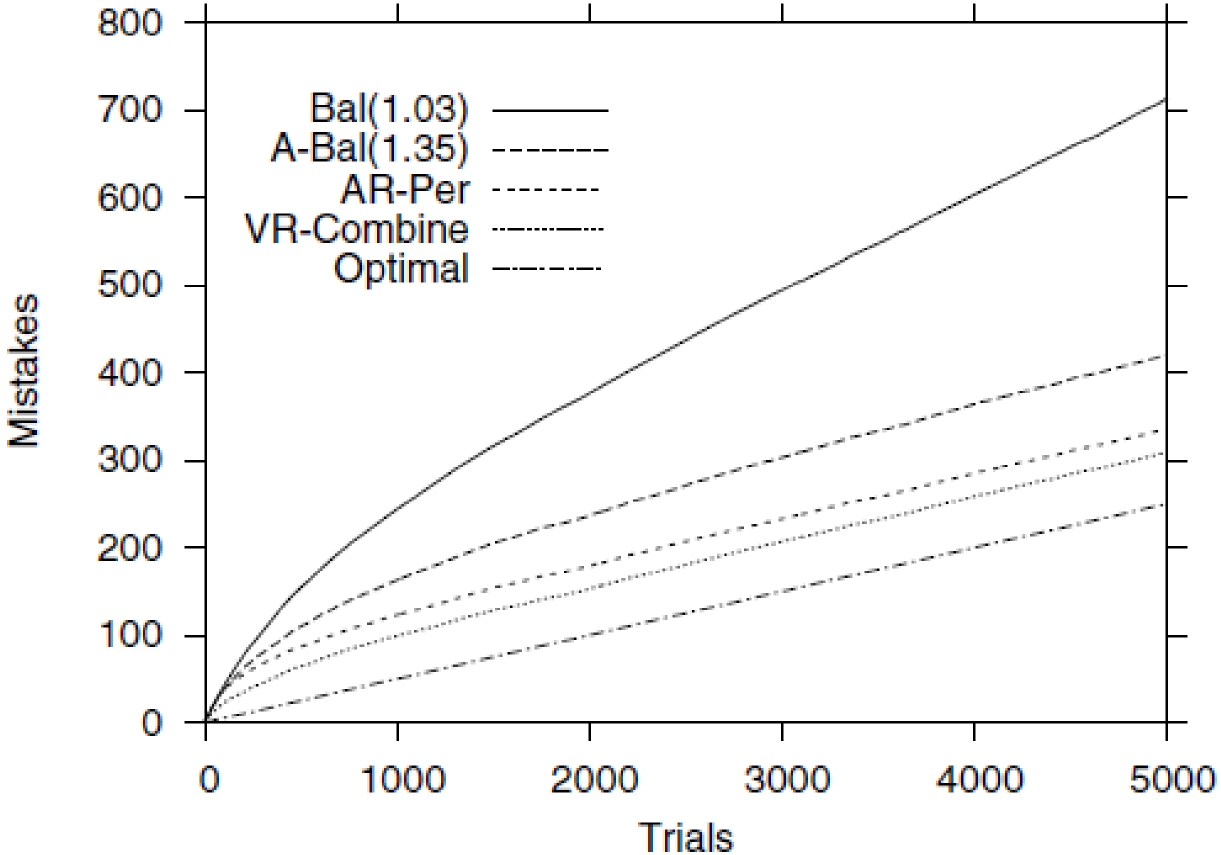
(e) On-line Learning

GOAL: The goal is to learn a linear combination of the classifier predictions that maximizes the accuracy on future instances.

* Sub-expert conversion
* Hypothesis voting
* Instance recycling

Ref: Mesterharm, C., Hsu, D.F. The 11th International Conference on Information Fusion, 2008. pp. 1117-1124

(e) On-line Learning



Mistake curves on majority learning problem with *r = 10, k = 5, n = 20, and p = .05*

(f) Classifier Ensemble

In regression, Krogh and Vedelsby (1995):

Ensemble generalization error: $E = \overline{E} - \overline{A}$

Weighted average of generalization errors$(\overline{E} = \sum_\alpha w_\alpha E^\alpha)$

Weighted average of ambiguities: $(\overline{A} = \sum_\alpha w_\alpha A^\alpha)$

In classification, Chung, Hsu, and Tang (2007):

$$\max\{\bar{P} - \bar{D}, p(\bar{P} + \bar{D}) + 1 - p\} \leq P^{\mathrm{m}} \leq \min\{\bar{P} + \bar{D}, p(\bar{P} - \bar{D})\}$$

Ref: Chung et al in Proceedings of 7th International Workshop on Multiple Classifier Systems (MCS2007), LNCS, Springer Verlag.

(f) Classifier Ensemble

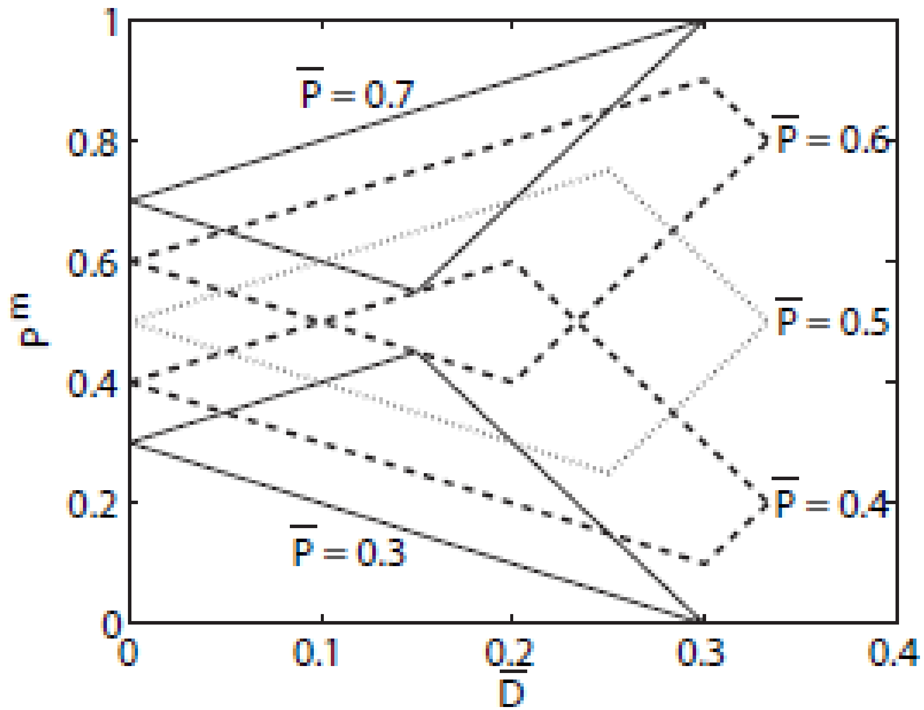


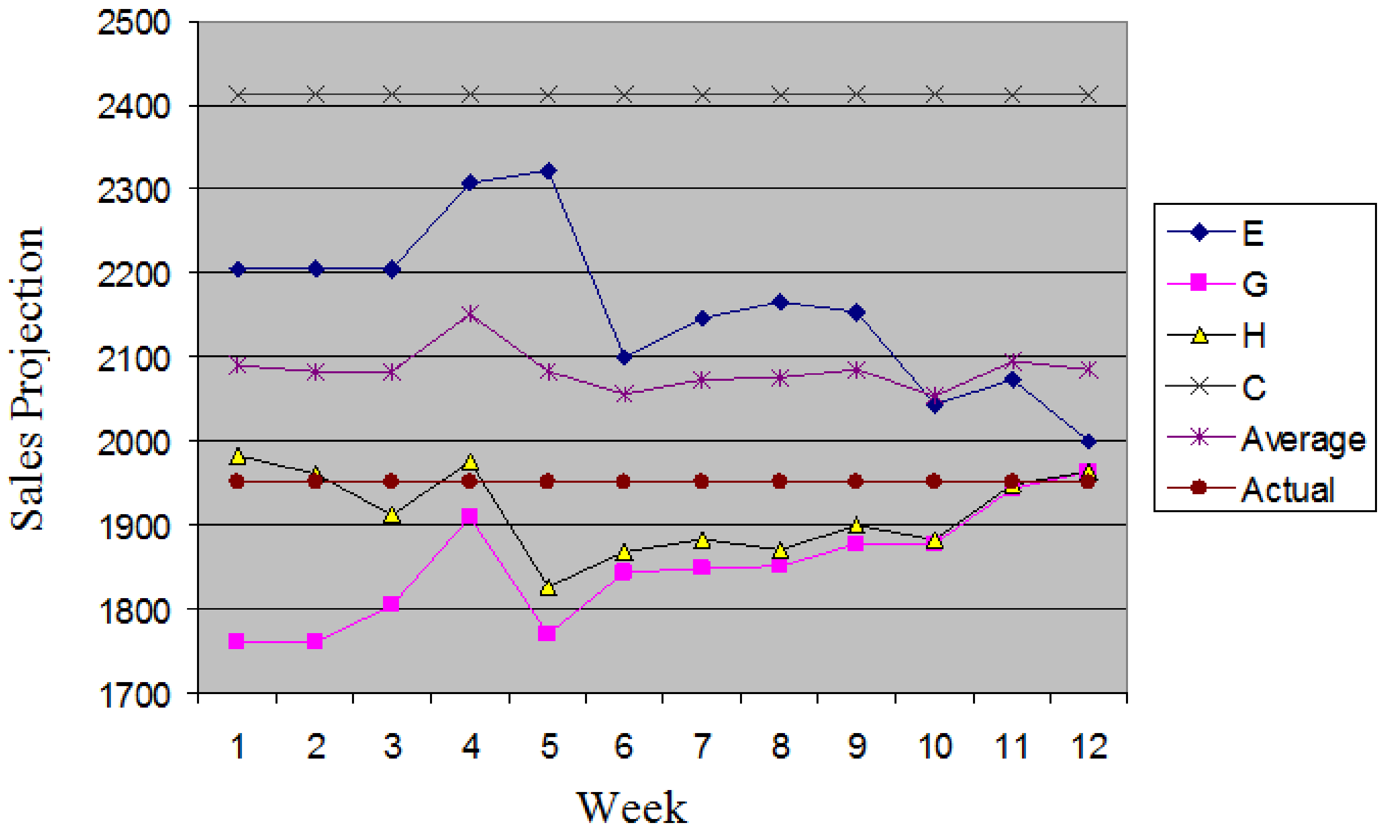
Fig. 1. Bounds for majority voting accuracy. Here $p = 3$. The dashed and dotted lines are for $\bar{P} \in (\frac{\ell}{p}, \frac{\ell+1}{p})$, while the solid lines are for $\bar{P} \notin (\frac{\ell}{p}, \frac{\ell+1}{p})$. The bounds have different spans in $\bar{D}$ since $\bar{D} \leq \min\{\bar{P}, 1 - \bar{P}, \frac{\ell}{p}\}$.

(C) Multiple Expert Systems Applications

Ref: Tsai, R., Schweikert, C., Yu, S., Hsu, D.F. Combining Multiple Forecasting Experts for Corporate Revenue Using Combinatorial Fusion Analysis. Global Business & Technology Association's Thirteenth Annual International Conference (GBATA 2011), "Fulfilling the Worldwide Sustainability Challenge: Strategies, Innovations, and Perspectives for Forward Momentum in Turbulent Times", 2011, pp. 986-995.

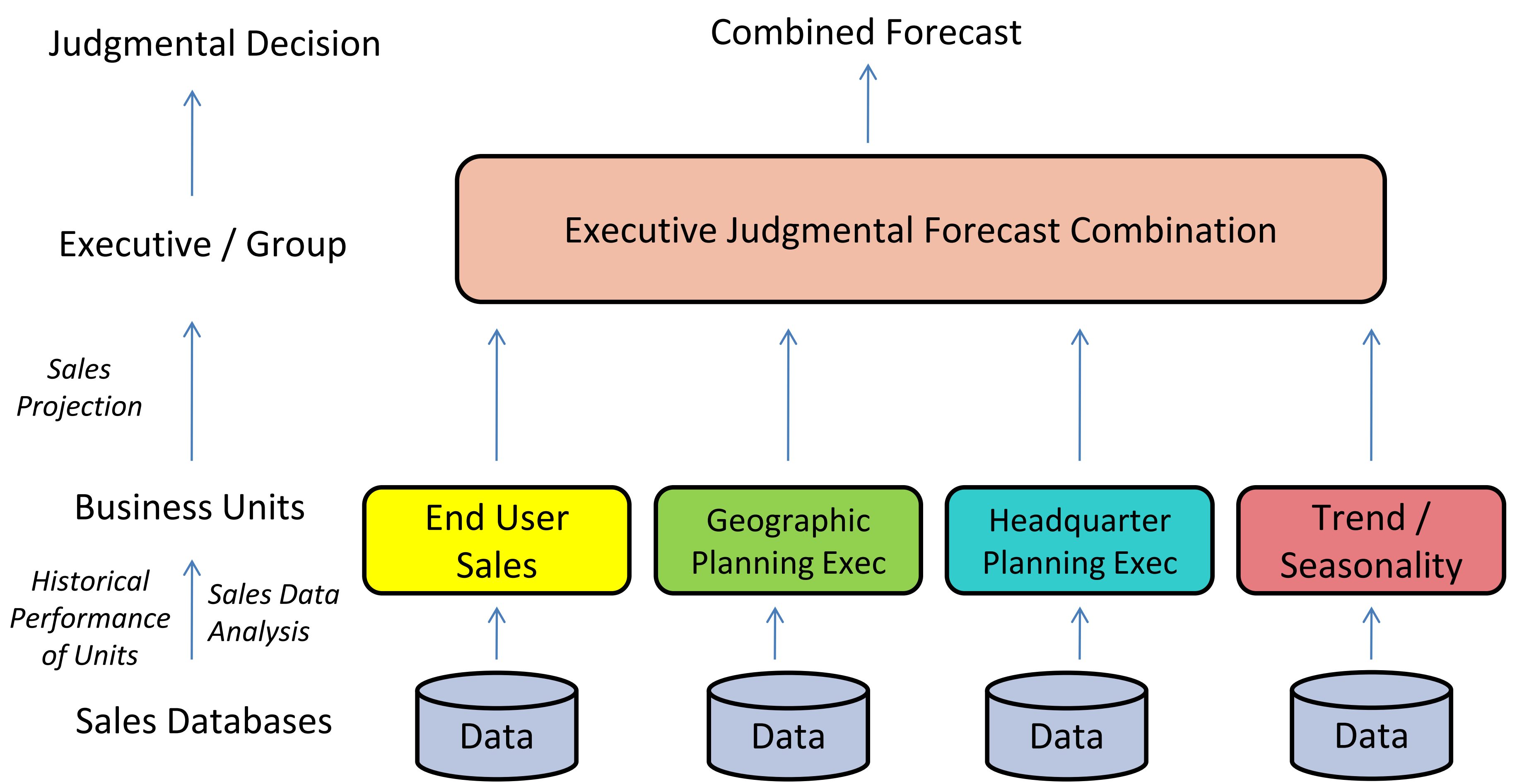
Combining Multiple Forecasting Experts for Corporate Revenue Using Combinatorial Fusion Analysis

Weekly sales projections from four functional business units.

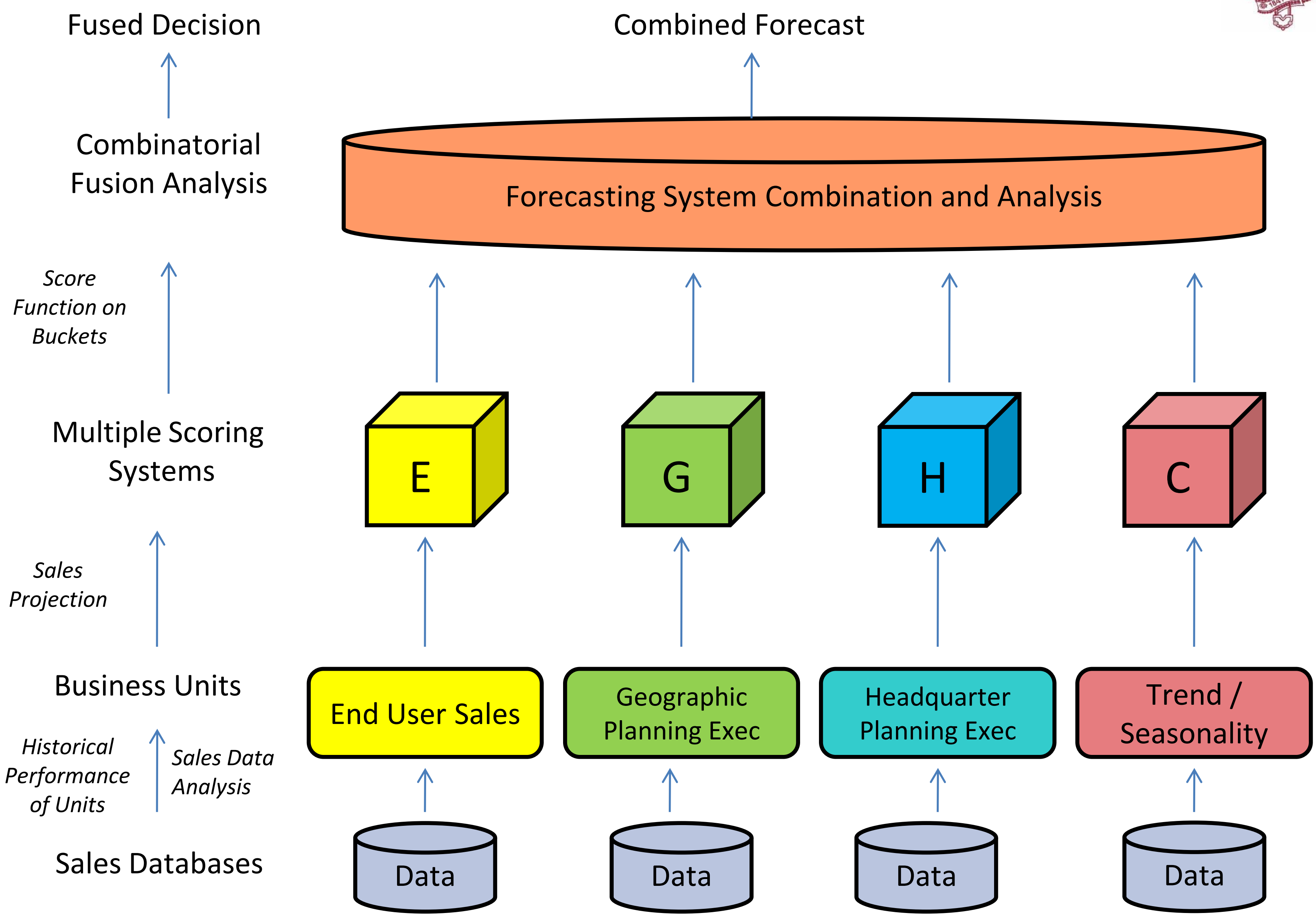


E: End User Sales
G: Geographic Planning Exec
H: Headquarter Planning Exec
C: Trend / Seasonality
Average: Average forecast inversely weighted by historical variance
Actual: Actual end of quarter sales

Traditional Business Approach to Forecast Combination



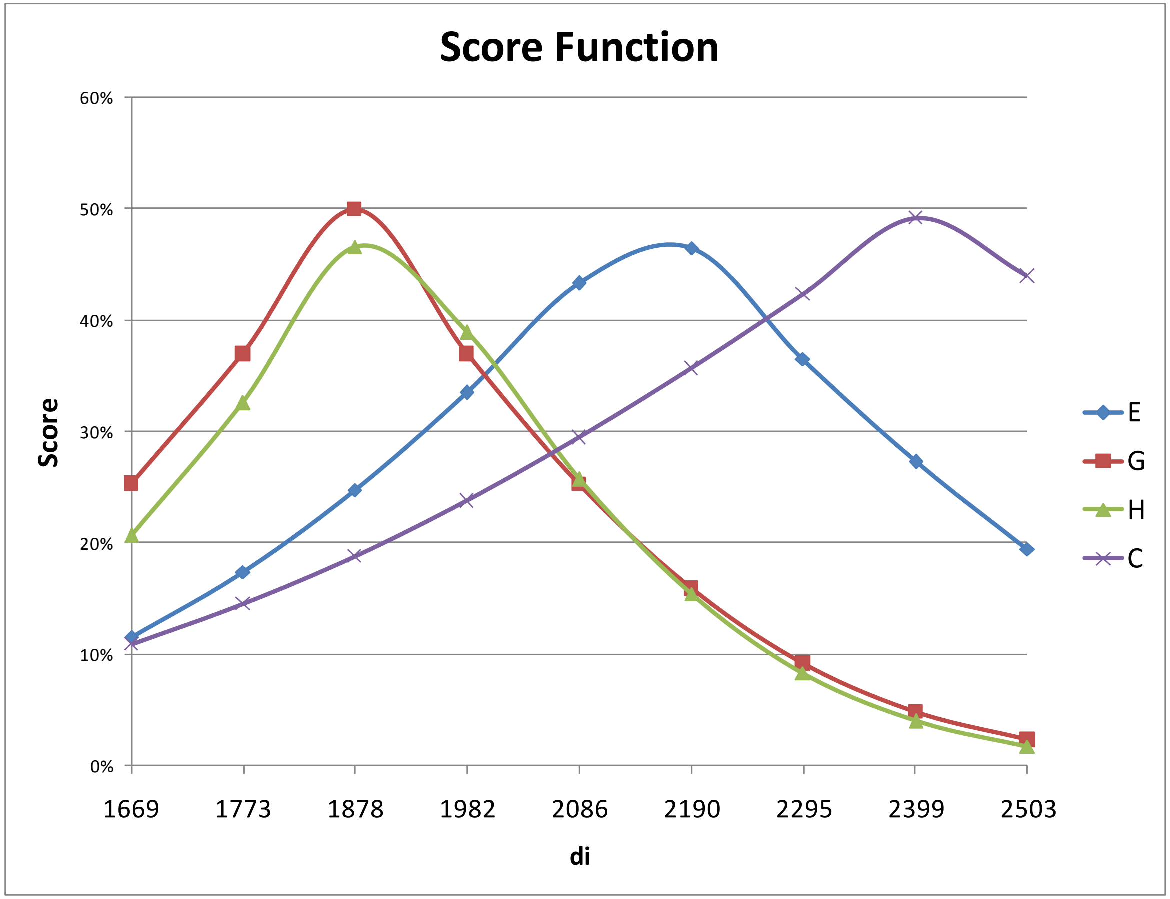
Forecast Combination with MSS and CFA



Individual score functions for week 9

forecast	2154	1877	1901	2411	2154	1877	1901	2411
	Original Score				Normalized Score			
buckets(di)	E	G	H	C	E	G	H	C
2503	19%	2%	2%	44%	23%	0%	0%	86%
2399	27%	5%	4%	49%	45%	5%	5%	100%
2294	36%	9%	8%	42%	71%	14%	15%	82%
2190	46%	16%	15%	36%	100%	28%	31%	65%
2086	43%	25%	26%	29%	91%	48%	53%	48%
1981	33%	37%	39%	24%	63%	73%	83%	34%
1877	25%	50%	47%	19%	38%	100%	100%	21%
1773	17%	37%	33%	14%	17%	73%	69%	9%
1669	12%	25%	21%	11%	0%	48%	42%	0%

Score functions constructed based on each unit's sales projection for week 9



Judge	E	G	H	C
Sigma	405	313	283	603
mean	2154	1877	1901	2411

$$Combined\ mean = \frac{\sum \frac{1}{\sigma_i^2} m_i}{\sum \frac{1}{\sigma_i^2}} = \overline{m}$$

Score and Rank Combinations

The score combination for c systems: X_1, X_2, … , X_c :

$$s_{SC}(d_i) = \frac{1}{c}\sum_{j=1}^{c} s_{X_j}(d_i)$$

The rank combination for c systems: X_1, X_2, … , X_c :

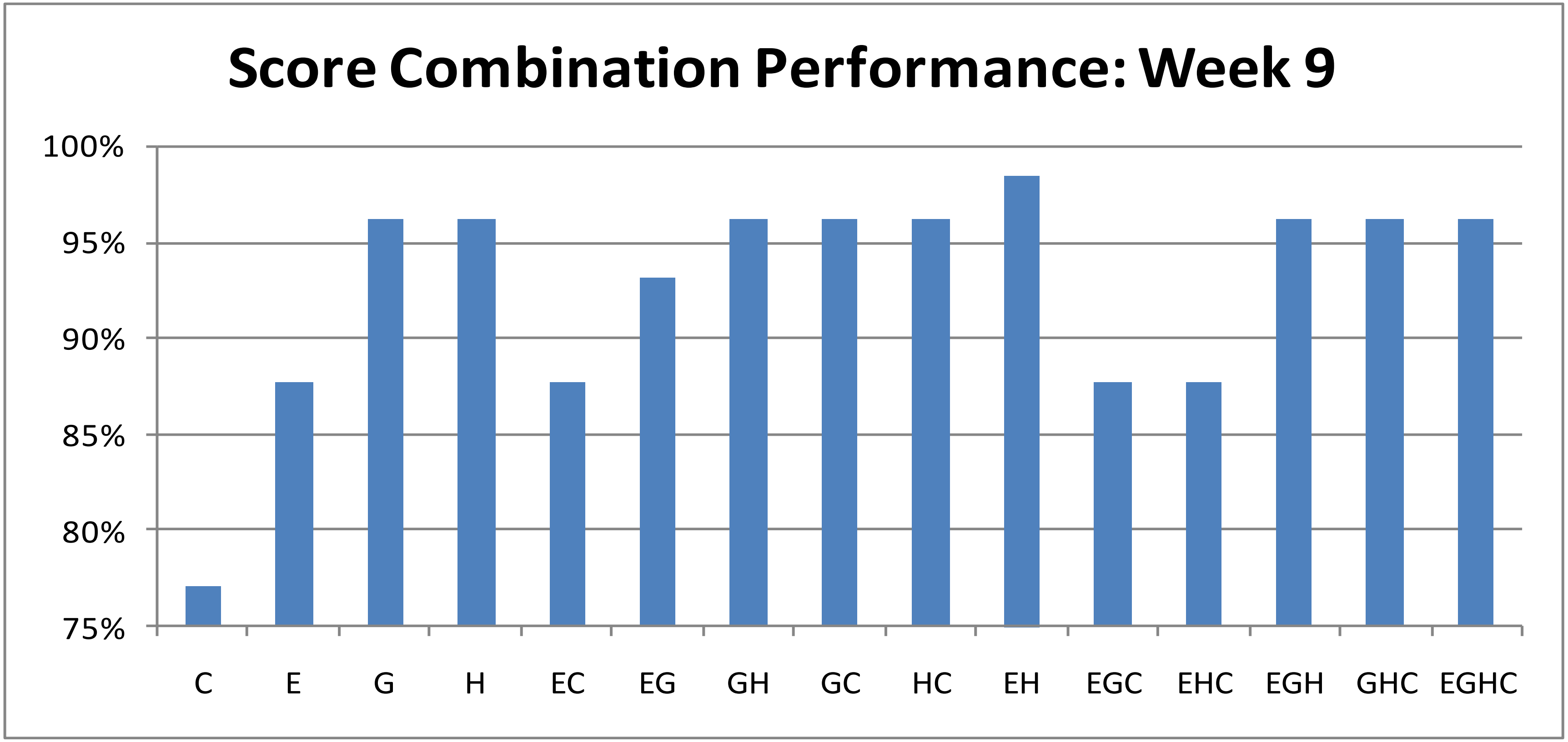
$$s_{RC}(d_i) = \frac{1}{c}\sum_{j=1}^{c} r_{X_j}(d_i)$$

Combination by Score

Rank Function of the Averaged Score Combination

buckets(di)	E	G	H	C	EG	EH	EC	GH	GC	HC	EGH	EGC	EHC	GHC	EGHC
2503	7	9	9	2	9	9	5	9	7	7	9	7	7	9	8
2399	5	8	8	1	7	7	3	8	3	3	8	6	6	7	7
2294	3	7	7	3	6	5	2	7	5	5	6	4	4	6	5
2190	1	6	6	4	4	4	1	6	6	6	4	1	1	5	4
2086	2	5	4	5	1	2	4	4	4	4	3	2	2	4	3
1981	4	2	2	6	3	1	6	2	2	2	2	3	3	2	2
1877	6	1	1	7	2	3	7	1	1	1	1	5	5	1	1
1773	8	2	3	8	5	6	8	3	8	8	5	8	8	3	6
1669	9	4	5	9	8	8	9	5	9	9	7	9	9	8	9
forecast	2190	1877	1877	2399	2086	1981	2190	1877	1877	1877	1877	2190	2190	1877	1877
performance	88%	96%	96%	77%	93%	98%	88%	96%	96%	96%	96%	88%	88%	96%	96%

Score combination performance for week 9



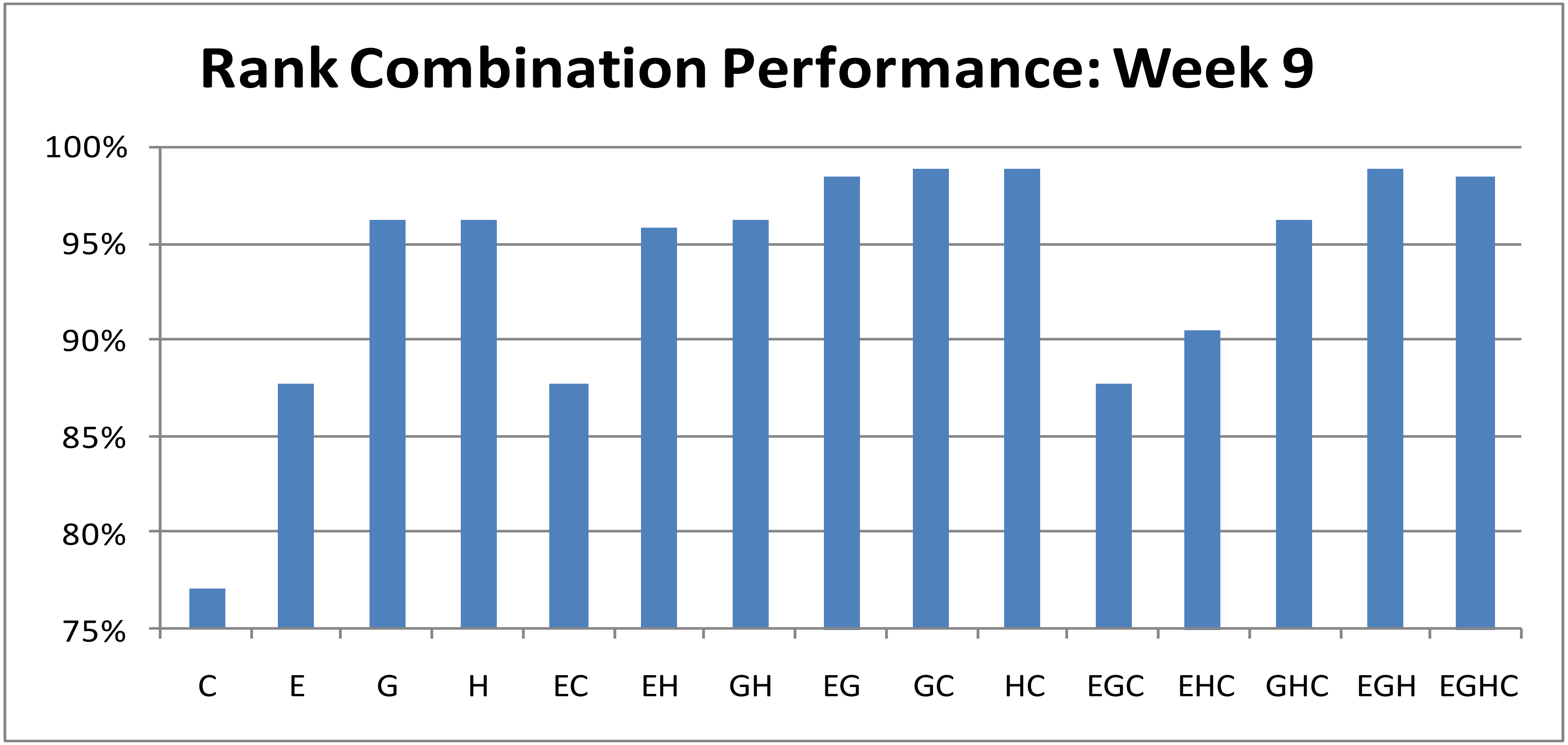
C	E	G	H	EC	EG	GH	GC	HC	EH	EGC	EHC	EGH	GHC	EGHC
77%	88%	96%	96%	88%	93%	96%	96%	96%	98%	88%	88%	96%	96%	96%

Combination by Rank

Rank Function of the Averaged Rank Combination

buckets(di)	E	G	H	C	EG	EH	EC	GH	GC	HC	EGH	EGC	EHC	GHC	EGHC
2503	7	9	9	2	9	9	5	9	8	8	9	8	7	9	9
2399	5	8	8	1	8	7	3	8	3	4	8	6	6	7	7
2294	3	7	7	3	6	5	3	7	7	6	6	4	4	7	5
2190	1	6	6	4	4	4	1	6	7	6	5	1	2	5	4
2086	2	5	4	5	4	2	4	5	7	4	3	3	2	4	3
1981	4	2	2	6	1	2	6	2	2	2	2	3	3	2	1
1877	6	1	1	7	4	4	7	1	2	2	2	6	6	1	2
1773	8	2	3	8	6	6	8	3	7	8	5	8	8	3	6
1669	9	4	5	9	8	8	9	5	9	9	7	9	9	8	9
forecast	2190	1877	1877	2399	1981	2033	2190	1877	1929	1929	1929	2190	2138	1877	1981
performance	88%	96%	96%	77%	98%	96%	88%	96%	99%	99%	99%	88%	90%	96%	98%

Rank combination performance for week 9



C	E	G	H	EC	EH	GH	EG	GC	HC	EGC	EHC	GHC	EGH	EGHC
77%	88%	96%	96%	88%	96%	96%	98%	99%	99%	88%	90%	96%	99%	98%

Test results with four quarters, using Score Combination

Week		E	G	H	C	EG	EH	EC	GH	GC	HC	EGH	EGC	EHC	GHC	EGHC
Q1	9	88%	96%	96%	77%	93%	98%	88%	96%	96%	96%	96%	88%	88%	96%	96%
Q2	9	97%	93%	93%	78%	98%	98%	97%	93%	93%	93%	93%	98%	98%	93%	93%
Q3	9	94%	94%	94%	86%	94%	94%	94%	94%	99%	94%	94%	94%	94%	94%	94%
Q4	9	93%	88%	88%	93%	88%	88%	93%	88%	88%	88%	88%	93%	93%	88%	88%
Average of week 9 performance		93%	93%	93%	84%	93%	95%	93%	93%	94%	93%	93%	93%	93%	93%	93%
Average Forecast Error		7%	7%	7%	16%	7%	5%	7%	7%	6%	7%	7%	7%	7%	7%	7%
Reduction of Error for the best single judge						**-6%**	**-25%**	2%	0%	**-18%**	0%	0%	-3%	-3%	0%	0%

Average Reduction of Error	**-5%**

Test results with four quarters, using Rank Combination

Week		E	G	H	C	EG	EH	EC	GH	GC	HC	EGH	EGC	EHC	GHC	EGHC
Q1	9	88%	96%	96%	77%	98%	96%	88%	96%	99%	99%	99%	88%	90%	96%	98%
Q2	9	97%	93%	93%	78%	98%	96%	97%	93%	96%	93%	93%	99%	99%	93%	96%
Q3	9	94%	94%	94%	86%	94%	94%	94%	94%	97%	97%	94%	94%	94%	94%	94%
Q4	9	93%	88%	88%	93%	88%	88%	93%	88%	88%	88%	88%	90%	90%	88%	88%
Average of week 9 performance		93%	93%	93%	84%	95%	93%	93%	93%	95%	94%	94%	93%	94%	93%	94%
Average Forecast Error		7%	7%	7%	16%	5%	7%	7%	7%	5%	6%	6%	7%	6%	7%	6%
Reduction of Error for the best single judge						**-25%**	**-7%**	2%	0%	**-27%**	**-18%**	**-9%**	1%	**-8%**	0%	**-17%**

Average Reduction of Error	**-10%**

(D) Remarks

Forecasting Combination Remarks

- CFA application to sales forecasting is more robust because it takes advantage of the strengths and compensates for the weakness of different scoring functions
- Outperforms each individual judge as well as average performance for the quarter

Our Future Research

- Optimize the methodology
  - -- more judgers
  - -- more buckets
- Score function transformation and diversity
- Analyze historical data, acquire new data