# Supplementary Material – Learning to Rank Based on Choquet Integral: Application to Association Rules

Charles Vernerey[1], Noureddine Aribi[2], Samir Loudni[1] ✉, Yahia Lebbah[2], and Nassim Belmecheri[3]

[1] TASC – DAPI, IMT-Atlantique, LS2N – CNRS, 44307 Nantes, France
{samir.loudni,Charles.Vernerey}@imt-atlantique.fr
[2] Lab. LITIO, Université Oran1, 31000 Oran, Algeria
{aribi.noureddine,lebbah.yahia}@univ-oran1.dz
[3] Simula Research Laboratory, Oslo, Norway nassim@simula.no

## A1) Association rule interestingness measures

| Family | Measures |
|---|---|
| 1 | **Yule's Q**, Yules Y, Odds Ratio |
| 2 | **Cosine**, Jaccard |
| 3 | Laplace, Support |
| 4 | $\phi$-Coefficient, Collective Strength, Piatetsky-Shapiro |
| 5 | **Goodman Kruskal's**, Gini Index |
| 6 | Interest factor, **Added Value**, Klosgen K |
| 7 | **Certainty factor**, Mutual Information, Kappa Coefficient |

(a) Groups of objective measures with similar properties .

| Name | Formula |
|---|---|
| Yule's Q | $\frac{n_{XY}.n_{\overline{X}.\overline{Y}} - n_{X\overline{Y}}.n_{\overline{X}Y}}{n_{XY}.n_{\overline{X}.\overline{Y}} + n_{X\overline{Y}}.n_{\overline{X}Y}}$ |
| Cosine | $\frac{n_{XY}}{\sqrt{n_X.n_Y}}$ |
| Goodman Kruskal's | $\frac{max(n_{XY}, n_{X\overline{Y}}) + max(n_{\overline{X}Y}, n_{\overline{X}.\overline{Y}}) - max(n_Y, n_{\overline{Y}})}{n - max(n_Y, n_{\overline{Y}})}$ |
| Added Value | $\frac{n_{XY}}{n_X} - \frac{n_Y}{n}$ |
| Certainty factor | $\frac{n.n_{XY} - n_X.n_Y}{n_X(1 - n_Y)}$ |

(b) Objective measures selected for association rules.

Table 1: Association rule interestingness measures.

Table 1 shows the different measures used in our evaluation for association rules. These measures are based on the frequency counts. Their formulas are expressed using a notation where $X$ denotes the rule antecedent; $Y$ denotes the rule consequent; $n$ denotes the total number of examples; $n_X$ (resp. $n_Y$) denotes the number of examples satisfying $X$ (resp. $Y$); $\overline{X}$ and $\overline{Y}$ denote the logical negation of $X$ and $Y$; $n_{XY}$ denotes the number of examples satisfying both $X$ and $Y$, while $n_{X\overline{Y}}$ denotes the number of examples satisfying $X$ but not $Y$.

| Notation | Significance |
|---|---|
| $\mathcal{X}$ | The set of alternatives |
| $\mathcal{P}$ | The set of user preferences |
| $N$ | The set of criteria |
| $T$ | Subset of criteria ( $T \subseteq N$) |
| $k$ | $k$-additivity value of the Choquet integral |
| $t \in [L]$ | Iteration index |
| $\delta_C$ | The minimal indifference threshold used in equation (3) |
| $\mathcal{X}^t$ | User query |
| $\mathcal{U}^t$ | User-defined feedback on $\mathcal{X}^t$ |
| $m_\mu^L$ | Learned Möbius function |

Table 2: Notations.

Table 2 summarises the different notations used in this paper.

### A2) Complementary results for research query Q1-B (passive learning)

Figure 1 plots a detailed comparison in terms of recall and average precision between `ChoquetRank`, `ListNet`, `RankBoost` and `RankNet` on two datasets.

| Learning Algorithm | Training Size fold | | | |
|---|---|---|---|---|
| | 100 | 200 | 500 | 1000 |
| | Time (s) | Time (s) | Time (s) | Time (s) |
| `AHPRank` | **0.0092** | **0.0018** | **0.0018** | **0.0026** |
| `RankingSVM` | 0.0108 | 0.0078 | 0.0152 | 0.0384 |
| `ChoquetRank` | 1.19 | 2.01 | 13.6 | 95.9 |

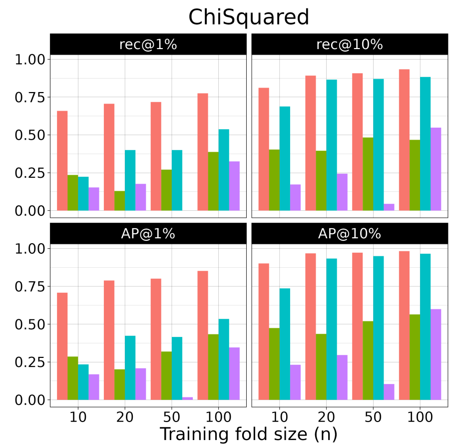
Table 3: Comparison in terms of running time between `ChoquetRank`, `AHPRank` and `RankingSVM` using CHOQUETPEARSON as user ranking function on RETAIL dataset and for different training fold sizes.

### A3) Complementary results for CPU-times comparison (research query Q1)

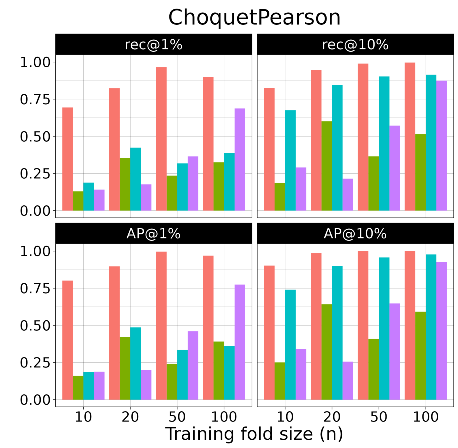
Table 3 reports a detailed comparison in terms of CPU times using CHOQUETPEARSON for different training fold sizes.

### A4) Eisen rules analysis

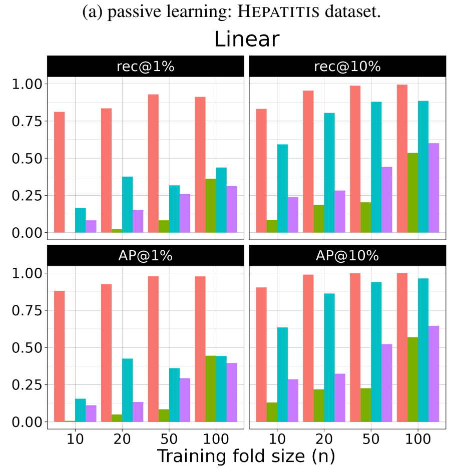
Here, we analyze the 100 top association rules ranked by `ChoquetRank` and `AHPRank` on the test set using CHOQUETPEARSON as the user feedback for the learning step. The top 100 rules (i.e. rules having the highest global interestingness values) ranked by the two approaches are similar but `ChoquetRank` tends to provide a better ranking. Table 4 shows examples of some extracted rules that are mostly well ranked by `ChoquetRank`.



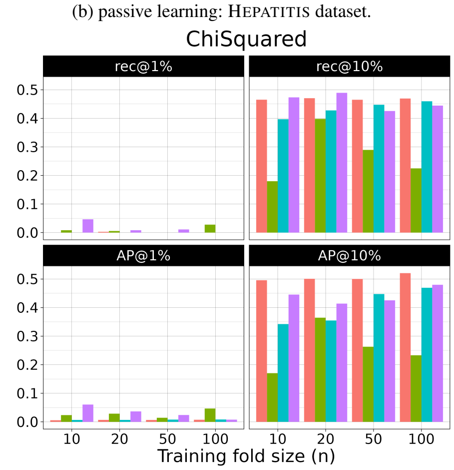

(a) passive learning: HEPATITIS dataset.



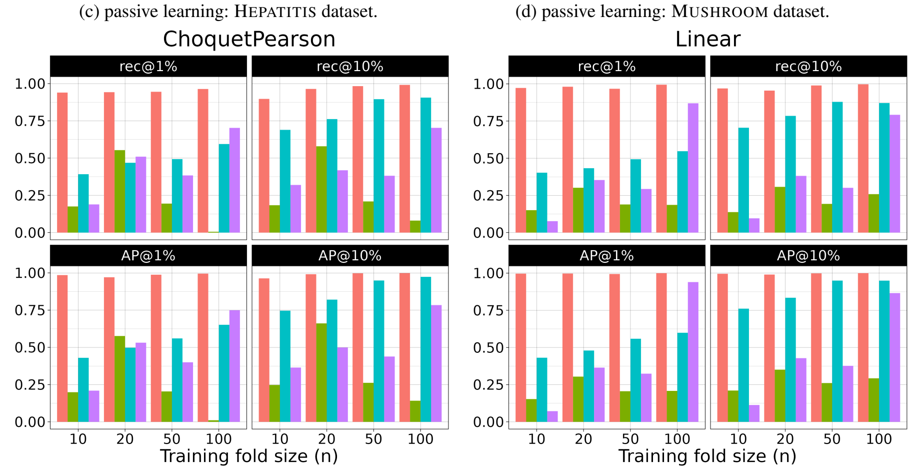

(b) passive learning: HEPATITIS dataset.




(c) passive learning: HEPATITIS dataset.

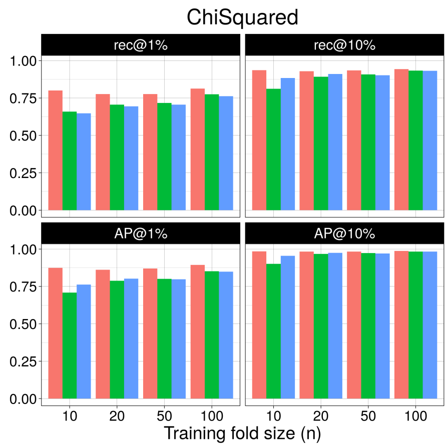



(d) passive learning: MUSHROOM dataset.

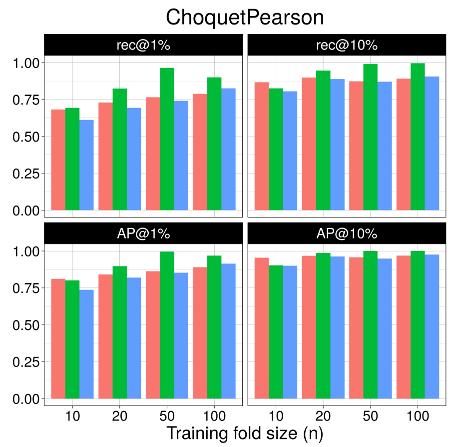



(e) passive learning: MUSHROOM dataset.

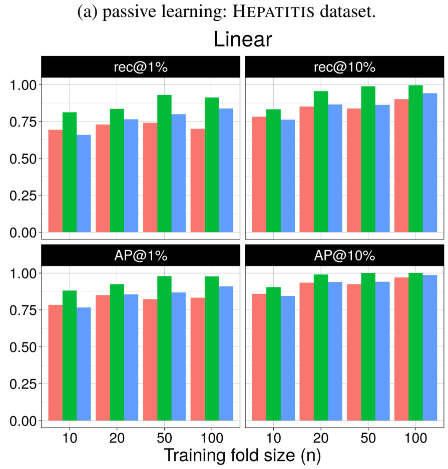

Fig. 1: Accuracy comparison between `ChoquetRank` and state-of-the-art Learning-to-Rank methods on HEPATITIS and MUSHROOM datasets (passive learning).



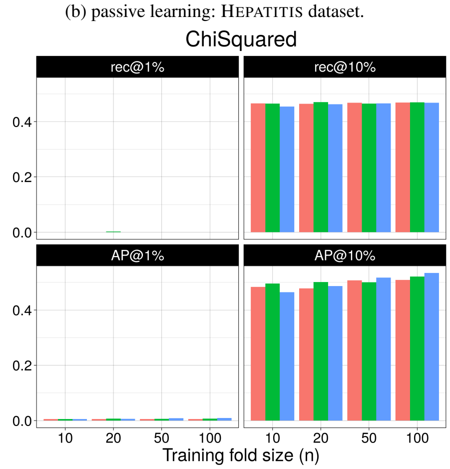

(a) passive learning: HEPATITIS dataset.



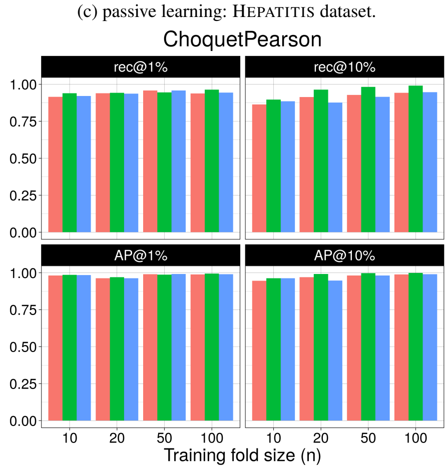

(b) passive learning: HEPATITIS dataset.



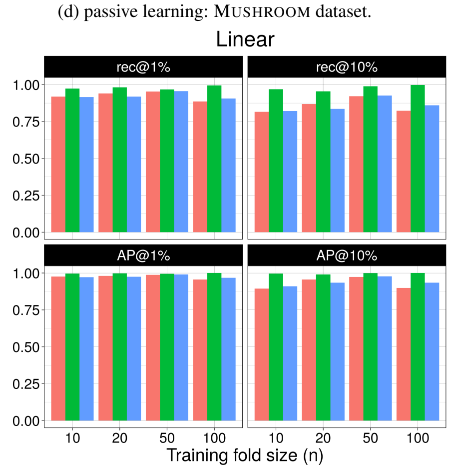

(c) passive learning: HEPATITIS dataset.




(d) passive learning: MUSHROOM dataset.




(e) passive learning: MUSHROOM dataset.




(f) passive learning: MUSHROOM dataset.

AHPRank ChoquetRank RankingSVM

Fig. 2: Learning accuracy comparison between `ChoquetRank`, `AHPRank` and `RankingSVM` on different training fold size on HEPATITIS and MUSHROOM datasets (passive learning).

| Rule | Antecedent | Consequent |
|---|---|---|
| 1 | path:03050 | pmid:16922378, go:0030163 |
| 2 | path:03020 | go:0016070, go:0005634, go:0006350, go:0016740, go:0016779, path:00230, path:00240 |
| 3 | go:0016779, path:00240 | go:0005634, go:0016740, path:00230 |
| 4 | go:0016779, path:00230 | go:0005634, go:0016740, path:00240 |
| 5 | pr:RAP1, go:0005840 | go:0005737, go:0006412, pr:FHL1 |
| 6 | PR:FHL1, pmid:11433365 | go:0005737, go:0005198, go:0005840, go:0006412, path:03010 |
| 7 | phenot:inviable,go:0016779, path:00240 | go:0005634, go:0016740, path:00230 |
| 8 | phenot:inviable, go:0016779 path:00230 | go:0005634, go:0016740, path:00240 |
| 9 | phenot:”shortened telomeres” | go:0006996, pmid:15161972 |
| 10 | phenot:exhibits growth defect on a non-fermentable (respiratory) carbon source, | go:0005737, go:0005739, go:0005198, go:0005840, go:0006412 |
| 11 | pmid:17024709 | go:0005737, go:0005783, go:0006464 go:0016740 |
| 12 | go:0005739, pmid:11433365 | go:0005737, go:0005198, go:0005840 go:0006412, pmid:16823961 |
| 13 | heat3↓, path:03010 | go:0005737, go:0005198, go:0006412, go:0005840, pmid:11433365 |
| 14 | heat3↓, pmid:12490706 | go:0005737, go:0005198, go:0005840, go:0006412, path:03010, pmid:5542014 |

Table 4: Examples of association rules. heat3 refers to the time points of the heat shock and sporulation experiments, respectively. ↓ denotes an under-expression. The prefixes `go`, `path`, `pmid`, `pr` allow us to identify GO terms, KEGG pathways, PubMed IDs and names of transcriptional regulators, respectively.

### A5) Complementary results for research query Q2 (passive learning)

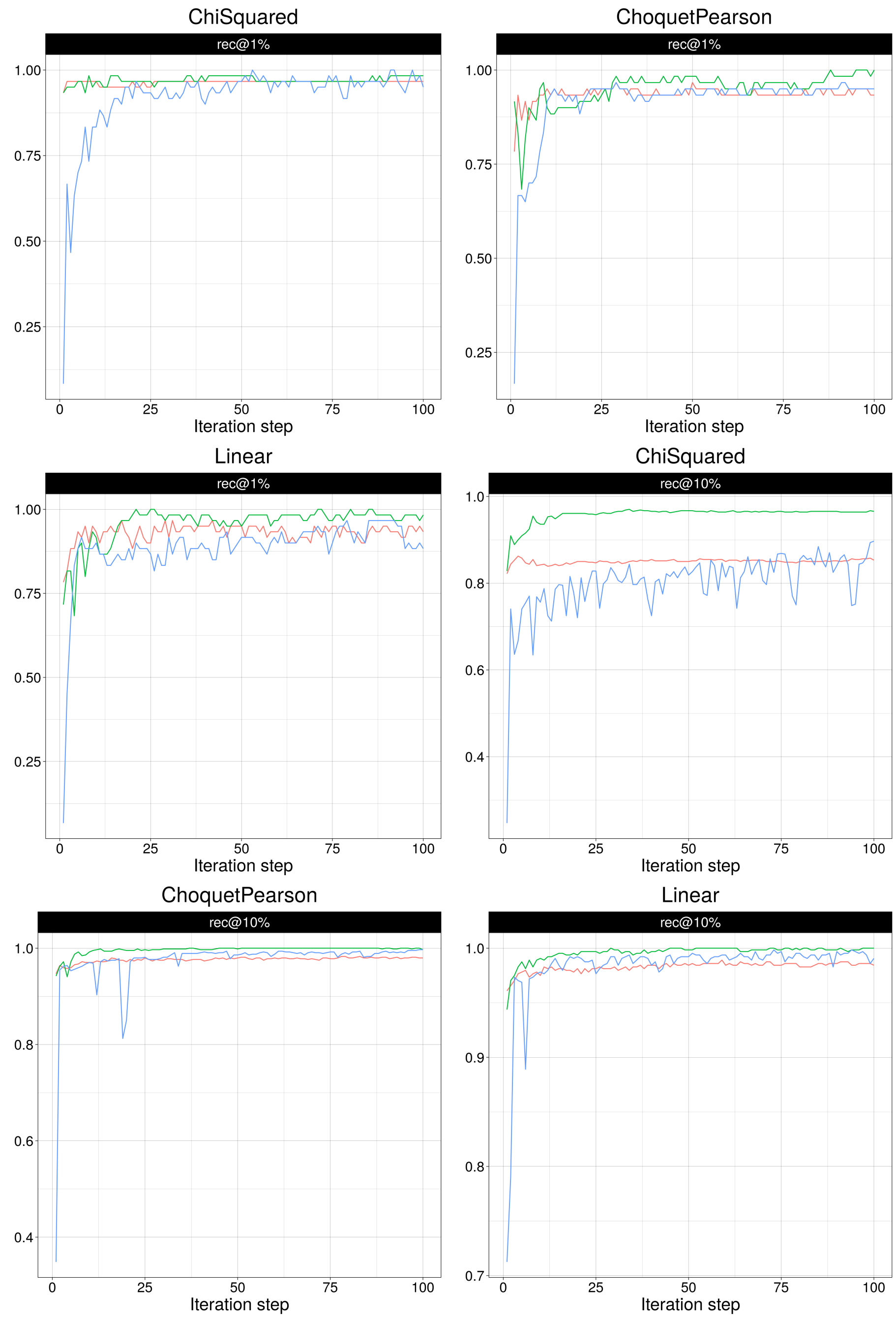
Figure 2 plots a detailed comparison in terms of recall and average precision between `ChoquetRank`, `AHPRank` and `RankingSVM` on two datasets for different training fold sizes (passive mode).

### A6) Complementary results for research query Q3 (active learning)

Figures 3, 4 and 5 plot detailed comparison in terms of recall between `ChoquetRank`, `AHPRank` and `RankingSVM` for active learning setting.

### A7) Complementary results for research query Q4 (preference inconsistencies)

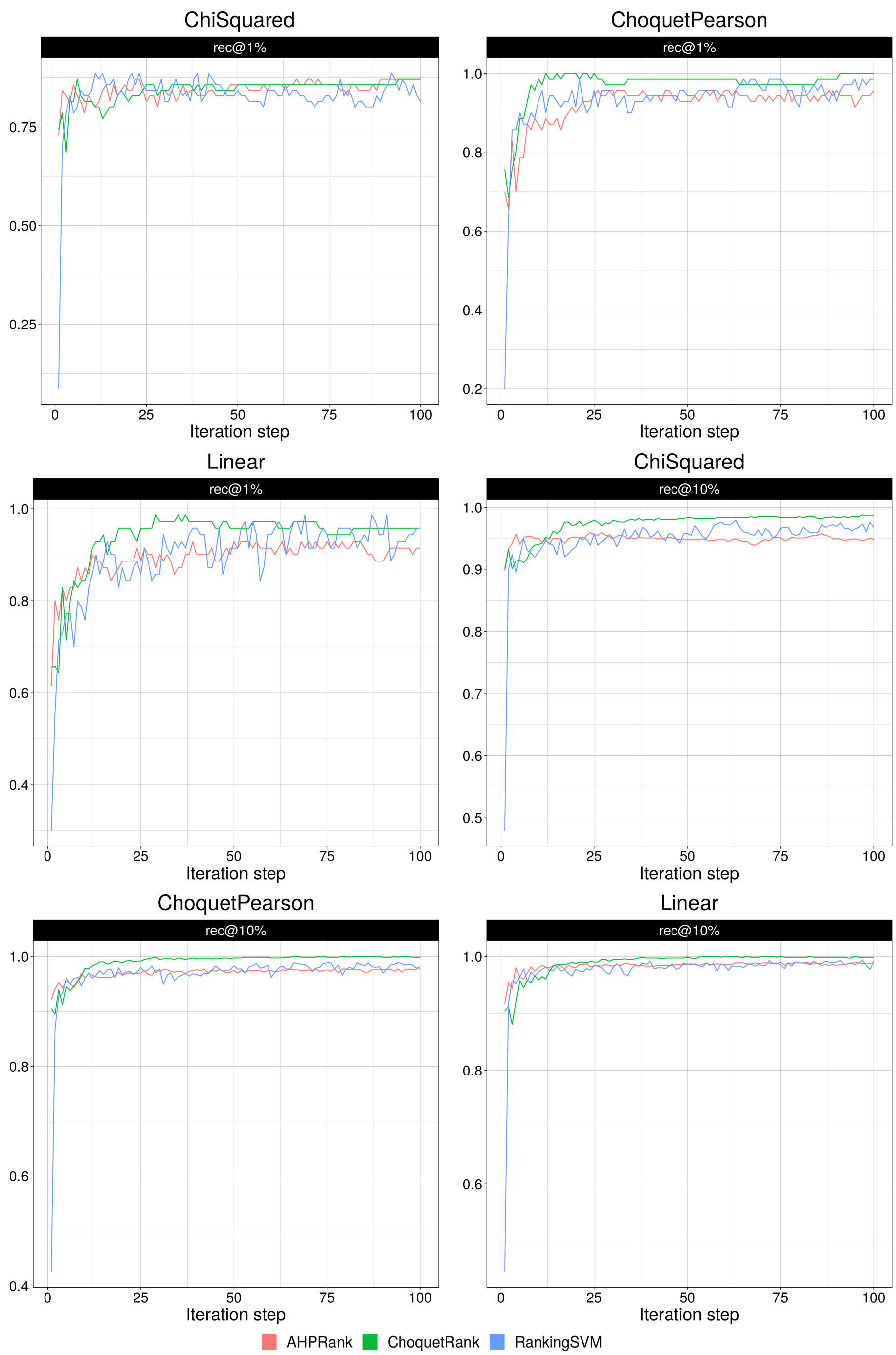
Figures 7, 8 and 6 show the evolution of the recall over 100 learning iterations of CHOQUETPEARSON when the user is prone to errors with different error probabilities.




(a) RETAIL dataset.

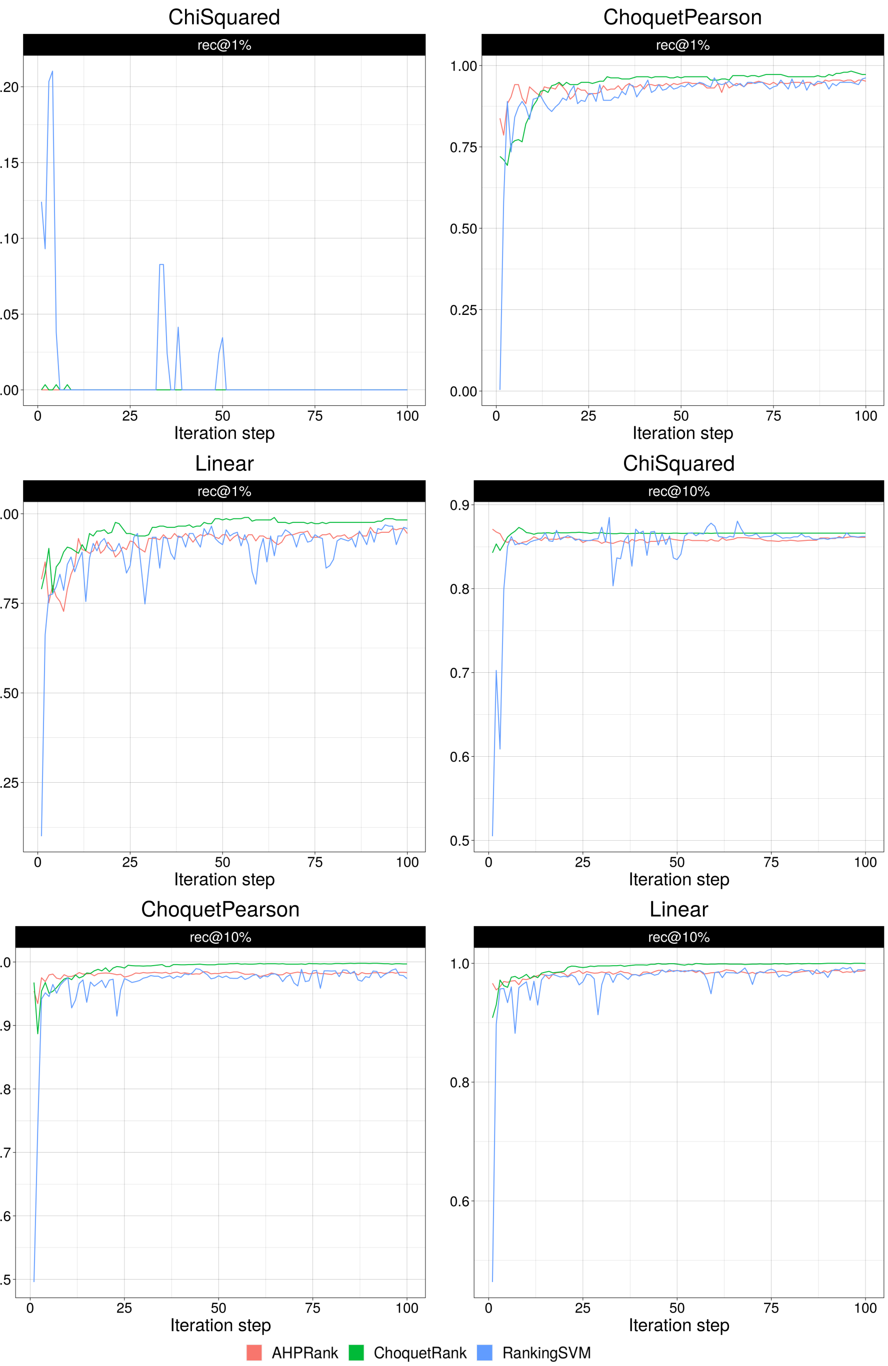
AHPRank ChoquetRank RankingSVM

Fig. 3: Active learning results: comparison between `ChoquetRank`, `AHPRank` and `RankingSVM`.




(a) HEPATITIS dataset.

Fig. 4: Active learning results: comparison between `ChoquetRank`, `AHPRank` and `RankingSVM`.




(a) MUSHROOM dataset.

Fig. 5: Active learning results: comparison between `ChoquetRank`, `AHPRank` and `RankingSVM`

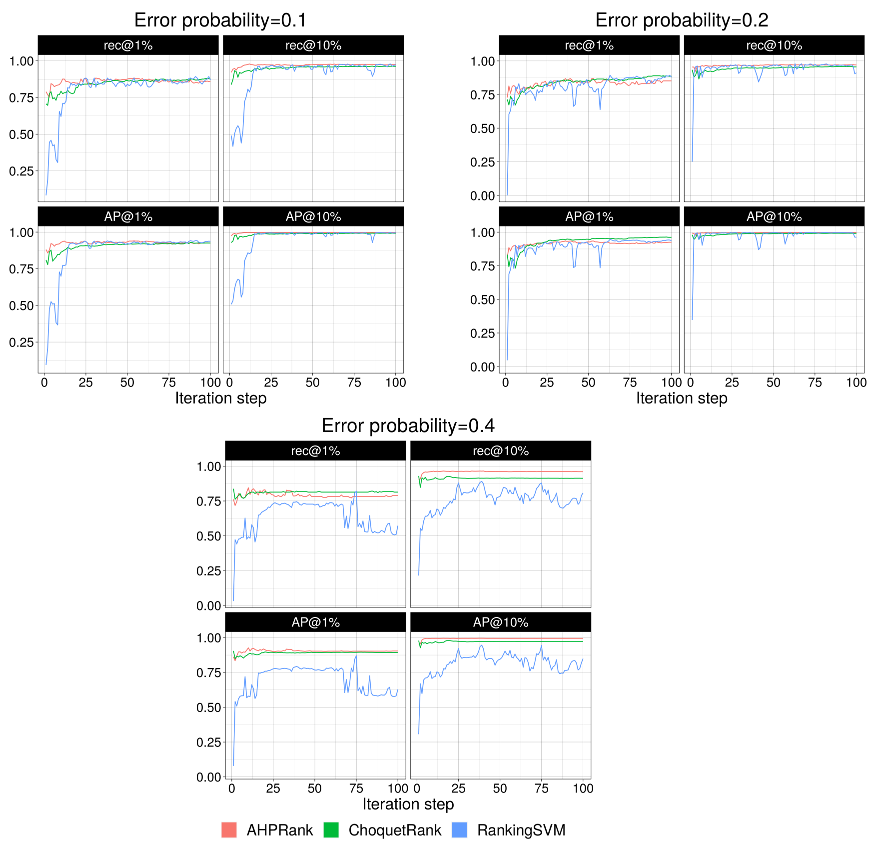



Fig. 6: Active learning results with different error probabilities on MUSHROOM dataset.

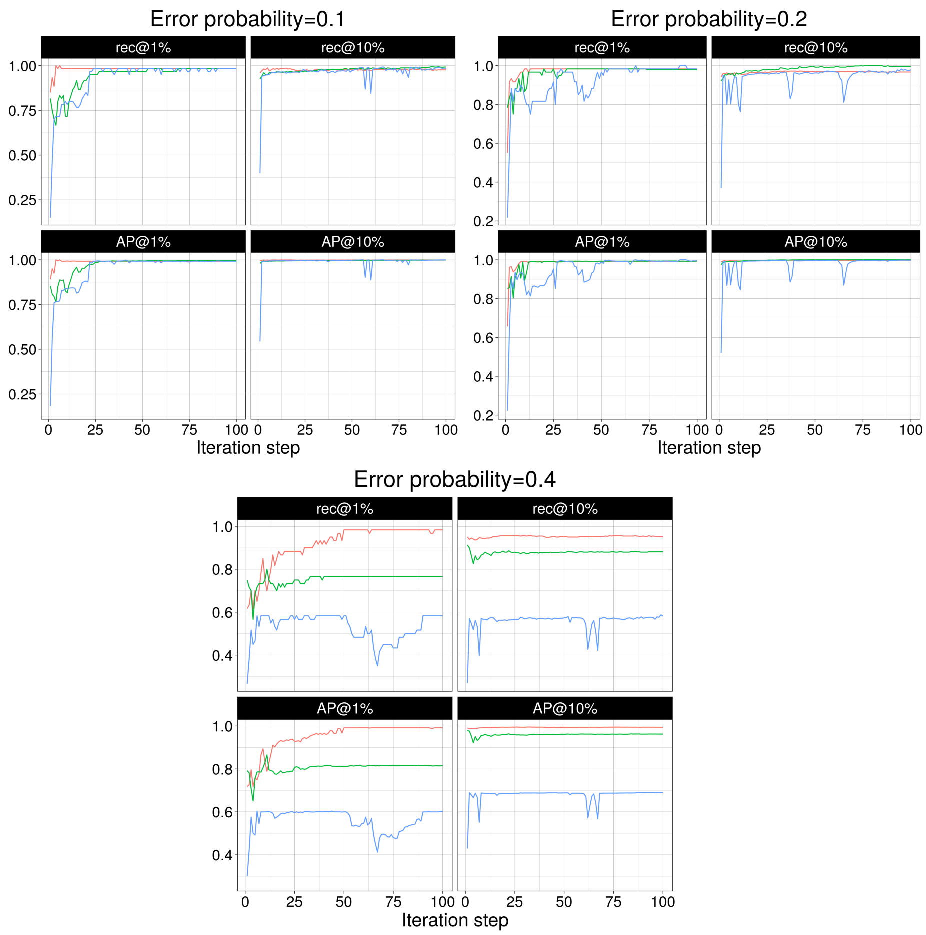



Fig. 7: Active learning results with different error probabilities on RETAIL dataset.

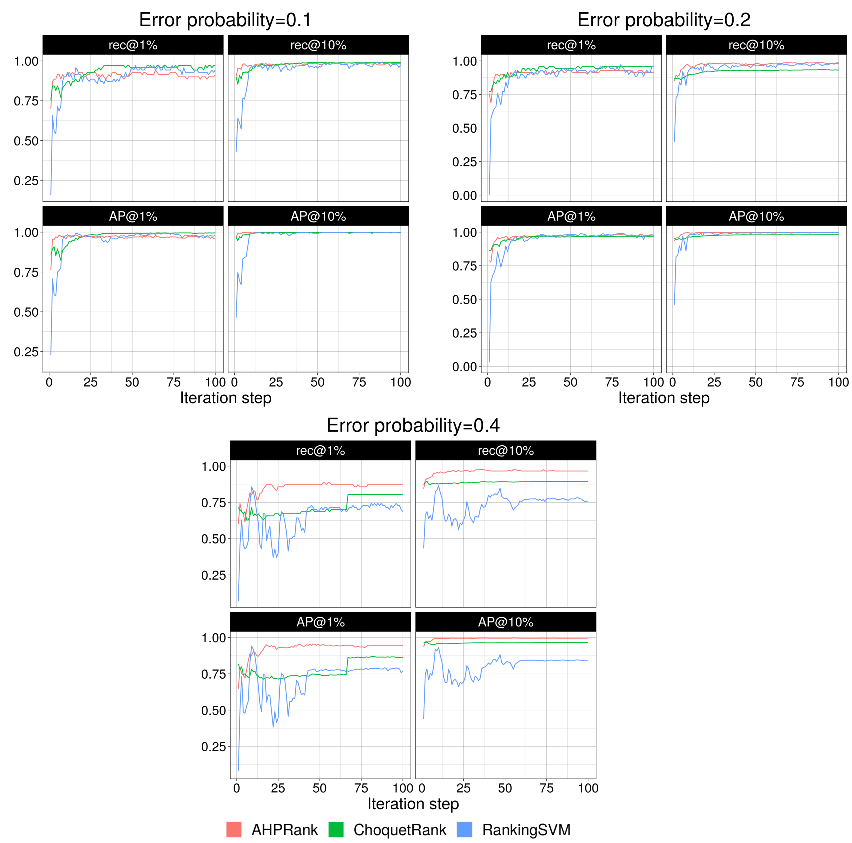



Fig. 8: Active learning results with different error probabilities on HEPATITIS dataset.