## **Experimental Evidence Extraction System in Data Science** with Hybrid Table Features and Ensemble Learning E THE V Wenhao Yu<sup>+</sup>, Wei Peng<sup>+</sup><sup>‡</sup>, Yu Shu<sup>+</sup><sup>\$</sup>, Qingkai Zeng<sup>+</sup>, Meng Jiang<sup>+</sup> CONFERENCE <sup>†</sup>University of Notre Dame, USA <sup>‡</sup> Zhejiang University, China <sup>\$</sup> Sichuan University, China ▼ Introduction/Motivation (Given massive PDF papers, can we automate the integration of tables into a database of experimental results?) (ACM TOIS 2011) (ACM TIST 2011) Table III. MAE Comparison with Other Approaches on Epinions Dataset D Comparisons (A Smaller MAE or RMSE Value Means a Better Performance Method Dataset Metric Score Source Methods 90% Training 80% Training 70% Training 60% Training CSUserMeanItemMean NMF PMF TCF Trust SoPee PSTF 10 UserMean Epinions MAE 0.9319 TOIS11-paper7-table3 UserMean,Item.Nam NMF FMF ICF Trust Sofkee K81E MAE 0.9184 0.9768 0.8778 0.8876 0.9005 0.9004 0.8442 0.8377 RMSE 1.1688 1.2375 1.1649 1.1575 1.1607 1.1959 1.1333 1.1109 MAE 0.9285 0.9913 0.8975 0.8981 0.9021 0.8638 0.8554 RMSE 1.1817 1.2584 1.1861 1.1826 1.1761 1.2140 1.1530 1.1346 Hear Maan 0.0204 0.9353 0.9384 90% 11 UserMean Epinions MAF 0.9285 TIST11-paper3-table3 Item Mean 0.8936 0.9316 0.9528 80% 12 UserMean Epinions MAE 0.9285 WSDM11-paper12-table Trust 0.9005 0 9044 0.9082 0.9153 IOM6 LINI / LINI / LINI Line <thLine</th> <thLine</th> <thLine</th> NMF 0.8938 0.8975 0.9229 0.9430 RSTE 109 ItemMean Epinions RMSE 1.1973 TOIS11-paper7-table4 SUD 0.9214 0.9421 110 ItemMean Epinions RMSE 1.2584 TIST11-paper3-table3 PMF 0.8678 0.8946 0.9127 0.9350 111 ItemMean Epinions RMSE 1 2584 WSDM11-naner12-tables SoRec 0.8442 0.8751 0.8948 0.8638 112 Trust Eninions DAACE 1.2132 TIST11-paper3-table3 NMF 0.9211 (WSDM 2011) 113 NMF Epinions RMSE 1.1832 TOIS11-paper7-table4 SVD 0.8921 0.9189 0.9382 10D114 NMF Epinions RMSE 1.1832 TIST11-paper3-table3 PMF 0.8651 0.0002 0.0328 115 NMF RMSE 1.1832 WSDM11-paper12-table Epinions SoRec 0.8404 0.8722 0.8921 116 SVD RMSE 1.1812 TOIS11-paper7-table4 (ACM TOIS 2011) 0.7026 0.7022 0.6992 0.6988 Epinions 117 TCF RMSE 1.1761 TIST11-paper3-table3 Table IV BMSE C Epinions 60% Improve RMSE Table Methods 90% Training 80% Training 70% Training 60% Training Table 118 PMF Epinions RMSE 1.1760 TOIS11-paper7-table4 DB 1.1968 1.1973 1.2014 0.5706 0.5702 0.5690 0.5685 119 PMF Epinions RMSE 1.1760 TIST11-paper3-table3 User Mear 1.1927 0.9854 0.0317 0.3859 17.06% 10.00% 3.63% 0.8567 0.7971 0.7482 16.83% 10.61% 4.77% 0.9134 0.9768 0.8712 40% Improve RMSE 1.1760 WSDM11-paper12-table5 Extraction 120 PMF RMSE Epinions 0.7172 0.7169 0.7129 0.7125 Integration 1.1697 1.1761 1.1797 Trust 1.1894 121 SoRe Operation Epinions RMSE 1.1492 TOIS11-paper7-table4 1.1649 1.1635 1.1861 0.9134 0.9768 0.8712 0.8851 0.8367 0.8290 0.8287 0.8258 0.8256 9.61% 15.48% 5.23% 4.57% 1.33% 0.8290 0.8287 0.8258 0.8256 1.1688 1.2375 1.1621 1.1544 1.1094 1.0792 1.0790 1.0744 1.0739 1.2090 122 RSTE Epinions RMSE 1.1256 TIST11-paper3-table3 SVD 1.1845 90% Improve RMSE 123 RSTE Epinions RMSE 1.1256 WSDM11-paper12-tables CoDoo 1.1333 1.1530 1 1 6 0 0 1 1 9 0 9 537 0.8493 0.8491 0.8447 0.8443 124 SR1VSS RMSE 1.1016 WSDM11-paper12-table AAAI Epinions 80% Improve RMSE 1.1621 1.1832 1.2073 1.2011 1.2294 1.1016 1.1013 1.0958 1.0954 125 SR1PCC Epinions RMSE 1.1013 WSDM11-paper12-table5 126 SR2VSS Epinions RMSE 1.0958 WSDM11-paper12-table5 MAE on Epinions (80% Training) Best baseline vs the proposed PMF 1.1544 1.1968 1.0954 WSDM11-paper12-table5 1.1492 127 SR2PCC Epinions RMSE SoRec 1.1293 RMSE on Epinions (80% Training) Conflicting between papers 169 SoRed Moviel en ▼ Proposed System (Table Extraction → Table Unification [Method Contribution] → DB Operation ) **Proposed Method (Hybrid Features and Learning)** Assumption1 (A1): Row/column header indication. $\min_{\phi,\psi} J_1(\phi,\psi) =$ $\sum_{\Gamma = [\mathcal{R}. C, \dots]} \sum_{(w, P) \in \mathcal{R} \cup C} \sum_{l \in \mathcal{L}}$ **Question 1: Find related methods, metrics, and datasets.** Assumption2 (A2): Row/column type consistency. **Data Science Paper** Table 4: Performance on the <u>Twitter</u> testing data $\sum \phi(p \in P^{(l)}) - |P| \cdot \psi(w \in W^{(l)})$ set by different approaches. PDF Collection Query: How many methods were used for the Epinions dataset? Assumption3 (A3): Cell context completeness. $\max_{\phi} J_2(\phi) =$ $\sum_{T=|\mathcal{R},C,\dots|} \sum_{P\in\mathcal{R}\cup C} \sum_{p\in P} \phi(p \in P^{(l^*(P))})$ 0.693 0.727 0.722 0.561 0.573 0.553 0.8(·, ·, ·) 37 0.717 Ensemable1 (E1): Structural concept embeddings. select count(distinct Method) from ERD where Dataset="Epinions" 0.730 PDF Table $\sum_{T=[..., B(B_1, B_2, B_3)]} |$ $\max_{\pm} J_3(\phi) =$ Ensemable2 (E2): Semantic concept embeddings. $| \cup_{k \in \{1,2,3\}} l_k^* |$ 0.831 0.805 0.818 0.809 Cropping 36. ("UserMean", "ItemMean", "Trust", "NMF", "SVD", "TCF" ...) 6 TableUni-R = A1+A2+A3 ; TableUni-L = E1+E2 exp(S(u, C))P(u|C) = $\overline{\sum_{p \in \mathcal{P}} exp(S(p, C))} \quad \texttt{Method}$ TableUni-(R+L) = TableUni-R + TableUni-L Question 2: Find top-performing methods on a dataset. Table Table Template B(p<sup>(R)</sup>, p<sup>(C)</sup>, d) P<sup>(R)</sup> B(p<sup>(R)</sup>, p<sup>(C)</sup>, d) Recognition Extraction Query: What are the top 3 methods on <u>Amazon</u> in terms of <u>F1</u>? P(C1) P(C1) Method Precision Recall F1-Score Ma-AUC Accuracy P(R) B(p(R), p(C1), p(C1)) TableUni-R 0.6908 0.6479 0.6807 0.6542 0.8601 select Method, Score from ERD where Dataset = "Amazon" and Table p(c) p(c) ÷., Metric = "F1" order by Score limit 3; TableUni-L 0.6333 0.5921 0.6187 0.6072 0.7264 B(p<sup>(R<sub>1</sub>)</sup>, p<sup>(R<sub>2</sub>)</sup>, p<sup>(C)</sup>) Cleaning 6 TableUni-(R+E1) 0.7505 0.7007 0.7443 0.7115 0.8705 "LEMON" (0.953), "LEMON-auto" (0.91), "LC" (0.815). TableUni-(R+E2) 0.8175 0.7821 0.7777 0.7798 0.8920 Row and Column Rule-based classifiers Question 3: Find conflicting reported numbers

Question 5. 1 mu connecting reported numbers.					
Dataset	(%)	SLEEC	FastXML	PfastreXML	PDSparse
AmazonCat	P@1	90.56/89.19	94.02/93.10	86.06/89.94	87.43/89.31
-13K	P@3	76.96/75.17	79.93/78.18	86.06/77.24	87.43/74.03
	P@5	62.63/61.09	64.90/63.38	63.65/63.53	56.70/60.11
Delicious	P@1	47.78/47.03	48.85/43.20	26.66/37.62	37.69/34.37
-200K	P@3	42.05/41.67	42.84/38.68	23.56/35.62	30.16/29.48
	P@5	39.29/38.88	39.83/36.21	23.21/34.03	27.01/27.04
WikiLSHTC	P@1	58.34/55.57	50.01/49.75	57.17/58.10	60.70/61.26
-325K	P@3	36.70/33.06	32.83/33.10	37.03/37.61	39.62/39.48
	P@5	26.45/24.07	24.13/24.45	27.19/27.69	29.20/28.79

Name Type

Classification

Experimental

Evidence DB

Construction: Table Integration

DB Operations for QA on

**Experimental Evidence** 

Table

Unification

(three assumptions)

Learning-based classifiers

(two embedding methods)

**Acknowledgement** 

0.6980

0.7567

0.6474

0.8307

TableUni-(A1+L)

TableUni-(A2+L)

TableUni-(A3+L)

TableUni-(R+L)

This work was advised by Dr. Meng Jiang. This work was supported by NSF Grant IIS-1849816.

0.6531

0.7123

0.6052

0.8195

0.6756

0.7250

0.6306

0.8053

0.6612

0.7179

0.6129

0.8104

0.8123

0.8633

0.7443

0.9000