1 Scores for models trained on all features

Model Name	Dataset	NDCG@3	NDCG@5	NDCG@10	MAP
MART	Train	0.42266197	0.43219629	0.45540289	0.61773220
	Validation	0.41018495	0.42002789	0.44378358	0.61134855
	Test	0.40878214	0.41958235	0.44310115	0.61132830
LambdaMART	Train	0.29391117	0.30653134	0.33549083	0.57249296
	Validation	0.29505739	0.30721627	0.33530975	0.57209706
	Test	0.29457428	0.30806687	0.33525631	0.57218601
XGBoost	Train	0.32926949	0.32926949	0.39625375	–
	Validation	0.34710293	0.37224366	0.38739088	–
	Test	0.38739088	0.36108438	0.37953240	–

2 Scores for models trained only on content features

Model Name	Dataset	NDCG@3	NDCG@5	NDCG@10	MAP
MART	Train	0.40739356	0.41564600	0.43769764	0.61104852
	Validation	0.39362988	0.40190963	0.42464190	0.60399381
	Test	0.39238632	0.40193676	0.42430687	0.60480526
LambdaMART	Train	0.29391117	0.30653134	0.33549083	0.33549083
	Validation	0.29505739	0.30721627	0.33530975	0.57209706
	Test	0.29457428	0.30806687	0.33525631	0.57218601
XGBoost	Train	0.57218601	0.46964361	0.49289840	–
	Validation	0.46040863	0.47144130	0.48663724	–
	Test	0.45649342	0.46673989	0.48397843	–

3 Scores for models trained on weak labels

Model Name	Dataset	NDCG@3	NDCG@5	NDCG@10	MAP
MART	Train	0.30944851	0.31258516	0.32182022	0.50268692
	Validation	0.30968153	0.31274478	0.32147090	0.50236534
	Test	0.30765263	0.31162291	0.32038378	0.50251630
LambdaMART	Train	0.29083979	0.30228472	0.32947254	0.53670266
	Validation	0.29157301	0.30334596	0.33067105	0.53717103
	Test	0.28940686	0.30178648	0.32966530	0.53666492
XGBoost	Train	0.28063948	0.26455529	0.19490686	–
	Validation	0.28077775	0.26468902	0.19504947	–
	Test	0.28074822	0.26465959	0.19502795	–