Improving Multimodal Social Media Popularity Prediction via Selective Retrieval Knowledge Augmentation

Xovee Xu, Yifan Zhang, Fan Zhou, and Jingkuan Song

University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China

xovee.xu@gmail.com, yifanzhang@std.uestc.edu, fan.zhou@uestc.edu.cn, jingkuan.song@gmail.com

Motivation

Existing: Existing approaches treat user-generated content (UGC) prediction as an isolated process, overlooking interconnected nature of UGCs.

Direction: Using retrieval-augmented technique to enhance UGC contextual learning is a promising direction.

However: (1) A simple retrieval strategy that relies solely on semantic similarities cannot fully reflect the contextual information of complex social UGCs. (2) Not all retrieved UGCs may be truly relevant to the query UGCs, inevitably introducing noises.

Solution

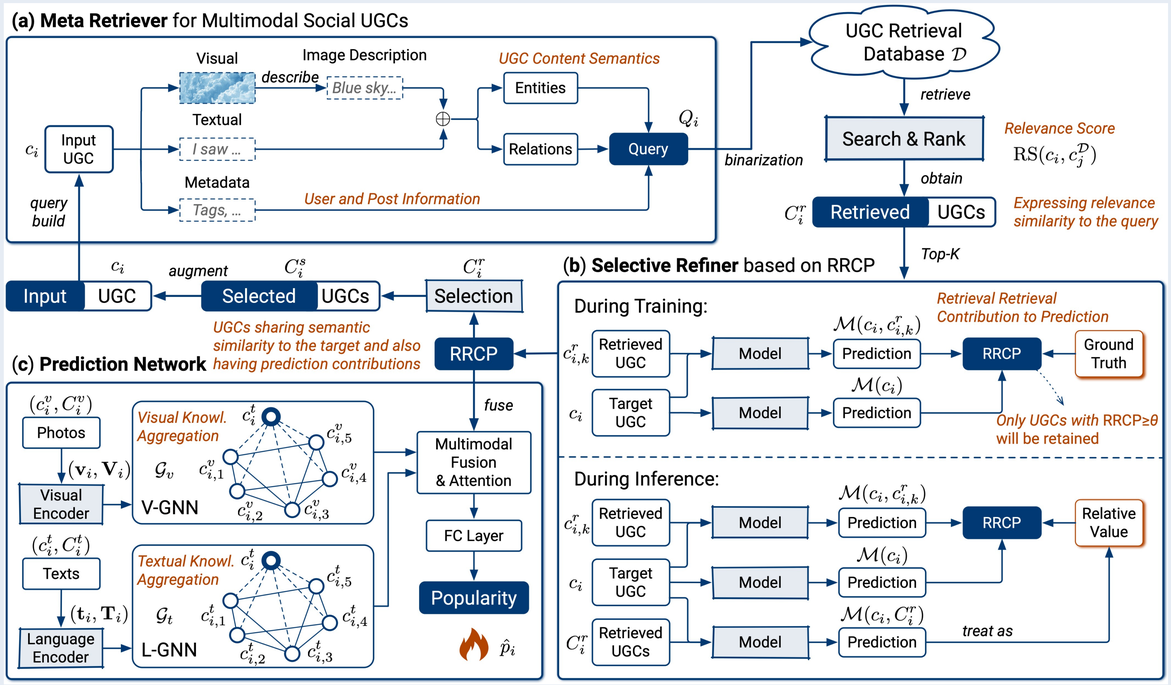
Retrieval Knowledge Augmentation: We choose to retrieve relevant UGCs to enhance the contextual information for the query UGC for multimodal social media popularity prediction.

Meta Retriever: We not only consider multimodal UGC semantics, but also social contexts of UGCs by incorporating diverse metadata.

Selective Refiner: We design a new measure, termed Relative Retrieval Contributions to Prediction (RRCP), to quantifies the gains in prediction of the retrieved UGCs.

VL-GNNs: To effectively aggregate the retrieved knowledge, we introduce a vision-language graph neural networks module, coupled with an RRCP-Attention-based prediction network.

Proposed Method: SKAPP



(a) **Meta Retriever**: constructs query by integrating UGC content semantics with metadata information.

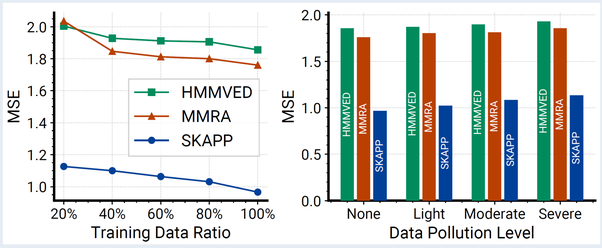
(b) **Selective Refiner**: employs a new relative retrieval contribution to prediction (RRCP) measure, which is inspired by the conditional cross-mutual information, to select UGCs that have positive gains in prediction, filtering out irrelevant and noisy UGCs.

(c) **Prediction Network**: leverages vision-language graph neural networks to aggregate contextual knowledge from selected UGCs with an RRCP-Attention-based module for accurate prediction.

Ablation Study

Variant	ICIP	SMPD	Instagram
Ablation of SKAPP's Modules			
w/o Retrieval	1.5614	4.0443	3.2734
w/o Meta Retriever	1.9006	4.1353	5.2537
w/o Selective Refiner	1.1004	2.0854	2.6332
w/o VL-GNN	1.1223	2.1056	2.7178
w/o RRCP-Attention	1.0761	1.9606	2.1636
Ablation of UGC modalities			
w/o Visual	1.1770	2.3567	2.4851
w/o Textual	1.1829	2.7037	2.3582
w/o Metadata	1.8188	4.0359	5.2537
Ablation of Retrieving Strategies			
retrieval based on Photo	1.9006	4.1353	5.7644
retrieval based on Texts	1.9653	3.9958	4.7259
retrieval based on Metadata	1.6280	2.6945	3.8679
retrieval based on FLICO	1.8255	3.8562	5.4786
retrieval based on NIPA	1.9321	4.1687	5.2468
retrieval based on MMRA	1.9627	4.0507	4.6693
SKAPP (Full)	0.9662	1.8196	2.0936

Robustness



Main Results on Three Social UGC Datasets

Significant Performance Improvements: Combining the meta retriever, selective refiner, and VL-GNN-based prediction module, our proposed SKAPP surpasses the baselines by a large margin.

Method	Type	ICIP			SMPD			Instagram		
		MSE	MAE	SRC	MSE	MAE	SRC	MSE	MAE	SRC
SVR	Feature	1.9009	0.8941	0.5241	6.2996	2.0208	0.2163	7.0534	1.9695	0.4035
HyFea	Feature	1.9013	1.0181	0.4497	4.7429	1.7080	0.4677	4.7132	1.6924	0.4708
MFTM	Feature	1.8970	0.9772	0.4156	4.0222	1.5481	0.5849	4.3073	1.6132	0.5321
CLSTM	Deep	1.8724	0.9823	0.4654	3.9143	1.5005	0.5888	4.2431	1.5882	0.5396
HMMVED	Deep	1.8556	0.9497	0.4524	3.7154	1.3636	0.6352	4.2461	1.6017	0.5385
DLBA	Deep	2.2290	1.0097	0.3614	4.8693	1.7021	0.4387	5.1425	1.7527	0.4007
MASSL	Deep	1.9446	0.9278	0.4499	5.5670	1.8427	0.5271	7.8583	2.2274	0.5188
BLIP	Deep	2.0646	0.9961	0.3603	4.3884	1.6340	0.5269	5.2436	1.8058	0.3762
CBAN	Deep	1.8098	0.9309	0.4727	4.0443	1.5123	0.5754	4.2808	1.5894	0.5426
NIPA	Retrieval	1.9999	0.9980	0.3989	4.2538	1.6532	0.4086	4.0209	1.5565	0.5696
MMRA	Retrieval	1.7600	0.8684	0.5439	3.5119	1.3730	0.6423	3.9456	1.5070	0.5806
SKAPP	Retrieval	**0.9662**	**0.6367**	**0.6965**	**1.8196**	**0.8249**	**0.8414**	**2.0936**	**1.0369**	**0.8272**
(improv.)		39.61%↑	26.68%↑	28.06%↑	48.19%↑	39.51%↑	31.00%↑	46.94%↑	29.06%↑	42.47%↑

Table 2: Social media popularity prediction performance comparison between our proposed SKAPP model and eleven baselines on three large-scale real-world datasets. The best results are marked in bold and the second best are underlined.

Key Findings

I. **We Need Diverse UGC Similarities**: complex social UGCs cannot be compared solely by semantic similarities.

II. **Quality Over Quantity**: Retrieving quality UGCs is more important than retrieving more (potentially noisy) UGCs.

III. **Effective Aggregation**: After retrieving and selecting UGCs for augmentation, effectively aggregating UGCs furthers boosts performance.

Future Work

I. **Define UGC Similarities**: What types of UGC contexts is more useful for prediction? We may need further investigations & new definitions.

II. **New Ways to Select UGCs**: Can we design a new lightweight but powerful selection algorithm?

III. **Improve Efficiency**: Dynamically determine the # of retrieved UGCs for each target.

IV. **End to End**: Running and improving the retrieval algorithm during model training.

AAAI-25 / IAAI-25 / EAAI-25

FEBRUARY 25 – MARCH 4, 2025 | PHILADELPHIA, USA

University of Electronic Science and Technology of China

UESTC ICDM Lab

Source Codes