[IS3-164] A Comprehensive Benchmark for Graphical Abstract Recommendation and Evaluation

川田 拓朗, 北田 俊輔, 根本 颯汰, 彌冨 仁 法政大学大学院 理工学研究科 {takuro.kawada.3g@stu.,iyatomi@}hosei.ac.jp

Graphical Abstract (GA)とは？

- 🖼️ 論文誌に提出される「研究内容を要約した画像 / 動画」
- 😄 論文の注目度, SNSにおける拡散力を高める
- 🤔 近年, 論文内の図1 (Teaser) が事実上の GA として働く
- ➜ **これらの視覚的資源を活用し, 科学的伝達の効率を高めたい**

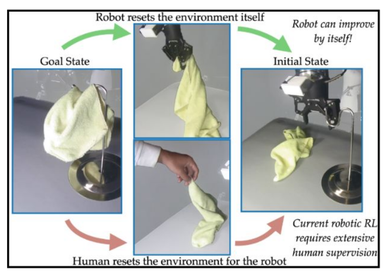
GA設計支援の基盤を構築

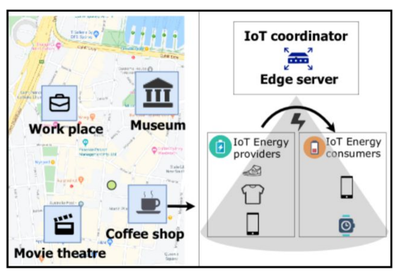
■GA / Teaser を含む初の論文データセット SciGA-145k

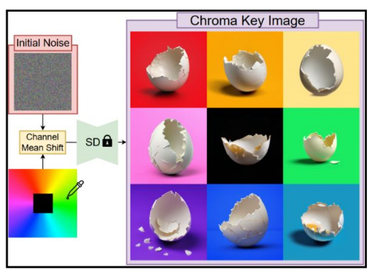
A Dataset of Papers, Figures, and Visual Summaries

- 145k Full Text
- 1.1M Figures
- 30k **GA / Teaser**
- Title
- Authors
- Abstract
- Research Fields
- Accepted Conference
- DOI

etc.

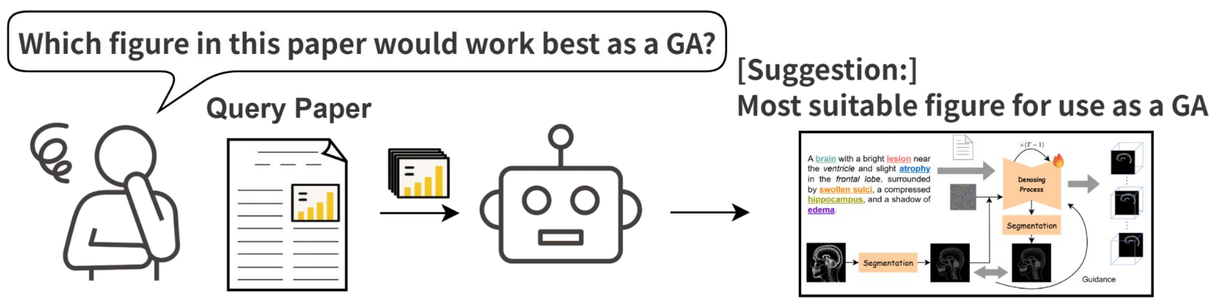




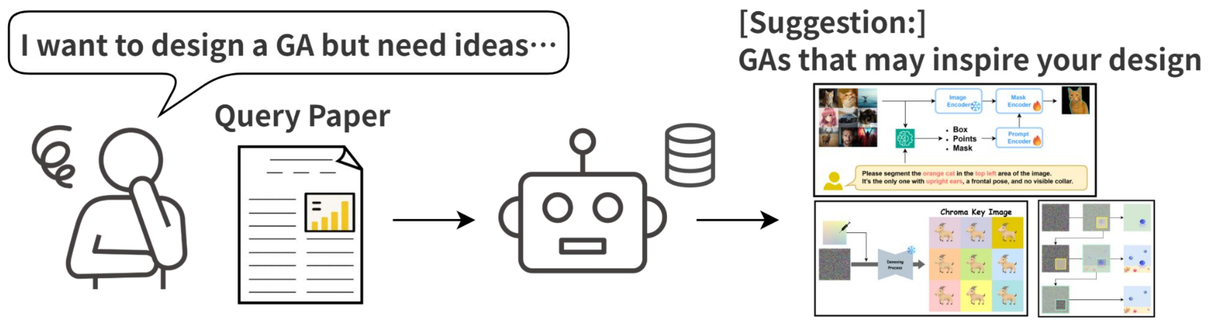


■SciGA-145k の活用で実現が期待される新たなタスク

1) Intra-GA Recommendation



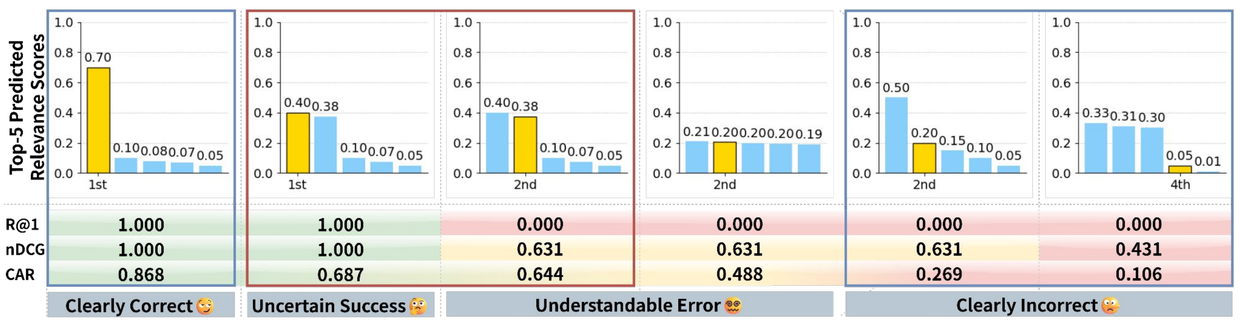
2) Inter-GA Recommendation



■推薦モデルが持つ自信を考慮した新たな指標 CAR

Intra-GA Recommendation Taskにおいて, 論文内にはラベルの付いた GA 以外にも GA として尤もらしい図が複数存在し, モデルはこれら候補間で悩み推論を誤る

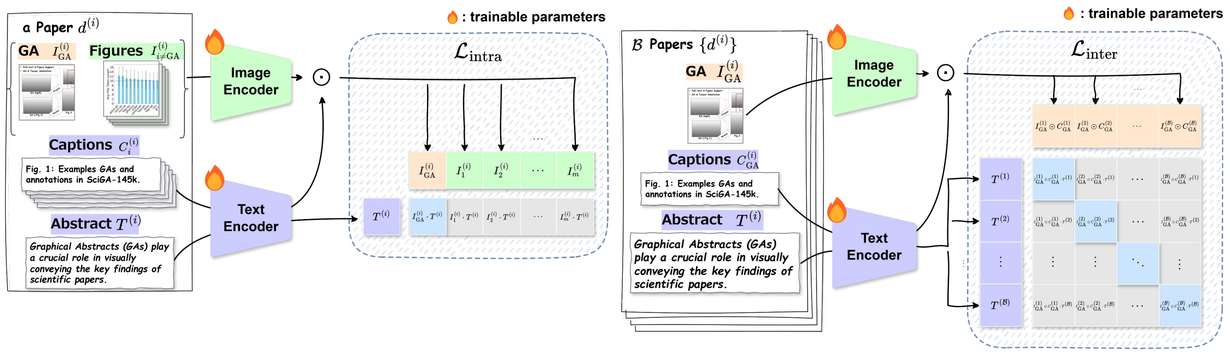
$$\mathrm{CAR@}k = \frac{p_{\mathrm{GT}}}{p_{\text{top-1}}}\left[1 - \frac{1}{2}\max\left(0, \frac{2H(\boldsymbol{P}) - \log k}{\log k}\right)\right]$$



➜ **GT の順位が同じでも, 分布を考慮して連続値を与える**

ベンチマーク手法

CLIP ベースの Text2Image Retrieval モデルを採用
Abstract をクエリとして, GA を検索・推薦



評価実験

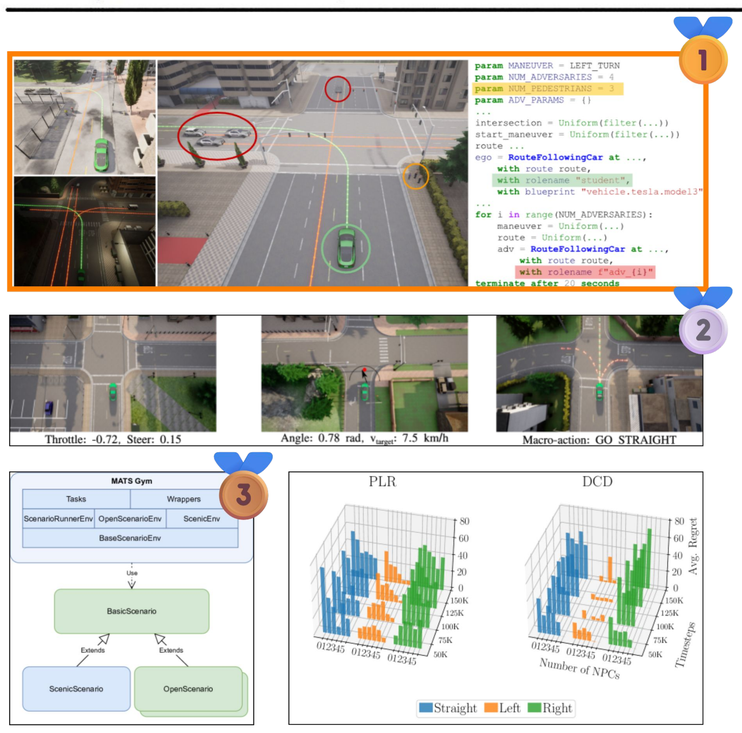
■Intra-GA Recommendation

↓異なるバックボーンに基づくIntra-GA推薦性能の比較

Backbone († Max Token Length)	R@1	R@2	R@3	MRR	nDCG@5	CAR@5 Mean	CAR@5 0.5↑
CLIP († 77)	0.628	0.822	0.902	0.771	0.816	0.610	0.689
X²-VLM († 40)	0.538	0.757	0.857	0.709	0.763	0.546	0.618
OpenCLIP († 77)	0.621	0.817	0.905	0.767	0.813	0.603	0.681
BLIP-2 († 512)	0.557	0.767	0.863	0.721	0.773	0.557	0.626
Long-CLIP († 248)	**0.637**	**0.826**	**0.914**	**0.778**	**0.824**	**0.615**	**0.691**

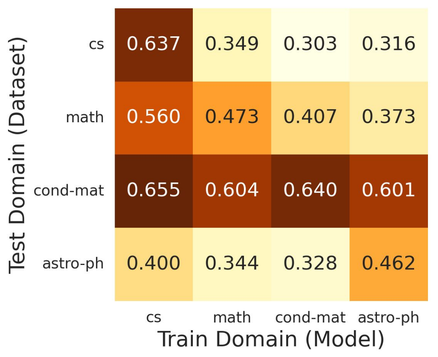
↓異なる研究分野におけるIntra-GA推薦性能の比較

Research Fields	Data Size	R@1	R@2	R@3	MRR	nDCG@5	CAR@5 Mean	CAR@5 0.5↑
cs	20,520	0.637	0.826	0.914	0.778	0.824	0.615	0.691
math	1,498	0.473	0.663	0.763	0.643	0.684	0.493	0.493
cond-mat	3,323	0.640	0.805	0.871	0.767	0.805	0.639	0.700
astro-ph	1,949	0.462	0.651	0.764	0.633	0.679	0.494	0.533



↑推薦結果の例 (GT はオレンジ枠)

↑研究分野ごとの GA の CLIP 埋め込み



↑Cross-Domain特性

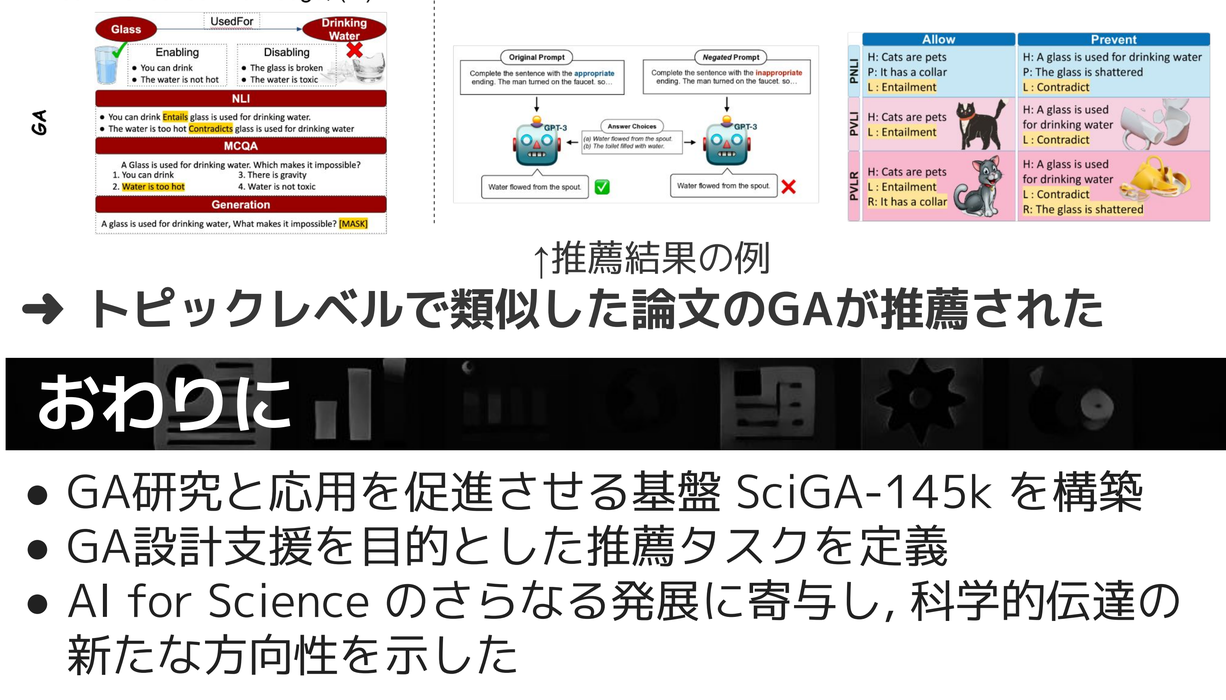
➜ **Teaser 文化が浸透している CS 分野の転移性能が高い**
➜ **CAR@*k*により, 詳細なモデルの振る舞いの分析が可能に**

■Inter-GA Recommendation

↓異なるバックボーンに基づくInter-GA推薦性能の比較

Backbone	Field-P@k top-5	Field-P@k top-10	Abs2Abs SBERT@k top-5	Abs2Abs SBERT@k top-10	GA2GA CLIP-S@k top-5	GA2GA CLIP-S@k top-10
CLIP	**0.755**	**0.742**	0.493 ± 0.098	0.479 ± 0.101	0.614 ± 0.067	0.611 ± 0.071
X²-VLM	0.415	0.399	0.254 ± 0.114	0.250 ± 0.119	0.555 ± 0.067	0.552 ± 0.072
OpenCLIP	0.749	0.737	0.489 ± 0.097	0.475 ± 0.100	0.615 ± 0.066	0.611 ± 0.069
BLIP-2	0.647	0.639	0.390 ± 0.105	0.382 ± 0.109	0.597 ± 0.067	0.596 ± 0.068
Long-CLIP	0.753	0.737	$\mathbf{0.498 \pm 0.098}$	$\mathbf{0.482 \pm 0.103}$	$\mathbf{0.614 \pm 0.070}$	$\mathbf{0.611 \pm 0.073}$

- **Field-P@*k***:「入力 Abst.」と「推薦された GA」の研究分野の一致度
- **Abs2Abs SBERT@*k***:「入力 Abst.」と「推薦された GA の論文の Abst.」の SBERT 埋め込みの cos 類似度
- **GA2GA CLIP-S@*k***:「著者が描いたGA」と「推薦されたGA」の CLIPScore



↑推薦結果の例

➜ **トピックレベルで類似した論文のGAが推薦された**

おわりに

- GA研究と応用を促進させる基盤 SciGA-145k を構築
- GA設計支援を目的とした推薦タスクを定義
- AI for Science のさらなる発展に寄与し, 科学的伝達の新たな方向性を示した