中国人民大學
RENMIN UNIVERSITY OF CHINA

GeWu-Lab
Gaoling School of Artificial Intelligence
Renmin University of China

Efficient Quantification of Multimodal Interaction at Sample Level

Zequn Yang, Hongfa Wang, Di Hu*

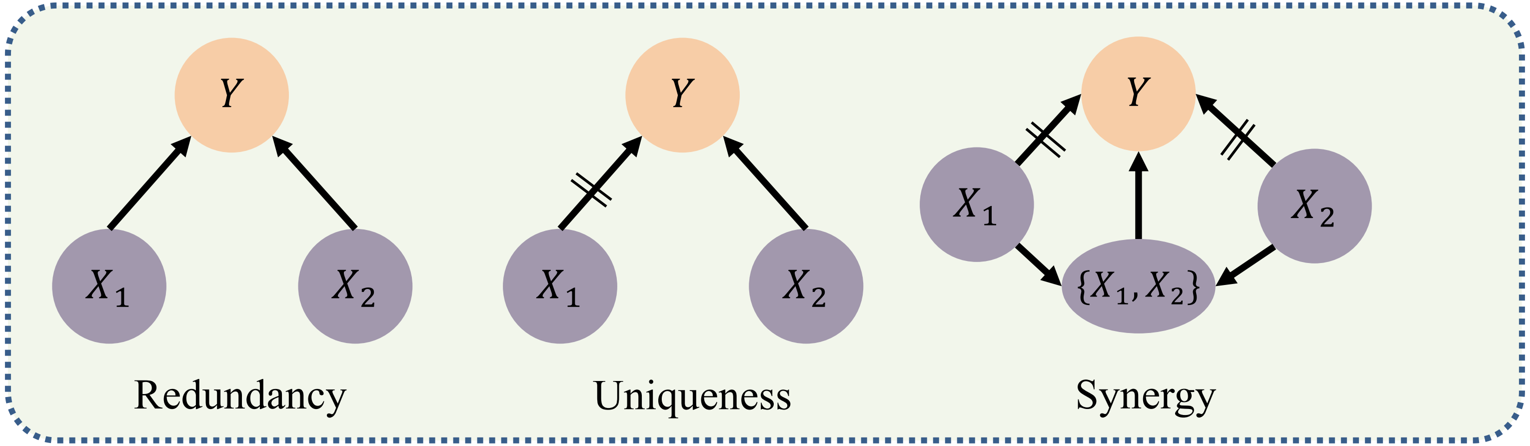
zqyang@ruc.edu.cn

2025-06-16

GeWu-Lab
Gaoling School of Artificial Intelligence
Renmin University of China

Multimodal Interaction

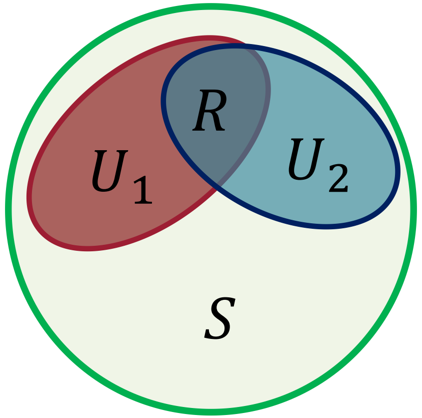
Multimodal interaction describe the way information contains in each modalities or their integration, including Redundancy, Uniqueness and Synergy.



GeWu-Lab
Gaoling School of Artificial Intelligence
Renmin University of China

Partial Information Decomposition [1]

According to Partial Information Decomposition, multimodal information $I(X_1, X_2; Y)$ can be divided into four distinct and positive components.



$I(X_1, X_2; Y)$

$$I(X_1; Y) = R + U_1$$

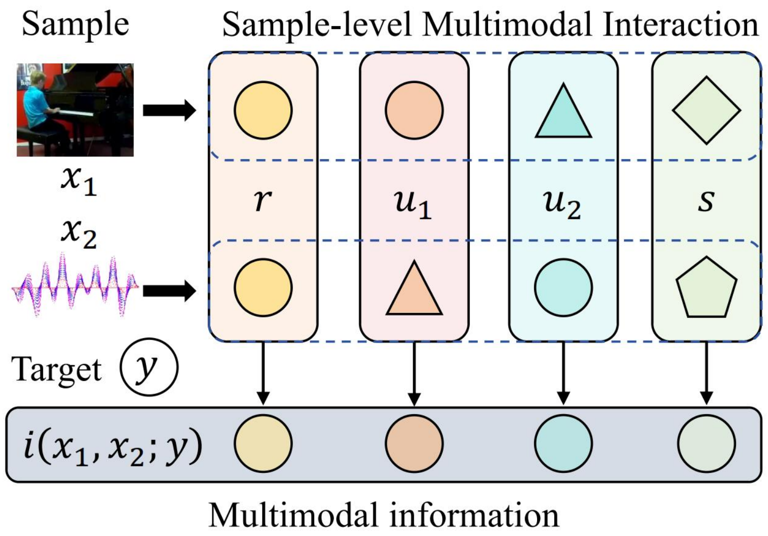
$$I(X_2; Y) = R + U_2$$

$$I(X_1, X_2; Y) = R + U_1 + U_2 + S$$

[1] P. L. Williams and R. D. Beer, "Nonnegative decomposition of multi-variate information," *arXiv preprint arXiv:1004.2515, 2010*.

Sample-level Interaction

Interaction within each sample can vary significantly. By contrast to dataset-level interaction [2,3], sample-level interaction provides fine-grained information and enhances interpretability for multimodal learning [3,4].



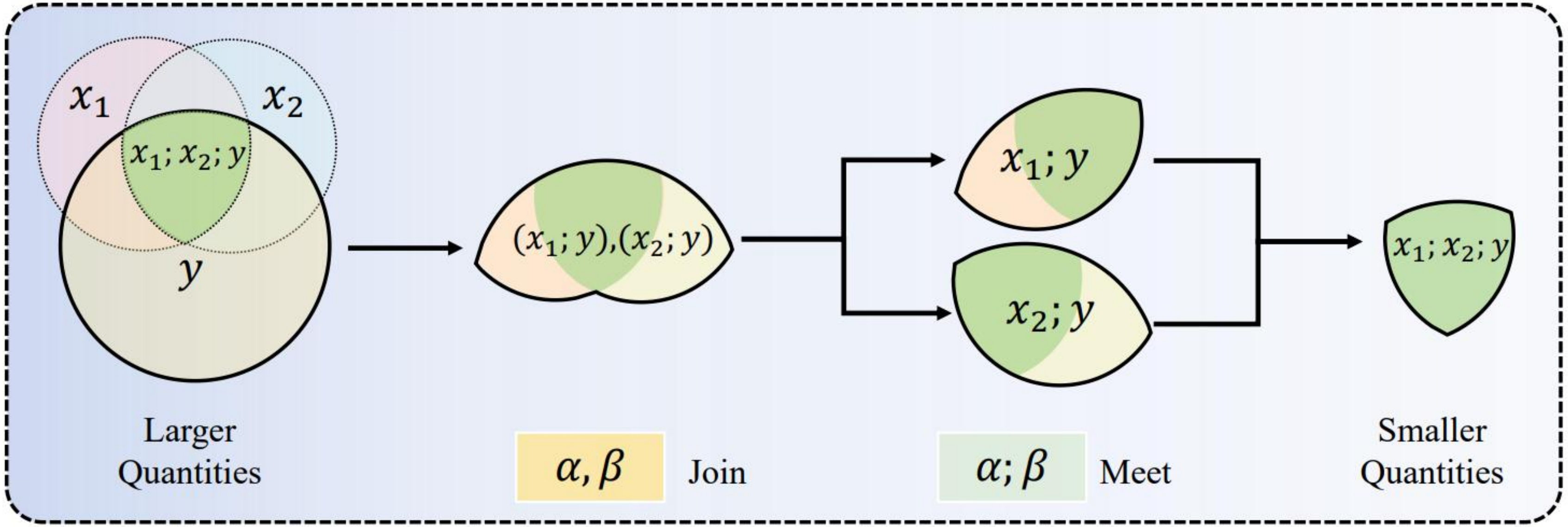
[2] N. Bertschinger, J. Rauh, E. Olbrich, J. Jost, and N. Ay, "Quantifying unique information," *Entropy*, vol. 16, no. 4, pp. 2161–2183, 2014.

[3] P. P. Liang, Y. Cheng, X. Fan, C. K. Ling, S. Nie, R. Chen, Z. Deng, F. Mahmood, R. Salakhutdinov, and L.-P. Morency, "Quantifying & modeling multimodal interactions: An information decomposition framework," in *Advances in Neural Information Processing Systems*, 2023.

[4] J. T. Lizier, B. Flecker, and P. L. Williams, "Towards a synergy-based approach to measuring information modification," in *2013 IEEE Symposium on Artificial Life (ALIFE)*. IEEE, 2013, pp. 43–51.

Method

■ Interaction Decomposition Framework



We propose the interaction decomposition framework and apply reasonable measure to ensure the information quantities monotonically decrease along the path.

Method

GeWu-Lab
Gaoling School of Artificial Intelligence
Renmin University of China

■ Lightweight Estimation over Continuous Distribution

We use the KNIFE estimator [5] to compute continuous entropy, then derive information components and measure sample-level interactions.

Algorithm 1 Lightweight Sample-wise Multimodal Interaction Estimation (LSMI) Algorithm

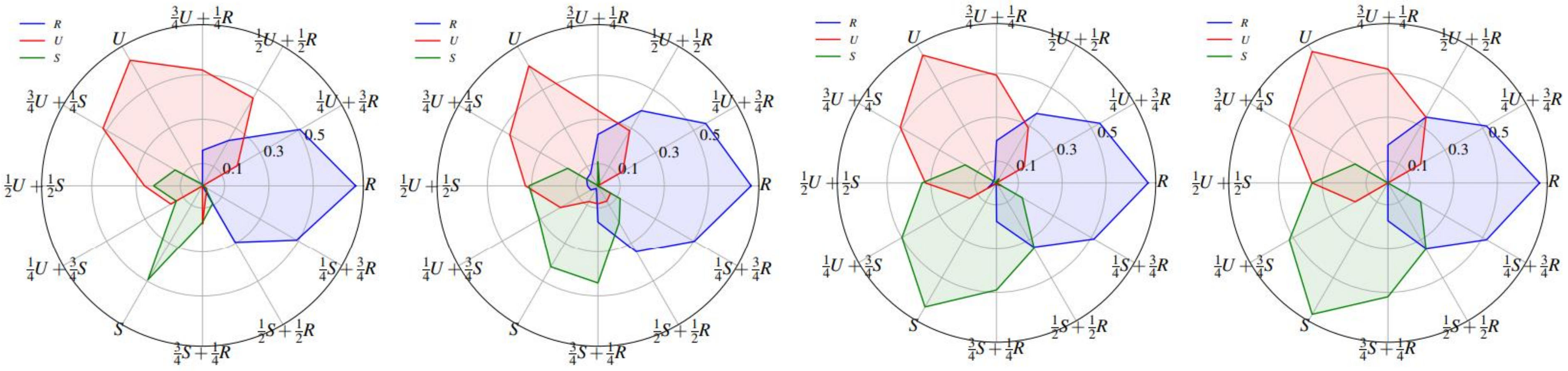
1: **Input:** Bimodal data x_1, x_2, target y; discriminative models $p(y|x_1, x_2), p(y|x_1), p(y|x_2)$.
2: **Initialize:** Entropy estimators $h_{\theta_1}(\cdot), h_{\theta_2}(\cdot)$.
3: Train entropy estimators $h_{\theta_1}, h_{\theta_2}$ using Equation 7 on data from $p(x_1), p(x_2)$ respectively.
4: Compute sample-wise $h(x_1), h(x_2)$ using $h_{\theta_1}, h_{\theta_2}$; then compute $h(x_1|y), h(x_2|y)$ via Equation 8.
5: Compute pointwise redundancy indicators r^+, r^- via Equation 5; then redundancy $r \leftarrow r^+ - r^-$.
6: Compute pointwise $i(x_1; y), i(x_2; y), i(x_1, x_2; y)$ using $p(y|x_1), p(y|x_2), p(y|x_1, x_2)$; then derive interactions u_1, u_2, s via Equation 2.
7: **Output:** Sample-wise interactions r, u_1, u_2, s.

[5] G. Pichler, P. J. A. Colombo, M. Boudiaf, G. Koliander, and P. Piantanida, "A differential entropy estimator for training neural networks," in *International Conference on Machine Learning*. PMLR, 2022, pp. 17 691– 17 715.

Experiment

■ Validation

We validate the precision of our method over sythetic dataset with preset interaction.



(a) PID-CVX (b) PID-Batch (c) LSMI (d) GT

Experiment

Estimation

Dataset / Interaction	KS R	KS U_1	KS U_2	KS S	Food-101 R	Food-101 U_1	Food-101 U_2	Food-101 S	UR-Funny R	UR-Funny U_1	UR-Funny U_2	UR-Funny S	CMU-MOSEI R	CMU-MOSEI U_1	CMU-MOSEI U_2	CMU-MOSEI S
PID-Batch	3.16	0.02	0.19	0.01	4.23	0.24	0.00	0.14	0.02	0.03	0.01	0.06	0.18	0.34	0.02	0.03
LSMI	3.28	0.11	0.00	0.03	4.19	0.34	0.00	0.08	0.02	0.12	0.01	0.24	0.13	0.22	0.01	0.00
Human	2.32	1.61	1.45	0.48	4.06	0.92	0.05	0.00	2.30	2.73	2.33	2.50	3.27	3.37	2.87	1.03

Table 2: Comparison of average interaction over various real-world datasets.

We apply LSMI to estimate dataset interactions and compare those learned by different multimodal methods.

Method	R	U_1	U_2	S
Feature-level fusion				
Joint	3.165	0.143	0.000	0.122
MMIB	3.284	0.113	0.000	0.030
Bilevel	2.604	0.552	0.000	0.277
Decision-level fusion				
Additive	3.397	0.006	0.000	0.029
Weighted	3.399	0.010	0.000	0.024
QMF	3.400	0.002	0.000	0.032
Additional Regulation				
Mod-drop	3.163	0.134	0.000	0.116
Alignment	3.372	0.015	0.000	0.040
Recon	2.984	0.311	0.000	0.139

Table 4: Comparison of interaction components across different multimodal learning methods on the KS dataset.

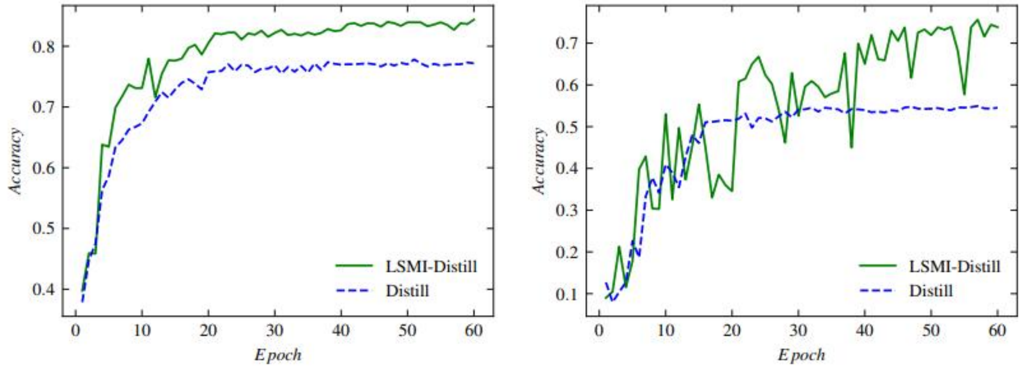
Experiment

GeWu-Lab
Gaoling School of Artificial Intelligence
Renmin University of China

Application

Data	KS			CREMA-D		
	V+A	V	A	V+A	V	A
All	0.854	0.818	0.727	0.795	0.684	0.725
Low	0.850	0.805	0.729	0.782	0.702	0.715
High	0.877	0.824	0.726	0.801	0.688	0.728

Table 6: Performance comparison of ImageBind model fine-tuned on complete dataset (All), low-redundancy subset (Low), and high-redundancy subset (High) across unimodal and multimodal settings.



(a) KS dataset (b) UCF dataset

Figure 5: Validation on LSMI-based distillation approach.

Partitioning Redundant data suitable for specific learning paradigm (ImageBind).

Distillation in different ways according to data specific multimodal interaction.

中国人民大學
RENMIN UNIVERSITY OF CHINA

GeWu-Lab
Gaoling School of Artificial Intelligence
Renmin University of China

Thank You for listening!

Zequn Yang, Hongfa Wang, Di Hu*

zqyang@ruc.edu.cn

2025-06-16