

Table 5: Statistics of the experimental datasets with multi-modal item Visual(V), Acoustic(A), and Textual(T) contents.

Dataset	TikTok			Amazon-Baby		Amazon-Sports	
Modality	V	A	T	V	T	V	T
Embed Dim	128	128	768	4096	1024	4096	1024
User	9319			19445		35598	
Item	6710			7050		18357	
Interactions	59541			139110		256308	
Sparsity	99.904%			99.899%		99.961%	