

# Improving Hierarchical Text Clustering with LLM-guided Multi-view Cluster Representation for Interaction Drivers in Contact Centers

Anup Pattnaik, Cijo George, Rishabh Kumar Tripathi<sup>1</sup>, Sasanka Vutla<sup>1</sup>, Jithendra Vepa

## Empowering Contact Centers to drive Insights from the Agent-Customer Interactions

### Motivation

- ◆ Contact Center Interaction Drivers are hierarchical in nature: L1, L2 levels
- ◆ Current SOTA methods rely heavily on LLM APIs, increasing overall costs.
- ◆ Lack of adaptability for incorporating different perspectives at L1 & L2 levels.
- ✓ **Proposed solution: LLM-guided Multi-View Clustering**

### Takeaways

- ◆ **Improved quality of top-level clusters** on average **Silhouette Score by upto 70%** and **Human Preference Scores by 36.7%** compared to standard agglomerative clustering for the business use-case.
- ◆ **Achieved SOTA on public datasets for non-hierarchical clustering use-cases** based on NMI and ACC scores, with minimal number of LLM queries.
- ◆ **Contributed two newly labeled datasets** for hierarchical clustering to support advancements within the research community.

### Methodology

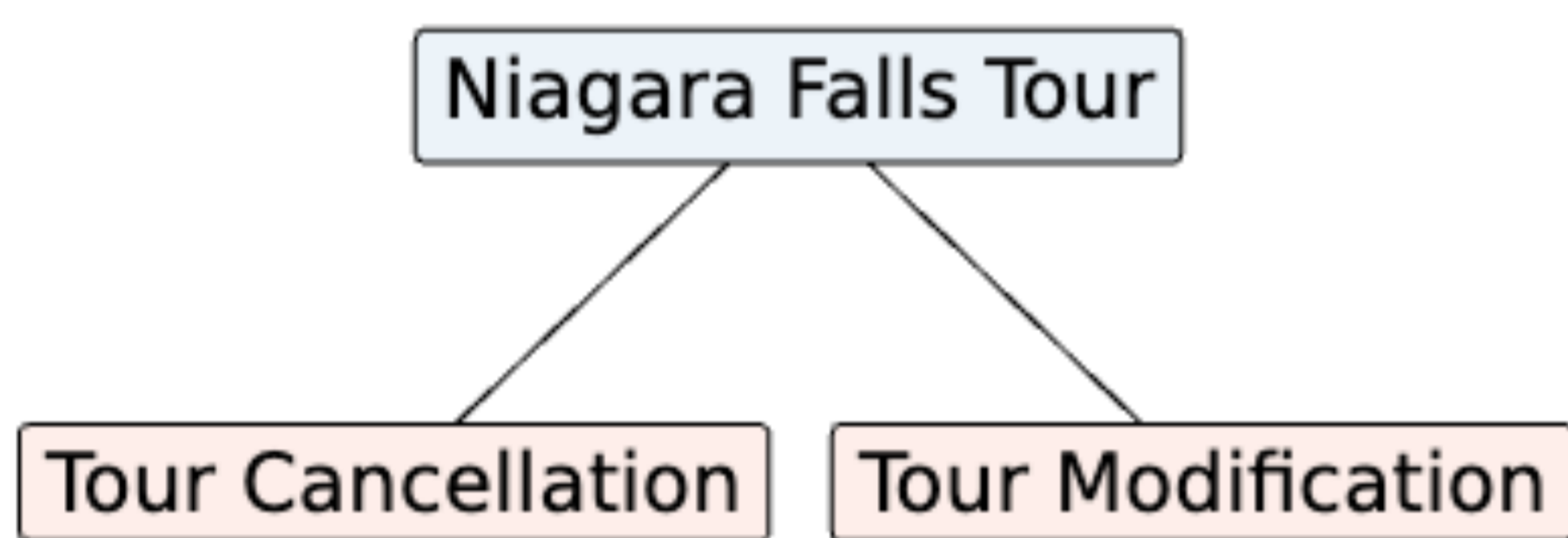


Figure 2: L1 and L2 level Interaction Drivers

Driver Name	Driver Description
Niagara Falls Tour	Queries for tour offered to Niagara Falls.
Tour Cancellation	Tour cancellation due to some reasons.
Tour Modification	Changes made to existing tour bookings.

Table 1: L1 and L2 level Interaction Drivers

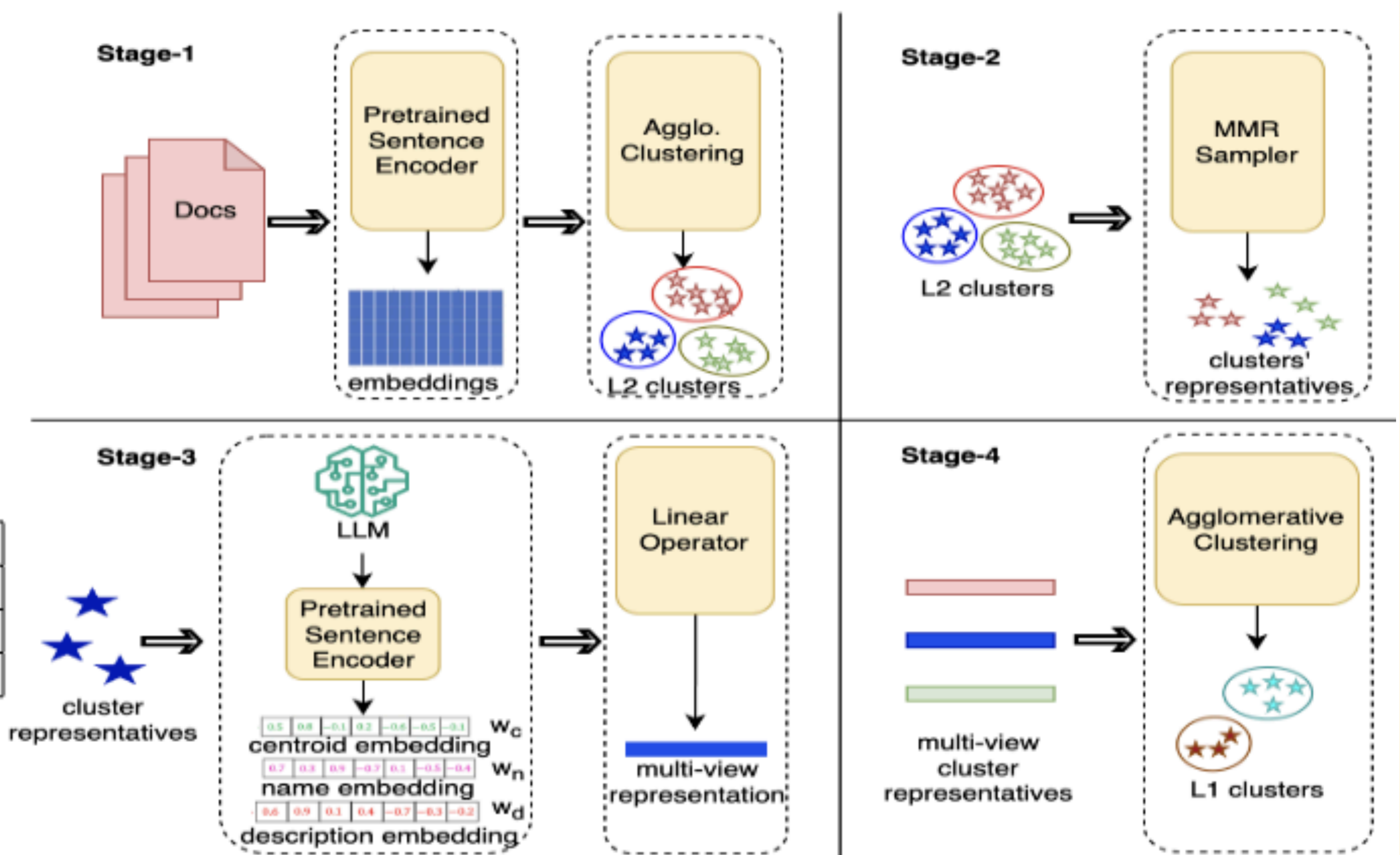


Figure 1: The proposed Approach

### Results

Approaches	Quick Commerce		Education		Travel	
	Silhouette	HPS	Silhouette	HPS	Silhouette	HPS
Std. Agglomerative w/ MPNet	0.035	3.262	0.038	3.39	0.039	3.411
Std. Agglomerative w/ Instructor	0.044	3.423	0.040	3.445	0.043	3.484
Proposed Approach w/ MPNet	<b>0.053</b>	4.412	<b>0.059</b>	4.563	<b>0.064</b>	4.57
Proposed Approach w/ Instructor	<b>0.065</b>	4.682	<b>0.068</b>	4.711	<b>0.071</b>	4.728

Table 2: Silhouette and Human Preference Scores (HPS) of L1 clusters across different approaches and domains. Note: HPS is computed on the basis of 5-point Likert scale.

- ◆ Removing centroid view significantly reduces average silhouette scores across all domains.
- ◆ Name view contributes more significantly to the clustering quality than the description view.

Approach	Banking77			CLINC150		
	NMI	ACC	Silhouette	NMI	ACC	Silhouette
Std. Agglomerative w/ MPNet	73.2	58.6	0.072	81.2	74.2	0.083
Std. Agglomerative w/ Instructor	76.4	60.1	0.085	84.5	76.1	0.092
IDAS	82.84	67.43	-	93.82	85.48	-
ClusterLLM w/ Instructor	<b>85.15</b>	<b>71.2</b>	-	94	83.8	-
Proposed Approach w/ MPNet	82.9	67.5	0.108	92.9	82.6	0.12
Proposed Approach w/ Instructor	84.9	69.6	<b>0.12</b>	<b>94.2</b>	<b>86.2</b>	<b>0.145</b>

Table 4: Evaluation on Public Intent Classification Datasets

$w_c$	$w_n$	$w_d$	Q. Comm.	Education	Travel
1.0	0.0	0.0	0.0458	0.049	0.054
0.0	1.0	0.0	0.032	0.04	0.051
0.0	0.0	1.0	0.03	0.038	0.044
0.5	0.5	0.0	0.046	0.042	<b>0.064</b>
0.0	0.5	0.5	0.034	0.039	0.046
0.5	0.0	0.5	0.044	0.04	0.048
0.34	0.33	0.33	0.05	<b>0.059</b>	0.056
0.5	0.25	0.25	<b>0.053</b>	0.053	0.058

Table 3: Impact of different views