AdaFlow: Non-blocking Inference with Heterogeneous Multi-modal Mobile Sensor Data

Fengmin Wu[a], Sicong Liu[a,*], Bin Guo[a], Xiaocheng Li[a],Yuan Gao[a], Zhiwen Yu[a,b]
[a] School of Computer Science, Northwestern Polytechnical University, Xi'an, China
[b] Harbin Engineering University, Harbin, China
{fenny}@mail.nwpu.edu.cn, {scliu, guob}@nwpu.edu.cn
{xiaochenli, rudygao}@mail.nwpu.edu.cn, {zhiwenyu}@nwpu.edu.cn

***Abstract*—Multi-modal deep learning offer conceptual advantages by integrating information from different vantage points. However, data flow blocking and corruption due to modalities asymmetric which make exist methods can hardly balance latency and accuracy still pose many challenges. To address these challenges, this paper proposes AdaFlow to establish a multi-modal mapping mechanism using an affinity matrix, achieving non-blocking data flow in multi-modal systems.**

***Index Terms*—Multi-modal, Affinity Matrix, Non-blocking Data Flow.**

I. Introduction

With the widespread of IoT devices, multi-modal inference has achieved success in numerous fields and often exhibits higher accuracy compared to single modality, applications like speech recognition, self-driving vehicles, and medical image segmentation. Multi-modal inference have advantages over single modal due to: 1) Enhanced environmental perception, 2) Synthesis of physical properties, 3) Independent cross-validation. The above points discussed represent concrete instances of implicit data contribution, the corresponding is explicit data importance estimation which has comprehensive existing works. However a few works in implicit data contribution,because quantifying this implicit contribution between heterogeneous modalities is challenging due to dynamic data collection and task characteristics.

Efficient multi-modal inference relies on well-constructed data, yet issues persist with data collected by multi-modal systems, the reasons are two-fold: 1) The one is inconsistent data arrival rates among heterogeneous mobile sensors, causing delays at end devices. For instance, video streams often lag behind audio streams due to bandwidth variations. Existing solutions struggle to balance accuracy and latency: high-precision methods block data streams, while others sacrifice accuracy for speed through frame sampling. 2) The another one is data size differences and mobile sensor state fluctuations, resulting in partial or corrupted data reception. For instance, Waymo's autonomous vehicles integrate diverse sensors with varying data rates and potential for corruption. Balancing latency and mode sensitivity while handling data corruption poses challenges. Some methods predefine interpolation sequences for modal sets but lack adaptability. In tasks like medical image segmentation, sensitivity to missing data outweighs latency concerns.

This paper introduces AdaFlow to establish non-blocking inference with heterogeneous multi-modal mobile sensor data. Implementing these ideas faces two main challenges:

- It is challenging to resample non-redundant sensor data by quantifying implicit data contribution in real-time on the mobile end without global data, and resource constraints. Existing methods generally vaguely represent implicit data contribution or ignore it.
- When distributed mobile sensor data are missing due to the asynchrony and data corruption, it is challenge in how to achieve non-blocking inference. Existing methods can hardly balance latency and accuracy.

To ensuring non-blocking data flow in multi-modal systems while addressing challenges, AdaFlow's primary approach involves two modules:

- Implicit Data Contribution-aware Data Resampling Module: It uses t-Distributed Stochastic Neighbor Embedding (t-SNE) to assess latency-aware modality affinity, then utilizes Analytic Hierarchy Process (AHP) to construct an affinity matrix based on modality affinity.
- Decision Graph-based Non-blocking Inference Module: It establishes a multi-modal mapping mechanism using the affinity matrix within a distributed multi-modal cooperative perception framework.

II. System Design

This section outlines AdaFlow designed to achieve the goal of quantifying implicit data contribution between heterogeneous modalities, constructing an affinity matrix, and enabling adaptive interpolation. AdaFlow consists of two modules as shown in Figure 1.

A. Implicit Data Contribution-aware Data Resampling Module

This module focuses on quantifying implicit data contribution between heterogeneous modalities. It employs t-SNE to assess the consistency and complementarity between modalities, crucial for understanding their implicit data contribution. Unlike previous approaches that use t-SNE for representing correlations in a vague manner, this system utilizes average cosine similarity in t-SNE graphs to denote consistency

*Corresponding author: scliu@nwpu.edu.cn

979-8-3503-6338-8/24/$31.00 ©2024 IEEE
DOI 10.1109/CSCAIoT62585.2024.00006

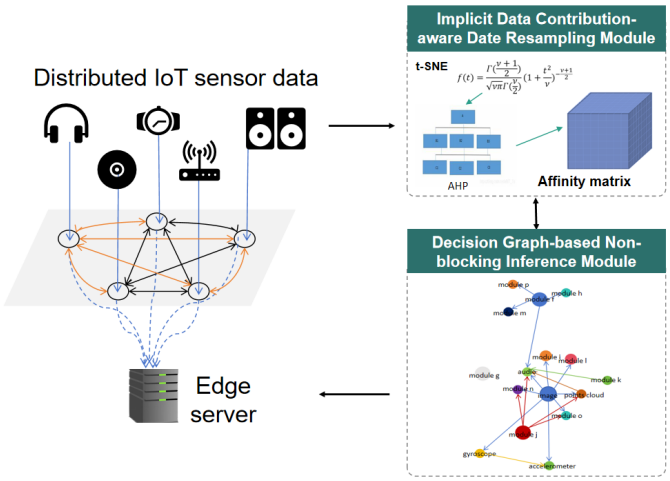

Fig. 1. System Framework

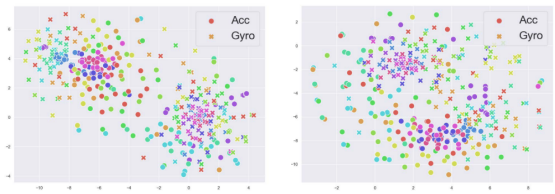

Fig. 2. Visualization of Acc and Gyro features

and complementarity between modalities. This idea different from traditional linear dimensionality reduction techniques, ensuring preservation of high-dimensional information. We use validation experiments to prove this idea, we have selected two different fusion models: the feature concatenated-based model and the self-attention-based model. Specifically, we evaluated these two models' performance using multi-modal data from 14 subjects in the public USC dataset. The task is to classify 12 human activities using accelerometer (Acc) and gyroscope (Gyro) data.

Figure 2a illustrates Acc and Gyro features from the feature concatenated-based model using t-SNE. Different colors denote various activities, while different point shapes represent modalities. Notably, features from both modalities are well-aligned and nearly symmetrical along the diagonal. Moreover, the average cosine distance between Acc and Gyro features is 0.7288, indicating high consistency. Concatenating features enhances the model's ability to capture consistent information across modalities, thus improving robustness to noisy multi-modal data.

Figure 2b depicts features from the self-attention-based model, showing less alignment among sensor data compared to the feature concatenated-based model. The average distance (0.7685) between Acc and Gyro features is slightly higher in the self-attention-based model. This indicates that the self-attention model integrates both consistent and complementary information from different modalities by combining features with varying weights.

Having quantified the consistency and complementarity between modalities, we now assign numerical values to represent their affinity. Drawing from prior research on multi-modal systems, modal complementary information is recognized for providing additional insights. Thus, in selecting the model, suitable values are chosen to accurately reflect the affinity between modalities. Then we use affinity to construct an affinity matrix using AHP, AHP is chosen due to the problem's resemblance to a graph planning problem in multi-modal systems, necessitating comparisons of average cosine values between modalities in t-SNE, this involves multiple levels of comparison and trade-offs between modalities.

B. Decision Graph-based Non-blocking Inference Module

This module aims to achieve adaptive interpolation between heterogeneous modalities in order to ensure non-blocking data flow without compromising accuracy. The system defines the problem as follows: a set of heterogeneous modals M = (m_1,m_2, ..., m_n), where M has two subsets, S and T, and M = S ∩ T. S represents the set of source tasks, while T represents the set of target tasks. For each task $s_i \in$ S,$t_i \in$ T, a_i represents the transformation time from s_i to t_i. Based on the arrival status of data streams and the corruption situation at each moment, the system adaptively determines which tasks are considered as source tasks S and which are target tasks T using the affinity matrix, in order to minimize the interpolation time $\sum a_i$.

III. Conclusion

This paper presents a framework for low-latency MLLM inference using spatiotemporal heterogeneous distributed multimodal data.The system is divided into three main modules: the distributed sensor data spatiotemporal alignment module, the cross-modal missing data imputation module, and the inference performance profiler for feedback module. These modules optimize the distributed multimodal system to overcome the heterogeneity and asynchronous arrival of distributed sensor data to achieve high-accuracy, low-latency MLLM inference.

Acknowledgment

This work was supported in part by the National Science Fund for Distinguished Young Scholars (No. 62025205), the National Key RD Program of China (No. 2021YFB2900100), and the National Natural Science Foundation of China (No. 61960206008, No. 62032020, No. 62102317).

References