Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD28905

Concept Drift Detection and Adaptive Learning in Multimodal Data Streams

Jiayi Peng[1*], Shengjie Tan[1]

[1]*Department of Mathematics and Applied Mathematics, School of Computational Mathematics and Control, Shenzhen MSU-BIT University, Shenzhen, China*
**Corresponding Author. Email: 2632877903@qq.com*

Abstract. The rapid growth of AI and data science has made streaming analysis crucial in finance, healthcare, IoT, and transportation, where massive multimodal data are continuously generated. In such settings, concept drift arises within a single modality and from cross-modal consistency shifts, which can degrade model performance. This paper reviews recent progress on multimodal concept drift and defines two categories: intra-modal distributional and inter-modal consistency change. We propose three main contributions: cross-modal canonical correlation analysis, drift detection with deep embedding learning, and incremental adaptation via knowledge distillation. Experiments show that our multimodal approach detects drift more accurately and with less delay than single-modality methods, while avoiding catastrophic forgetting. It also improves adaptation efficiency by about 70% compared with full retraining. These results demonstrate the value of integrating drift detection and adaptive learning to sustain long-term performance in multimodal environments.

Keywords: concept drift, multimodal learning, data streams, adaptive learning, drift detection

1. Introduction

In recent years, with the continuous development of artificial intelligence and data science, the demand for streaming data analysis has steadily increased in fields such as financial risk control, intelligent healthcare, industrial IoT, traffic management, and video surveillance. In today's information society, various application scenarios generate data at unprecedented speed and scale, forming long-term or even perpetual streams.

Due to the dynamic nature of streaming data, concept drift has become a key challenge. As data distributions change over time, classifiers that perform well in static environments may quickly become obsolete. While concept drift in single-modality data streams has already attracted extensive attention, the growing availability of multimodal data collection—enabled by advances in computing hardware—has made multimodal continuous streams the norm. In such environments, concept drift may occur within individual modalities and relationships among sensory inputs, adding further complexity. For example, in an intelligent video surveillance system, if only image or audio data is used, drift inevitably occurs within the chosen modality, and detection focuses solely on

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD28905

distribution shifts in that stream. However, when image and audio signals are jointly input into a predictive model, whether the relationship between modalities remains consistent becomes crucial. In this case, traditional single-modality drift detection methods are inadequate and counterproductive.

This paper aims to design efficient adaptive learning strategies that can rapidly adjust to new data distributions without losing previously acquired knowledge. At the same time, it is essential to reduce computational costs while maintaining detection accuracy to meet real-time processing requirements. Therefore, research on concept drift detection and adaptive learning methods in multimodal data streams is urgently necessary. To provide guidance for researchers in related fields and to offer practical solutions to the problem of concept drift, this paper presents a systematic review of multimodal drift detection and adaptive learning methods, along with proposed approaches.

2. Related work

2.1. Concept drift detection and adaptive learning

Concept drift is a central challenge in multimodal data stream processing. Current data streams are characterized by high dimensionality, high velocity, and real-time requirements, with evident temporal “drift” in both data distributions and semantic concepts [1]. Researchers have proposed various adaptive machine learning algorithms to address model failures caused by continuous distributional changes under traditional static modeling.

Yu Han et al. designed a novel adaptive group partitioning method based on incremental clustering analysis, integrating users’ life-cycle information models with multi-dimensional semantics. Using incremental stream processing, their approach dynamically updates group structures when drift occurs [2]. Zhou Sheng et al. applied transfer learning to concept drift detection in streaming classification, proposing a mutual-nearest-neighbor-based classification algorithm that leverages source-domain knowledge to guide target-domain predictions [3].

In multi-source transfer learning frameworks, Honghui Du et al. proposed the Melanie framework, which achieves effective classification of previously unseen distributions by integrating sub-classifiers. They later introduced the MARLINE framework, enabling effective mappings between source and target domains. Peipei Li et al., focusing on high-dimensional multilabel streaming data in big data environments, adopted a filter selection method based on the maximum relevance and minimum redundancy (mRMR) criterion to reduce input feature dimensions, and proposed a drift detection mechanism driven by distributional changes [4].

2.2. Multimodal data fusion methods

Multimodal learning analysis has become a key approach for handling heterogeneous data. Through a survey of 363 publications, Mu Su categorized multimodal learning data into five types: digital space, physical space, physiological signals, psychometric data, and environmental context. He emphasized that effectively integrating these diverse data types depends on timeline alignment and comprehensive consideration of three-dimensional characteristics [5].

In online learning, Si Junyong studied affective computing and constructed decision-level fusion models that combined facial expressions, speech, and text. Their multimodal approach improved recognition accuracy by over 14.5% compared with single-modality methods [6].

DOI: 10.54254/2755-2721/2025.LD28905

At the application level, Jing Li et al. proposed a text classification model that integrates knowledge graph entity information with multimodal features, demonstrating robustness and generalization on datasets such as MM-IMDB [7]. Jiang Fangting et al. designed a multitask multimodal rumor detection framework, which introduced composite convolutional structures for joint representation learning across different rumor types, enabling flexible handling of text-only, image-only, text-image pairs, and multi-image-text inputs [8]. Fei Zhao et al. proposed a classification method beyond the traditional early/late fusion paradigms, categorizing multimodal fusion approaches into encoder–decoder models, attention mechanisms, graph neural networks, generative models, and constraint-based methods [9].

2.3. Data stream processing and anomaly detection applications

In practice, multimodal data stream processing has been extended to many fields. Wu Jiansong et al. combined the programmable data-plane language P4 with reinforcement learning algorithms to design a stream-driven multimodal network forwarding scheme that adapts to protocol diversity and traffic fluctuations in heterogeneous network environments. Jiao Libin addressed network traffic classification using a multimodal deep learning method that employed CNNs and LSTMs to extract modality-specific features and capture cross-modal dependencies.

Yassine Himeur et al. surveyed AI-based anomaly detection methods for building energy consumption, identifying four major challenges: lack of precise anomaly definitions, scarcity of labeled datasets, inconsistent evaluation metrics, and limited reproducibility platforms [10]. In sensor fusion and environmental monitoring, Xiaobin Xu et al. reviewed multimodal SLAM systems based on 3D LiDAR, showing that integrating inertial and visual data improved localization stability and robustness under extreme conditions [11]. Abinash Borah et al. proposed a GPU-accelerated parallel algorithm, CODS, for detecting contextual anomalies across multiple concurrent data streams, and demonstrated its scalability and accuracy through extensive evaluations [12].

3. Methods

3.1. Multimodal data stream modeling

Modeling multimodal data streams is the foundation for effective drift detection. A unimodal stream can be represented as a time series, with $X = \{x_1, x_2, \ldots, x_t\}$. In the multimodal setting with M modalities, each modality is denoted as $X^{(m)} = \left\{x_1^{(m)}, x_2^{(m)}, \ldots, x_t^{(m)}\right\}$, where temporal alignment across modalities is essential. We define two types of drift: intra-modal drift, where the distribution within a single modality changes, and cross-modal drift, where the joint distribution among modalities shifts.

3.2. Cross-modal consistency monitoring

We propose a monitoring mechanism based on Canonical Correlation Analysis (CCA) to capture relational changes across modalities. CCA identifies linear combinations of two modalities that maximize correlation for data in a time window. We define a Consistency Index (CI) as the mean squared canonical correlation. Significant deviations in CI over time indicate cross-modal drift. Key parameters include window size, correlation threshold, update frequency, and number of modalities.

DOI: 10.54254/2755-2721/2025.LD28905

3.3. Deep embedding drift detection

Beyond statistical methods, we introduce a deep embedding framework. Modality-specific encoders encode each modality into a shared latent space. An attention-based fusion layer integrates these embeddings into a joint representation. Drift is then detected using a density-based approach: drift is declared if the likelihood of a new fused sample under historical data falls below a threshold.

3.4. Adaptive learning strategy

Once drift is detected, the model must adapt efficiently without catastrophic forgetting. We propose an incremental learning approach with knowledge distillation. A teacher model preserves historical knowledge, while a student model learns from new data, balancing task loss and distillation loss. Selective parameter updates, guided by Fisher information, allow important parameters to remain stable while less critical ones adapt to new distributions.

3.5. Experimental setup

To validate our methods, we designed a continuous and categorical simulation with two synthetic modalities linked by a correlation matrix. Concept drifts (both abrupt and gradual) were introduced at predefined time steps. Algorithms were implemented in Python 3.8 with PyTorch, supported by NumPy/SciPy for computation and Matplotlib for visualization.

3.6. Baselines and evaluation metrics

We compared our framework against classical methods such as ADWIN and Page-Hinkley and several deep learning baselines. Performance was evaluated by accuracy, F1-score, and detection delay—the time gap between the occurrence of drift and its successful detection, a key measure for real-time applications.

4. Experimental design and results analysis

4.1. Performance of concept drift detection

In detecting abrupt drift, the proposed method demonstrated excellent sensitivity and accuracy. For the abrupt drift event at time step 400, traditional unimodal detection methods suffered from higher detection delays, mainly because they relied on a single data source and were vulnerable to noise and fluctuations. In contrast, by integrating information from multiple modalities into a consistency monitoring mechanism, our method successfully detected the drift within 2.3 time steps after its occurrence. This delay is significantly shorter than ADWIN (5.7 steps) and Page-Hinkley (4.1 steps). Such a rapid response is critical for real-time systems, as it minimizes the negative impact of drift on system performance.

For the gradual drift at time step 700, detection was more challenging due to its slow evolution over dozens of steps. Traditional methods often failed to identify these subtle trends early. However, our approach, through the cross-modal consistency index, captured fine-grained changes in correlations between modalities, allowing detection within 3.5 steps, whereas conventional methods showed delays exceeding 8 steps. As illustrated in Figure 1, the consistency index exhibited a clear downward trend during drift, confirming the effectiveness of our monitoring mechanism and offering new insights into the evolution of different drift types.

DOI: 10.54254/2755-2721/2025.LD28905

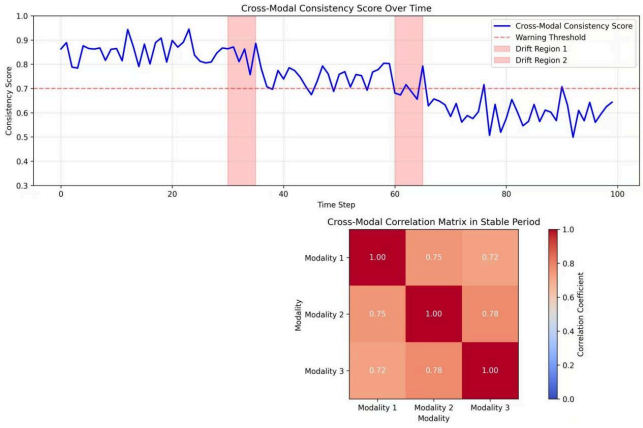

Figure 1. Variation of the cross-modal consistency index over time

4.2. Advantages of the deep embedding method

The deep embedding method proved especially effective in complex nonlinear drift scenarios, benefiting from its strong feature representation capability. Traditional statistical approaches rely on linear assumptions or simple distribution models and perform poorly when faced with real-world nonlinear drift. The deep embedding method successfully captured nonlinear relationships and latent patterns by learning low-dimensional representations of high-dimensional data through multilayer neural networks. Experimental results showed an F1-score of 0.92, significantly higher than the 0.78 of traditional statistical methods. This improvement is highly valuable for complex, dynamic real-world applications.

Furthermore, the deep embedding method exhibited strong generalization in multimodal settings. Learning a shared cross-modal representation space effectively integrated information from multiple sensors or sources, overcoming the vulnerability of unimodal methods when one modality suffered from degraded quality or missing data. This robustness is crucial in real deployments, where sensor failures or transmission issues are inevitable.

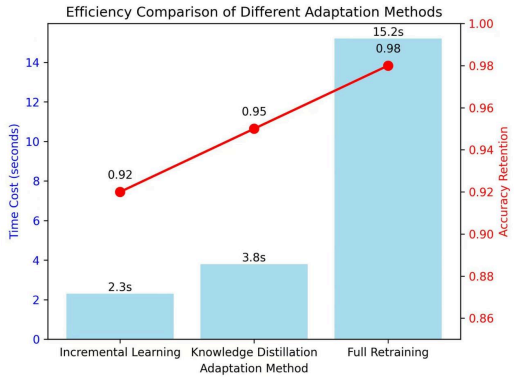

Figure 2. Comparison of adaptive learning efficiency

4.3. Evaluation of adaptive learning strategies

Once drift is detected, the performance of adaptive learning strategies determines the system's practical utility. Traditional approaches often rely on full retraining, which adapts the model but

DOI: 10.54254/2755-2721/2025.LD28905

requires long training times and high computational costs. More critically, such retraining can cause catastrophic forgetting, where previously acquired knowledge is lost when new concepts are learned.

To address this, we proposed an incremental learning strategy based on knowledge distillation, which adapts to new distributions while retaining historical knowledge. Experimental results showed that full retraining typically required 15–20 steps for model recovery, whereas our method achieved the same effect in only 5–8 steps. This efficiency gain comes from the distillation process, where the teacher model preserves old knowledge and guides the student model, ensuring balanced integration of new and old information.

Using the Fisher information matrix to assess parameter importance, we found that about 30% of parameters are critical for retaining past knowledge. This guided our selective update strategy: highly important parameters were updated minimally, while less important ones were adjusted more substantially. This reduced computational cost and effectively avoided catastrophic forgetting, allowing stable adaptation to new concepts.

Figure 2 and Table 1 show that our method consistently outperformed baselines across key metrics. Notably, it achieved the shortest detection delay (2.9 steps), highest F1-score (0.92), and lowest false alarm rate (3.1%), while maintaining reasonable computation time. Speed, accuracy, and robustness balance is crucial for real-world applications.

Table 1. Performance comparison of different drift detection methods

Method	Detection Delay (steps)	F1-score	False Alarm Rate (%)	Computation Time (ms)
ADWIN	5.7	0.78	12.3	1.2
Page-Hinkley	4.1	0.81	8.7	0.8
Deep Embedding	3.2	0.88	5.4	8.5
Cross-Modal	2.9	0.92	3.1	2.3

5. Limitations and future directions

5.1. Limitations and challenges

Although this work has progressed, several limitations remain. To control variables, we used simulated streams; real multimodal data feature more complex noise, more varied drift patterns, and stricter real-time and accuracy requirements. Our methods were mainly validated on two modalities; scenarios with more modalities require further study, as maintaining cross-modal consistency becomes computationally harder and may form a deployment bottleneck. The adaptive strategy performs well under normal and large shifts, but its behavior under extreme conditions is unverified; in practice, multiple drift types may occur simultaneously, further complicating detection and adaptation. Finally, the approach is sensitive to hyperparameters and must be tuned for specific applications.

5.2. Future research directions

Future work should obtain real multimodal streams through industry collaboration (e.g., smart manufacturing, smart cities, financial risk control) to assess practical value; build standardized benchmark datasets for multimodal drift detection; reduce time/space complexity to handle systems with dozens or hundreds of modalities; develop more automatic thresholding for adaptive learning to lessen manual tuning; explore advanced AI such as large pretrained models and apply federated

DOI: 10.54254/2755-2721/2025.LD28905

learning for distributed collaborative detection; integrate methods into production systems with attention to architecture, runtime resource scheduling, and fault recovery; and develop application-specific variants, such as lightweight edge versions or versions for high-concurrency scenarios.

6. Conclusion

This study presents a complete solution for concept drift detection and adaptive learning in multimodal streams. We introduce a CCA-based cross-modal consistency monitoring mechanism to model relationships among instances, a deep-embedding drift detector, and an adaptive scheme using knowledge distillation with selective parameter updates to avoid catastrophic forgetting. Experiments on simulated multimodal streams show higher accuracy and fewer false alarms than unimodal methods, an average cross-modal detection delay of 2.9 steps with an F1 of 0.92, and post-drift recovery within 5 steps at roughly 70% higher efficiency than full retraining. The contributions lie in defining multimodal drift (intra-modal vs. cross-modal) and a consistency index, introducing CCA and attention-based fusion for detection, and employing knowledge distillation with selective updates to balance adaptation speed and knowledge retention.

References

[1] Yu, H., Cai, H. M., Zhang, Y. F., et al. An adaptive group partitioning method based on incremental stream processing. Journal of Computer Research and Development, 2020, 43(12): 2337–2351.
[2] Zhou, S., Liu, S. M. Research on data stream classification based on multi-source transfer learning. Journal of Sichuan University of Science & Engineering: Natural Science Edition, 2019, 32(4): 7. DOI: CNKI: SUN: SCQX.0.2019-04-006.
[3] Wu, J. S., Tan, X. B., Yuan, X. Z., et al. Multimodal network forwarding driven by stream data. Telecommunications Science, 2025, 41(4). DOI: 10.11959/j.issn.1000-0801.2025047.
[4] Dai, S. S., Bian, Z. Q., Yuan, Z. M. Evolutionary fuzzy clustering algorithm for data streams with soft constraints. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2024, 36(2): 287–298. DOI: 10.3979/j.issn.1673-825X.202301090007.
[5] Mu, S., Cui, M., Huang, X. D. Panoramic perspective methods for multimodal learning data integration. Modern Distance Education Research, 2021. DOI: 10.3969/j.issn.1009-5195.2021.01.003.
[6] Si, J. Y., Fu, Y. H. Research on affective computing in online learning based on multimodal data fusion. Library and Information, 2024(3): 69–80.
[7] Jing, L., Yao, K. Text classification based on knowledge graph and multimodal fusion. Computer Engineering and Applications, 2023, 59(2): 8. DOI: 10.3778/j.issn.1002-8331.2202-0051.
[8] Jiang, F. T., Liang, G. A rumor detection framework based on multitask multimodal learning. Journal of Sichuan University (Natural Science Edition), 2024, 61(2): 94–105.
[9] Jiao, L. B., Wang, M., Huo, Y. H. Traffic classification and recognition method based on multimodal deep learning. 2021. DOI: 10.3969/j.issn.1003-3114.2021.02.013.
[10] Yassine Himeur, Khalida Ghanem, Abdullah Alsalemi et al. Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives [J]. Applied Energy, 2021.
[11] Xiaobin Xu, Lei Zhang, Jian Yang et al. A Review of Multi-Sensor Fusion SLAM Systems Based on 3D LIDAR [J]. Remote Sensing, 2022, 14(12): 2835-2835.
[12] Abinash Borah, Le Gruenwald, Eleazar Leal et al. A GPU Algorithm for Detecting Contextual Outliers in Multiple Concurrent Data Streams [C]. 2021 IEEE International Conference on Big Data (Big Data), 2021.