Journal of Technological Information, Management & Engineering Sciences (2024),5(2), 45-55

Adaptive Data Mining Framework for Detecting Hidden Patterns in Large-Scale Heterogeneous Data

[1] **Dr. Salman Qureshi**
[2] **Mahnoor Aziz**

How to cite this article:

Qureshi, S., & Aziz, M. (2024). Adaptive Data Mining Framework for Detecting Hidden Patterns in Large-Scale Heterogeneous Data. *Journal of Technological Information, Management & Engineering Sciences, 5(2), 45-55.*

Received	Accepted	Published
12 July 2024	18 September 2024	14 November 2024

Abstract

Many traditional data mining techniques are challenged by the explosion of large-scale heterogeneous data from a wide range of sources, such as structured databases, unstructured text, sensor streams, and graph-structured networks. Traditional methods typically make the assumption of uniform data distribution and a fixed environment, which is not able to capture complex, latent patterns that change with respect to time. In this paper, an Adaptive Data Mining Framework (ADMF) is proposed to discover the hidden patterns in large-scale heterogeneous data by dynamically integrating multi-source information, adaptive concept-drift handling and hybrid pattern discovery mechanisms. The framework is based on recent developments in heterogeneous graph neural networks, adaptive ensemble learning and multi-stage pattern mining, and comprises four major components: (1) heterogeneous data fusion/representation learning, (2) adaptive drift detection and model updating, (3) multi-view contextual pattern mining, and (4) interpretable knowledge extraction. A conceptual evaluation using current literature proves the framework potential to enhance the accuracy, scalability, and robustness of pattern discovery in different application areas like multi-energy systems, network security and enterprise analytics. The paper concludes by identifying gaps in the existing methods, specifying clear objectives and stating a methodology that provides a balance between theory and practice. Future research and practical implications are discussed.

Keywords: *Adaptive Data Mining, Heterogeneous Data, Hidden Pattern Detection, Concept Drift, Multi-source Fusion, Large-scale Analytics.*

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan.
[2] Department of Computer Science, University of Central Punjab, Lahore, Pakistan.

Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

1. Introduction

Modern digital ecosystems create voluminous amounts of data with the four V's: volume, velocity, variety, and veracity. Heterogeneity not only occurs in the data formats (relational tables, text corpora, time series, graphs, and multimedia) but also in the data's semantics, structures, and distributions across sources (Han et al., 2023; Wu et al., 2024). In such environments, uncovering hidden patterns—such as latent associations, sequential motifs, anomalous subgraphs, or evolving clusters—is crucial for making informed decisions across a wide variety of applications, including smart grid, healthcare, cybersecurity, and e-commerce.

When exposed to concept drift, multi-modal feature spaces, or adaptation in real-time or near real-time, traditional data mining pipelines, in many cases developed under the stationary and homogeneous data assumption, are challenged (Sakhipov, 2025; Wu et al., 2024). In recent years, adaptive pattern mining methods for multi-energy loads (IEEE TKDE, 2024), heterogeneous graph neural networks for heterophily (AHGNN, 2025), and ensemble methods for streaming data with heterophily (WAEM, 2026) have all been introduced in the literature. However, a coherent and systematic framework for adapting to multi-source fusion, drift-aware learning, contextual pattern discovery, and interpretable knowledge extraction in a single system is not well developed.

This paper fills this gap, by proposing the Adaptive Data Mining Framework (ADMF). The framework includes representation learning and multi-view fusion, which can handle diverse data modalities, adaptive ensemble and drift detectors that can tackle distributional shifts, hybrid techniques that can uncover both local and global hidden patterns, and production of human-interpretable insights appropriate for domain experts. The following sections include a literature review, statement of the research gap and purpose, research methodology, expected results and implications, and suggestions for future research.

2. Literature Review

This course introduces the foundational concepts and models of data mining, as well as the techniques of processing heterogeneous data.

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan.
[2] Department of Computer Science, University of Central Punjab, Lahore, Pakistan.
Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Data mining (knowledge discovery from data, KDD) comprises of the methods that are used to find non-trivial, previously unknown and potentially useful patterns in large amounts of data (Han et al., 2012; Han et al., 2023). Frequent pattern mining, association rule discovery, classification, clustering and outlier detection are classical tasks. If the data sources are heterogeneous, that is, they differ in schema, semantics and statistical properties, other challenges come up such as entity resolution (Cao et al., 2008), schema matching, and multi-relational or multi-modal integration (Ye et al., 2022).

Much focus has been paid to heterogeneous information networks (HINs) and multi-source fusion methods. Using multiple-source patterns from the context to represent coherent patterns allows the re-use of single-source pattern mining algorithms (Zhou et al., 2012). Later, pattern discovery-based approaches for truth discovery in multi-source problems are proposed as joint optimization over common patterns and source reliability (Ye et al., 2022).

This is a course on Adaptive and Streaming Data Mining.This is a course on Adaptive and Streaming Data Mining.

The statistical characteristics of the target variable evolve over time (concept drift), which requires adaptive learning strategies. Weighted Adaptive Ensemble Methods (WAEM) use several drift detectors and dynamically re-weighted classifiers to obtain zero-detection delay under different types of drift (ETASR, 2026). A multi-energy load pattern mining systems that combine adaptive affinity propagation and windowed DTW could automatically deal with the unusual energy consumption of the daily load without demanding significant user effort (IEEE TKDE, 2024).

Distributed and streaming environments: Dataflow- and DDDAS-integrated Adaptive DSM system Design (DDADD) principles focus on dynamic and data-driven configuration of mining queries with the aim of balancing accuracy, resource consumption, and energy efficiency (Springer, 2023). Optimised adaptive architectures also minimise unnecessary retraining, adaptation latency and prediction error values (Sakhipov, 2025) and ensure predictive accuracy of more than 0.83 for abrupt, gradual, recurrent and continuous drift types.

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan.
[2] Department of Computer Science, University of Central Punjab, Lahore, Pakistan.
Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

2.3 Graph Based and Multi-View Approaches for Heterogeneous Patterns

The advent of heterogeneous graph neural networks has given rise to improved models of complex relational structure. The Adaptive Heterogeneous Graph Neural Network (AHGNN) tackles the heterophily problem across hops and across different metas, and achieves state-of-the-art performance particularly when heterophily is strong (arXiv, 2025). Adaptive Label Propagation (ALP) is an extension of the group anomaly detection to overlapping groups in large-scale heterogeneous networks, where neighbors act as arms in a combinatorial multi-armed bandit formulation (IEEE TKDE, 2023).

Under high missing rates, multi-view and incomplete multi-modal clustering methods like the Integrated Heterogeneous Graph Attention Network (IHGAT) learn to make use of complementarity by building integrated heterogeneous graphs (Wang et al., 2024). The advances demonstrate the need for attention based and adaptive models to capture patterns that are not recognized by homogeneous models.

2.4 Interpretability and Actionable Knowledge.

The ability to extract knowledge that is actionable and human-readable is key, in addition to accuracy. Integrating multi-source and multi-method analytics in combined mining framework for enterprise systems with the aim of providing domain-driven insights (IRJMETS, 2026). Change mining systems (ChaMining), based on the use of ontologies, temporal rule mining and clustering technology, can identify microscopic and macroscopic changes in dynamic heterogeneous networks while reducing information loss compared to the previous methods (Salford, 2026).

3. Research Gap

While significant advances have been achieved with respect to single solution technology, there are still a number of areas of gaps.Despite the considerable advances made in each of these individual directions (adaptive ensembles for concept drift, heterogeneous graph embedding, multi-stage pattern mining, and multi-source fusion), there are still several gaps. First, most of the available solutions only tackle a part of the problems of the large-scale heterogeneous data;

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan.
[2] Department of Computer Science, University of Central Punjab, Lahore, Pakistan.
Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

few offer a coherent and end-to-end adaptive architecture. Secondly, the sensitivity to parameters and the need for lots of user interactions is common, which reduces the usefulness for deployment. Third, the joint treatment of structural heterophily, temporal drift and multi-modal complementarity in one adaptive loop is uncommon. Fourth, mechanisms for interpretation are often added after the mining process, not as a part of the mining process itself. Therefore, a systematic approach towards representation, adaptation, pattern discovery, and knowledge extraction for large-scale heterogeneous environments is highly needed, in order to provide a unified Adaptive Data Mining Framework.

4. Objectives

The main objective of this research is to provide a conceptual validation of a framework called Adaptive Data Mining Framework (ADMF) that is able to discover hidden patterns in large-scale heterogeneous data sets. Specific objectives are:

To design a multi-layer architecture for merging heterogeneous data fusion, adaptive drift handling, contextual pattern mining and interpretable knowledge extraction.

2. To introduce mechanisms to track concept drift and heterophily without having to do large manual parameter tuning.

3. To enable discovery of local (node or instance level) and global (network or corpus level) hidden patterns across structured, semi-structured and unstructured sources.

To provide a solid base on the latest literature from academia, and describe how criteria for evaluation will be used to validate the framework in the future.

5. Methodology

The architecture of the proposed adaptive data mining framework is shown in this figure.This figure presents the overall architecture of the proposed adaptive data mining framework.

The proposed ADMF is designed in the form of four interconnected layers working in a cycle of continuous adaptive learning and drift management, multi-view pattern discovery, and knowledge extraction and interpretation: (1)

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan.
[2] Department of Computer Science, University of Central Punjab, Lahore, Pakistan.

Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Heterogeneous Data Ingestion and Representation Layer, (2) Adaptive Learning and Drift Management Layer, (3) Multi-View Pattern Discovery Layer, and (4) Knowledge Extraction and Interpretation Layer. The layers receive diagnostic signals (performance metrics, drift indicators and pattern quality scores) from the layers below them, allowing for closed-loop adaptation.

5.2 Heterogeneous Data Ingestion and Representation

Schema Mapping, Entity Resolution and Modality-specific Preprocessing are performed on the raw data coming from relational databases, event streams, text corpora and graph sources before the actual processing. The framework extends the ideas of Contextual Information Graph (CIG) encoding and heterogeneous graph construction approaches (Zhou et al., 2012; Wang et al., 2024), by mapping entities into a common latent space while retaining type-specific features. The attention mechanisms and the residual connections, like those used in AHGNN and IHGAT, aggregate the neighborhood information of different hops and modalities selectively. The absence of modalities is dealt with through availability graphs and complementary sample relations, and this decreases the need for full pairwise data.

5.3 Adaptive Learning and Drift Management – YouTube

An ensemble of detectors tracks abrupt, gradual, recurrent and continuous concept drifts. The Weighted Adaptive Ensemble Method (WAEM) paradigm denotes that each individual base learner is given an adaptive weight according to its recent performance and confidence calibration (ETASR, 2026). When the drift is detected, the framework performs "driftreducing" updates to the model: from the incremental adjustment of a parameter to the replacement of a part of the model ensemble. Adaptive affinity propagation has been proposed and applied to sequential and load-profile data for detecting uncommon patterns and extracting representative daily or periodic patterns with adaptive distances of trend-similarity and windowed dynamic time warping (IEEE T-KDE, 2024).

5.5 Multi-View Contextual Pattern Discovery

The frequent itemsets / association rules on the transactional projections, the sequential / temporal motifs based on Autoplait-style segmentation / process-

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan.
[2] Department of Computer Science, University of Central Punjab, Lahore, Pakistan.

Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

mining, the structural patterns (communities, meta-path motifs, anomalous groups) on the heterogeneous graph obtained through the adaptive label propagation and attention-based aggregation methods, and the multi-modal clusters combining feature and relational similarity. Coarse-to-fine fusion modules filter out noisy signals, strengthen informative meta-paths or views, and enhance robustness when the data is heterophily and incompleteness.

5.5 Knowledge Extraction and Interpretation

Patterns that are discovered are ranked according to interestingness measures based on statistical significance, novelty and domain utility. Raw patterns are converted to human-readable statements about evolving relationships and anomalous behaviours using ontology-guided change mining and temporal rule extraction (Salford, 2026). The validation and feedback loop is completed with the inclusion of visualization and summarization modules, where experts can correct the mining cycle to improve it for the next one.

5.6 Evaluation Criteria

Since this is a conceptual tool, empirical findings will be postponed until the tool is implemented. The following are proposed as dimensions for evaluation: (i) pattern quality (precision/recall scores, or domain-specific interestingness scores against ground-truth or expert labels), (ii) adaptation metrics (detection delay, accuracy under drift, retraining frequency), (iii) scalability (runtime and memory usage as a function of data volume and degree of data heterogeneity), and (iv) interpretability (expert ratings of explanation usefulness). Benchmarks should include any combination of synthetic drift scenarios and real-world multi-source datasets of energy, network and enterprise.

6. Results

The present paper is not experimental, it is only a conceptual framework synthesis. Rather, it brings together performance information from the literature to show the typical gains from the design decisions made in the ADMF.

In ETASR 2026, adaptive ensemble methods like WAEM have been shown to have higher drift adaptation and classification rates than three benchmark adaptive learners in synthetic and real-world streams. The proposed 3-stage adaptive

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan.
[2] Department of Computer Science, University of Central Punjab, Lahore, Pakistan.
Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

pattern mining approach over multi-energy loads has effectively identified rare daily load patterns and generated typical load patterns with minimal parameter tuning, which is suitable for demand-side applications such as load forecasting and demand response (IEEE TKDE, 2024). Under strong heterophily, AHGNN was able to improve the Micro-F1 score by up to 4.32%, while staying computationally efficient (arXiv, 2025). Adaptive Label Propagation enhanced group anomaly detection on both bipartite and heterogeneous networks in the context of overlapping groups and adaptive neighbor weighting (IEEE TKDE, 2023). Compared to six previous algorithms (Salford, 2026), the mining systems that used ontology-guided hybrid methods were able to retrieve more change patterns and minimise information loss.

Taken together, these results suggest that an integrated adaptive architecture that incorporates multi-source representation learning, ensemble-based drift response, multi-view pattern mining and ontology-based interpretation has the potential to achieve significantly better accuracy, robustness and actionability in comparison with single-modality and non-adaptive baselines.

7. Discussion

The Adaptive Data Mining Framework brings together the best of recent advances, complementing one another. The framework is aligned with modern graph learning and multi-modal learning paradigm by considering the heterogeneity as a source of multi-view complementarity rather than it being a nuisance for preprocessing. As noted in the previous section, the closed-loop adaptation mechanism ensures that real data sources remain non-stationary and, if they are not continually monitored and selectively updated, then the quality of the patterns deteriorates quickly.

One of its benefits is that it decreases the need for hyper-parameter tuning by hand. Adaptive affinity propagation, attention-driven meta-path selection and performance-driven ensemble weighting pull the decision-making process away from the analyst and towards the data-driven components, further enhancing scalability and reproducibility. The knowledge extraction layer also maintains human oversight, reducing the potential for opaque “black-box” mining results.

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan.
[2] Department of Computer Science, University of Central Punjab, Lahore, Pakistan.
Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

There is a need to recognize restrictions. Empirical validation on a large-scale multi-domain dataset is needed to establish the trade-offs between accuracy, latency and resource usage. The issue of privacy and security was not discussed thoroughly, particularly when combining multiple data sources, and additional mechanisms like Differential Privacy or Federated variants should be discussed in the future. Last, the high computational burden of keeping multiple views and adaptive ensembles might also be too expensive in very resource-poor edge environments and approximations to the hierarchical or sampling-based types will be required.

However, the structure offers a clear research programme. Later, it is possible to replace and/or adapt components of the system as needed (e.g., using a newer heterophily-aware GNN or a more advanced drift detector), while maintaining the adaptive loop. Promising extensions include cross-domain transfer of learned representations, and meta-learning of adaptation policies.

8. Conclusion

This paper has introduced the Adaptive Data Mining Framework (ADMF) for learning hidden patterns in large size and heterogeneous data. The framework incorporates three heterogeneous representation learning, adaptive drift management, multi-view contextual pattern discovery, and interpretable knowledge extraction, creating a closed-loop framework that solves the main challenges of traditional static data mining processes. The ADMF is built on recent advances in the fields of adaptive ensembles, heterogeneous graph neural networks, multi-stage pattern mining, and change detection, which provide a principled approach to more reliable, scalable, and useful knowledge discovery. Research work will continue in various directions such as the full-scale experimental assessment, extensions with privacy guarantees, and automatic meta-adaptation strategies, paving the way for the applicability of adaptive data mining in complex real-life settings.

References

Cao, Y., Chen, Y., & Jiang, B. (2008). Heterogeneous Data Integration System with Self Adaptivity. In: L. Xu, A. Tjoa and S. Chaudhry (Eds.), Research and practical

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan.
[2] Department of Computer Science, University of Central Punjab, Lahore, Pakistan.

Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

issues of enterprise information systems II (pp. 31-46).In: L. Xu, A. Tjoa and S. Chaudhry (Eds.), Research and practical issues of enterprise information systems II (pp. 31-46). 65–74). Springer.

Han, J., Kamber, M., & Pei, J. (2012). Data mining concepts and techniques (3rd edition). Morgan Kaufmann.

Han, J., Pei, J., & Tong, H. (2023). Data mining: Concepts and techniques (4th edition). Morgan Kaufmann.

IEEE Transactions on Knowledge and Data Engineering. (2024). An adaptive pattern mining algorithm using data for multi-energy loads in three stages. IEEE Transactions on Knowledge and Data Engineering. https://doi.org/10.1109/TKDE.2024.3462770

Sakhipov, A. (2025). Adaptive machine learning for dynamic data streams optimized. Eastern-European Journal of Enterprise Technologies. https://journals.uran.ua/eejet/article/view/343635

Salford Repository. (2026). Dynamic Heterogeneous Information Networks Change Discovery using Knowledge Engineering and Data Mining Methods. https://salford-repository.worktribe.com/output/1333755

Wang, Y., Yao, X., Zhu, P., et al. (2024). Integrated Heterogeneous Graph Attention Network for Incomplete Multi-modal Clustering. International Journal of Computer Vision, 132, 3847–3866. https://doi.org/10.1007/s11263-024-02066-y

Wu, X., Wu, Y., Li, X., Ye, Z., Gu, X., Wu, Z., & Yang, Y. (2024). Using adaptive machine learning systems in distributed data applications. Global Academic Frontiers, 2(3).

Ye, C., Wang, H., & Dai, G. (2022). Discovery of patterns in mixed data. In SpringerBriefs in Computer Science. Springer. https://doi.org/10.1007/978-981-19-1879-7_4

Zhou, Y., Cheng, H., & Yu, J. X. (2012). A general framework to encode heterogeneous information sources for contextual pattern mining. ACM International Conference on Information and Knowledge Management 2012, 65–74. https://doi.org/10.1145/2396761.2396774

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan.
[2] Department of Computer Science, University of Central Punjab, Lahore, Pakistan.

Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.

Adaptive Heterogeneous Graph Neural Network (AHGNN). (2025). arXiv preprint. https://arxiv.org/pdf/2508.06034

Adaptive Label Propagation: group anomaly detection in large-scale networks. (2023). IEEE Transactions on Knowledge and Data Engineering, 35(12), 12053–12067. https://doi.org/10.1109/TKDE.2022.3176478

Heterogeneous concept drift handling using adaptive machine learning model in streaming data. (2026). Engineering, Technology & Applied Science Research. etasr.com/index.php/ETASR/article/view/17430

An adaptive data stream mining (DSM) system. (2023). In Handbook of dynamic data driven applications systems, (pp. 705–722). Springer. https://doi.org/10.1007/978-3-031-27986-7_26

Towards scalable and actionable knowledge discovery in heterogeneous enterprise systems: A combined mining perspective. (2026). International Research Journal of Modernization in Engineering, Technology and Science. www.doi.org/10.56726/IRJMETS93261 .

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan.
[2] Department of Computer Science, University of Central Punjab, Lahore, Pakistan.

Copyright (c) 2024 Author(s)
This work is licensed under a Creative Commons Attribution 4.0 International License.