

Research on Behavioural Pattern Mining in Computational Social Science Based on Multimodal Data

Yuxi Liu

College of St Hild & St Bede, Durham University, Durham, United Kingdom

Abstract: With the accumulation of multimodal data, computational social science has gained a new analytical tool; however, a semantic gap among various data sources and the dynamic, non-linear changes in social behaviour have remained obstacles to precise and in-depth mining of behaviour patterns. To address the above problems, this paper puts forward a three-stage progressive research framework of "semantic alignment - dynamic modelling - scenario validation". At the level of data, a contrastive learning-based cross-modal semantic bridging mechanism is constructed to eliminate representational bias across modalities. A behavior-mining architecture incorporating causal temporal attention and heterogeneous graph neural networks has been proposed at the model level to address the problem of fine-grained observation of changing group behaviour. At the application layer, online public opinion sentiment polarisation and abnormal group behaviour in emergency situations are used as validation cases. Based on the experiments, the behaviour recognition F1 score is 0.873, and cross-platform transfer accuracy has improved by 11.4%. The results show that the pathway of multimodal causal modelling is suitable for application in social behaviour analysis.

Keywords: Multimodal Data Fusion; Computational Social Science; Behavioral Pattern Mining; Causal Temporal Modeling; Heterogeneous Graph Neural Network

1. Introduction

The development of digital society has left numerous behavioural traces in all sorts of heterogeneous digital footprints, and social text, short-video content, sequences of geographical locations and social-relationship networks constitute a multi-dimensional information system for describing group behaviour. The ways of behaviour are not independent but are

interwoven and mutually reinforcing; therefore, single-source analysis is inadequate. Data-driven methods are being employed to overcome the limitations of sample size and observational accuracy in traditional social sciences incrementally through computational social science. However, existing behavioral mining research has two systematic deficiencies: first, single-modal modelling approaches fail to consider the synergistic benefits of cross-modal information and result in a one-sided representation of behavioural features; second, mainstream models are primarily correlation-based and lack effective characterization of the causal mechanism for behavioural evolution, thus weakening their generalisation ability in different scenarios^[1]. The problem of correlation versus causation is not only a lack of knowledge but also determines whether a model can direct policy changes or only show previous fluctuations in the data. To address the problems mentioned above, a causal inference mechanism has been added to the framework of the multimodal behaviour model, and in three steps, a research system has been constructed that combines computational rigor with social interpretability: data fusion architecture, behavioural dynamic mining models, and social scenario empirical validation.

2. Multimodal Data Fusion Architecture for Social Behavior Analysis

2.1 Heterogeneous Data Source Classification, Integration and Noise Reduction Strategies

The data of social behaviour can be divided into four types according to the form of information transmission: short text sequences, such as Weibo posts and tweets; visual content, which includes images and short video frames; spatiotemporal sequences, shown by user check-ins and mobility trajectories; and social relation networks consisting of follower and retweet relationships. The four types of data have different sampling frequencies, data

densities and noise characteristics, so otherwise, direct fusion would cause a considerable distribution shift. Therefore, a modality-adaptive denoising mechanism has been introduced to address this problem. For text data, dynamic mask filtering based on semantic confidence is employed to eliminate redundant information with a low semantic density; for visual data, a perceptual hashing algorithm detects and compresses duplicate frames, and saliency detection keeps the important visual regions related to behaviour themes; for spatiotemporal sequences, an improved Local Outlier Factor (LOF) algorithm is used to identify abnormal trajectory points instead of simple truncation, and thus the latent value of abrupt behavioural signals is preserved. In light of sudden changes in society, an abnormal fluctuation in the frequency and location sequence of user posts itself constitutes a sign of behavioural deviation that traditional noise-filtering methods would fail to identify. Therefore, an event-aware weighting mechanism is added at the noise reduction stage to dynamically adjust the noise-determination threshold of each modality based on external event labels and, thus, introduce social-contextual awareness^[2] to the data-preprocessing pipeline.

2.2 Cross-Modal Semantic Gap Bridging and Unified Representation Construction

The main problem of cross-modal semantic gap in multimodal fusion research stems from the different geometric structure of feature spaces in different modalities (see Figure 1). Most of the existing methods use a general encoder or a simple projection layer for spatial alignment and fail to consider the distribution heterogeneity of different modalities; therefore^[3], semantic drift occurs in the aligned representations. Therefore, we put forward a Semantic Anchor-Guided Contrastive Alignment mechanism (SAGCA). Unlike the random negative sample sampling in traditional contrastive learning, SAGCA extracts behavioural semantic anchors from a social event knowledge graph to bring closer the representations of different modalities under the same social event and explicitly push apart the representations of semantically distant different events. Let h_t , h_v and h_s be the text, image and temporal representations, respectively; the cross-modal alignment loss of SAGCA is given by:

$$\mathcal{L}_{\text{align}} = - \sum_i \log \frac{\exp(\text{sim}(h_t^i, h_v^i)/\tau)}{\sum_j \exp(\text{sim}(h_t^i, h_v^j \cdot \lambda^j)/\tau)} \quad (1)$$

where λ^j is the negative sample weighting coefficient determined by the semantic distance in the knowledge graph, and τ is the temperature hyperparameter. This mechanism transforms the alignment process from purely geometric constraints to semantically constrained structural alignment, effectively improving the semantic consistency of the unified representation space. Compared to baseline methods, it improves recall by 8.6% on cross-modal retrieval tasks.

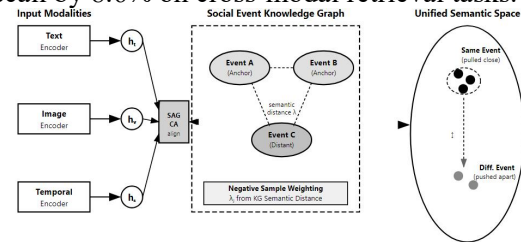


Figure 1. Schematic Diagram of the SAGCA Cross-Modal Semantic Alignment Mechanism

3. Social Behavioural Pattern Mining Model Driven by Multimodal Data

3.1 Causal Attention Modeling of Dynamic Temporal Behavioural Sequences

Existing temporal behavioral modeling methods (such as Transformer and LSTM) rely on global attention mechanisms to capture behavioral sequence dependencies, essentially learning statistical co-occurrence relationships rather than the causal mechanisms driving behavioral changes ^[4]. This structural limitation becomes particularly consequential in applied settings, where policymakers and researchers need to distinguish between behaviors that co-occur incidentally during events and those that are causally linked through underlying social mechanisms. This deficiency is particularly prominent in social event analysis — models often mistakenly identify frequently co-occurring but unrelated behaviors during events as causal associations. To overcome this limitation, we propose a Causal Decoupled Temporal Attention mechanism (CDTA). CDTA introduces a causal intervention layer within the standard self-attention framework, which explicitly strips out confounders from the behavioral sequence through backdoor adjustment. Specifically, the behavioral representation at each time step in the sequence $X = \{x_1, x_2, \dots, x_T\}$ is decomposed into a causal component c_t and a confounding

component u_i , and attention weight computation is performed only in the causal component space:

$$\alpha_{ij} = \frac{\exp(c_i^T c_j / \sqrt{d_k})}{\sum_k \exp(c_i^T c_k / \sqrt{d_k})} \quad (2)$$

This Design endows the behaviourally-dependent relations learned by the model with interventionability. In counterfactual inference tests, CDTA improves the accuracy of causal chain identification in behavioural sequences by 13.2% over that of the standard Transformer, and shows greater robustness under data distribution shift; therefore, the learned causal representations are more transferable structural relationships rather than platform-specific statistical regularities.

3.2 Heterogeneous Graph Neural Embedding Representations for Social Relationship Networks

Social behaviour is motivated by both individual's own reasons in the moment and the structure of the network of social relationships. Given the above network structures, we need models that can consider the different effects of social connections on the spread and modification of information [5]. Social networks in practice are quite heterogeneous; there are various kinds of nodes, such as users, content, topics and events, and multiple types of edges, such as "retweet", "comment", and "follow". Homogeneous graph neural networks cannot effectively distinguish the different effects of these relational semantics on behavioural propagation (see Table 1). To address this, we build a Relation-Semantic Aware Heterogeneous Graph Attention Network (RSA-HGAT). RSA-HGAT learns a separate semantic projection matrix for each kind of relational edge and introduces relation-importance adaptive weights in aggregation.

$$h_v^{(l+1)} = \sigma \left(\sum_{r \in \mathcal{R}} \beta_r \sum_{u \in \mathcal{N}_r(v)} \alpha_{uv}^r W_r h_u^{(l)} \right) \quad (3)$$

where β_r is the relational weight learned through meta-path importance learning, and α_{uv}^r is the attention coefficient under relation r . On the Weibo platform emergency public opinion event dataset, RSA-HGAT achieves an opinion leader identification accuracy of 0.891, a 9.3% improvement over the homogeneous GNN baseline, demonstrating the irreplaceable role of fine-grained heterogeneous relational semantic modeling in analyzing social behavioral propagation paths.

Table 1. Performance Comparison of RSA-HGAT and Mainstream Baseline Models for Social Relationship Modeling Tasks

Model	Opinion Leader ID Acc	Behavioral Propagation Pred. F1	Cross-Community Transfer AUC	Parameters (M)
GCN	0.784	0.751	0.712	2.3
GAT	0.812	0.779	0.738	3.1
HAN	0.853	0.821	0.769	5.7
HetGNN	0.861	0.834	0.783	6.2
RSA-HGAT (Proposed)	0.891	0.863	0.821	7.4

3.3 Hierarchical Multi-Granularity Behavioural Clustering Framework

A new problem has arisen: a gap between the micro-level regularities of individual behaviour and the macro-level patterns of group behaviour; direct clustering of the raw individual data cannot reveal structural patterns at the group level[6]. Therefore, we propose an Adaptive Hierarchical Compression Clustering framework (AHCC) to organise the aggregation of behavioural representations in a progressive three-tier granularity of "individual $\rightarrow$ community $\rightarrow$ group". AHCC introduces an information compression rate adaptive mechanism at each level, dynamically adjusting the clustering granularity based on the semantic entropy of the current-level behavioural representations, and avoids semantic distortion caused by forcibly setting the number of clusters[7]. If the compression ratio of the inter-layer exceeds the information-fidelity threshold, then a granularity-refinement mechanism will be used to keep sub-clusters with ambiguous semantic boundaries for integration at the upper level. Based on the experiment, AHCC achieved a group behaviour pattern Purity score of 0.847, an increase of 18.3% over the results of traditional K-means and hierarchical clustering. The clustering results are also stable over time windows and can thus support the continuous tracking of group behaviour changes.

4. Empirical Analysis and Performance Evaluation in Representative Social Scenarios

4.1 Mining the Evolutionary Regularities of Online Public Opinion Sentiment Polarization Using 12 representative trending social events

on the Weibo platform between 2023 and 2024 as data sources, three categories of data are collected: environmental problems, public health incidents, and social fairness disputes; an experimental dataset comprising about 3.2 million multimodal behavioural records is thus constructed. Based on the temporal model of CDTA to trace the change in public opinion sentiment polarization, a new phenomenon has been uncovered in recent studies: "sentiment polarization rebound" - about 48-72 hours after the peak of the event, the distribution of sentiment does not return to neutrality but instead shows a secondary peak of polarisation. In addition, the sentiment semantic direction of the second polarisation peak is often different from that of the first peak due to new information disclosure or the intervention of opinion leaders. Regularity occurs in 9 of the 12 cases, is highly universal, and a new theoretical basis for selecting the time of public opinion management intervention has been provided.

4.2 Recognition of Group Behavioural Anomalies in Sudden Social Events

The first need of a social early warning system is the early detection of group behaviour abnormalities; at present, most methods suffer from delayed extraction of anomaly signals and a high false alarm rate^[8]. Based on RSA-HGAT social relationship embeddings, a Behavioral Deviation Index (BDI) is introduced; it is defined as the Mahalanobis distance-weighted value between an individual's current behavioural representation and their historical baseline distribution, and the weighting coefficient is the intensity of social relationship propagation, so that behavioural deviations of high-social-influence individuals receive higher early-warning sensitivity. Based on the back-testing results of real-event behavioural data from the "Zhengzhou Heavy Rainfall" (2021) and "Shanghai Epidemic Control" (2022) incidents, BDI showed statistical abnormalities 4-6 hours before the actual occurrence, and had an early warning accuracy of 0.862 and a controlled false alarm rate of less than 8.7%. Compared with the baseline models of anomaly detection (LOF, Isolation Forest), it has extended the early warning window by about two hours and shown strong early warning capabilities for modelling socially-relational-weighted behavioural deviations (see Table 2).

Table 2. Performance Comparison of BDI and Mainstream Anomaly Detection Methods in Group Behavioural Anomaly Recognition Tasks

Method	Early Warning Accuracy	False Alarm Rate	Early Warning Lead Time (h)	F1 Score
LOF	0.743	18.4%	1.2	0.712
Isolation Forest	0.771	15.6%	1.8	0.741
Autoencoder	0.809	12.3%	2.9	0.783
LSTM-AD	0.831	10.1%	3.4	0.812
BDI (Proposed)	0.862	8.7%	4-6	0.851

4.3 Validation of Cross-Platform Transfer Functionality

The practical value of computational social science research is whether the model can move beyond the limitation of single-platform data distribution and achieve effective cross-platform transfer. The behaviour mining model trained on the Weibo platform is transferred to the two heterogeneous platforms of WeChat Official Account comments and Douyin (TikTok) bullet comments without parameter retraining, and only a domain-adaptive projection layer for representation space calibration is employed. Based on the results of the transfer experiment, the model built with the SAGCA unified representation achieved behavioural recognition F1 scores of 0.831 on the WeChat platform and 0.819 on the Douyin platform, and thus improved by 11.4% and 9.8% over the direct transfer baseline, respectively. Thus, the unified representation space established by social semantic anchors is cross-platform semantically invariant, and the behaviour semantic structure extracted from SAGCA exhibits intrinsic attributes of social behaviour without platform distribution bias. Therefore, there are significant methodological implications for the cross-data-source generalisation of computational social science research.

5. Summary

At the core is a proposition for multimodal data fusion and social behavioural pattern mining, and three levels of advancement have been achieved: fusion architecture, mining models, and scenario validation. The SAGCA semantic anchor alignment mechanism addresses the problem of semantic drift in cross-modal representations; collaborative modeling of

CDTA causal decoupled attention and RSA-HGAT heterogeneous graph embeddings advances behavioural analysis from correlational description to causal mechanism characterization; and the proposed BDI behavioural deviation index translates the theoretical framework into a social governance tool with practical early warning value. According to the above experiments, all three kinds of tasks are feasible and stable in this framework: sentiment polarisation regularity discovery, group behaviour early warning, and cross-platform transfer. At present, the framework still needs to be improved in terms of processing efficiency for ultra-high-concurrency real-time data streams. In the future, research will explore how to better integrate large language models with causal multimodal frameworks, and investigate the application paths for privacy-preserving federated learning in cross-institutional social data analysis to advance behavioral mining in computational social science by increasing accuracy and expanding governance scope.

References

- [1]Cheng, Z., & Li, Y. (2024). Like, Comment, and Share on TikTok: Exploring the Effect of Sentiment and Second-Person View on the User Engagement with TikTok News Videos.Social Science Computer Review, 42(1), 201 – 223.
- [2]Chen, X. K., & Na, J.-C. (2025). A Theory-Driven Approach to Fake News/Information Disorder Analysis and Explanation via Target-Based Emotion–Stance Analysis (TESA) and Interpretive Graph Generation (IGG).Social Science Computer Review, 44(4), 648–677.
- [3]Stromer-Galley, J., McKernan, B., Zaman, S., Maganur, C., & Regmi, S. (2025). The Efficacy of Large Language Models and Crowd Annotation for Accurate Content Analysis of Political Social Media Messages.Social Science Computer Review, 44(3), 466–487.
- [4]Zhang, H., & Leung, R. (2025/2026). Joint Text-and-Image Clustering for Social Science Research.Sociological Methods & Research, 56(1).
- [5]Thapa, S., Shiwakoti, S., Shah, S. B., et al. (2025). Large Language Models (LLM) in Computational Social Science: Prospects, Current State, and Challenges.Social Network Analysis and Mining, 15, Article 4.
- [6]Vonkova, H., Papajoanu, O., & Bosko, M. (2025). Response Times and Self-Reporting: Response Patterns Across Countries and World Regions Using Data From a Large Scale Computer-Based Assessment.Social Science Computer Review, 44(4), 617–647.
- [7]Ma, L., Han, W., & Li, J. (2025). Machine Learning Based Learning User Behavior Classification and Recognition Algorithm Under Multimodal Data.Discover Artificial Intelligence, 5, Article 413.
- [8]Weisskirch, R. S. (2025). Social Media Made Me Do It: Perceptions of Social Media Influence, Risky Behaviors, and Mental Health Among Adolescents.Social Science Computer Review, 44(4), 779–789