# Research Statement

MA Yunshan
School of Computing and Information Systems, Singapore Management University
Email: ysma@smu.edu.sg

## Background

Forecasting what will happen next given historical observations is a long-standing yet increasingly critical capability in modern society. Accurate and timely forecasting not only promotes business and commerce, but also plays an essential role in public safety, financial stability, and cyber and national security. Existing research on forecasting has primarily focused on well-structured time series data, while real-world forecasting problems are predominantly driven by large-scale, multimodal, and unstructured data, such as text, images, graphs, and interaction logs. Leveraging such data for forecasting remains a fundamental challenge.

My research focuses on developing data-driven and computational methods for reliable forecasting from multimodal data, with forecasting targets ranging from macro-level events (e.g., global fashion trends and international political events) to micro-level behaviors (e.g., user activities on online platforms and attackers' behaviors during cyber-attacks). While multimodal data provides richer signals and broader context, it also introduces new challenges related to noise, sparsity, temporal misalignment, and uncertainty, which directly affect the reliability and trustworthiness of forecasting outcomes.

Across these forecasting objectives, several core challenges must be addressed.

- Real-world multimodal data are unstructured, noisy, and often lack high-quality labels, making it infeasible to directly apply forecasting models. Therefore, acquiring temporally grounded and task-relevant information from multimodal data is a key prerequisite for downstream forecasting.
- Macro-level event forecasting typically suffers from data sparsity, stochasticity, and rapidly evolving contexts. Effectively leveraging external information while maintaining robust and reliable predictions is crucial for these high-impact scenarios.
- Micro-level behavior forecasting involves heterogeneous entities and complex interaction patterns. Capturing these fine-grained temporal and relational dynamics is essential for accurate and explainable forecasting.

Considering these challenges, my research vision is to lay the technical foundations for **reliable multimodal forecasting** in real-world critical problems, spanning both macro-level events and micro-level behaviors. To pursue this vision, my research agenda is centered on establishing a unified and general framework that consumes large-scale real-world multimodal data and produces forecasting results that are not only accurate, but also robust, interpretable, and decision relevant. Consistent with this goal, my work is organized around three interconnected research problems: 1) Forecasting-oriented Multimodal Information Acquisition; 2) Macro-level Temporal Forecasting; and 3) Micro-level User Behavior Modeling. These research problems are motivated and validated through practical applications, including fashion trend analysis, finance forecasting, temporal political event forecasting, recommender systems, and cyber-attack analysis. Moving forward, the fundamental techniques developed in multimodal forecasting will be further employed in high-stake scenarios such as cybersecurity and national security, where forecasting reliability is especially critical.

**Research Areas**

**1. Forecasting-oriented Knowledge Acquisition**

Knowledge acquisition is the pre-requisite step for forecasting since the raw data in real world are highly unstructured, noisy, and more importantly, in large scale. Moreover, with the remarkable success of Large Language Models (LLMs), the reliability of internal knowledge within LLMs is an important source of knowledge. I explore this problem of knowledge acquisition as follows:

**1) Fashion Knowledge Extraction from Social Media.**

Even though multimodal information extraction has been studied in the literature, the information required by various forecasting tasks is highly domain- and task-specific, suffering from low-quality datasets with noisy or even no labels. Especially for the social media data, it is highly expensive and labor-intensive to manually annotate the dataset. Considering such constraints, I leverage machine-generated labels as the base dataset, then propose a weak label learning module to calibrate the biased machine-generated labels. Based on this framework [9], I develop a demo platform FashionKE [8] for fashion knowledge acquisition, which has been used to construct fashion trend forecasting dataset FIT [10].

**2) Complex Event Detection and Extraction using LLMs**.

Complex event refers to the super event that consists of a series of correlated atomic events along the timeline. Detecting and extracting the structured representation plays a pivotal role in understanding and forecasting events in critical scenarios, such as civil unrests and geopolitical conflicts. However, it is also infeasible to manually annotate large-scale datasets since recruiting well-trained qualified domain experts is extremely difficult and expensive. Fortunately, the great success of LLMs shed light on this problem, and I propose a framework to utilize LLMs to detect and extract complex events and build a large-scale dataset MidEast-TE [13]. It is the first dataset that endows all the features of Structured, Complex, and Time-Complete, which are essential and critical properties for temporal event forecasting.

**2. Macro-level Temporal Forecasting**

Macro-level temporal forecasting aims to predict the evolving trends or major events that reflect by groups of people and have salient impacts to the society. The main challenge of this problem lies in how to leverage various information thus to achieve precise and reliable forecasting. I conduct several studies as follows.

**1) Knowledge-aware Fashion Trend Forecasting.**

Forecasting the trend of each fashion element separately ignores the potential relations among multiple time series. To address this limitation, I propose to integrate various knowledge, including both internal knowledge endowed by time series and external knowledge of fashion ontology, to a neural time series forecasting model [10][17]. Our methods largely boost the SOTA performance of fashion trend forecasting. In addition, we curate a large-scale dataset FIT [10], which is featured with more fine-grained fashion elements and longer time span compared with previous datasets. It fills the gap of lacking high-quality dataset in the community of fashion trend analysis, which even attracts much attention from the fashion community.

**2) Context-aware Temporal Event Forecasting.**

Temporal event forecasting confronts the challenge of large set of candidate entities. Previous approaches mainly account for the structured representation of events, while overlooking the rich textual information that provides critical contexts to the events. Inspired by such motivation, I propose a context-aware event forecasting method through graph disentanglement [12]. In addition to the coarse-grained categorical contexts, we propose a novel formulation to introduce the concept of complex event into the structured temporal event forecasting problem, thus, to achieve more fine-grained context modeling. Based on such formulation, we design an effective forecasting model by leveraging both local and global contextual information, which significantly improve the forecasting performance [13].

**3) Explainable and Robust Financial Forecasting.**

Financial forecasting is a high-stake macro-level forecasting problem where robustness to uncertainty and interpretability are critical. Traditional time-series models often struggle with stochastic market dynamics and provide limited insight into the reasoning behind predictions. To address these limitations, I study explainable and robust financial forecasting by integrating probabilistic modeling, time-series reasoning, and large language models. Specifically, I propose a diffusion variational autoencoder to explicitly model stochasticity in multi-step stock price prediction, improving robustness under volatile conditions [5]. More recently, I investigate self-reflective large language models that generate stock predictions together with natural-language explanations grounded in historical market patterns, enhancing transparency and decision support [6].

## 3. Micro-level User Behavior Modeling

Micro-level forecasting seeks to model the fine-grained human-level behaviors. Different from the macro-level events that are results of groups or the whole society, micro-level forecasting requires to grasp the trivial and fine-grained patterns in terms of both users and forecasting objectives. I study this problem from two main aspects:

**1) User Behavior Forecasting in Sequential and Bundle Recommendation.**

Recommender system has become the foundational architecture to assist users in information seeking, where the key is to capture the patterns of users' historical behaviors and make accurate forecasting. Specifically, for sequential recommendation, most previous approaches cannot achieve optimal forecasting performance due to cross-sequence and latent intent modeling capacity. Worse still, the modeling of users' instant intent, which is the behind true factors that affect users' behaviors, has long been ignored. To bridge these gaps, we propose to leverage two types of global graphs, i.e., user-item interaction and item-item transition graphs, for sequential recommendation [1]. Furthermore, we explicitly model users' instant intent and integrate such modeling into sequential recommendation, which improve both accuracy and explainability of the results [2]. For bundle recommendation, it aims to recommend users a set of pre-defined items, i.e., a bundle, under certain product bundling strategies. Thereafter, the problem gets to be more complicated due to the introduction of two auxiliary relations: bundle-item affiliation and user-bundle interaction. The key challenge is how to fully utilize the heterogeneous relations and maximize their complementary effects. To address this problem, we employ contrastive learning to strengthen the purely graph learning-based approaches [4][11][14][16]. Moreover, instead of just recommending pre-defined bundles, we propose to automatically construct bundles by leveraging multimodal features and item-level user feedback [15].

**2) Attackers' Behavior Forecasting for Cyber Attack Analysis.**
Cyber-attacks are increasingly complex and diverse, where the attackers often leverage complicated tactics and techniques that threaten the security of cyber space. To combat these attacks, security practitioners actively summarize and report their witnessed attacking processing in the form of CTI reports, which systematically record the attackers' behaviors during the cyber-attack. Such attack behaviors or actions follow certain pattern and could be identified. Motivated by this intuition, we propose to forecast the attackers' behaviors during cyber-attacks. Specifically, we formulate the attack process as a temporal graph [7], and we propose to leverage both structural information and textual information to conduct collaborative reasoning, thus, to achieve accurate forecasting of the attacker's next behavior.

## 4. Future Works

Even though we have achieved great progress in the problem of macro- and micro-level forecasting on multimodal data, we are still far from solving the problems due to an extensive number of challenging problems yet to be explored. Moving forward, I plan to advance my research on multimodal forecasting to tackle the challenges in Framework, Input Data and Forecasting Objectives, especially in the domain of Cyber Security and National Security.

**First, it is crucial to unify multimodal extraction and forecasting into a holistic framework.** Even though the current extraction system is designed to satisfy various forecasting needs, it is still independent with the downstream forecasting model. Thereafter, the errors induced by the extraction process would inevitably propagate to and severely affect the forecasting module. For example, new types of entities and relations appear with the event unfolding, and such newly emerging information is critical for the forecasting performance while being missed by the outdated information extraction system [12][13]. How to unify the extraction and forecasting into a holistic framework and enable the extraction module to swiftly adjust catering to the requirement of the forecasting model is a key challenge.

**Second, I will explore effective forecasting methods by leveraging long-context and multimodal data.** Even though several forecasting works have utilized multimodal data, there is still a large gap in encapsulating all the accessible modalities and maximizing multimodal complementary effects. For example, in political event forecasting, textual and structural data are the two widely used modalities. Meanwhile, the visual modality, such as image and video that represent more concrete and detailed information, has not been exploited. On the other hand, current forecasting works are restricted to short-term context and cannot handle long context. For example, in complex event forecasting, one complex event may involve hundreds or even thousands of historical news articles, and how to make full use of the information in such a long context remains to be solved.

**Third, I will investigate forecasting problems in early-stage and long-term scenarios.** Existing works only focus on next day or next action prediction, while ignoring the more challenging settings of early stage and long-term forecasting. First, some critical scenarios, such as civil unrest and geopolitical conflicts, require swift response to emergency incidents. However, it is difficult to make accurate forecasting given the extremely short context and limited information at the very early stage of the event. Second, in some applications such as macro economy, people need relative long leading time to prepare for countering potential negative or even catastrophic outcomes. Thereafter, enhancing the long-term forecasting induces another new challenge. However, it is non-trivial to make long-term forecasting due to the increasing volatility and uncertainty with the forecasting horizons increasing.

To tackle the above challenges, the main research directions include: **1) Leverage the powerful Large Language Models (LLMs) and Large Multimodal Models (LMMs) for multimodal data extraction and forecasting [6][13].** Even though there some preliminary studies of using LLMs for extraction and forecasting tasks, they are far from fully unleashing the potential power of LLMs. **2) Explore the retrieval augmented generation (RAG) to tackle the problems of long-context and early-stage forecasting.** RAG can retrieve concise and crucial information from large-scale historical data and external knowledge, based on which the forecasting model would be able to make sound and reliable forecasting. **3) Investigate robust and stable learning to reduce the stochasticity in early-stage and long-term forecasting.** In addition to the conventional adversarial learning, it is interesting and promising to employ LLM-based multi-agent system with the self-reflection and mixture-of-experts (MoE) mechanisms.

**Most importantly, I will ground these fundamental research ideas to the scenarios of cyber security and national security.** First, I will explore the problem of cyber threat and pertinent alleviation technology trend forecasting, through analyzing multimodal multi-source online information. Second, I will investigate the problem of multimodal event extraction, forecasting, and intervention, thus, to achieve enhanced national security. Empowered by the multimodal data forecasting technology, I aim to upgrade the conventional reactive and passive approaches to predictive and proactive solutions. Therefore, security practitioners and governments would have sufficient leading time to make preparation for potential risks or even apply interventions to prevent catastrophic disasters.

## Selected Publications and Outputs


[1] Yujuan Ding, **Yunshan Ma**, Wai Keung Wong, and Tat-Seng Chua. Leveraging two types of global graph for sequential fashion recommendation. In International Conference on Multimedia Retrieval, pages 73–81. ACM, 2021. Best Student Paper.

[2] Yujuan Ding, **Yunshan Ma**, Wai Keung Wong, and Tat-Seng Chua. Modeling instant user intent and content-level transition for sequential fashion recommendation. IEEE Transactions on Multimedia, 24:2687–2700, 2022.

[3] Yujuan Ding, P. Y. Mok, **Yunshan Ma**, and Yi Bin. Personalized fashion outfit generation with user coordination preference learning. Information Processing and Management, 60:103434, 2023.

[4] Xiaoyu Du, Kun Qian, **Yunshan Ma**, and Xinguang Xiang. Enhancing item-level bundle representation for bundle recommendation. ACM Transactions on Recommender Systems, 2023.

[5] Kelvin J. L. Koa, **Yunshan Ma**, Ritchie Ng, and Tat-Seng Chua. Diffusion variational autoencoder for tackling stochasticity in multi-step regression stock price prediction. In ACM International Conference on Information and Knowledge, pages 1087–1096. ACM, 2023.

[6] Kelvin J. L. Koa, **Yunshan Ma**, Ritchie Ng, and Tat-Seng Chua. Learning to generate explainable stock predictions using self-reflective large language models. The Web Conference, abs/2402.03659, 2024.

[7] Yongheng Zhang, Tingwen Du, **Yunshan Ma**, Xiang Wang, Yi Xie, Guozheng Yang, Yuliang Lu, Ee-Chien Chang. AttacKG+: Boosting attack graph construction with Large Language Models. Computers & Security, 2024.

[8] **Yunshan Ma**, Lizi Liao, and Tat-Seng Chua. Automatic fashion knowledge extraction from social media. In ACM Multimedia, pages 2223–2224. ACM, 2019.

[9] **Yunshan Ma**, Xun Yang, Lizi Liao, Yixin Cao, and Tat-Seng Chua. Who, where, and what to wear?: Extracting fashion knowledge from social media. In ACM Multimedia, pages 257–265. ACM, 2019.

[10] **Yunshan Ma**, Yujuan Ding, Xun Yang, Lizi Liao, Wai Keung Wong, and Tat-Seng Chua. Knowledge enhanced neural fashion trend forecasting. In International Conference on Multimedia Retrieval, pages 82–90. ACM, 2020.

[11] **Yunshan Ma**, Yingzhi He, An Zhang, Xiang Wang, and Tat-Seng Chua. Crosscbr: Cross-view contrastive learning for bundle recommendation. In ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 1233–1241. ACM, 2022.

[12] **Yunshan Ma**, Chenchen Ye, Zijian Wu, Xiang Wang, Yixin Cao, and Tat-Seng Chua. Context-aware event forecasting via graph disentanglement. In ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 1643–1652. ACM, 2023.

[13] **Yunshan Ma**, Chenchen Ye, Zijian Wu, Xiang Wang, Yixin Cao, Liang Pang, and Tat-Seng Chua. Structured, complex and time-complete temporal event forecasting. CoRR, abs/2312.01052, 2023.

[14] **Yunshan Ma**, Yingzhi He, Xiang Wang, Yinwei Wei, Xiaoyu Du, Yuyangzi Fu, and Tat-Seng Chua. Multicbr: Multi-view contrastive learning for bundle recommendation. ACM Transactions on Information Systems, 2024.

[15] **Yunshan Ma**, Xiaohao Liu, Yinwei Wei, Zhulin Tao, Xiang Wang, and Tat-Seng Chua. Leveraging multimodal features and item-level user feedback for bundle construction. ACM International Conference on Web Search and Data Mining, abs/2310.18770, 2024.

[16] Yinwei Wei, Xiaohao Liu, **Yunshan Ma**, Xiang Wang, Liqiang Nie, and Tat-Seng Chua. Strategy-aware bundle recommender system. In International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 1198–1207. ACM, 2023.

[17] Yujuan Ding, **Yunshan Ma**, Lizi Liao, Wai Keung Wong, and Tat-Seng Chua. Leveraging multiple relations for fashion trend forecasting based on social media. IEEE Transactions on Multimedia, 24: 2287–2299, 2022.