Combining Graph Neural Network with Attention Mechanism in Sequence-Based Music Recommendation System

1st Rama Aulia Gemilang
School of Computing
Universitas Telkom
Bandung, Indonesia
gemilangrama@student.telkomuniversity.ac.id

2nd Agung Toto Wibowo
School of Computing
Universitas Telkom
Bandung, Indonesia
agungtoto@telkomuniversity.ac.id

***Abstract*—Music recommendations have become an important part of everyday life. With the increasing amount of music available, it has become increasingly difficult to find suitable music content. To address this issue, recommendation systems have emerged as a solution to personalize song suggestions for users. Traditional recommendation systems often fail to capture sequential patterns of user interaction. To understand these patterns, recommendation systems are based on a sequential approach. However, traditional sequential recommendation system models do not sufficiently capture dependencies on long-term temporal patterns. In this study, we apply a combination of graph neural networks with attention mechanisms to the Music4All dataset to address sequential music recommendations. This approach, which adapts the GASM framework, is applied for the first time to this music dataset, which has greater diversity and user distribution, providing new insights into the effectiveness of this model in the music recommendation domain. Graph neural networks capture the relationships between users and items, as well as between items themselves, while the attention mechanism captures user interest patterns over time within a session. We also incorporate long-term interests and dynamic interests to enrich the context of user preferences. By applying this model to the Music4All dataset, the model can recommend the next music track with a hit ratio evaluation ranging from 32.69 to 49.16 and a mean reciprocal rank ranging from 23.34 to 30.71. These results show that our approach can effectively capture sequential patterns in music in large-scale datasets.**

***Index Terms*—Sequential recommendation system, music, graph neural network, attention mechanism, preference.**

I. Introduction

In this digital age, music has become almost inseparable from everyday life. This is followed by the growing music production. The development of music production makes the number of songs available to consumers also become more and more [1]. This makes it more difficult for consumers to find music content that suits their preferences because they will spend a lot of time searching for music that suits their tastes, which can reduce the user experience in finding music [2]. To reduce the burden on consumers in finding music content, recommendation systems play an important role in this regard.

A recommendation system is a system that can personalize with the aim of providing product suggestions that may be of interest to users [3]. The main goal is to improve the user experience. Improved user experience in recommendation systems can be achieved from user interactions over time. In this case the recommendation can be based on a time sequence approach called a sequential recommendation system.

Sequential recommendation systems are tasked with analyzing user interaction history to predict consumer preferences [4]. User interactions can be a history of accessing sites, listening to songs, watching videos, etc. The interaction history will usually form a time sequence pattern following the transition from one item to the next. Using interaction time sequence patterns, the recommendation system will accurately understand user preferences. One of the implementations of sequential recommendation systems is through traditional approaches, such as markov models that provide item predictions by considering the last interaction. However, such models cannot handle long-term sequence dependency [5]. In addition, there are latent factor models that work by learning latent representations to analyze transitions between items. However, transition models often vary across different data, which makes them in a position to capture complex transition relationships [5].

To overcome sequence dependency, researchers have conducted several experiments, one of which uses a neural network graph representing user interactions with items and attention mechanisms that provide important weights in the order of user history. Research by Chang et al. [6], has integrated a graph neural network model as a graph of user relationships with items and attention mechanisms as weights in assessing preferences in short-term transitions. The integration of the model showed better results than the baseline in handling the dependence of time sequence [6].

In this study, we adapt the Graph-based Attentive Sequential model with Metadata (GASM) approach [7]. The model works by combining a graphical neural network with an attention mechanism, where the graphical neural network forms connections between users and items, as well as between items and items, while the attention mechanism forms short-term preferences by weighting items based on changes that occur over time during a session. In addition, other preferences are built dynamically over the long term, creating complex

preferences.

The main contribution of this study lies in the application of a recommendation model that has been adapted to the different data characteristics from previous studies. The previous GASM models used datasets with relatively limited numbers of users, such as LastFm with 896 users, 30Music with 2,965 users, and Xiami with 3,982 users, while this study used the Music4All dataset with a much larger number of users, namely 7,998 users. This significant difference allows the model to be tested on a scale that is more representative of real-world use cases. This approach distinguishes our research from the GASM model, which was previously tested on data with a different scope. Thus, this study provides a new perspective on music recommendation modeling based on more diverse user preferences.

II. Related Work

A. Sequential Recommendation System

A sequential recommendation system (SRS) is a recommendation system with the aim of processing information from a sequence of interactions that are useful in predicting the next item [8]. This system has proven to be very effective in suggesting items or products to users because user interactions with items or products usually occur successively in a time sequence. In contrast to user-based, item-based, and content-based recommendation systems that cannot properly capture user interaction patterns [9]. Some research on sequential recommendation has been done, for example research by Fan et al. [10] uses a stochastic model as a representation of items and self attention to capture the relationship between items in sequence. Another example is research from by Zhang et al. [11], uses a deep learning model with a dynamically formed graph neural network to model the sequence of user behavior dynamically and link prediction to predict the next item.

B. Graph Neural Network

Graph Neural Network is a model of deep learning that works on graphs. GNN processes information sets from its neighbor relationships with a message-passing mechanism [12]. In SRS, GNN works to form interaction patterns between users and items. Several research methods regarding GNN in SRS have been conducted with superior results compared to traditional methods. For example, research by Chang et al. [6] showed that the GNN model achieved the highest performance in SRS with a mean reciprocal rank value of 0.4 to 0.9.

C. Attention Mechanism

Attention mechanism is a mechanism for the selection process by giving weight to existing inputs [13]. The weight given is formed like a selective focus, where the weight will be grouped at different levels of importance [14]. In SRS, the attention mechanism works to shape user preferences by capturing interaction changes over time. Several research methods regarding attention mechanism in SRS have been carried out, such as the example of research by Kang et al [15] giving satisfactory results by outperforming the baseline on the improvement hit ratio value around 5% to 21%.

III. System Design

This section will explain how the model was developed in this study. The model process is illustrated in Fig. 1, which shows the overall flow of the research design of the Combining Graph Neural Network with Attention Mechanism in Sequence-Based Music Recommendation System model. Meanwhile, the core design of the model, namely graph construction to preference building, is illustrated in Fig. 2.

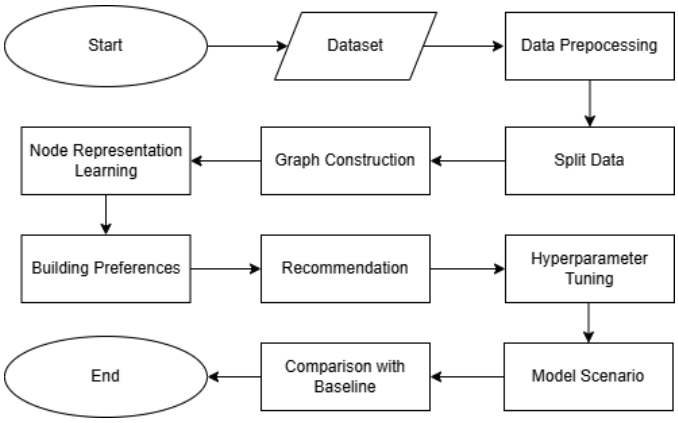

Fig. 1. Research Design on Combining Graph Neural Network with Attention Mechanism in Sequence-Based Music Recommendation System

A. Dataset

This research uses the Music4All [16] data which is obtained directly from the Google Sites Music4All site: A New Music Database. The Music4All data has many dataset and those used in this research are the id_information and listening_history sections. In the id_information, music metadata can be formed consisting of albums and artists. Meanwhile, listening_history will form information about users with item access to music items. We chose this dataset because it has a large and complex amount of user data and music information, allowing the model to learn many patterns of user history and preferences.

B. Data Prepocessing

Data preprocessing is done to process raw data into ready to use data for sequential recommendation. In data preprocessing, we take the first step by forming input data that provides metadata (artist and album) and user information about song listening history. Next, we select the data to compact the data and keep the computing resources stable. In this case, the top 8000 users are selected by ignoring the frequency of music appearing less than 10. Then, we perform index mapping between users and music items with album and artist information. Next, from the item index, time-based sequential data is formed. The formation was done by setting the listening length in the range of 200-500 to maintain the consistency of data from the top users. Finally, we divide the data into 80% train data which contains the previous listening history and 20% test data which contains the latest listening history. From data preprocessing, it produces user statistics of 7998, music