ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Research on Intelligent recognition of music theory evolution law for music history research

Wenqi Ma[1,*]

[1] College of Music, Michigan State University, 333 W Circle Dr, East Lansing, MI 48824, The United States of America

SUMMARY: *An intelligent recognition framework combining text computing, symbolic music modeling and historical context analysis is constructed to solve the problems of low efficiency of theoretical evolution recognition and dependence on manual induction across periods in music history research. Based on theoretical literature, symbol genealogy examples and metadata, this paper establishes a multi-dimensional label database, and designs a collaborative recognition model that integrates term features, harmony-mode features and historical context features, which is used to distinguish historical periods, distinguish theoretical schools and extract evolution paths. On the historical data set of music theory, which contains theoretical texts, symbolic music examples and historical metadata, the experiment divided the training set, validation set and test set according to 8:1:1. The results show that the proposed model achieves 89.8% Accuracy, 88.9% Macro-F1 Accuracy, and 96.4% Top-3 accuracy in the music theory evolution recognition task. In the music history case verification, the average recognition accuracy reaches 90.6%, and the interpretation consistency score is 4.6. The results show that the proposed method can stably capture the linkage relationship between theoretical concepts, music example structure and historical context, and provide a computable, traceable and interpretable analysis tool for the research of music theory history.*

KEYWORDS: *Music history research; Evolution of music theory; Intelligent recognition; Symbolic music analysis*

1 Introduction

With the continuous development of digital humanities research, music history research is gradually shifting from the traditional path of text interpretation and case research to the comprehensive paradigm of data-driven and computational analysis. Music theory is not a static knowledge item, but a historical structure that has been continuously generated, modified and expanded in different periods of creative practice, notation systems, modal concepts, and terms of harmony organization and analysis. For researchers, the real difficulty is not only to identify the meaning of a theoretical concept in a specific period, but also to reveal how these concepts have undergone association reorganization, category transfer and functional deformation in the long-term evolution. Relying on manual reading through the literature, comparing the music examples and summarizing the experience can form a detailed judgment, but this method has revealed obvious limitations in processing efficiency, scale expansion and rule extraction when facing cross-period, multi-lingual, multi-version documents and large-scale symbolic music data.

*mawenqicello@163.com
https://doi.org/10.65102/is2026221
Published: 30 April, 2026

In recent years, music information retrieval, knowledge graph, symbolic music modeling and deep learning methods have provided a new technical foundation for the research of music theory history. On the one hand, the establishment of harmony corpus, functional harmony ontology, music knowledge graph and multi-modal symbolic representation have enabled the elements of music theory to obtain computable, comparable and traceable digital expressions. On the other hand, the development of sequence modeling, graph structure learning and representation learning techniques also enables the implicit connections between theoretical concepts, style boundaries and historical transition characteristics to be identified and verified in a wider scope. However, existing researches focus more on harmony analysis, style classification, melody modeling or generation tasks, and pay insufficient attention to the more historical problem of "the evolution law of music theory". Existing methods tend to emphasize local classification accuracy, but rarely deal with the problems such as the semantic drift of theoretical terms across periods, the inconsistent expression of different schools, and the difficulty of unified modeling of historical knowledge and musical score evidence. As a result, even if the model can recognize some surface features, it may not be able to explain the continuity, discontinuity and path structure behind theoretical changes.

Based on this, this paper regards the theoretical evolution recognition in music history research as a composite task that integrates literature computing, symbolic modeling and historical law extraction, and attempts to construct an intelligent recognition framework for music theory historical data. Starting from the digital representation of music theory evolution elements, this paper integrates theoretical terms, modal organization, harmony progression, termination form, texture structure and historical context labels into a unified data system. On this basis, multi-layer feature extraction and adaptive optimization mechanism are introduced to form a recognition model that takes into account both classification performance and historical interpretation ability. Different from the technical route solely oriented to music content classification, this paper focuses on the ability of the model to describe the process of theoretical change, focusing on the identification of stable characteristics, transition interval and mutation direction of theoretical characteristics in different historical periods, and further characterizing the differentiation trajectories and correlation between theoretical schools.

The goal of this paper is not only to improve the recognition accuracy, but also to establish a more stable docking relationship between music history research and computational methods. On the one hand, through standardized data modeling and label design, it provides a reusable scheme for the digital collation of music theory historical materials. On the other hand, through intelligent identification and path extraction, it provides an analysis tool for the study of music theory evolution, which can take into account both macroscopic law discovery and microscopic case interpretation. Therefore, this paper integrates model performance evaluation, case verification and visual interpretation into the same research framework, and tries to make the calculation results not only "computable", but also "readable" and "provable". In this sense, the intelligent identification of the evolution law of music theory for the study of music history is not a simple technical packaging of traditional musicology problems, but a methodological extension of the organization of historical materials, the method of theoretical comparison and the mechanism of evidence generation.

2 Music theory digital modeling and data system construction for music history research

After the research problem is clarified in the previous section, whether the evolution law of music theory can be stably identified depends to a large extent on the structural quality of the

underlying data system. Different from general music classification tasks, music history research is not faced with a single audio object, but a composite material composed of theoretical terms, score structure, analytical discourse, historical period and genre context. Without a unified data modeling framework, even if the subsequent models have strong fitting ability, it is difficult to make reliable judgments on the time sequence of theoretical changes, concept differentiation and path extension. Therefore, this chapter focuses on the digital expression of the evolution elements of music theory, the collection and standardized processing of literature samples, and the design of multi-dimensional labels and database structures, which provide a computable, traceable and scalable data foundation for subsequent intelligent recognition models.

2.1 Analysis and digital representation of evolution elements of music theory

The historical evolution of music theory is not only manifested in the replacement of term names, but also reflected in the continuous reorganization of mode centers, harmonic organization, termination mode, texture writing, voice movement and analysis category. In order to enable these changes to enter the computational framework, this paper divides the evolution factors into three categories. The first is the structural factors, including interval distribution, chord type, termination position, mode stability and voice part progression mode. The second is the conceptual elements, including theoretical terms, functional interpretation, category attribution and school expression differences; The third is the historical context elements, including time, region, author, edition, educational communication path and related ideological background. For structural elements, we use symbolic music coding, harmony event segmentation and sequence vectorization methods to convert the spectral information into discrete representations that can be input to the model. For the conceptual elements, the term standardized mapping and semantic embedding technology were used to uniformly represent the synonymous, near-synonymous and transfer expressions in different literatures. For the historical context elements, timestamps, genre nodes and associated edges are constructed to form the historical relationship structure that can be called by the subsequent graph model. Through the above processing, the evolution of music theory is no longer a text description, but transcribed into a multi-level data object that can be analyzed.

2.2 Music history literature collection, labeling and standardization processing

In order to construct a representative research sample, this paper collects relevant materials from music theoretical literature, textbooks, composer reviews, score compilations and digital music resources, covering Baroque, classicism, Romanticism and several important stages since the twentieth century. Considering the differences in terminology system, notation and version quality between different source texts, the data should be uniformly preprocessed before entering the system. In the text part of the literature, the process of "segment extraction-term identification-manual review" was used to extract paragraphs involving mode, harmony, counterpoint, structure analysis and other core contents. The code cleaning, repeated sample elimination and segment segmentation are carried out to retain the effective fragments that can reflect the theoretical characteristics. The annotation mechanism adopts a double-layer method of "theoretical feature annotation + historical attribute annotation": the former records the content of harmony function, termination type, mode shift, syntactic organization, etc., and the latter records the period, theoretical school, source and version information of the literature. In order to reduce the interference caused by foreign terms and old-style notation, we further

introduce term merging, spelling normalization and cross-version alignment strategies to collate the original materials into a unified data input format, so as to ensure the stability and comparability of the model in the training phase.

2.3 Multi-dimensional label system and database design

At the data organization level, it is difficult to support the deep association analysis in the evolution rule recognition if only using the common label structure of "literature name-author-year". Based on this, this paper designs a multi-granularity and multi-level tag system. The first is theoretical feature labels, including mode categories, harmony structure, termination style, texture morphology, syntactic hierarchy and typical interval patterns. The second is historical semantic tags, including era sections, theoretical schools, regional traditions, author identities and knowledge source types. The third is the evolution relation label, which is used to record dynamic information such as concept inheritance, category differentiation, term substitution and structural transformation. The database structure is a hybrid scheme combining relational database and nested document storage. The relational table is responsible for managing document primary key, period index and label mapping, and the document layer stores instance encoding, term vector, context fragment and time correlation information. This structure not only supports conditional retrieval for history stage, but also facilitates batch reading and parallel training of deep learning models. At the same time, the system sets the label consistency check and abnormal sample detection mechanism to prevent the mismatch or drift of the same theoretical concept in different versions of data. The resulting data system not only provides stable input for subsequent intelligent recognition, but also lays the information foundation for intertemporal comparison and visual analysis in music history research.

3 Construction and optimization of intelligent recognition model of music theory evolution law

3.1 Intelligent recognition model design and feature extraction

After completing the digital modeling and label organization of music theory historical materials, the key of the recognition task turns to how to extract the core information that can truly represent the "theory evolution" from the heterogeneous materials. Different from general music style classification, the object we face in this paper contains three types of data: theoretical terms, symbolic examples and historical context, which are inconsistent in time scale, representation and semantic density. If only a single network is used for end-to-end classification, the model can only capture local co-occurrence features, and it is difficult to identify the functional transfer of a theoretical concept in different periods and the deep connection between it and the spectral structure. Based on this, this paper constructs a multi-branch intelligent model for music theory evolution recognition, which expands the three pathways of "term encoding - music example pattern extraction - historical context modeling" in parallel, and completes cross-feature alignment and weighted aggregation in the fusion layer. The overall structure is shown in Figure 1.

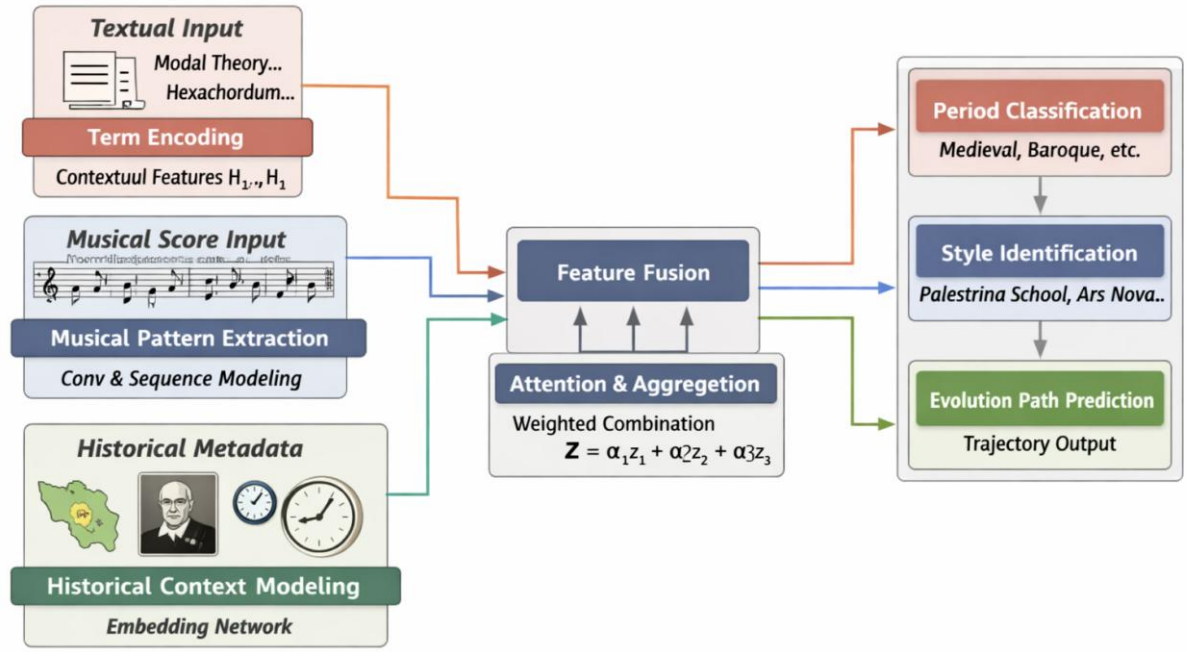

Figure 1: Structure diagram of intelligent recognition model for evolution law of music theory

Among them, the term encoding branch mainly deals with the core concept sequence in theoretical works, textbook fragments and analysis texts. After word segmentation, term merging and position encoding, the text is input into the context encoder to obtain the dynamic representation of different concepts in specific historical context. Let the sequence of document fragments be $X = \{x_1, x_2, \ldots, x_n\}$, whose context features can be expressed as:

$$H_t = \text{Encoder}(x_t, p_t), \quad t = 1,2, \ldots, n \tag{1}$$

where, x_t is the t-th term or lemma embedding, p_t is the corresponding position vector, and H_t is the encoded semantic representation. This branch focuses not on the frequency of single term occurrence, but on the semantic position changes of terms in definition, comparison, revision and inheritance relationships.

The spectrum pattern branch is used to extract structural information such as mode organization, harmony connection, termination form and voice part movement. The input data is composed of standardized symbolic music fragments, which are mapped into discrete pitch, duration and function joint vectors after event segmentation. The repeated, transposed and transitional theoretical patterns are identified by combining local convolution and sequence modeling. In order to highlight the contribution degree of different features in historical discrimination, this paper introduces the attention weighting mechanism in the fusion stage. Let the outputs of the three types of branches be z_1, z_2, z_3 respectively, then the comprehensive representation is:

$$Z = \sum_{i=1}^{3} \alpha_i z_i, \alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{3} \exp(e_j)} \tag{2}$$

Here, α_i represents the weight of the ith class feature in the current sample, and e_i is the relevance score learned by the fusion layer. This mechanism enables the model to dynamically adjust the focus of attention in different task scenarios, such as increasing the weight of historical context features in period recognition, and enhancing the synergy between terms and

genealogy examples in genre discrimination.

The historical context branch is responsible for absorbing metadata such as age, region, author, version and knowledge transmission chain, and transcribe the originally scattered background information into time-sensitive embedding vectors. The significance of this treatment is that the evolution of music theory does not always appear in explicit terms, and many changes are actually implicit in spectral shifts, analytical perspective shifts, and intra-genre reinterpretations. Only when texts, musical examples and historical labels are integrated into the unified representation space, the model can recognize the common phenomenon that "similar structures have different theoretical meanings in different eras".

3.2 Optimization algorithm and parameter adaptation strategy

After the structure design of the recognition model is completed, the performance improvement of the model no longer depends solely on the increase of the depth of the network, but more depends on whether the parameter configuration can match the heterogeneous characteristics of the music theory historical data. The input of this paper contains term sequence, pedigree structure and historical metadata at the same time, and the sample distribution has problems such as unbalanced periods, overlapping genre boundaries and inconsistent label granularity. If the combination of fixed learning rate and static parameters is still used, the phenomenon of too fast convergence in the early stage, obvious oscillation in the later stage and insufficient recognition of minority categories are easy to occur in the training process. Based on this, an optimization strategy combining "hyperparameter search + dynamic adjustment" is introduced in the model training phase to jointly control the learning rate, batch size, hidden layer dimension, attention head number and regularization strength.

In the aspect of parameter optimization, Bayesian optimization is used to update the search space iteratively, and the macro-average F1 value on the validation set and the path identification consistency index are jointly used as the objective function to reduce the misdirection of the model direction by a single accuracy index. Let the vector of parameters to be optimized be θ and the objective function be denoted as $f(\theta)$, then the optimal parameters of each round of search can be expressed as:

$$\theta^* = \arg\max_{\theta \in \Omega} f(\theta) \tag{3}$$

Here, Ω represents the preset hyperparameter space. This strategy can preferentially explore more potential parameter regions under a limited training budget, and avoid the computational redundancy of traditional grid search in high-dimensional space. In the experiment, the learning rate search range was set from 1×10^{-5} to 5×10^{-4}, Dropout was controlled between 0.1 and 0.5, and the number of attention heads was set to 3 levels: 4, 8, and 12, so as to balance model capacity and training cost.

In order to enhance the stability in the later stage of training, this paper further uses cosine annealing mechanism to adaptively update the learning rate. Its expression is:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_0 - \eta_{min})\left(1 + \cos\frac{\pi t}{T}\right) \tag{4}$$

where η_0 is the initial learning rate, η_{min} is the minimum learning rate, t is the current training round, and T is the total number of rounds. This strategy makes the model maintain sufficient parameter update amplitude in the early stage of training, and gradually turn to fine-grained convergence in the middle and later stages, which is more suitable for handling the task characteristics of "local features are easy to learn, and deep rules are difficult to learn" in music

theory evolution recognition.

At the same time, considering the difficulty of different recognition subtasks is not consistent, we introduce dynamic loss weights to the three outputs of period recognition, genre discrimination and evolution path representation, and automatically adjust the training center according to the error change of each task in the current round. After this process, the model is no longer dominated by a certain type of easy-to-learn labels, and can allocate more update ability to samples with fuzzy boundaries and stronger historical transitions. On the whole, the optimization scheme adopted in this section is not simply for parameter tuning, but tries to make the model maintain a more stable convergence trajectory and stronger generalization ability in complex historical materials, so as to provide a reliable computational basis for subsequent rule extraction and case verification.

3.3 Analysis of model output mechanism and deployment adaptability

After the model structure and parameters are optimized, whether the recognition system can really serve the music history research depends on whether the output results are readable, traceable and deployable. Based on this, the output layer is designed as a three-in-one linkage mechanism of "classification results, path representation and explanation information". For a single input sample, the system generates the probability distribution of the historical period, the classification result of the theoretical school and the evolution path vector representation simultaneously, and returns the corresponding key terms, genealogy fragments and historical labels together. This design avoids the information compression caused by only giving a single class label, so that researchers can not only see "which category is classified into", but also further determine "why it is classified into this category" and "how is it related to the previous and subsequent stages".

For the output calculation, this paper adopts the two-layer mapping strategy. The classification task obtains the prediction probability of period or genre through the fully connected layer and Softmax, and the evolution path task uses the low-dimensional embedding vector to preserve the transfer direction of theoretical features. Its classification output can be expressed as:

$$\hat{y} = \mathrm{Softmax}(W_o Z + b_o) \tag{5}$$

where Z is the global representation after fusion, W_o and b_o are the output layer weights and biases, respectively, and y represents the posterior probability distribution of the target class. In order to enhance the interpretability of the results, the system further extracts the term nodes with higher attention weight and the pattern of music examples to form a "label-evidence-path" corresponding chain, so that the recognition results can be backtracked and checked by music history researchers, rather than remaining at the level of uninterpretable black-box judgment.

Considering that music history materials are used in inconsistent ways in research scenarios, this paper sets three modes of lightweight inference, batch analysis and database linkage at the deployment level. Lightweight reasoning is suitable for instant analysis of a single document or a single group of examples, batch analysis is suitable for large-scale period comparison, and database linkage is for long-term accumulation of digital humanities platform construction. The prototype system uses Python and PyTorch to complete the back-end reasoning, and the front-end displays the period probability, the school adjacency relationship and the path evolution graph through the visual interface, and supports JSON result writeback and secondary retrieval. Table 1 presents the main configurations for adaptation of model output to deployment.

Table 1: Model output mechanism and deployment adaptation configuration

Output Module	Output Form	Storage/Interface Method	Average Response Time	Applicable Scenario
Historical Period Identification	Probability vector and Top-3 labels	JSON/API	0.18 s/sample	Rapid judgment of a single document
Theoretical School Classification	Category label and confidence score	JSON/API	0.21 s/sample	Comparison of differences among theoretical schools
Evolution Path Representation	Low-dimensional embedding vectors, relationship chains	Vector database/graph database	0.34 s/sample	Diachronic evolution tracking
Visual Explanation Output	Keywords, score-example evidence, weight heatmaps	Web interface/exported report	0.42 s/sample	Academic interpretation and review

4 Design of music theory evolution law recognition method

4.1 Historical period music theory feature recognition task design

In order to make the recognition results of music theory evolution law can truly correspond to the study of music history stages, this paper designs the recognition of music theory features in historical periods as a discrimination task driven by multi-source information collaboration. This task is not satisfied with giving the period label of a literature fragment or pedigree sample, but requires the model to output theoretical evidence, structural basis and historical context clues supporting the judgment at the same time. Based on this goal, the system constructs a task process of "input normalization-period feature extraction-multi-dimensional fusion discrimination-evidence backtracking output", and its structure is shown in Figure 2.

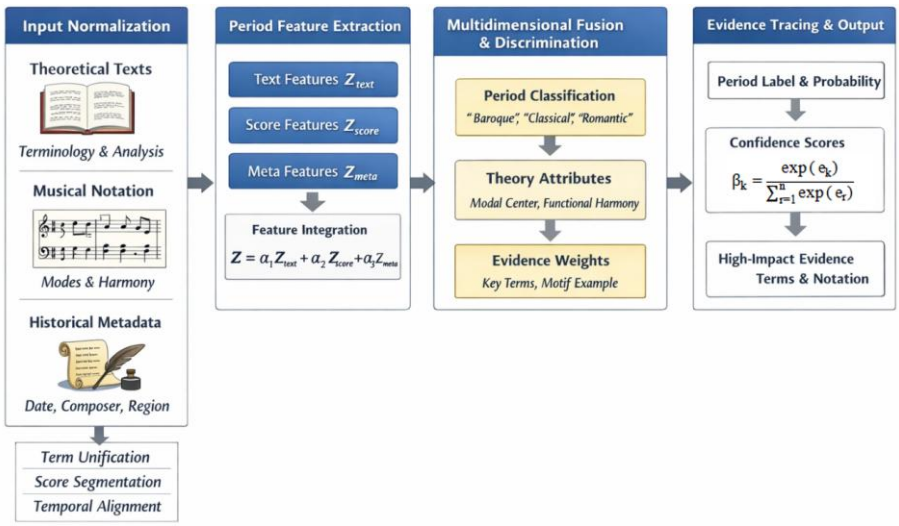

Figure 2: Design diagram of the feature recognition task for music theory in historical periods

In the input layer, the recognition objects are divided into three categories: theoretical text, symbolic pedigree and historical metadata. The theoretical text contains term definitions, category boundaries and analysis discourse, the symbolic notation reflects mode layout, harmony connection, termination form and voice part movement, and the historical metadata records the date, author, edition, region and background of knowledge dissemination. Since the

granularity and representation of the three types of data are significantly different, the system first performs term merging, pedigree segmentation and time alignment to map the heterogeneous materials into a unified representation space. Let the text feature, pedigree feature and metadata feature be Z_{text}, Z_{score} and Z_{meta}, respectively, then the period representation after fusion can be written as

$$Z = \alpha_1 Z_{text} + \alpha_2 Z_{score} + \alpha_3 Z_{meta}, \alpha_1 + \alpha_2 + \alpha_3 = 1 \tag{6}$$

Here, α_i is the weight adaptively learned by the model according to the sample content. After this process, the period identification is no longer dependent on a single term or a single spectrum, but is completed under the joint action of multi-dimensional evidence.

In the task label design, this paper uses a three-level structure to organize period information. The first level is the macro-period label, such as Baroque, classicism, Romanticism and modern transition stage. The second level is the meso theory feature labels, including mode center, functional harmony, termination mode, texture morphology and syntactic organization. The third level is evidential labels, which correspond to specific term fragments, genealogy patterns and annotation sources. Based on this label system, the model outputs the probability distribution of each historical period:

$$\hat{y}_i = \frac{\exp(w_i^T Z + b_i)}{\sum_{j=1}^{m} \exp(w_j^T Z + b_j)}, i = 1,2,\dots,m \tag{7}$$

where, m is the number of period categories and $\hat{y}_i$ represents the predicted probability that the sample belongs to the ith historical period. This expression preserves the competition relationship between adjacent periods, and can deal with the transitional samples and boundary fuzzy phenomena commonly seen in the evolution of music theory.

In order to enhance the interpretability of the results, the system further constructs an evidence tracking mechanism at the output side. For the KTH fragment of theoretical evidence in the input sample, the weight of its contribution to period discrimination is defined as

$$\beta_k = \frac{\exp(e_k)}{\sum_{r=1}^{n} \exp(e_r)} \tag{8}$$

Here, e_k is the relevance score of the evidence fragment in the fusion space, and β_k represents its contribution to the judgment of the final period. With the help of this mechanism, the model can synchronously return period labels, confidence scores, and high-weight terms and pedigree evidence, so as to form an explanation chain of "period-feature-evidence" integration.

The significance of the design of this task is that it transcribe the judgment process that originally relied on empirical induction in the staging of music history into a calculable, retrospective, and comparable recognition process. Therefore, the identification of historical periods is no longer just a static classification link, but becomes the basic interface for the subsequent discrimination of theoretical school differences and the extraction of evolution paths, and also provides a more stable computational support for the comparison of large-scale samples in the research of music theory history.

4.2 Theoretical school difference identification and category discrimination mechanism

In the study of the history of music theory, the differences between different schools are often

not manifested in the simple distinction of names, but hidden in the habit of using terms, the framework of harmonic interpretation, the concept of mode organization and the continuous shift of analysis focus. Even though some theoretical literatures discuss the same pedigree, they may give completely different functional attribution and structural explanation due to different school positions. If we only rely on keyword matching or single tag retrieval, it is easy to misjudge materials with surface similarity as the same theoretical category. Based on this, this paper designs the identification of theoretical school differences as a joint discrimination process that integrates text semantics, pedigree pattern and historical context, and completes the genre representation, similarity calculation and category output in a unified representation space.

At the feature organization level, each genre candidate is represented as a prototype vector p_c, which gathers the stable features of the genre in terms of terminology system, harmony interpretation, termination mode, mode tendency and representative literature. For any input sample, the model generates its comprehensive representation vector h, and then measures how close the sample is to the prototypes of each genre through similarity calculation:

$$s_c = \frac{h \cdot p_c}{\|h\|\|p_c\|} \tag{9}$$

where, s_c represents the cosine similarity between the input sample and the theory school of class c. This calculation method can not only maintain the vector direction information, but also weaken the interference of sample length difference on the discrimination results, which is more suitable for dealing with the problems such as uneven length and short text and different scale of music examples common in music history materials.

Static similarity alone is still not enough to support robust classification, because some schools are highly similar in terms of surface level, but have obvious differences in historical stage and analysis intention. To this end, this paper further introduces the historical consistency correction term to incorporate the period label, regional background and knowledge source into the discriminant function:

$$\hat{c} = \arg\max_c (s_c + \lambda r_c) \tag{10}$$

Here, r_c is the consistency score between the input sample and the CTH genre in historical context, λ is the adjustment parameter, and $\hat{c}$ is the final discrimination result. In this way, the model no longer only makes decisions based on local term similarity, but takes into account whether similar expressions are in similar historical structures, thereby reducing the recognition bias caused by cross-temporal misjudgment and cross-genre drift.

4.3 Music theory evolution path identification and law extraction method

After the identification of historical features and the discrimination of theoretical schools, the research focus is further turned to the analysis of diachronic structure level, that is, how to connect the local features scattered in different periods, different documents and different music examples, and then identify the evolution path of music theory, and extract stable laws with explanatory value. Different from static classification tasks, evolutionary path recognition emphasizes "forward and backward association", "transfer direction" and "stage transition". Its goal is not only to determine which period or genre a material belongs to, but also to reveal how theoretical concepts inherit, offset, recombine and differentiate in the progress of history. Based on this requirement, this paper constructs a four-level method process of "time graph generation-path search-rule extraction-evidence backtracking", and its overall structure is

shown in Figure 3.

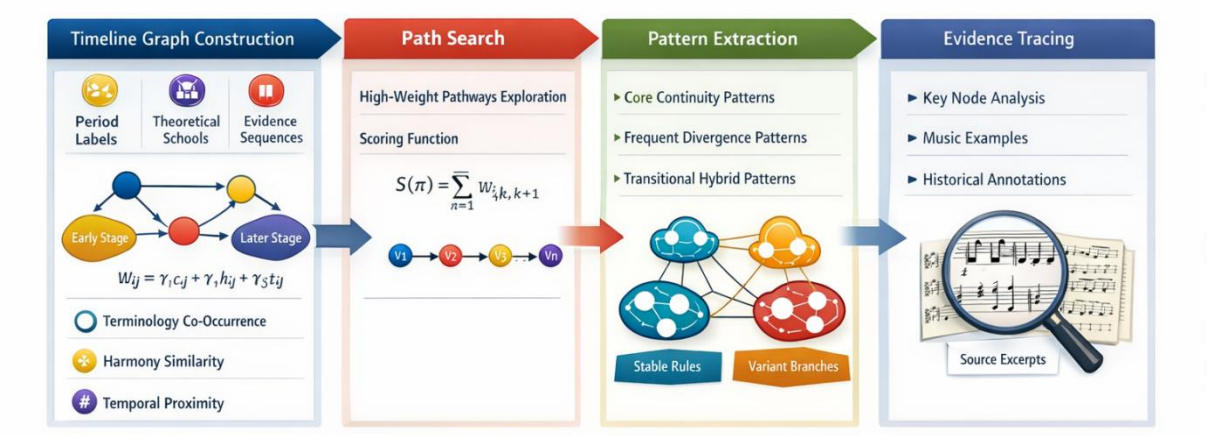

Figure 3: Method diagram of music theory evolution path identification and law extraction

In the graph structure construction stage, the system mapped period labels, genre features and theoretical evidence sequences into a time directed graph. In the graph, the nodes represent the theoretical states in a specific historical stage, and the edges represent the possible inheritance relationship or transformation relationship between adjacent states. Considering that different transitions do not have the same strength, this paper constructs the edge weight function according to the degree of term co-occurrence, the continuity of harmony pattern, the amplitude of mode center shift and the temporal proximity of documents. Let the transition weight between nodes v_i and v_j be w_{ij}, then its expression is:

$$w_{ij} = \gamma_1 c_{ij} + \gamma_2 h_{ij} + \gamma_3 t_{ij} \tag{11}$$

Here, c_{ij} represents the continuity between term and concept level, h_{ij} represents the similarity at the pedigree structure level, t_{ij} represents the temporal proximity score, and $\gamma_1, \gamma_2, \gamma_3$ are the normalized weight parameters. This expression makes the path in the graph not only be determined by the chronological order, but also reflect the continuity and mutation of the theoretical content itself.

In the candidate path search phase, starting from the initial node, the system traverses and screens the transition chain composed of high-weight edges to find the most representative evolution path. For a sequence of nodes $\pi = (v_1, v_2, \ldots, v_n)$, and its overall score is defined as:

$$S(\pi) = \sum_{k=1}^{n-1} w_{k,k+1} \tag{12}$$

Here, $S(\pi)$ represents the cumulative evolution strength of path π. The higher the score, the stronger consistency of the path in three aspects: concept continuation, structural similarity and time cohesion. In this way, the system can filter out a number of dominant evolution paths from a large number of local transitions, rather than staying in a scattered pairwise comparison.

In order to further extract the evolution rules from the candidate paths for music history research, this paper introduces a joint judgment mechanism of path frequency and stage stability. If a transition pattern appears repeatedly in multiple high-scoring paths and its node characteristics maintain high consistency in adjacent stages, it is summarized as a stable law. On the contrary, if a transition only appears in a few samples, or its edge weight is high but the

persistence is insufficient, it is regarded as a local variant or temporary branch. In this way, the final output of the system is not only the path sequence itself, but also includes "core continuation pattern", "high-frequency differentiation pattern" and "transitional mixed pattern" and other law categories.

5 Experimental implementation and research evaluation

5.1 Experimental environment and parameter setting

In order to verify the feasibility and stability of the intelligent recognition method of music theory evolution law proposed in this paper in real research scenarios, the experimental part constructs a unified implementation environment around four links of "data processing, model training, result output-path analysis". The overall system adopts a hierarchical deployment method, the bottom layer is the literature and pedigree data management module, the middle layer is the period identification, school discrimination and path extraction model, and the upper layer is the result visualization and evidence traceback interface. The back-end uses Python 3.11 to complete data preprocessing, model training and inference service encapsulation. The deep learning framework uses PyTorch 2.1, and the text vector processing relies on Transformers and SentencePiece. The graph structure operation uses PyTorch Geometric, the database part uses PostgreSQL to save the structural label and index information, and Neo4j stores the evolution path graph and node relationship. In order to ensure the reproducibility of the experimental process, the parameter configuration, training log, validation results and model weights are uniformly written into the version experimental record module.

In terms of hardware configuration, the experimental server is equipped with Intel Xeon Gold 6330 processor, 128 GB memory and NVIDIA RTX 4090 GPU with 24 GB memory capacity. This configuration can meet the computational requirements of multi-source feature joint training and batch path search, while avoiding significant bottlenecks in the long text encoding and graph structure traversal stages. Considering the characteristics of music history materials, such as large differences in text length, unbalanced density of music examples and skewed class distribution, the experiment adopts a staged training strategy. In the preprocessing stage, the theoretical text is merged by terms, sentence segmentation and time alignment, and the symbolic genealogy is encoded by event coding and harmony fragment extraction. Then the period label, genre label and evidence label are synchronously written into the sample index. In the model training stage, the training set, validation set and test set are divided into 8:1:1, and a low intra-batch repetition rate is set for the samples in the transition period to reduce the coverage effect of a single large class of samples.

In terms of parameter setting, the maximum length of text encoding is set to 512, the maximum length of pedigree event sequence is set to 256, and the dimension of fusion representation is unified to 384. AdamW was selected as the optimizer, the initial learning rate was set to 2×10^{-4}, the weight decay coefficient was 1×10^{-2}, the batch size was 32, the maximum training round was 40, and the early stopping mechanism was triggered when the verification metrics did not improve for five consecutive rounds. To suppress local overfitting in multi-task training, Dropout is set to 0.3 and gradient clipping threshold is set to 1.0. In the path identification stage, only candidate connections with edge weight greater than 0.35 are retained, and the maximum search depth of a single path is limited to 6 to balance the law extraction accuracy and graph search overhead. The output of the experimental results adopts the double writing mechanism of JSON and graph database, which is convenient for subsequent statistical analysis and visualization calls. Table 2 shows the main experimental environment and key parameter configuration of this study.

Table 2: Experimental environment and key parameter Settings

Category	Configuration Item	Parameter/Description
Operating Environment	Operating System	Ubuntu 22.04 LTS
Programming Environment	Language and Framework	Python 3.11, PyTorch 2.1
Text Processing	Related Tools	Transformers, SentencePiece
Graph Computing	Graph Learning Library	PyTorch Geometric
Database	Structured and Graph Data	PostgreSQL 15, Neo4j 5
Hardware	CPU / Memory / GPU	Xeon Gold 6330 / 128 GB / RTX 4090 24 GB
Data Split	Training:Validation:Test	8:1:1
Input Length	Text / Score Sequence	512 / 256
Fusion Layer	Feature Dimension	384
Optimizer	Training Strategy	AdamW
Learning Rate	Initial Value	2×10^{-4}
Batch Size	Batch Size	32
Training Epochs	Max Epoch	40
Regularization Control	Dropout / Gradient Clipping	0.3 / 1.0
Early Stopping Strategy	Patience	5
Path Search	Edge Weight Threshold / Maximum Depth	0.35 / 6

5.2 Performance test of music theory evolution recognition

In order to test the actual performance of the proposed method in the music theory evolution recognition task, this section conducts performance tests around period discrimination, genre discrimination and comprehensive recognition stability, and compares the constructed multi-source fusion model with several single-channel baseline models. The test objects include: the text encoding model using only theoretical text features, the pedigree encoding model using only symbolic pedigree features, the context discrimination model using only historical metadata, the text-instance dual-channel fusion model and the full feature collaborative recognition model proposed in this paper. Accuracy, Macro-F1 and Top-3 Accuracy are used to evaluate the overall classification ability, class balance performance and coverage ability of adjacent periods. The experimental data is still divided by 8:1:1 as described above, and the results are averaged by repeating 5 times in a unified training environment.

The test results show that the single feature model can complete the basic discrimination, but the stability is weak in the historical transition samples. Among them, the text coding model has a good effect on the recognition of term-intensive materials, with an overall accuracy of 82.6%, indicating that theoretical discourse is still an important basis for period judgment. The accuracy of the spectrum encoding model is 79.8%, which is stable in the samples with distinct mode center and termination mode, but it has insufficient discrimination ability when facing the theoretical discussion with strong text dependence. The metadata model only achieves an accuracy of 74.9%, which indicates that the age and author information can provide external constraints, but it is not enough to undertake the task of theory evolution identification independently. In contrast, the two-channel fusion model has shown strong advantages, with the accuracy improved to 86.7% and Macro-F1 reached 85.9%, indicating that there is an obvious complementary relationship between text and pedigree.

The collaborative recognition model with full features proposed in this paper achieves the

best results on three indicators, with Accuracy of 89.8%, Macro-F1 of 88.9%, and Top-3 Accuracy of 96.4%. This result indicates that when theoretical texts, symbolic examples and historical context are incorporated into the unified representation space, the model can more effectively deal with the problems of boundary blurring and intertemporal overlap common in the history of music theory. Especially between late Baroque and early classicism, and between late Romantic and modern transition stages, although the model still has a small amount of cross prediction, the first three candidate categories can usually cover the true labels, indicating that it has good recognition resilience to evolving continuous bands. The detailed results are shown in Table 3.

Table 3: Performance comparison of different models in the music theory evolution recognition task

Model Architecture	Accuracy / %	Macro-F1 / %	Top-3 Accuracy / %
Text Encoding Model	82.6	81.4	91.8
Score Encoding Model	79.8	78.7	89.9
Historical Metadata Model	74.9	73.5	85.6
Dual-Channel Fusion Model (Text + Score)	86.7	85.9	94.2
Proposed Model (Text + Score + Context Collaboration)	89.8	88.9	96.4

5.3 Comparative experiment and ablation analysis

In order to further explain the source of the performance improvement of the proposed model, this section carries out comparative experiments and ablation experiments in a unified data division and training environment. The contrast experiment focuses on the influence of different input channels on recognition results, and the ablation experiment investigates the actual contribution of historical context modeling and adaptive weight mechanism in the overall framework. In addition to Accuracy and Macro-F1, Path Consistency is added to measure the consistency between the evolution path recognition results and the manually labeled historical chains. This is handled in this way because this study does not only pursue classification accuracy, but pays more attention to the ability of the model to grasp the historical continuity of music theory.

As shown in Table 4, the single text model has better performance in the samples with dense theoretical terms and clear conceptual boundaries, with an Accuracy of 82.6% and Macro-F1 of 81.4%, indicating that literature expression is still an important basis for the discrimination of periods and genres. The Accuracy of the single spectrum example model is 79.8%, which is slightly lower than that of the text model, but it is stable on the samples with prominent mode center, termination habit and harmony structure, indicating that the spectrum evidence can effectively supplement the structural information outside the text. After the dual-channel fusion of text and pedigree, the Accuracy is improved to 86.7%, and the Path Consistency is 84.9%, which shows that after the collaboration of multi-source features, the model is significantly better than the single-channel scheme in identifying the theoretical evolution chain.

Table 4: Results of comparative experiments and ablation experiments

Model/Scheme	Accuracy / %	Macro-F1 / %	Path Consistency / %
Text Model	82.6	81.4	78.9

Score Model	79.8	78.7	75.6
Text + Score Fusion Model	86.7	85.9	84.9
Without Historical Context Branch	87.4	86.3	85.7
Without Adaptive Weighting Mechanism	88.1	87.0	86.4
Full Model	89.8	88.9	88.2

In the ablation experiment, after removing the historical context branch, the model Accuracy is reduced to 87.4%, and Macro-F1 is reduced to 86.3%. This change shows that age, author, version and knowledge dissemination background are not ancillary information, but important constraints affecting the boundaries of theoretical categories. After removing the adaptive weight mechanism, the Accuracy is 88.1%, which is still higher than the two-channel model, but there is a visible gap compared with the full model, indicating that the information density of different samples in terms, pedigrees and contexts is not consistent, and the fixed fusion method is difficult to fully describe the complex differences in historical materials. The complete model achieves the optimal results on the three indicators, showing better stability and stronger comprehensive discrimination ability. Figure 4 shows the change trend of each scheme in the two indicators of Accuracy and Macro-F1 more intuitively. It can be seen that the advantage of the complete model is not the local fluctuation, but the overall performance improvement.

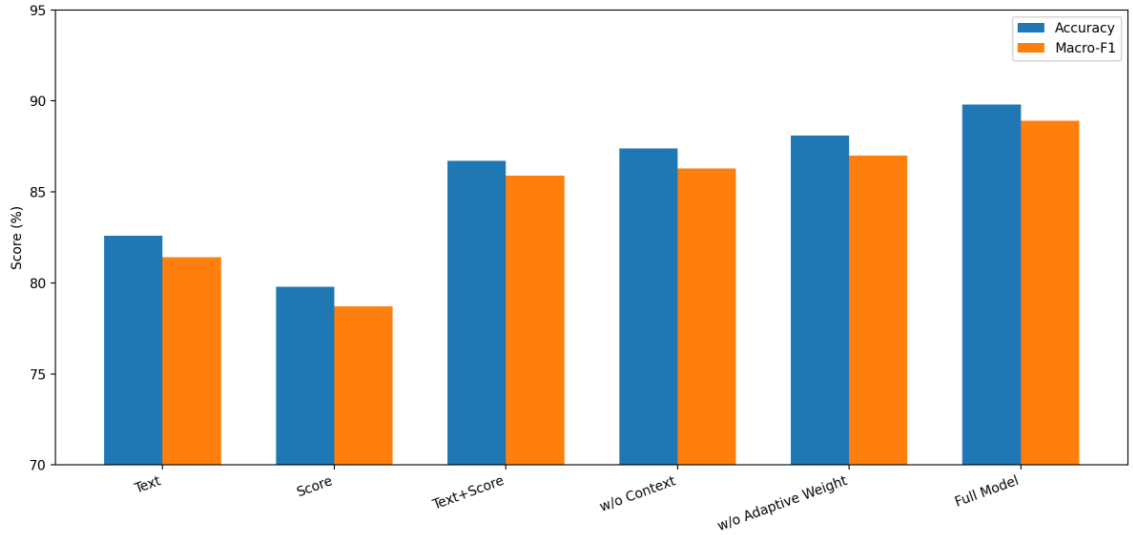

Figure 4: Comparison and ablation result plots for the music theory evolution recognition task

Taken together, the improved model performance is not brought by simple stacking of parameter scales, but comes from more effective collaborative modeling among text, genealogy, and historical context. This is particularly critical for the study of the evolution of music theory, because the real differences in historical materials often do not exist in a single dimension, but are scattered at the intersection of conceptual expressions, structural features, and historical contexts.

5.4 Music history Case verification and visual interpretation analysis

In order to verify the applicability of this model in the real music history research situation, this section selects three representative cases for comprehensive analysis, corresponding to the classical sample with gradually stable functional harmony system, the late Romantic sample

with coexistence of harmony expansion and mode migration, and the modern transformation sample with reorganization of theoretical terms. In the experimental setting, the support vector machine, random forest, single Transformer model and the multi-source fusion model proposed in this paper are placed on the same data set and the same preprocessing conditions, and are evaluated from three dimensions: recognition accuracy, average inference time of a single sample and explanation consistency. Interpretation consistency was scored on a 5-point scale by five researchers with a background in music analysis according to the degree of agreement between the evidence chain given by the model and the historical judgment, and the final average was taken.

The results show that although traditional machine learning methods can make basic judgments on small-scale samples with relatively single characteristics, they are prone to the problems of fuzzy category boundaries and scattered evidence directions when facing complex materials with historical transition stages and polysemy terms. The average recognition accuracy of support vector machine on the three groups of cases is 78.9%, and that of random forest is 81.7%. The average inference time of a single sample is 0.012 s and 0.019 s, respectively, and the interpretation consistency score is 2.9 and 3.3. The single Transformer model significantly improves the text understanding ability through context modeling, with an average accuracy of 86.8% and an interpretation consistency of 4.1, but it still has the problem of insufficient capture of the termination mode, mode center of gravity transfer, and voice part connection details in music examples. In contrast, the average accuracy of the fusion model proposed in this paper on the three cases reaches 90.6%, the average inference time of a single sample is 0.041 s, and the interpretation consistency score is increased to 4.6, indicating that it has more obvious advantages in balancing performance and interpretability.

From the specific case, in the classical sample, the model focused on the stable use of functional terms, standardized termination patterns and clear host-specific relationships, so it could be determined that it belongs to the stage where the functional harmony system tends to be mature with high confidence. In the late Romantic samples, the system gives higher weights to mode shift, extended chord and voice part delay structures, so as to avoid misattributing them to the early classical framework. In the modern transformation sample, the model combines contextual metadata and genre tags to identify the theoretical center of gravity corresponding to the redefinition of terms and the change of structure interpretation. This shows that the model does not rely on a single text label to make judgments, but forms a relatively stable evidence linkage among terms, genealogy and historical context.

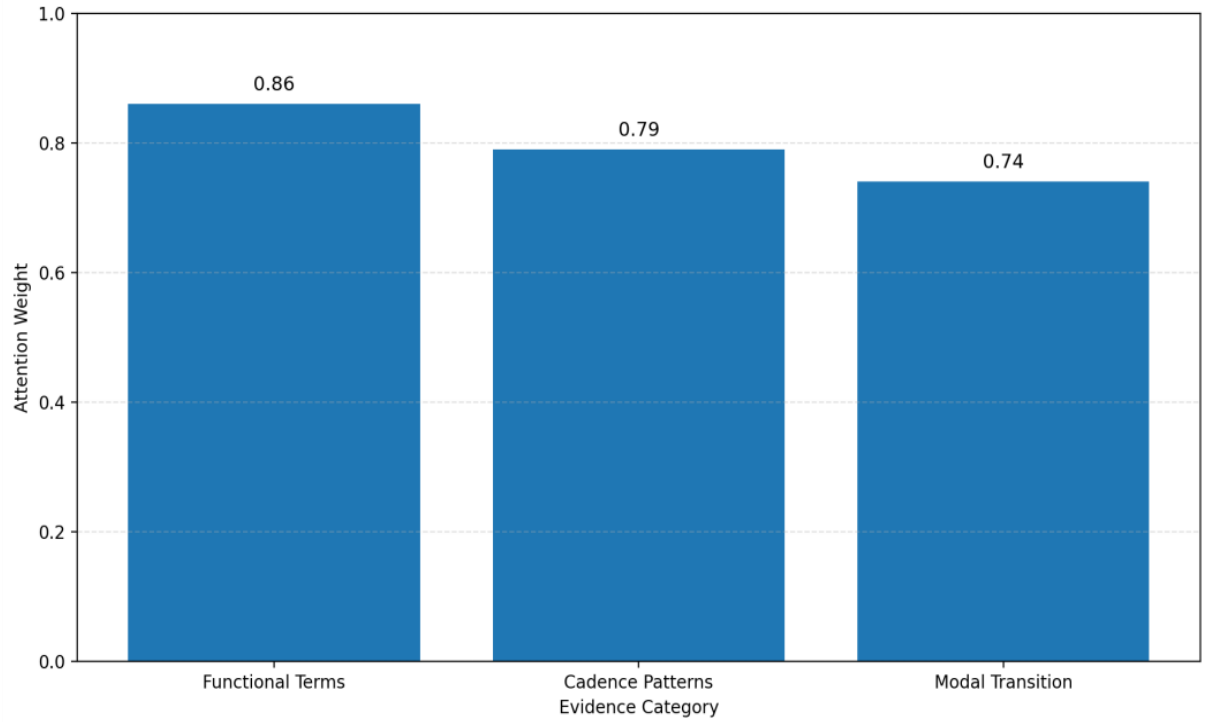

Figure 5: Visual interpretation of key evidence in the music history case

In order to further show the interpretation process of the model, this paper introduces an evidence visualization module based on attention weight, which maps the key judgment basis into the evidence contribution map. Figure 5 shows the weight distribution of evidence in a typical case. It can be seen that functional terms, termination modes, and modal transitions receive the highest attention, with attention weights of 0.86, 0.79, and 0.74, respectively, while metadata alone does not dominate. This result is basically consistent with the common sense judgment in music history research, and also indicates that the recognition process of the model has better academic readability. Overall, the case verification not only proves the applicability of the fusion model in complex music history materials, but also shows that the visual interpretation mechanism can provide more transparent judgment basis for researchers, thereby enhancing the credibility and use value of intelligent recognition results in the research of music theory history.

6 Conclusion and Prospect

Focusing on the problem of theoretical evolution recognition in music history research, this paper constructs a computational analysis framework consisting of data system, recognition model, path extraction and visual interpretation. In this study, theoretical texts, symbolic examples and historical metadata are incorporated into a unified representation space, and mode organization, harmony function, termination mode, term transfer and genre context are structurally encoded, so as to transform the historical analysis process that relies on manual comparison and empirical judgment into a computable, traceable and verifiable identification process. The experimental results show that the proposed model achieves 89.8% Accuracy, 88.9% Macro-F1 and 96.4% Top-3 Accuracy in the music theory evolution recognition task. In the case verification of music history, the average recognition accuracy reaches 90.6%, and the interpretation consistency score is 4.6, which shows that the method not only has good classification ability, but also can provide researchers with relatively clear evidence chain and

path basis. At the same time, there are still some deficiencies in this paper. First, the sample coverage is still dominated by the more representative stages and documents in the history of western music theory, and the absorption of marginal lineages, cross-regional theoretical exchange texts and small sample historical materials is still insufficient, which will affect the accuracy of the model in judging long-tail categories and transitional materials. Second, although the multi-source fusion model improves the recognition stability, it still has a strong dependence on computing power and annotation quality. Especially in the stage of complex path search and graph structure inference, the inference overhead is still higher than that of the lightweight method. Third, the current explanation mechanism mainly relies on attention weight and evidence backtracking, which can provide a certain degree of transparency, but the deep historical logic of "how the theory turns" still needs to be corrected by combining artificial musicological analysis.

The follow-up research can be carried out in three directions. First, expand the language, period and geographical coverage of music theory historical data, and establish a larger scale cross-version and cross-genre knowledge graph to improve the adaptability of the model to complex historical sections. Secondly, graph neural network, contrastive learning and small sample transfer strategies are introduced to enhance the recognition ability of the model for concept drift, boundary samples and implicit structural changes. Thirdly, the interactive analysis environment for researchers is further improved, and the period discrimination, genre comparison, path evolution and evidence visualization are integrated into a sustainable iterative digital music history tool. Therefore, the research on the evolution of music theory is expected to gradually move from static induction to dynamic interpretation supported by data, and also provide a more stable technical path for the deep integration of musicology and computational methods.

References

[1] Gotham M, Micchi G, López N N, et al. When in Rome: A meta-corpus of functional harmony[J]. Transactions of the International Society for Music Information Retrieval, 2023, 6(1).

[2] de Berardinis J, Meroño-Peñuela A, Poltronieri A, et al. Choco: a chord corpus and a data transformation workflow for musical harmony knowledge graphs[J]. Scientific Data, 2023, 10(1): 641.

[3] Kantarelis S, Dervakos E, Kotsani N, et al. Functional harmony ontology: Musical harmony analysis with Description Logics[J]. Journal of Web Semantics, 2023, 75: 100754.

[4] de Berardinis J, Meroño-Peñuela A, Poltronieri A, et al. The harmonic memory: a knowledge graph of harmonic patterns as a trustworthy framework for computational creativity[C]//Proceedings of the ACM Web Conference 2023. 2023: 3873-3882.

[5] Arthur C, Condit-Schultz N. The Coordinated Corpus of Popular Musics (CoCoPops): A Meta-Corpus of Melodic and Harmonic Transcriptions[C]//ISMIR. 2023: 239-246.

[6] Zhang H, Karystinaios E, Dixon S, et al. Symbolic music representations for classification tasks: A systematic evaluation[J]. arXiv preprint arXiv:2309.02567, 2023.

[7] Fradet N, Gutowski N, Chhel F, et al. Impact of time and note duration tokenizations on

deep learning symbolic music modeling[J]. arXiv preprint arXiv:2310.08497, 2023.

[8] Weiß C, Müller M. Studying Tonal Evolution of Western Choral Music: A Corpus-Based Strategy[C]//CHR. 2023: 687-702.

[9] Lustig E, Temperley D. The FAV Corpus: An Audio Dataset of Favorite Pieces and Excerpts, With Formal Analyses and Music Theory Descriptors[C]//ISMIR. 2023: 335-342.

[10] Hentschel J, Rammos Y, Moss F C, et al. An annotated corpus of tonal piano music from the long 19th century[J]. Empirical Musicology Review, 2024, 18(1).

[11] Moss F C, Lieck R, Rohrmeier M. Computational modeling of interval distributions in tonal space reveals paradigmatic stylistic changes in Western music history[J]. Humanities and Social Sciences Communications, 2024, 11(1): 1-11.

[12] Morales Tirado A, Carvalho J, Ratta M, et al. Musical Meetups Knowledge Graph (MMKG): a collection of evidence for historical social network analysis[C]//European Semantic Web Conference. Cham: Springer Nature Switzerland, 2024: 110-127.

[13] Zeitler J, Weiß C, Arifi-Müller V, et al. BPSD: A coherent multi-version dataset for analyzing the first movements of Beethoven's piano sonatas[J]. Transactions of the International Society for Music Information Retrieval, 2024, 7(1).

[14] Qin Y, Xie H, Ding S, et al. Score Images as a Modality: Enhancing Symbolic Music Understanding through Large-Scale Multimodal Pre-Training[J]. Sensors, 2024, 24(15): 5017.

[15] Ji S, Yang X, Luo J. A survey on deep learning for symbolic music generation: Representations, algorithms, evaluations, and challenges[J]. ACM Computing Surveys, 2023, 56(1): 1-39.

[16] Yuan R, Lin H, Wang Y, et al. Chatmusician: Understanding and generating music intrinsically with llm[C]//Findings of the Association for Computational Linguistics: ACL 2024. 2024: 6252-6271.

[17] Tian J, Li Z, Li J, et al. N-gram unsupervised compoundation and feature injection for better symbolic music understanding[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(14): 15364-15372.

[18] Kosta K, Lu W T, Medeot G, et al. A deep learning method for melody extraction from a polyphonic symbolic music representation[C]//ISMIR. 2022: 757-763.

[19] Karystinaios E, Widmer G. Cadence detection in symbolic classical music using graph neural networks[J]. arXiv preprint arXiv:2208.14819, 2022.

[20] Couturier L, Bigo L, Levé F. A dataset of symbolic texture annotations in mozart piano sonatas[C]//International Society for Music Information Retrieval Conference (ISMIR 2022). 2022.