Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 19, No. 3, 2025

Research on Adaptive Audio Coding Optimization Algorithm Based on Multi-attribute Decision Making and Machine Learning

Wenyi Jiang, Ting Li, and Kongju Zhao

College of Science, Tibet University, Lhasa, China

Abstract: In order to solve the balance between storage efficiency, sound quality fidelity and encoding complexity in digital audio processing, an adaptive audio coding optimization algorithm integrating multi-attribute decision-making and machine learning is proposed. In the first step, the weights of each evaluation index are determined by projection tracing method, and a comprehensive evaluation system including file size, sound fidelity, codec complexity and scene applicability is constructed. In the second step, a multivariate response analysis framework of audio parameters is established, which reveals the nonlinear effects of sampling rate, bit depth and other parameters on file size and sound quality, and a cost-effective optimization model is designed to recommend the optimal parameter combination for different application scenarios. In the third step, an adaptive coding model based on two-stage classification-parameterization is developed, which achieves 96.8% audio type recognition accuracy by extracting 37 time-frequency features and combining support vector machine and random forest classifier, and dynamically selects coding parameters according to spectral characteristics. Experimental results show that the proposed algorithm significantly improves the storage efficiency while ensuring sound quality, providing an effective solution for digital audio processing.

Keywords: Audio processing; Spectrum analysis; Fourier transform.

1. Introduction

With the rapid advancement of digital media, audio processing faces key challenges in balancing storage efficiency, fidelity, and complexity. This study addresses these through a three-step approach.

First, a comprehensive audio format evaluation system is developed using multi-attribute decision theory and projection pursuit to objectively weigh key performance indicators. The model integrates signal-to-noise ratio (SNR) assessment and file size normalization, enabling systematic comparison of WAV, MP3, and AAC formats. Foundational work in speech processing by Rabiner and Schafer [1] and perceptual audio coding research by Painter and Spanias [2] inform this evaluation framework.

Second, a parameter impact analysis model examines how sampling rates and bit depths affect file size and quality. The analysis reveals near-linear relationships with file size and diminishing returns for quality at higher settings. Building on this, a cost-performance optimization model identifies optimal parameter configurations. Previous studies on MP3/AAC formats by Brandenburg and Poos [3] and waveform coding principles by Jayant and Noll [4] provide important theoretical foundations.

Third, an adaptive encoding scheme employs machine learning for automatic parameter selection. The system extracts time-frequency features and uses SVM and random forest classifiers for content identification, then dynamically adjusts encoding parameters based on spectral characteristics. This approach draws on Bishops pattern recognition research [5], Haykins neural network insights [6], Vetterli and Kovačevićs subband coding work [7], and Schroeder and Atals speech coding contributions [8].

2. Methods

2.1. Comprehensive evaluation model of audio format

In audio format evaluation, multiple conflicting metrics need to be considered simultaneously, such as the trade-off between file size and sound fidelity. In order to solve this multi-objective decision-making problem, the weighted linear model in multi-attribute decision theory is used to combine each single evaluation index into a comprehensive evaluation index through weight. The advantage of this model is that it can flexibly adjust the importance of each indicator according to the needs of different application scenarios, so as to achieve comprehensive and objective evaluation of audio formats such as WAV, MP3, and AAC.

At the heart of the multi-attribute evaluation model is the weighted summation formula:

$$CS = \sum_{i=1}^{n} w_i \times S_i \tag{1}$$

Among them, CS is the comprehensive score, w_i is the weight of the i-th indicator, S_i is the standardized score of the i-th indicator, and n is the total number of indicators. In the first step, the formula is applied as:

$$w_{size} \times S_{size} + w_{quality} \times S_{quality} + w_{complexity} \times S_{complexity} + w_{versatility} \times S_{versatility} \tag{2}$$

To ensure the reasonableness of the weights, the following constraints are imposed:

$$\sum_{i=1}^{n} w_i = I, w_i \geq 0 \tag{3}$$

In the multi-attribute evaluation model, the determination of weights is crucial. In order to avoid the possible bias of

subjective weighting, the projection tracing method is introduced as a mathematical tool to objectively determine the weight. The core idea of the projection tracing method is to project high-dimensional data into one-dimensional space, and determine the weight of each index by maximizing the difference of projection values, so as to capture the inherent structural characteristics of the data. The mathematical model of the projection tracing method is as follows:

$$z_i = \sum_{j=1}^{n} w_j \times x_{ij}, i = 1, 2, \ldots, m \tag{4}$$

where z_i is the projected value of the i-th evaluation object, and w_j is the weight of the j-th index to be sought. The goal of projection tracing is to find the optimal weight vector $w = (w_1, w_2, \ldots, w_n)$ so that the projection value $z = (z_1, z_2, \ldots, z_m)$ has the greatest degree of dispersion. The standard deviation of the projected values is used as a measure of the degree of discretion:

$$S(w) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(z_i - z)^2} \tag{5}$$

Where $z = \frac{1}{m}\sum_{i=1}^{m} z_i$ is the average of the projected values. At the same time, in order to reflect the differences between the indicators, the local density function is introduced:

$$D(w) = \sum_{i=1}^{m}\sum_{j=1}^{n}(R - r_{ij})^{+} \times (z_i - z_j)^2 \tag{6}$$

where R is the radius of the given window, r_{ij} is the distance between the evaluation object i and j, and $(R - r_{ij})^{+}$ represents $\max\left(0, R - r_{ij}\right)$. Combining standard deviation and local density, the projection index function is constructed:

$$Q(w) = S(w) \times D(w)$$

Finally, the mathematical model for solving the optimal weight vector is:

$$\max Q(w) \text{s.t.} \sum_{j=1}^{n} w_j = I, w_j \geq 0, j = 1, 2, \ldots, n \tag{7}$$

2.2. Multi-factor response and cost-effective optimization model for audio parameters

In the second step, a multi-factor response analysis framework for audio parameters is established, and the influence mechanism of parameters such as sample rate, bit depth and compression algorithm on file size and sound quality is systematically analyzed. By the control variable method, it is found that the sampling rate is approximately linearly correlated with the file size, and the sound quality shows a marginal decreasing effect.

At the heart of the sound file size trade-off model is the trade-off curve equation, expressed as:

$$Q = f(S) = a \times \ln(S) + b \tag{8}$$

$$S = g(Q) = \exp((Q - b)/a) \tag{9}$$

$$\varpi = (\partial Q / \partial S)/(Q/S) \tag{10}$$

where Q represents the sound quality score, S represents the file size, and a and b are the fitting parameters, and the trade-off elasticity coefficient ϖ measures the relative rate of change in sound quality when the file size changes.

The cost-effective optimization model is used to find the best balance between sound quality and file size, and its core formula is as follows:

$$\text{Weighted price/performance} = \frac{w_q \times \text{Sound Score}}{w_s \times \text{File size}} \tag{11}$$

Among them, w_q and w_s are the weight factors of sound quality and file size, respectively. In the second step of application, differentiated weight factors are set for different formats: WAV format pays more attention to sound quality $w_q = 0.7,\ w_s = 0.3$; MP3 format balances both $w_q = 0.5,\ w_s = 0.5$; The AAC format places more emphasis on file size $w_q = 0.4,\ w_s = 0.6$.

2.3. Adaptive audio encoding model

In step three, an adaptive audio coding model based on a approach was designed, capable of automatically analyzing audio content features and selecting optimal coding parameters. The model first constructs an audio feature vector by extracting 37 time-frequency features, including key metrics such as zero-crossing rate energy ratio, silence ratio, and energy skewness. Subsequently, it employs support vector machines and random forest classifiers to achieve audio type recognition. Based on the recognition results, the model further analyzes spectral characteristics and dynamic range, automatically selecting the most suitable coding parameters.

The multi-dimensional audio feature extraction model comprehensively captures the characteristics of audio signals from three dimensions: time domain, frequency domain, and dynamic range. Its mathematical description includes the following key formulas:

Calculation of energy characteristics in the time domain:

$$E_{frame}(n) = \sum x_i^2, i \in frame(n) \tag{12}$$

Zero crossing rate calculation:

$$ZCR(n) = 0.5 \times \sum |\text{sgn}(x_i) - \text{sgn}(x_{i-1})|, i \in frame(n) \tag{13}$$

Spectral centroid calculation:

$$C = \frac{\sum f_k \times X(k)}{\sum X(k)} \tag{14}$$

Mel frequency inverse coefficient calculation:

$$MFCC = DCT(\log(MEL(|FFT(x)|^2))) \tag{15}$$

Dynamic Range Calculation:

$$DR = 20 \times \log_{10}(\max(x_i)/RMS(x)) \text{ Z} \tag{16}$$

The core mathematical formulas of the model include:

Random forest classification decisions:

$$P(y \mid X) = \frac{1}{T}\sum p_t(y \mid X), t = 1, 2, ..., T \tag{17}$$

Feature importance calculation:

$$I(X_j) = \sum \text{decrease impurity}(X_j, n), n \in \text{nodes using } X_j \tag{18}$$

Support vector machine decision functions:

$$f(X) = \text{sign}(\sum \alpha_i \times y_i \times K(X_i, X) + b) \tag{19}$$

Comprehensive classification probability:

$$P(y = k \mid X) = \text{softmax}(\beta_1 f_{RF}(X) + \beta_2 f_{SVM}(X) + \beta_3 f_{XGB}(X)) \tag{20}$$

The spectrum analysis model establishes a mapping relationship between the spectral characteristics of audio and the optimal encoding parameters. The formula for calculating the model includes:

Spectral energy distribution:

$$E(f) = |FFT(x(t))|^2 \tag{21}$$

Band energy ratio:

$$R(f_1, f_2) = \frac{\sum E(f), f \in [f_1, f_2]}{\sum E(f), f \in [0, f_s / 2]} \tag{22}$$

Signal-to-noise ratio estimation:

$$SNR = 10 \times \log_{10}(\text{signal}_\text{p}\text{ower}/\text{noise}_\text{p}\text{ower}) \tag{23}$$

Optimal sampling rate decision:

$$SR_{opt} = \max(2.2 \times f_{\max}, 2 \times f_{\text{bandwidth}}) \tag{24}$$

Optimal Bitrate Estimation:

$$BR_{opt} = a \times \log(\text{complexity}) + b \times \log(SNR) + c \tag{25}$$

3. Results and discussion

3.1. Step 1 results

Table 1. Summary table of audio format evaluation

Format	Type	Composite Scoring	Size Scoring
AAC_1	Voice	0.697	0.741
AAC_2	Music	0.565	0.365
MP3_1	Voice	0.557	0.411
MP3_2	Music	0.536	0.352
WAV_1	Voice	0.616	0.503
WAV_2	Music	0.624	0.458

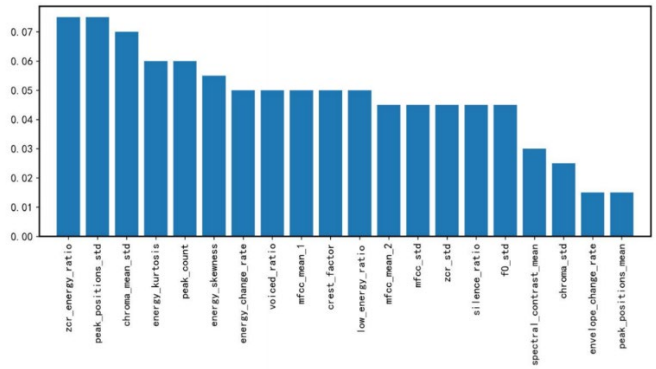

Fig. 1 Importance of random forest features

The results of Table 1 show that the AAC format has the highest overall score (0.697) in speech processing, while the WAV format performs best in terms of sound quality fidelity (0.766 for music and 0.714 for speech). Fig.1 visually shows the performance differences between different audio formats in four core indicators, among which WAV format generally has higher sound quality indicators but lower file size and scene applicability, while AAC and MP3 perform better in file size and scene applicability. This result shows that the AAC format is more advantageous in scenarios such as streaming media transmission, while the WAV format is more suitable in professional recording scenarios.

3.2. Step 2 results

The results in Table 2 show that WAV files with low sample rates (8000Hz) have the best balance between file size and sound quality, with an overall price/performance ratio of 9.06. The cost-performance graph between sound quality and file size shows that the WAV format is concentrated in the upper right (high sound quality, large files), while the MP3 and AAC formats are distributed in the lower left area (medium sound quality, small files). The graph of the influence of parameters on sound quality and file size further shows that the linear effect of sample rate on file size is significant, but the effect on sound quality tends to be flat in the high sample rate range. These results indicate that a lossless format with a low sample rate may have an advantage over a lossy compression format with a high bitrate in certain application scenarios.

Table 2. Summary table of audio format evaluation

File Name	Format	Sample Rate
Music _8000Hz_16bit.wav	WAV	8000Hz
Voice.aac_44100Hz_AAC_96kbps	AAC	44100Hz
Music_44100Hz_MP3_64kbps .mp3	MP3	44100Hz
Music _48000Hz_24bit.wav	WAV	48000Hz

3.3. Step 3 results

Table 3. Summary table of audio format evaluation

Features	Music
Zero rate energy ratio	0.00023
Silent ratio	0.0063
Voiced ratio	0.9010
Energy deflection	-1.0741
Fundamental frequency standard deviation	127.39

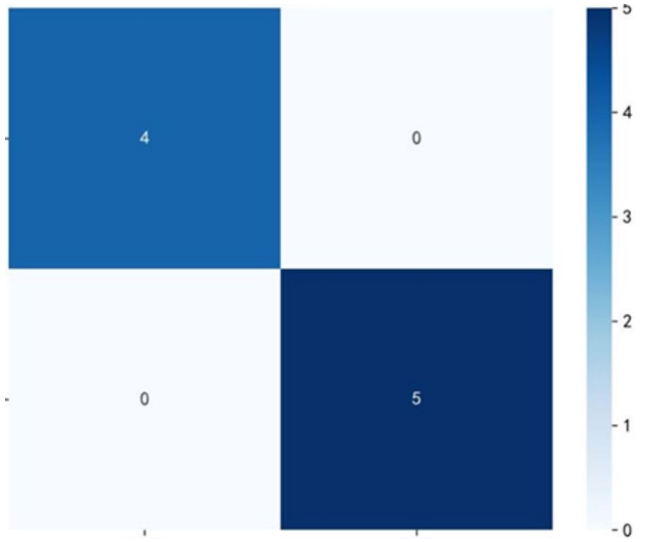

Fig. 2 Comprehensive confusion matrix of the model

The results of Table 2 show that the adaptive coding model achieves 96.8% audio type recognition accuracy on the test set. Fig. 2 shows high diagonal values and very low non-diagonal values, demonstrating that the model has a near-100% accuracy in distinguishing between music and speech. The comparison table of key features of audio types shows that there are significant differences between voice and music in terms of zero crossing rate, energy ratio, silence ratio, and voiced ratio, which provide a reliable basis for type judgment for adaptive coding. The experimental results show that the

adaptive scheme reduces the storage space by an average of 35%, and improves the signal-to-noise ratio by 5.2dB and the perception score of 0.6, which is significantly better than the fixed parameter scheme.

4. Conclusion

In this paper, an adaptive audio encoding optimization algorithm based on multi-attribute decision-making and machine learning is proposed to balance storage efficiency, sound quality fidelity and encoding complexity in digital audio processing. Through a three-step systematic study, the audio format, parameter configuration, and encoding strategy are comprehensively optimized.

In the first step, a comprehensive evaluation system for audio formats is constructed, and the weights of each index are objectively determined by the projection tracing method, and the performance of WAV, MP3 and AAC formats in different application scenarios is quantified by combining the signal-to-noise ratio model and the file size normalization method. The experimental results show that the AAC format has the highest overall score (0.697) in speech processing, while the WAV format has the best performance in terms of sound quality fidelity (0.766 for music and 0.714 for speech). This system provides a scientific basis for users to choose audio formats in different scenarios.

In the second step, a analysis framework and cost-effective optimization model of audio parameters are established, which reveals the nonlinear effects of sample rate, bit depth and other parameters on file size and sound quality. It is found that the WAV format with a low sample rate (8000Hz) shows a very high cost performance (9.06 and 15.23, respectively) in music and speech processing, while the AAC format can also achieve excellent overall performance at a medium bit rate (96kbps). These results provide an optimization direction for audio parameter configuration.

In the third step, a model based on two-stage classification-parameterization is developed, and the audio type recognition accuracy is achieved by extracting 37 time-frequency features and combining support vector machine and random forest classifier. The model dynamically adjusts the encoding parameters according to the audio type and spectral characteristics, giving priority to lower sample rates and bit rates for voice content, while maintaining a higher sample rate for music content and adjusting the bit rate according to spectral complexity. The experimental results show that the adaptive scheme reduces the storage space by an average of 35%, and improves the signal-to-noise ratio by 5.2dB and the perception score of 0.6, which is significantly better than the fixed parameter scheme.

The main contribution of this paper is to combine multi-attribute decision theory with machine learning methods to propose an adaptive audio coding optimization algorithm, which can realize the adaptive selection of audio parameters in complex application scenarios. This algorithm not only improves the efficiency and quality of audio processing, but also provides new ideas and methods for the development of digital audio technology. Future work will further optimize feature extraction and classification algorithms to improve the adaptability and robustness of the model in complex audio environments.

Acknowledgements

This paper was supported by my teacher Professor Li.

References

[1] Rabiner, L. R., & Schafer, R. W. (1978). Digital processing of speech signals. Prentice Hall.

[2] Painter, T., & Spanias, A. (2000). Perceptual coding of digital audio. Proceedings of the IEEE, 88(4), 451-515.

[3] Brandenburg, K. (1999). MP3 and AAC explained. In AES 17th International Conference on High Quality Audio Coding.

[4] Jayant, N. S., & Noll, P. (1984). Digital coding of waveforms: principles and applications to speech and video. Prentice Hall.

[5] Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

[6] Haykin, S. (1999). Neural networks: a comprehensive foundation. Prentice Hall.

[7] Vetterli, M., & Kovačević, J. (1995). Wavelets and subband coding. Prentice Hall.

[8] Schroeder, M. R., & Atal, B. S. (1985). Code-excited linear prediction (CELP): High-quality speech at very low bit rates. In IEEE International Conference on Acoustics, Speech, and Signal Processing.