Artificial Neural Network Application on Determining Chord Composition for Melody Accompaniment

Mochammad Dikra Prasetya
Institut Teknologi Bandung
dikraprasetya@yahoo.co.id

Dr. Ir. Rinaldi Munir, M.T.
Institut Teknologi Bandung
rinaldi@informatika.org

Abstract— **Melody is a sequence of succession of tones and itself is the major part of a song composition. To accompany the melody, chord compositions will be prepared in accordance with the harmonization of tones within it. Composing chords is an unquantifiable process which may only be rated by subjective judgement. Variations are welcomed, however subjectively each person have their own preference thus making a generally likeable composition is a challenge to musicians since long. Simon et al. (2008) proposed a solution on composing chord accompaniment for a melody in real-time as an application named *MySong*. Machine learning serves a main part on the application, therefore it suggests that other machine learning variations may possibly applicable to the problem and possibly produce better result. Artificial neural network is considered as a potential alternative that may have an advantage on parameter customization and time-series applicable. Based on test results, it is proven that an artificial neural network solution is able to produce a generally both hearable and likeable chord composition.**

Key words—**Artificial Neural Network, Machine Learning, Melody, Chord Composition, Song Accompaniment.**

I. INTRODUCTORY

Mainly the experiment is conducted to examine artificial neural network's capability on solving the problem of chord composition for melody accompaniment with general acceptance of audience. Artificial neural network is configured in accordance with music theory in the process. Result's quality is assessed through cross-validation testing and survey from sample audiences. Some other methods such as greedy alternatives and manual by experts are present as the comparative method in the evaluation.

As a complex problem, there are several limitations brought in the conducted experiment. Input of the prototype is a music melody transcribed in MusicXML format and its chord composition is generated in MIDI format. Unfortunately, there is no source of downloadable melody-only MusicXML. Hence all of the song bank is transcribed manually through this experiment. Songs used in both training and testing are originated from national mars and hymnes due to the limitation of melody-chord source. There is no transposes of root scale within the songs, and used chords are limited to simple chord forms only (major, minor, diminished). The examination is also limited only to major-themed song, since processing major and minor should be parallel (according to literature study)

and examining only in one theme is applicable to the other. This means that by examining only major themed song, the proposed solution is also applicable for minor themed song by a parallel similar model. In spite of these limitations, the proposed solution are configured as general as possible to let the possibility of further development in the future. All of the codes of prototype can be accessed through author's Git repository: https://github.com/dkrprasetya/Chordman.

II. MUSIC THEORY AND RELATED RESEARCHES

Literature study is conducted on music theory and researches related to chord composing. [10-11] Simon et al. (2008) have contributed on solving this problem with their product, *MySong*. They are using *Hidden Markov Model* on their implementation. Other researches are not in same objective as current experiment, but lies in the same domain and adaptable to this chord composing problem. [12] Yaremchuk et al. (2008) proposed a solution on guessing played chords in an MP3 using artificial neural network, while [9] Mauch et al. (2009) proposed that segmentation in a song beforehand could result significant improvement on chord guessing. From above researches it is concluded few things to be adapted in the implementation of artificial neural network.

As [9] Mauch et al. (2009) suggests, it is preferably to break down the whole melody beforehand. However since the parts identification (prelude, riff, elude, etc.) are not given, the segmentation should be done in different relevant term. [10-11] Simon, et al. (2008) processed the song in a function of time, which is adaptable to this segmentation process. Measure beat is a preferable parameter for segmenting the melody. These segments will be translated into input parameters of the artificial neural network. Chords resulted from each segment are parts of the whole composition that will be reconstructed in the end.

[10-11] Simon, et al. (2008) also suggests on many adaptable processes that is necessary to be done in order to improve the quality of processing. Melody as sequence of tones should be transposed into uniform root scale before being used as a training data. This is done to increase the accuracy of predicting, since similar root scale means more relevance in the similarity matching of determining probability of playing a chord in a segment. Major and minor songs should also be parallel in its process. However, this observation is not relevant anymore since it has been declared that only major-themed song will be used in the experiment.

Observation parameter is an important feature of a predictor. Tones from each segment are obviously become a deciding parameter on chord composing. Its occurrence and density are different in terms since one of it a true-false variable while the other is calculated by its frequency per total of tones in the segment. There are a total of 12 tones exist in music (C, C#, D, D#, E, F, F#, G, G#, A, A#, B), each reserves two (occurrence and density) input node in the designed network.

Chord composition is a harmonious sequence, therefore it is natural that observing the previous chord is considered as necessary. This hypothesis is supported by chord progression study in music theory which stated that each chords have its tendency to move only a few chord candidates as the next played chord. In the implementation, this could be translated as the time-series input where we will observe previous results for our current observation. However it is not known the best range of time-series implementation in the network. This limitation suggests the experimentation should include examination of time-series range.

III. CHORD COMPOSING PROCESS DESIGN

A. Artificial Neural Network Design

Designed artificial neural network should consider all of the suggested features as part of the observation. Melody as the input is processed beforehand to produce segments. These segments are mapped into values for input nodes in the designed network. From the analysis, it is found that the observation candidates including: (1) Occurrence of each tones; (2) Density of each tones; and (3) Previous segments' result chords, which range is going to be examined through the experiment. With this configuration, feed-forward multilayer perceptron architecture is used. Backpropagation algorithm and sigmoid activation function is applied to the network with a fixed iteration. Training until some number of error is not considered as it is not applicable for such non-deterministic (not necessarily exact match) problem. However, the number of fixed iteration is not defined hence it will be examined also through the experiment.
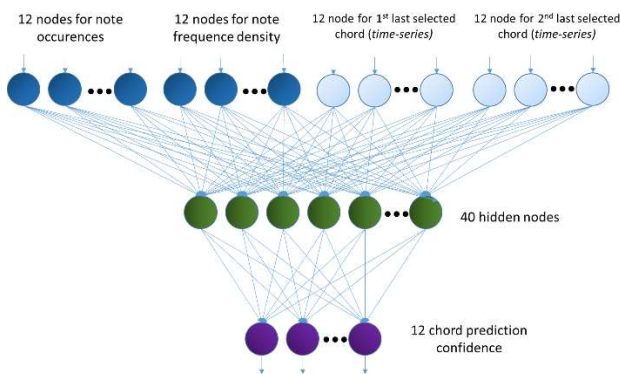

Figure 1. Artificial Neural Network Model

Observed parameters are mapped into 24 main input nodes (12 occurrence nodes and 12 density nodes) and 12 output nodes. The time-series parameter are adding a number of 12 x N input nodes, where N is the range of the time-series. One hidden-layer is present with its number of nodes ¾ of the sum of input and output nodes. This number is taken from empiric study references on artificial neural network. Designed model as a whole is illustrated on Figure 1.

Each of 12 output nodes represent the confidence level of that tone being the suggested played chord on current segment. Result of the segment is decided by taking each of these values as the probability its chord is taken. This algorithm is described as in written Formula 1, where confidence values of each chord tone is notated as v_i.

$$p(nada_i) = \frac{v_i}{\sum v}$$

(1)

B. Process Scheme

The whole process consist of three processes: Pre-process, training process, and testing process.

Pre-process takes role on preparing inputted melody into class models so that further processing could be conducted with more flexibility on tweaking its variables. This includes transposing tones into C major as the uniform root scale and segmenting the melody into parts and mapping it according to the network's design which the whole scheme illustrated on Figure 2. As it is been discussed on the analysis, segmentation is done per half measure. Therefore, if the song used 4 beats per measure, it will be segmented into 2 segments (per 2 beats).
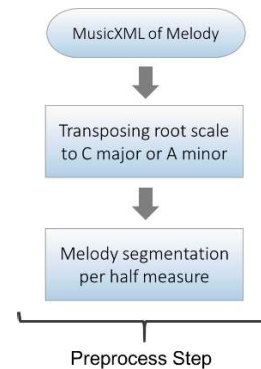

Figure 2. Segmentation Result

Song bank are divided into training data set and testing data set by ratio of 10:3 respectively. Data sets are pre-processed to output the segments that will be used in further process. In training process, implemented artificial neural network will be trained to each provided data set. Weights resulted after training is stored in a file to be used in testing process. Testing process consists of processing each pre-process' output segments to generate chords for its position, then the final chord composition is constructed from those. The whole main process scheme are illustrated in Figure 3.
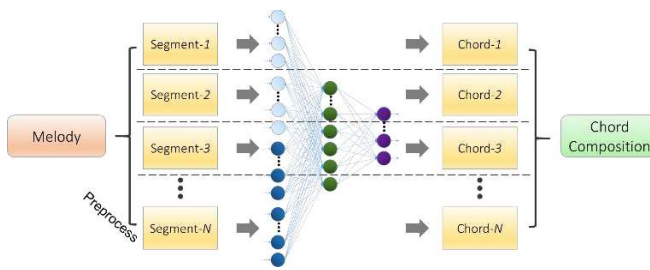
Figure 3. Process Scheme

$$Equivalence\ Score = \frac{\sum f(a,b)}{N}$$

$$f(a,b) = \begin{cases} 1, & if\ a = b \\ 0.75, & if\ a\ is\ relative\ \frac{major}{minor}\ of\ b \\ \sum \frac{g(x_i, b)}{3}, & otherwise \end{cases}$$

$$g(x_i, b) = \begin{cases} 1, & if\ tone\ x_i\ occured\ in\ chord\ b \\ 0, & otherwise \end{cases} \tag{2}$$

IV. IMPLEMENTATION AND EVALUATION

It has been succeeded to implement the previous design of artificial neural network that is successfully able to produce chord composition for melody accompaniment. From the song bank, it is produced 600 of data sets for training and testing. Each process are executed on each data sets and experiment configuration. In Figure 4, the resulted composition from the implemented artificial neural network (as chord A) and manually generated chords by expert (as chord B) are presented side by side. Both composition have different assigned chords on some parts, however, both are well-fit for its segment and generally acceptable by audience, which evaluation described in further section.

Figure 4. Composed Chords Sample

The first evaluation is conducted by cross-validation accuracy test on the result. It should be noted that the chord matching equivalence condition are modified to fit the relevance matching of two different chords, since they are not necessarily to be exact match to be stated as a well or badly predicted chord. The equivalence score follows Formula 2. If both have the exact same chord, it will score 1, or exactly equivalent. The second check is whether both chord is the relative major/minor of the other, which if true will score 0.75. Otherwise, these chords will be matched by each tones, as if it appeared in one and another it will score 1/3 for each tone. Number of segments present is notated by N. Both chords which currently being matched is notated as a and b in the formula. Notation $f(a, b)$ represents the whole equivalence function. Notation $g(x_i, b)$ tests whether tone x_i, which occurred in chord a, also occurred in chord b. Result of this accuracy test is shown on following Table 1.

Throughout this accuracy evaluation, all of the configuration candidates and alternative methods are experimented. Configuration with best accuracy concluded from this evaluation will be used on further testing. The experimented parameters are the range of time-series and the number of fixed iteration for training process. Alternative methods that are present to be compared with artificial neural network solution are: (1) Random function method; (2) Greedy by tone occurrence; and (3) Greedy by applying chord progression theory. Random function method is done by picking one of exist tones (12 candidates) randomly with a uniform probability of 1/12. Greedy by tone occurrence done by assigning the most frequent occurring tone's chord. The other greedy method is by applying chord progression theory, which the method will pick randomly one of the candidates decided by our current chord and its next tendentious chords. These alternative methods are compared with the best configuration of artificial neural network. The accuracy of each methods could be compared through the following Table 2.

Table 1. Cross-Validation Test on Experiments

Num of Iteration	No Time-series	1 Segment Time-series	2 Segment Time-series	Average
10	61.520 %	53.510 %	35.182 %	50.070 %
100	74.002 %	72.534 %	70.122 %	72.219 %
1000	74.755 %	68.420 %	68.727 %	70.634 %
10000	73.232 %	73.226 %	70.525 %	72.327 %
100000	74.815 %	70.593 %	66.925 %	70.777 %

Table 2. Accuracy of Methods

Method	Accuracy
No time-series	74.815 %
1 Segment time-series	73.226 %
2 Segment time-series	70.525 %
Random function	31.247 %
Greedy by tone occurrence	37.897 %
Greedy by chord progression	65.859 %

Based on above score, the best accuracy achieved by artificial neural network is 74.815%, with no time-series configuration and 100000 iteration on its training. It could be seen that after 100 iteration, the network shown a stability of accuracy movement with ± 2% of fluctuation. Even so, it is concluded that the combination of 100000 iteration and no time-series is the best configuration out of this experiment. Compared with other alternative methods, its best competitor is greedy by chord progression, with the

accuracy of 65.859%. Although at first the experiment's hypothesis was "observing previous played chords is naturally necessary to build harmonious chord composition", here it is found that no time series configuration produced the best accuracy. It is a bit contradictive since it is well-known that chord progression theory defined that the current played chord limits the candidates of chords to play next. This could come into two alternative conclusion: 1) Time-series parameter is considered as an excessive observation feature; 2) Limitation of data set affects the pattern learning that should have been strengthening the prediction. These two analysis is not yet concluded and need further experiment to decide. Despite the uncertainty, here it is solidly concluded that the implemented artificial neural network could predict quite satisfyingly accurate.

The second and last evaluation is sample audience survey. It is conducted to gain audience subjective judgement on composed melody accompaniment. Quality of the composition is scored by averaging ratings from audience. The selected three sample melody are accompanied by artificial neural network, then given to the audience to be rated. To filter audience responses, they are categorized into three category: Expert, semi-expert, and non-expert. Expert category respondents are audience that are musician and capable to determine chords to accompany a melody. Semi-experts are ones that are musician but not well-experienced in determining chords. Other than the recent category will join non-expert category respondents. Responses from expert category are obviously having a high priority for evaluation reference, since it is assumed that experts have the most objective scoring than the other category.

Gathered audience are 51 respondent with ratio of expert, semi-expert, and non-expert of 1:4:2 respectively. From each of three songs, it is presented three version of chord composition which two of them is produced by parallel artificial neural network and one manually assigned by experts. In this experiment, term ANN-1 and ANN-2 is used to differentiate these two artificial neural network. Recapitulated survey result of these scores is shown in following Table 3. It should be noted that the rating are scored with value range of 1 to 5.

Table 3. Survey Result

Respondent Category	Rating		
	ANN-1	**ANN-2**	**Manual**
Expert	3.2	3.067	4.2
Semi-expert	2.94	2.714	4
Non-expert	3.361	3.25	4.222
Average	3.167	3.01	4.141

Based on above result, manually assigned chords scored better than both of artificial neural network's composition. This is actually obvious since the manually composed is the one used as training data. However, this shown that the upper bound of the experiment's score is 4.141 out of 5. It is not surprising that perfect score is not achieved. Since it is well assumed, universally, that music is a very subjective

matter, even manually assigned chords by experts may sound not good enough to some audience. But there is also a possibility where this may happen because of expert's assigned chords are actually not good enough. Despite its imperfect upper bound, artificial neural network reached the score of 3.167 and 3.01, which actually can be concluded as good enough generally to audiences and achieved more than 75% of its upper bound.

There are still rooms for further development. One of the biggest limitation is data set limit. Current data sets are all manually transcribed. It is believed that if more variations of data set are able to be provided, the result could be improved significantly. The design of artificial neural network is good enough, however, there are still options of configurations available to be experimented on, i.e. recurrence network architecture. Observing responses from audience, it is also possible that the MIDI quality as the sample for the composition to be hearable, is decreasing the esthetics of chord composition. Even though instrumentation for playing the chords is not part of the evaluation, it could affect the songs rating. If possible, future experiments could handle this problem to gain higher quality of result.

V. CONCLUSION

It is concluded that artificial neural network is capable to produce a solid chord composition for melody accompaniment that is generally acceptable for hearers. Feed-forward multilayer perceptron architecture with 12 input nodes of tone occurrence, 12 input nodes of tone density, 12 output nodes of chord confidence, and one hidden layer with 2/3 of total nodes, backpropagation algorithm, 100000 fixed iteration, and sigmoid activation function for training are the best configuration proposed from the experiment. This is satisfyingly achieved the score of 3.167 with upper bound of 4.141 (out of 5).

REFERENCES

[1] Berkeley, I. S. N., Dawson, M. R. W., Medler, D. A., Schopflocher, D. P., & Hornsby, L. (1995). *Density plots of hidden value unit activations reveal interpretable bands*. Connection Science, 7, 167-186

[2] Curtis, M. E., Bharucha, J. J. (2010). *The minor third communicates sadness in speech, mirroring its use in music*. Emotion, 10, 335-348.

[3] Demuth, H. B., et al. (2014). *Neural Network Design 2nd Edition*. Paperback.

[4] Good, M. (2001). *MusicXML for Notation and Analysis*. MIT Press.

[5] Graupe, D. (1997). *Principles of Artificial Neural Networks 2nd Edition*. World Scientific Publishing.

[6] Hermawan, Arief. (2006). *Jaringan Saraf Tiruan: Teori dan Aplikasi*. Penerbit ANDI.

[7] Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press.

[8] Laden, B., Keefe, B. H. (1989). *The representation of pitch in a neural net model of pitch classification*. Computer Music Journal, 13, 12-26.

[9] Mauch, M., et al. (2009). *Using Musical Structure to Enhance Automatic Chord Transcription*. International Society for Music Information Retrieval Conference.

[10] Simon, I., et al. (2008). *MySong: Automatic Accompaniment Generation for Vocal Melodies*. ACM CHI Conference on Human Factors in Computing Systems.

[11] Simon, I., et al. (2008). *Exposing Parameters of a Trained Dynamic Model for Interactive Music Creation*. Association for the Advancement of Artificial Intelligence.

[12] Yaremchuk, V., et al. (2008). *Artificial Neural Networks that Classify Musical Chords*. International Journal of Cognitive Informatics and Natural Intelligence.