Music Recommendation System

Isabel Urrego Gómez
Universidad EAFIT
Ingeniería Matemática
Medellín, Colombia
iurregog@eafit.edu.co

***Abstract*—In the framework of the beginning of the investigation due to a work of the Artificial Intelligence undergraduate course, the author proposes a Music Recommendation System using a Fuzzy Inference System with four input variables and one output. The model was built from the beginning with the definition of the basic aspects: universe of discourse, linguistic categories, membership functions and decision rules. After that, the Mandami Fuzzy Inference System was implemented in python to finally evaluate the performance of the model. This document presents the model, results and conclusions achieved.**

***Index Terms*—Music Recommendation System, Fuzzy Logic, Mandami Fuzzy Inference System**

I. Introduction

Music plays a great role in peoples life and influences different parts of the daily routine such as the mood or the performance in sports, study, work, etc. Therefore, in the contemporary era of information many researchers have dedicated their work to improving music recommendation systems that enable people to find new music or allow artists to reach a wider public.

Since there are different approaches for Music Recommendation, the one chosen for this work was rather emotional than technical. This means, the selection of the input variables and its respective membership functions were not decided based on technical aspects but biased by the personal taste of the author, because it was considered that the emotional part influences more the perception of music than more specialized features. Besides, working with this second option would require a more rigorous study in signal analysis which was far from the purpose of this research.

Thus, this work aims to propose a Music Recommendation System based on the personal taste of the author to explore the scope of fuzzy logic and its wide range of applications by building a Fuzzy Inference System from the beginning. Starting from the selection of the variables. Then defining the universe of discourse, linguistic categories, membership functions and decision rules to finally evaluate the model and compare the results.

II. State of the art

The music recommendation process starts with the selection of an appropriate filter for the data. The two main approaches that have been discussed in literature are content-based filtering and collaborative filtering.

The first one explores the similarities between the users profile and the data obtained from the music in order to build a recommendation system. On the other hand, the collaborative approach groups together different user profiles and makes suggestions among members in the same group. However, a more robust idea has been developed by combining the advantages from both content-based filtering and collaborative filtering. The system collects the information of a single profile and users with similar information are grouped for collaborative recommendation [2]. For example, this approach has been used for TV-Programs recommendation [5] and websites [1].

Using this ideas, Chen and Chen built a website that based on music data grouping and user interests, makes musical recommendations to the user extracting six main features: Mean and Standard Deviation of the pitch values, pitch density, pitch entropy, tempo degree and loudness [2].

Alternatively, fuzzy logic has been used by some researchers to build music recommendation systems. By way of illustration, Lesaffre and Leman published a potential application of fuzzy logic to semantic music recommendation. The system they proposed consists in four parts: definition of the users profile, specification using semantic descriptors, recommendation of music and evaluation tasks. On the second part, the user chooses combinations between five music genres, eight emotion labels, four adjective pairs (soft-hard, clear-dull, rough-harmonious and void-compact) and another three adjective pairs (slow-quick, flowing-stuttering and dynamic-static). The desired output is an ordered list music titles ordered from most probable to like to the least one [4].

The article *A Fuzzy Inference-based Music Emotion Recognition System* presents a prototype music emotion recognition system called "FUMERS". They used the idea that music emotion could be useful in music understanding to be able to propose a music recommendation system. In particular, the key of this work was identifying the emotional factors hidden in music such as intensity, rhythm, scale and harmony and using them as inputs of the model. Finally, the six musical input features chosen were: tempo, loudness, mode, tonality, key and rhythm in order to obtain the output in terms of Arousal and the Valence. For instance, Valence was proposed with three membership functions negative, neutral and positive [3].

III. Mathematical Model

A. Input and Output Variables

Based on the literature review and exploring two Kaggle data sets for music classification, there were different possible features that can be extracted from songs and used as input variables for this model. For example, acousticness, danceability, energy, instrumentalness, tempo and length. However, using all of them would lead to a more complicated model which is not the purpose of this work.

Since the recommendation system proposed for this work is built based on the authors taste, the four final input variables were chosen based on what the author is more sensitive to: length, instrumentality, genre and a new variable called monotony level. The first variable is the length of the song, the second one refers to the level of presence of real instruments rather than computer created sounds and the last variable is a measure of how monotonous a song is.

Finally, the model will output a grade that shows how likely it is for the author to enjoy the song.

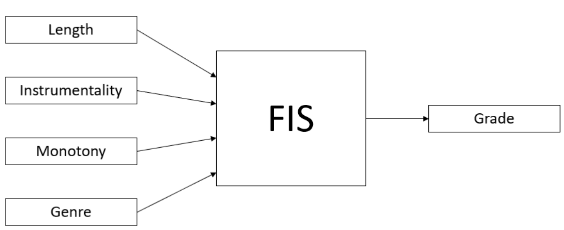

Fig. 1. Music Recommendation System

B. Universe of Discourse

To define the Universe of Discourse and knowing how extensive music is, it is important to establish some limits. For example, even if there are music pieces whose duration is 45 minutes, this system only accepts songs that last no more than 10 minutes. This means, the numeric value for length goes lays on the interval $[0, 10]$. On the other hand, the instrumentality will be measured from 0 to 1, being 0 the absence of real instruments like in some electronic music and 1 means that the song has no computer created sounds like La Vie en Rose the version of Louis Armstrong. The last input variable, the monotony level will also lay on the interval $[0, 1]$. To illustrate this feature, the song Bohemian Rhapsody by Queen presents a lot of changes in terms of rhythm , tone, lyrics and composition. Therefore, its monotony level would be very close to one, while the song Danza Kuduro presents the same beat, lyrics and rhythm for 3 minutes obtaining a probable monotony level closer to 0. The variable genre will receive a numerical value between 0 and 10.

$$length \quad \epsilon[0, 10]$$

$$instrumentality \quad \epsilon[0, 1]$$

$$monotony \quad \epsilon[0, 1]$$

$$genre \quad \epsilon[0, 10]$$

To conclude, the combination of the four input variables should output a grade between 0 to 10 of how likely it is for Isabel to like the song, being 10 total certainty that the song is a good recommendation for her and 0 total certainty that is not.

$$grade \quad \epsilon[0, 10]$$

C. Linguistic Categories

Classifying variables into linguistic categories is much more intuitive than assigning them numeric values. Fuzzy Logic, allows an approach that is very similar to human reasoning by using membership functions.

The variable length has three categories ”long”, ”short” and ”normal”. Additionally, the instrumentality is separated in ”low”, ”medium” and ”high”, whilst the monotony can be classified in ”not at all”, ”a bit” and ”very”. The selection of the genre classification was made according to the authors taste and the three resulting categories were: ”vallenato”, ”others” and ”rock”. The last one points to Isabel’s favorite genre and the first one is the worst musical genre in her opinion. Since the output will grade the songs according to the three 4 input variables, this grade is going to be split into 5 different linguistic categories: ” really bad”,”bad”, ”normal”, ”good” and ”amazing”.

D. Membership Functions

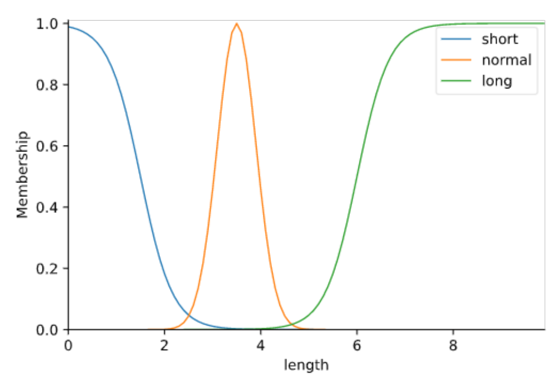

Fig. 2. Membership Function for length

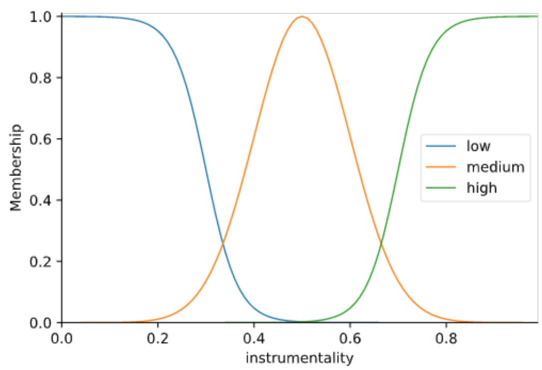

Fig. 3. Membership Function for instrumentality

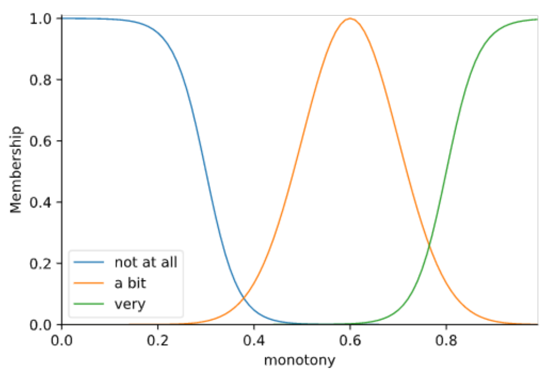

Fig. 4. Membership Function for monotony

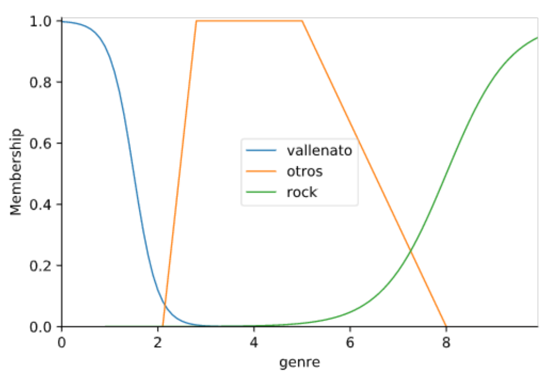

Fig. 5. Membership Function for genre

E. Decision Rules

1) IF the instrumentality is high and the song is not monotonous at all and is not vallenato, THEN the grade is amazing
2) IF the instrumentality is medium and the genre is not vallenato and either the length is not long or is not very monotonous, THEN the grade is good
3) IF the genre is vallenato, THEN the song is really bad
4) IF the instrumentality is low and is very monotonous and is a long song, THEN the grade will be really bad
5) IF genre is not vallenato and not rock and the instrumentality is not high, THEN the song is normal
6) IF genre is rock and instrumentality is high and (the song is not very monotonous or not long), THEN the song is amazing
7) IF the instrumentality is low and is a long song, THEN the song is bad
8) IF the length is normal, the instrumentality medium, is a bit monotonous and the genre is others, THEN the song is normal
9) IF the instrumentality is low and is very monotonous, THEN the grade is bad
10) IF the length of the song is short and the genre is others, THEN the grade is normal
11) IF the instrumentality is high and is very monotonous and the genre is others, THEN the grade is normal
12) IF the song is not monotonous at all and the length is short, THEN the grade is bad
13) IF the genre is rock and is very monotonous and has normal length, THEN the grade is good
14) IF the instrumentality is high and is very monotonous and the genre is rock and has long length, THEN the grade will be normal
15) IF the instrumentality is high and is a bit monotonous and is NOT vallenato, THEN the grade is normal
16) IF the instrumentality is low and is not monotonous at all and the genre is others, THEN the grade is normal
17) IF the instrumentality is high and is not monotonous at all and the genre is rock and has long length, THEN the song is amazing

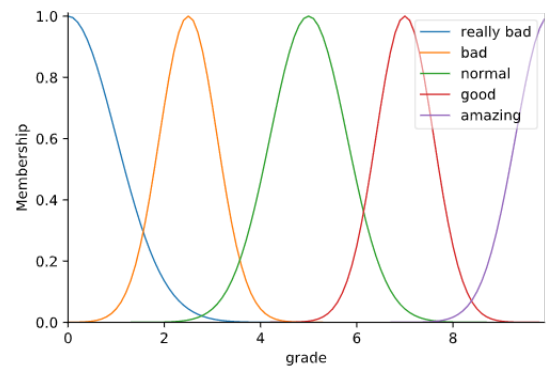

Fig. 6. Membership Function for grade

F. T-Norm and S-Norm

The T-Norms used were:

- Minimum: $T_{min}(a, b) = min(a, b)$
- Algebraic Product: $T_{ap}(a, b) = ab$

On the other hand, the S-Norms used were the following:

- Maximum: $T_{max}(a, b) = max(a, b)$
- Algebraic Sum: $T_{as}(a, b) = a + b - ab$

G. Defuzzification Methods

In order to evaluate the performance of the model, the following three defuzzification methods were applied

- centroid
- bisector
- max of maximum

IV. Results

The evaluation of the model was divided into two different phases. The first one consisted in using songs that the author proposed. On the contrary, the second phase was about checking on the performance of the model with songs proposed by other people. As a remark for the second phase, the author listened to the song and assigned it to one of the linguistic categories possible, then evaluated the song with the model and compared the results.

The Fuzzy Inference System implemented was Mandami-like using the Max-Min composition rule.

A. Phase 1

Song	Artist	Instrumentality	Monotony	Length	Genre
Weird Fishes/ Arpeggi	Radiohead	1	0.2	5	7
Olvídala	Binomio de Oro	1	0.5	5	0
Blanco	J Balvin	0.1	0.8	2:24	4
More Than Words	Extreme	1	0.4	5:30	6
Riders on the Storm	The Doors	0.8	0.6	7:12	7

TABLE I
SONGS PHASE 1

To illustrate the performance of the model, the following graphics show the results with the song *Riders on the Storm* based on the decision rule and the evaluation depending on the defuzzification method applied.

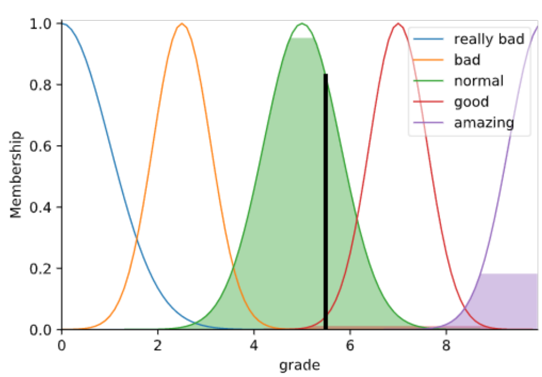

Fig. 7. Evaluation Using centroid method

B. Phase II

V. Conclusions

According to the results in both phases, changing the S-Norm and T-Norm did not lead to a significant change on the results. Therefore, it can be concluded that the combination minimum and maximum works as well as the algebraic product and the algebraic sum.

However, the accuracy obtained by the different defuzzification methods allows to summarize that the max of maximum is the one that gives the most realistic evaluation if the personal evaluation is above the one given by the model. For example, neither the centroid method nor the bisector classified *More than Words* as good but as normal. Nevertheless, the max of maximum was right with its evaluation.

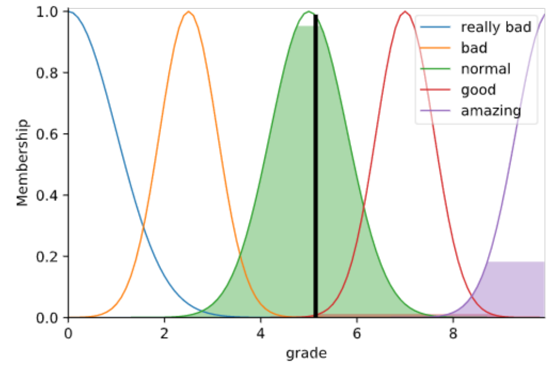

Fig. 8. Evaluation Using bisector method

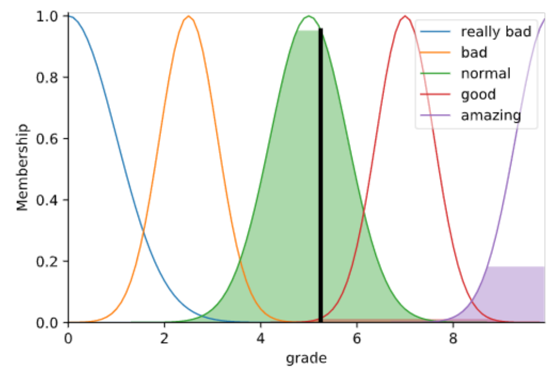

Fig. 9. Evaluation Using centroid max of maximum

Song	Defuzzification Method			Personal Evaluation
	centroid	bisector	max of maximum	
Weird Fishes	amazing 9.36	amazing 9.46	amazing 9.9	amazing
Olvídala	really bad 0.87	really bad 0.68	really bad 0.05	really bad
Blanco	normal 4.164	normal 4.5	normal 5.031	bad
Riders on the Storm	normal 5.49	normal 5.141	normal 5.244	normal
More than words	normal 5.56	normal 5.32	good 6.6	good

TABLE II
RESULTS PHASE 1 WITH T-NORM MIN AND S-NORM MAX

Song	Defuzzification Method			Personal Evaluation
	centroid	bisector	max of maximum	
Weird Fishes	amazing 9.361	amazing 9.46	amazing 9.9	amazing
Olvídala	really bad 0.824	really bad 0.68	really bad 0.05	really bad
Blanco	normal 4.161	normal 4.501	normal 5.032	bad
Riders on the Storm	normal 5.485	normal 5.134	normal 5	normal
More than words	normal 5.567	normal 5.323	good 6.6	good

TABLE III
RESULTS PHASE 1 WITH T-NORM ALGEBRAIC PRODUCT AND S-NORM ALGEBRAIC SUM

Song	Artist	Instrumentality	Monotony	Length	Genre
Charlie Brown	Coldplay	0.7	0.5	4:44	5
Ernie	Fat Freddy´s Drop	1	0.35	7:17	4
Remember the days of the old schoolyard	Yusuf/ Cat Stevens	0.8	0.6	2:44	6
Another Brick in the Wall	Pink Floyd	0.9	0.7	6	7
Baby´s On Fire	Die Antwoord	0.1	1	3:56	4

TABLE IV
SONGS PHASE 2

Song	Defuzzification Method			Personal Evaluation
	centroid	bisector	max of maximum	
Charlie Brown	normal 5.363	normal 5.215	normal 5.94	good
Ernie	good 7.38	really good 8.581	really good 9.9	really good
Remember the days of the old schoolyard	normal 5.16	normal 5.046	normal 5.244	normal
Another Brick In the Wall	normal 5.337	normal 5.124	normal 5.337	good
Baby´s On Fire	normal 3.948	normal 4.198	normal 5.031	really bad

TABLE V
RESULTS PHASE 2 WITH T-NORM MIN AND S-NORM MAX

Since the max of maximum tends to be more optimistic about the evaluation it has the worst performance in those cases where the model throws better results than the real ones. That is the case of the song *Baby´s On Fire*, also the song with the less accurate result. In particular, the model classified the song with all methods two categories better than the real one, but the method that was closest to the correct answer was the centroid, whilst the max of maximum was the wrongest among all.

Additionally, another important observation of this work is the presence of human bias. In the first place, the selection of the input variables was based on the features that the author considered more important in the process of music recommendation, but changing the person might result in different input variables. Besides, the evaluation of the model was also biased by the perception of the person proposing songs. For example the monotony level of a song is measured differently depending on the listener and that would lead to a different performance of the model. This means that the model works better if the users of the system have similar perceptions of the instrumentality, monotony and genre as Isabel.

Song	Defuzzification Method			Personal Evaluation
	centroid	bisector	max of maximum	
Charlie Brown	normal 5.35	normal 5.211	normal 5.94	good
Ernie	good 7.37	really good 8.581	really good 9.9	really good
Remember the days of the old schoolyard	normal 5.16	normal 5.045	normal 5.244	normal
Another Brick In the Wall	normal 5.455	normal 5.171	normal 5.8	good
Baby´s On Fire	normal 3.946	normal 4.195	normal 5.032	really bad

TABLE VI
RESULTS PHASE 2 WITH T-NORM ALGEBRAIC PRODUCT AND S-NORM ALGEBRAIC SUM

REFERENCES

[1] M. Balabanovic and Y. Shoham, “Fab: Content-based Collaborative Filtering Recommendation,” Communications of the ACM, vol. 40, number 3, pp 66–72, March 1997.

[2] H. Chen, and A.L.P Chen, “A music recommendation system based on music data grouping and user interests”Journal of Intelligent Information Systems, vol. 24, pp. 113—132, 2005.

[3] S. Jun, S. Rho, B. Han and E. Hwang, “A fuzzy inference-based music emotion recognition system” 5th International Conference on Visual Information Engineering, pp. 673–677, 2008.

[4] M. Lesaffre and M. Leman, “Using Fuzzy Logic to Handle the Users’ Semantic Descriptions in a Music Retrieval System ,” In: O, Castillo, P. Melin, O.M. Ross, R. Sepúlveda Cruz, W. Pedrycz and J. Kacprzyk “Theoretical Advances and Applications of Fuzzy Logic and Soft Computing” , vol. 42, Springer, Berlin, Heidelberg, pp 89– 98, 2007.

[5] B. Smith and P. Cotter, “A Personalized Television Listings Service,” Communications of the ACM, vol. 43, number 8, pp 107– 111, August 2000.