

Analisis Pengenalan Emosi Pada Musik dengan Sistem Berbasis Fuzzy

Emriliza Amarulhaq¹, Dede Rohidin,, Drs., MT.², Mahmud Dwi S., ST., MT.³

^{1,2,3}Departemen Teknik Informatika Universitas Telkom, Bandung

¹email.is.real@gmail.com, ²ddr@politeknikt Telkom.ac.id, ³10880696@ittelkom.ac.id

Abstrak

Penyimpanan musik digital membutuhkan mekanisme pencarian yang lebih mudah, fleksibel dan lebih maju menyesuaikan dengan kebutuhan individual user. Oleh karena itu dibutuhkan indeks retrieval yang lebih sesuai dan indeks retrieval yang paling baik adalah yang memfasilitasi pencarian yang sesuai dengan fungsi psikologis dan sosialnya. Indeks yang dimaksud secara khusus akan fokus pada informasi mengenai gaya, mood dan kesamaan musik.

Dari permasalahan diatas, telah banyak dikembangkan metode untuk *information retrieval* dari suatu musik dengan pemahaman yang telah didapat dari penelitian sebelumnya seperti pada genre classification dan speech recognition salah satunya adalah dengan pendekatan fuzzy. Subyektivitas persepsi manusia dalam mengklasifikasi emosi memberi kesan logika fuzzy adalah solusi kuat untuk permasalahan ini. Metode-metode berbasis *fuzzy system* memiliki kemampuan menghadapi masalah yang membutuhkan proses penalaran seperti pemodelan emosi dalam music emotion recognition. Akan digunakan fuzzy dengan dua metode learning berbeda untuk dianalisa hasilnya.

Kata kunci : *music emotion recognition, fuzzy inference system, music information retrieval, fuzzy k-nn, anfis*

Abstract

Music digital library needs more simple, flexible, and advance search mechanism to suit to user individual needs. Hence needed more befitting index retrieval method and the really good one is the one that can facilitate the search with its psychologic and social function. Indexes referred will specifically focus on information about style, mood, and music similarity.

Refer to the problems above, methods have been developed for information retrieval from music using knowledge from other previous researches in genre classification and speech recognition, one of them is using fuzzy approach. Subjective nature of human perception in classifying emotion suggest that fuzzy logic is a more appropriate tool for this problems. Fuzzy based methods have ability to face problem that require reasoning process such as emotion modeling in music emotion recognition. Thus, fuzzy with two different learning method will be used for the result of each will be analyzed.

Keywords: *music emotion recognition, fuzzy inference system, music information retrieval, fuzzy k-nn, anfis*

1. Pendahuluan

Penyimpanan musik digital membutuhkan mekanisme pencarian yang lebih mudah, fleksibel dan lebih maju menyesuaikan dengan kebutuhan individual user. "Fungsi dominan dari musik adalah dibidang sosial dan psikologi", dari pernyataan tersebut, maka indeks retrieval yang paling baik adalah yang memfasilitasi pencarian yang sesuai dengan fungsi psikologis dan sosialnya. Indeks yang dimaksud secara khusus akan fokus pada informasi mengenai gaya, mood dan kesamaan musik. [1]

Hal ini didukung dengan penelitian pada perilaku informasi musik yang telah mengidentifikasi mood musik sebagai patokan penting untuk *music retrieval* dan organisasi musik. [2] Mood atau emosi dari musik merupakan hal subyektif. Setiap orang sering merasakan emosi yang berbeda bahkan bertolak belakang dari satu lagu yang sama. Meskipun pendengar setuju akan jenis mood atau emosi dari suatu musik, masih ada kerancuan

mengenai deskripsi emosi tersebut. Mood atau emosi bisa direpresentasikan dengan berbagai cara, baik model kategori maupun model dimensi.

Metode yang digunakan untuk penelitian ini adalah metode machine learning berbasis *fuzzy*. Konsep *fuzzy* dipilih karena memiliki kemampuan memodelkan masalah yang rancu dan membutuhkan proses penalaran seperti emosi [3]. Akan digunakan *fuzzy* dengan dua metode learning berbeda yaitu dengan metode clustering dan feature learning untuk kemudian masing-masing dianalisa performansinya

2. Dasar Teori

2.1. Psikologi musik, mood dan emosi

2.1.1 Musik

Definisi musik, yang umum dikutip, adalah suara yang terorganisir. Musik adalah seni yang mediumnya adalah suara dan kesunyian. Elemen umumnya meliputi pitch, ritme, dinamika, dan kualitas timbre dan tekstur.

2.1.2. Mood dan Emosi

Peneliti MIR (*Music Information Retrieval*) menggunakan istilah mood dan emosi secara bergantian dan dapat dipertukarkan. Berbeda dengan pemahaman para psikolog, menurut mereka mood memiliki makna berbeda dengan emosi. Mood atau dalam bahasa baku Indonesianya suasana hati adalah suatu *state* atau keadaan emosional. Berbeda dengan emosi, mood bersifat kurang spesifik, kurang intens dan kecil kemungkinannya terpicu oleh stimulus atau peristiwa tertentu. Mood secara umum memiliki *valence* positif dan negative. Bisa dikatakan, mood dapat dibagi menjadi dua kategori yaitu *good mood* (suasana hati yang baik) atau *bad mood* (suasana hati yang buruk).

Banyak ilmuwan dan psikolog yang mencoba membuat model untuk merepresentasikan mood. Model tersebut bisa dibagi menjadi dua pendekatan : model kategori dan model dimensi. Model dimensional menggunakan beberapa sumbu untuk memetakan emosi ke suatu bidang. Dimensi yang umum digunakan adalah model menggunakan dua sumbu (*arousal-valence* atau *energy-stress*) dengan beberapa kasus adanya dimensi ketiga (*dominance*). Keuntungan dari model dimensional adalah mengurangi kerancuan jika dibanding dengan model pendekatan kategori

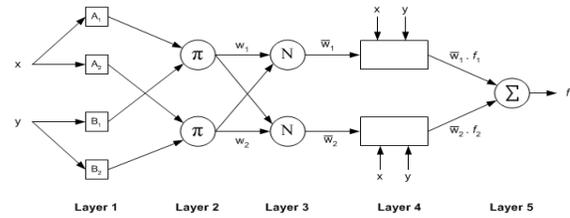


Gambar 2-1: Bidang Thayer untuk memodelkan emosi

2.2. Fuzzy Classifier

2.2.1 ANFIS

ANFIS merupakan gabungan antara Fuzzy System dengan Neural Network. ANFIS mampu menghadapi masalah yang membutuhkan penjelasan proses penalaran seperti *Fuzzy System* dan memiliki kemampuan learning dari data dan *feedback* seperti ANN (*Artificial Neural Network*)



Pada layer pertama, semua node adalah node adaptif. Output dari layer 1 adalah nilai keanggotaan fuzzy dari input. Biasanya, fungsi keanggotaan untuk himpunan fuzzy dapat berupa fungsi keanggotaan parameter, seperti segitiga, trapesium, Gaussian, atau Generalized Bell. Setiap node dalam lapisan kedua adalah node tetap berlabel M atau π , yang outputnya adalah produk atau multiplier sederhana semua sinyal yang masuk. Setiap node output mewakili kekuatan tembak sebuah aturan. Pada lapisan ketiga, node juga node tetap. Mereka diberi label dengan N, menunjukkan bahwa mereka memainkan peranan normalisasi terhadap kekuatan tembak dari lapisan sebelumnya.

Setiap simpul i pada lapisan 4 adalah simpul adaptif dan dinamai layer aturan yang diperoleh dari perkalian kekuatan tembak yang telah dinormalisasi (diperoleh di lapisan sebelumnya) dengan orde pertama aturan fuzzy Sugeno. Simpul tunggal di lapisan terakhir adalah node tetap berlabel S, yang menghitung output secara keseluruhan sebagai penjumlahan dari semua sinyal yang masuk.

2.2.2. Fuzzy K-Nearest Neighbor

Fuzzy k-NN classifier [11] (FKNN), sebuah kombinasi dari *fuzzy logic* dan *k-NN classifier*, didisain untuk menyelesaikan permasalahan di atas. FKNN terdiri dari dua langkah: *fuzzy labeling* yang menghitung *fuzzy vector* dari sampel training (pada *Model Generator*), dan *fuzzy classification* yang menghitung *fuzzy vectors* dari sampel masukan (dilakukan di *Emotion Clasifier*).

Pada klasifikasi dengan *fuzzy*, akan ditentukan sebuah nilai keanggotaan atau membership fuzzy μ_{uc} untuk satu sampel masukan x_u ke tiap kelas c sebagai kombinasi linear dari *fuzzy vectors k-nearest* sampel training:

Dimana μ_{ic} adalah nilai keanggotaan fuzzy dari sampel training x_i di class c , x_i adalah salah satu sampel *k-nearest*, dan w_i adalah bobot yang berbanding terbalik dengan jarak d_{iu} antara x_i dan x_u :

Dengan persamaan 1, akan didapat $C \times 1$ fuzzy vector μ_u yang menunjukkan kekuatan emosi musik ($C = 4$ dalam sistem ini) dari input sampel:

Pada fuzzy labeling akan dihitung μ_i , yaitu fuzzy vector dari sampel training. Beberapa metode telah dikembangkan sebelumnya ([11], [12]) dan bisa di generalisasikan sebagai berikut:

Dimana v adalah voted class dari x_i , nc adalah jumlah sampel yang berada di kelas c di K -nearest sampel training dari x_i , dan β adalah parameter bias yang mengindikasikan bagaimana v ambil bagian dalam proses labeling ($\beta \in [0,1]$). Bila $\beta=1$, ini adalah *crisp labeling* yang menetapkan tiap sampel training keanggotaan penuh pada *voted class* v . Bila $\beta=0$, nilai keanggotaan ditetapkan sesuai *K-nearest neighbors* (K mungkin berbeda dengan k yang digunakan pada *Emotion Classifier*).

Nilai Arousal dan Valence didapat dengan rumus berikut:

Valence dari
Arousal dari

3. Perancangan Sistem

3.1. Preprocessing

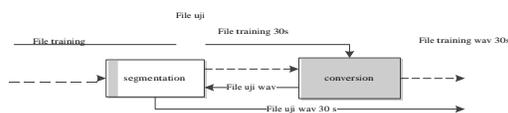
1. Segmentasi

Merupakan proses pemilihan bagian lagu sepanjang 30 detik. Untuk data training, pemilihan bagian dilakukan secara manual dengan menggunakan software audio editor. Detik awal pemotongan lagu pada data training disimpan untuk digunakan sebagai Roulette Wheel untuk segmentasi data uji.

Pada data uji pemotongan dilakukan secara random dengan Roulette Wheel yang telah didapat sebelumnya pada saat training.

2. Konversi

Merupakan proses konversi file mp3 menjadi file wav dengan sample rate 22050Hz dan kedalaman bit 16.



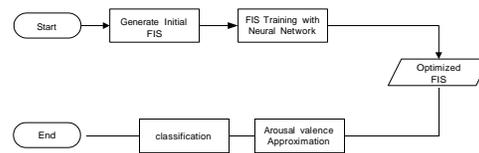
Gambar 3-1: proses preprocessing

3. Labeling

Merupakan tahapan dimana data yang sudah dipisahkan berdasarkan fitur diberi label sesuai kelas emosinya

3.2 ANFIS

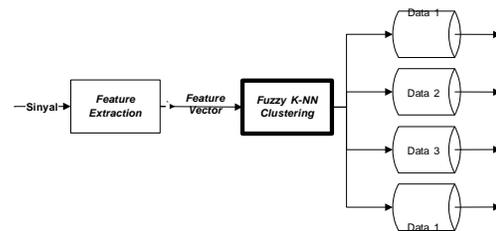
Data sinyal diproses, ANFIS men-tuning parameter-parameter FIS dengan menggunakan data input dan output. Hasil FIS yang sudah di optimisasi digunakan untuk mencari nilai Arousal dan Valence untuk diklasifikasikan.



Gambar 0-2 Alur proses ANFIS

3.3 FKNN

Data sinyal diproses, dilakukan *clustering* untuk mengelompokkan data berdasarkan kedekatan antar komponen fitur. Metode *Cluster* yang digunakan dalam pengujian ini adalah *Fuzzy k-NN Cluster*. Data input dibagi menjadi 4 buah kelompok, setiap kelompok terdiri atas komponen dengan kedekatan feature satu sama lain. Sedikit berbeda dengan k-NN biasa, pada algoritma Fuzzy k-NN sedikit dilakukan modifikasi, yakni membuat fuzzy vector untuk menunjukkan nilai keanggotaan data terhadap masing-masing kelas. Tiap nilai pada fuzzy vector menunjukkan kekuatan emosi yang relatif terhadap suatu kelas. Sebagai contoh, (0.0 0.1 0.8 0.1)^t menunjukkan fuzzy vector suatu data memiliki kekuatan emosi paling kuat untuk kelas 3, sehingga data tersebut masuk ke kelas 3.



Gambar 0-3 Alur proses FKNN

4. Analisa

4.1. Pengujian Validasi Sistem

Tabel 4-1: hasil validasi system dengan k-fold cross validation

Pengujian ke	Validasi Testing (%)								
	Dengan ANFIS (epoch 100)						Dengan FKNN		
	$\tau=0.01$ $\sigma_1=0.95$ $\sigma_2=1.05$	$\tau=0.001$ $\sigma_1=0.95$ $\sigma_2=1.05$	$\tau=0.0001$ $\sigma_1=0.95$ $\sigma_2=1.05$	$\tau=0.01$ $\sigma_1=0.85$ $\sigma_2=1.15$	$\tau=0.001$ $\sigma_1=0.85$ $\sigma_2=1.15$	$\tau=0.0001$ $\sigma_1=0.85$ $\sigma_2=1.15$	k=1	k=10	k=150
1	30,77	30,77	38,46	30,77	30,77	38,46	30,77	38,46	30,77
2	30,77	30,77	30,77	38,46	30,77	30,77	30,77	46,15	46,15
3	23,08	61,54	23,08	30,77	38,46	30,77	23,08	38,46	30,77
4	38,46	30,77	38,46	38,46	38,46	30,77	30,77	38,46	38,46
5	23,08	46,15	23,08	46,15	46,15	46,15	38,46	38,46	30,77
6	15,38	61,54	30,77	23,08	53,85	23,08	46,15	69,23	46,15
7	38,46	38,46	38,46	46,15	46,15	38,46	46,15	46,15	38,46
8	61,54	38,46	53,85	61,54	38,46	53,85	30,77	38,46	38,46
9	7,69	46,15	7,69	23,08	38,46	15,38	38,46	30,77	53,85
10	30,77	23,08	38,46	23,08	23,08	23,08	30,77	38,46	46,15
11	30,77	23,08	23,08	23,08	23,08	23,08	38,46	23,08	23,08
12	38,46	23,08	38,46	30,77	30,77	15,38	15,38	46,15	38,46
13	61,54	30,77	38,46	46,15	38,46	38,46	46,15	46,15	46,15
Rata-rata	33,14	37,28	32,54	35,50	36,69	31,36	34,32	41,42	39,05

Dari tabel 4-1 di atas dapat dilihat bahwa sistem menunjukkan akurasi terbaik pada percobaan ke-3 dan ke-6 untuk ANFIS dengan parameter *initial step size* 0,001 , *decrease rate* 0,95 dan *increase rate* 1,15 serta pada percobaan ke-8 dan ke-13 dengan parameter *initial step size* 0,01 , *decrease rate* 0,95 dan *increase rate* 1,15 yaitu sebesar 61,54%. Sedangkan untuk FKNN hasil terbaik didapat pada percobaan ke-6 dengan k=10 yaitu sebesar 69,23%. Hasil validasi menunjukkan parameter terbaik untuk ANFIS adalah *initial step size* 0,001 , *decrease rate* 0,95 dan *increase rate* 1,15 dengan rata-rata akurasi 37,28%. Sedangkan rata-rata akurasi terbaik untuk FKNN adalah 41,42% untuk nilai k=10

4.2 Pengujian Sistem

Pada pengujian ini dilakukan *training* menggunakan selain menggunakan data training yang telah dipartisi pada *K-Fold Cross Validation* juga menggunakan data gabungan data *training* dan data validasi. Data *testing* yang digunakan sebanyak 26 file dan data training yang digunakan sebanyak 156 file dan 169 file. Adapun hasil pengujian adalah sebagai berikut

Tabel 4-2: hasil akurasi untuk tiap metode

Observasi ke	FKNN					
	156 file			169 file		
	1	10	150	1	10	150
1	34,62%	34,62%	50,00%			
2	30,77%	42,31%	42,31%			
3	38,46%	38,46%	50,00%			
4	34,62%	38,46%	50,00%			
5	26,92%	38,46%	46,15%			
6	34,62%	30,77%	50,00%			
7	30,77%	46,15%	50,00%	30,77%	38,46%	50,00%
8	23,08%	30,77%	50,00%			
9	30,77%	42,31%	50,00%			
10	26,92%	26,92%	46,15%			
11	30,77%	30,77%	50,00%			
12	30,77%	38,46%	50,00%			
13	30,77%	38,46%	50,00%			
Rata-rata	31,04%	36,81%	48,90%	30,77%	38,46%	50,00%

Tabel 4-2 di atas menunjukkan pengujian untuk FKNN dengan nilai k=150 memiliki tingkat akurasi paling baik yaitu sebesar 50% baik untuk jumlah data 156 maupun 169 file. Rata-rata akurasi untuk hasil 13 kali pengujian menghasilkan nilai akurasi terbesar pada nilai k=150 yaitu sebesar 48,90%.

Tabel 4-3: hasil akurasi pengujian ANFIS

Percobaan ke	ANFIS											
	156 file						169 file					
	1	2	3	4	5	6	1	2	3	4	5	6
1	19,23	3,85	19,23	30,77	15,38	11,54	26,92	26,92	23,08	38,46	30,77	23,08
2	15,38	26,92	19,23	7,69	23,08	19,23						

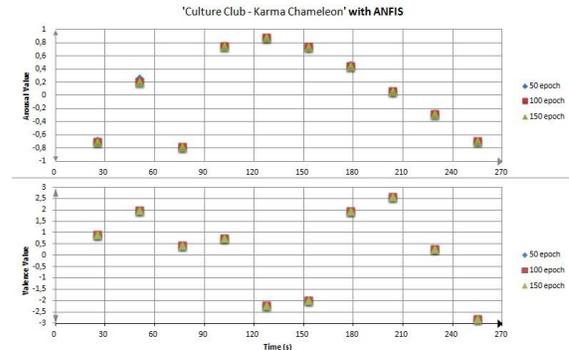
3	26,92	23,08	7,69	15,38	15,38	3,85													
4	19,23	15,38	30,77	34,62	19,23	19,23													
5	19,23	38,46	19,23	26,92	34,62	26,92													
6	26,92	26,92	15,38	23,08	23,08	15,38													
7	19,23	15,38	23,08	15,38	15,38	23,08													
8	26,92	15,38	23,08	19,23	26,92	26,92													
9	26,92	26,92	19,23	19,23	30,77	26,92													
10	19,23	26,92	30,77	23,08	26,92	26,92													
11	7,69	23,08	30,77	23,08	30,77	26,92													
12	19,23	26,92	23,08	23,08	34,62	30,77													
13	23,08	34,62	23,08	23,08	34,62	30,77													
Rata-rata	20,71	23,37	21,89	21,89	25,44	22,19	26,92	26,92	23,08	38,46	30,77	23,08							

Tabel 4-3 di atas menunjukkan ANFIS dengan parameter no 5 yaitu *initial step size* 0,001 , *decrease rate* 0,85 dan *increase rate* 1,15 memiliki tingkat akurasi paling baik yaitu sebesar 34,62% untuk jumlah data 156 pada percobaan ke-5,12 dan 13 dan parameter no 4 yaitu yaitu *initial step size* 0,01 , *decrease rate* 0,85 dan *increase rate* 1,15 untuk percobaan ke-4. Akurasi 38,4% diperoleh dengan parameter ANFIS ke-4 untuk 169 file. Dilihat dari rata-rata akurasi menggunakan 156 file, hasil terbaik didapat menggunakan parameter ke-5 dengan nilai rata-rata akurasi sebesar 25,44%.

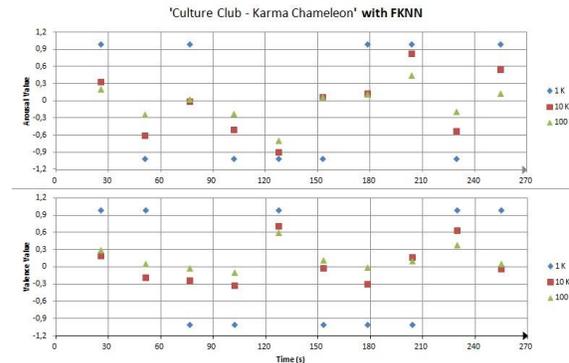
Untuk ANFIS selain pengujian diatas juga dilakukan pengujian hasil ANFIS yang ditraining secara terus menerus tiap percobaan menggunakan FIS hasil pengujian sebelumnya sebagai initial FIS untuk mentraining data berikutnya.

4.3 Variansi Emosi dalam Rentang Satu lagu

Berikut adalah hasil pengujian nilai arousal dan valence tiap metode untuk lagu ‘Culture Club – Karma Chameleon’ yang berdurasi 254 detik yang dibagi menjadi 10 potongan atau segmen lagu yang masing-masing segmen berdurasi 25,4 detik.



Gambar 4-1 Variansi nilai Arousal dan Valence dengan metode ANFIS



Gambar 4-2 Variansi nilai Arousal dan Valence dengan metode FKNN

Gambar diatas menunjukkan perubahan nilai Arousal dan Valence dalam rentang waktu tertentu dari suatu lagu. Dapat dilihat dengan menggunakan metode ANFIS tidak terlihat perbedaan nilai Arousal dan Valence yang besar untuk tiap nilai parameternya, sedangkan pada metode FKNN terlihat jelas perubahan nilai terjadi pada nilai K yang berbeda

5. Penutup

5.1 Kesimpulan

Berdasarkan hasil pengujian dan analisis yang telah dilakukan, dapat diambil beberapa kesimpulan sebagai berikut:

1. Percobaan menunjukkan penggunaan metode berbasis Fuzzy untuk mengenali mood pada musik masih belum menghasilkan akurasi yang besar pada saat pengujian, diperlukan data training lebih besar dan beragam serta pelabelan yang lebih objektif
2. Klasifikasi mood dengan FKNN menghasilkan akurasi rata-rata paling baik yaitu 41,42% di nilai k=10 pada saat validasi dan 48,9% pada k=150 pada saat pengujian .

3. Bila nilai fitur yang pada data uji melebihi nilai maksimum dan minimum fitur tersebut pada data training ANFIS maka keluaran akan menghasilkan nilai yang tidak diinginkan, pada kasus tersebut FKNN menunjukkan hasil yang lebih baik daripada ANFIS dalam menghadapi data baru.
4. Penentuan kelas emosi suatu lagu/musik dengan mengambil sampel beberapa detik dari lagu/musik tersebut belum bisa menunjukkan kelas emosi keseluruhan lagu/musik dengan baik.
5. Semakin besar dan objektif data training yang digunakan semakin baik dan akurat hasil yg didapat.

4.2 Saran

Hasil evaluasi dan analisa terhadap system menunjukkan bahwa sistem masih dapat dikembangkan. Beberapa saran pengembangan yang bisa dilakukan yaitu:

1. Perlu data training lebih banyak dengan pelabelan yang lebih obyektif dan akurat
2. Perlu dilakukan penelitian lebih lanjut mengenai variasi emosi dalam satu lagu/musik
3. Untuk ANFIS perlu dilakukan transformasi data uji agar tidak melebihi nilai batas minimum dan maksimum data latihnya.

Daftar Pustaka

- 19, no. 7, pp. 573-6, April 2009.
- [1] T. Fritz et al, "Universal Recognition of Three Basic Emotion in Music," *Current Biology*, vol. Elements of Expression in Music," *American Journal of Psychology*, vol. 48, no. 2, pp. 246-268, 1936.
 - [2] K. Heyner, "Experimental Studies of the Elements of Expression in Music," *American Journal of Psychology*, vol. 48, no. 2, pp. 246-268, 1936.
 - [3] Y.-H. Yang, C.-C. Liu and H. H. Chen, "Music emotion classification: A fuzzy approach", Santa Barbara, USA, Proc. ACM Multimedia, 2006, p. 81-84.
 - [4] J. Molino, "Musical Fact and the Semiology of Music," in *Music Analysis*, vol. 9, Wiley Blackwell Publishing, 1990, pp. 113-156.
 - [5] J. K. Kathi, "The Family Guide to Holistic Care for a Healthy Mind and Body," in *Mental Health, Naturally*, American Academy of Pediatrics, 2010.
 - [6] J. L. Schinnerer, "Guide to Self: The Beginner's Guide to Managing Emotion and Thought," AuthorHouse, 2006.
 - [7] N. H. Frijda, "The Emotions," Cambridge(UK), Cambridge University Press, 1986, p. 207.
 - [8] R. E. Thayer, "The Biopsychology of Mood and Arousal," USA, Oxford University Press, 1982.
 - [9] B. Schuller, C. Hage, D. Schuller and G. Rigoll, "Mister D.J., Cheer Me Up!': Musical and Textual Features for Automatic Mood Classification," *Journal of New Music Research*, vol. 39, no. 1, pp. 13-34, 2010.
 - [10] O. Lartillot and P. Toiviainen, "MIR in MATLAB (II): A Toolbox for Musical feature Extraction from Audio," in *International Conference on Music Information Retrieval*, Vienna, 2007.
 - [11] J. M. Keller, M. R. Gray and J. A. Givens, "A Fuzzy k- Nearest Neighbor Algorithm," *IEEE Trans. Syst. Man. Cybern.*, Vols. SMC-15, no. 4, pp. 580-585, 1985.
 - [12] J. H. Han et al, "A Fuzzy K-NN Algorithm Using Weights from the Variance of Membership Values," *CVPR*, 1999.
 - [13] J.-S. R. Jang, "ANFIS: Adaptive-Network-based Fuzzy Inference Systems,," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665-685, May 1993.
 - [14] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer, 2001.