

# Case-based Reasoning with Input Text Processing to Diagnose Mood [Affective] Disorders

Sri Mulyana, Sri Hartati, Retantyo Wardoyo, Edi Winarko

Department of Computer Sciences and Electronics  
Gadjah Mada University  
Yogyakarta, Indonesia

**Abstract**—Case-Based Reasoning is one of the methods used in expert systems. Calculation of similarity degree among the cases has always been an important aspect in CBR as the system will attempt to identify cases with the highest of similarity degree in a case-base to provide solutions for new problems. In this research, a CBR model with input text processing for diagnosing mood [affective] disorder is developed. It correlates with the increased tendency of mood disorder in accordance with the dynamics of the economic and political situation.

Calculation of similarity degree among the cases is one of the main focuses in this research. This study proposed a new method to calculate similarity degree between cases, Modified-Tversky. The analysis performed to assess the method used in measuring case similarity reveals that the Modified-Tversky Method surpasses the other methods. In the all tests conducted, the results of case similarity measures using the Modified-Tversky method is greater than or equal to the calculations performed using the Jaccard dan Tversky methods. The test results also provide an average level of performance in processing text input is 89.3 %.

**Keywords**—Case-Based Reasoning; mood disorder; case similarity; Jaccard Method; Tversky Method; Modified-Tversky Method

## I. INTRODUCTION

*Case-Based Reasoning (CBR)* is a method adopted from knowledge-based system in various domains. The method uses experiences from previous similar cases to solve new problems. The main idea is the assumption that similar cases have similar solutions [1].

The development of CBR possesses vast research opportunities including those related to case similarity calculation algorithms, indexing techniques to increase the efficiency of retrieval processes, case representation techniques, and methods to append new cases to the case base [2].

Researches could also be conducted by combining existing techniques through improvement or addition of new algorithms to achieve a need-fulfilling case-based computerized reasoning system. To test the validity of the techniques developed, implementation in a certain field is required.

In this research, a CBR model to diagnose mood [affective] disorder is developed. Mood [affective] disorder is caused by the inability of an individual to adapt to social

evolution. There is an increased tendency of mood disorder parallel to economic and political change.

That choice of field of implementation is based on the increased tendency of mood disorder due to the occurrence of economic and political change. The cause is the inability of an individual to adapt to social evolution.

In life, every single human being exhibits different kinds of emotions. Those expressions and feelings vary and is usually temporary, but could pose a problem if not immediately treated, resulting in emotional disorder including heavy depression rendering the subject devoid of motivation, happiness, and empathy.

Bitter experiences could cause an individual to feel insecure, void, depressed, and even hateful. That in turn results in uncontrollable anger planted deeply in one's inner conscience. What is felt as an unpleasant experience leads to extreme grief and all other forms of depression [3]. This condition could formulate as a chronic and repetitive problem interfering with an individual's ability to live out his everyday responsibilities.

## II. CBR RESEARCH IN HEALTH

The application of CBR has been vastly developed in various fields, particularly in health. Those is deployed to diagnose infection related diseases [4]. Explained in those researches is that such cases possess certain attributes such as temperature, headaches including its intensity and area of effect, coughing, bowel movement frequency, nausea, and urination. All those attributes have a numerical value in a scale of 1 to 4 to represent its intensity ranging from none or never, low, average, and high. Euclidean distance is employed in this research to measure similarity degrees between cases.

Current ongoing researches show that the implementation of CBR in the medical field has seen great development. Chakraborty at al. introduced a CBR system to detect swine flu dubbed SFDA (*Swine Flu Diagnostic Assistant*) [5], and Tomar at al. developed a CBR system to diagnose heart attacks [6]. Pant and Joshi have also done the same in the field of neurology in the form of NDS (*Neurology Diagnosis System*) [7]. The method used to measure similarity was the Nearest-neighbor method. Other CBR-based system was developed to advertise the dangers of smoking. It contained early warnings about the health effects of smoking broadcasted through mobile phones [8].

In the future, CBR systems could help provide better services in the medical field and perhaps have tighter integration with the clinical environment.

### III. MOOD [AFFECTIVE] DISORDERS

An individual's mood is the internal emotion that he or she dominantly feels, affects one's behavior and perception, and tends to be stable for a period of time. On the other hand, an affect is an external expression of a mood. Moods can be in a normal state, increasing, and depressive. A healthy mind has a wide range of moods, a good balance of affect expression, and is able to control mood and affects [9]. Mood disorders are triggered by extreme and uncontrollable mood fluctuation, resulting in a lack of adaptive ability, and could cause difficulty or discomfort.

Mood disorders are emotional illnesses accompanied by symptoms of mania and depression [10]. It is the result of the presence of complete negativity in an individual's mind, where said individual views himself, life, and the future as a chaotic mess of failure. Someone experiencing such a mood [affective] disorder would feel a lack of connection with other people or devoid of an influential role in their lives [3].

Mania is a certain mood disorder indicated by an abnormal increase of a certain mood. This condition is also accompanied by behavioral symptoms in the form of over-activity, talkativeness, great enthusiasm, euphoria, and even sexual deviation [10]. There exists two groups of maniacal symptoms, one being euphoric (elation, increased enthusiasm, heightened optimism), and the other being irritability (anger, aggressiveness, and violence). Hypomania is the condition of a slightly lower increase of mood compared to mania, and that usually does not reach the level where day to day functionality is threatened. Related to the time needed for diagnosis, hypomania can be identified as quickly as four days since the symptoms arrive [11].

Depression is a common emotional disorder indicated by constant sadness, loss of desire, a feeling of guilt, insecurity, sleeping problems or a decrease of appetite, low enthusiasm and energy, and lack of concentration [10]. It can also be associated with continuous grief in a dangerous amount, and is the most common emotional disorder. One in seven all around the globe have experienced an episode of depression at least once in their lives. It is the most common cause of death in developed nations and is the fourth most common cause in developing countries. Bipolar disorder is a chronic type of mood disorder with recurring episodes of mania or hypomania that strikes in turn with depression [11]. It also includes depression symptoms major or minor, usually identifiable by regression or even loss of ability to feel happy and positive during fun activities. This is constantly felt every day for at least two weeks [11]. People suffering from bipolar disorder can appear as overly optimistic individuals, frequently boasting about themselves, and even to the extent of making dangerous or risky decisions without calculating drawbacks and consequences. Sometimes those individuals can turn up brilliant ideas with the need for it to be immediately fulfilled. Unfortunately, those ideas tend to end in failure because of lack of thought, preparation, and some are just too unrealistic.

As a result of their easily irritable mood and the tendency to be easily offended in a state of either mania or depression, those with bipolar disorder often experience interpersonal communication problems with family and people around them. Sometimes they suspect that others are envious of them, and endeavor to prevent those bipolar individuals from achieving great things. During episodes of depression, they tend to blame family members and close friends as the root of their failure and suffering because in their minds they think that they are not provided with enough love, care, and support.

### IV. TEXT PROCESSING

Text processing is often required to simplify search process. The main objective is to convert various word forms into much more consistent index terms. An *Index term* is the representation of the essence of a document needed for search purposes. Text processing includes:

a) *Information Extraction: extracting important information within a document.*

b) *Text Summarization: to automatically produce a summary of a document.*

c) *Data Mining: A validation identification process to understand data patterns recorded in a structured database.*

d) *Text Mining: also known as text data mining or information searching.*

e) *Information retrieval: a document search.*

f) *Document Clustering: similar to document classification, but document class is not determined beforehand.*

The steps to process a document text is usually specific according to the process that needs to be carried out. One of the steps of processing a text document is indexing. The aim is to identify the best term to represent a certain document, so during document processing it could be obtained accurately. Automatic indexing consists of:

1) *Tokenization: a stage of processing where the input text is divided into units or "tokens" in the forms of words, numbers or punctuation marks.*

2) *Deletion of stop words: erasing less meaningful words. The choice of stop words are assessed using specific dictionaries, or lists of stops words.*

3) *Synonym identification: locating synonyms from words obtained from a document. This can be conducted with the use of a thesaurus or other similar methods. Similar words can be replaced with general terms.*

4) *Stemming: The process of text normalization within a document or changing an expression to its simplest form that in a literal view may not have a concrete meaning [12].*

5) *Frequency measures of stemming results: One of the methods that can be applied is the n-gram or a series of linguistic n item orders. Those items can be in the form of letters, vocabulary, or words. It is one of the implementations basic word features used in machine learning [13].*

6) *Selection or comparison of word frequency measure results.*

V. CALCULATION OF CASE SIMILARITY DEGREE

Let  $Sim(X, Y)$  denotes the similarity between two cases X dan Y where each have a finite number of features. The following are techniques to calculate the similarity degree:

1) Jaccard Method [14]

Case similarity using the Jaccard Method can be formulated as:

$$Sim(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} \tag{1}$$

which also means :

$$Sim(X, Y) = \frac{\#SAME(X, Y)}{\#SAME(X, Y) + \#DIFFER(X, Y)} \tag{2}$$

Where:

#SAME(X,Y): The number of same features between X and Y

#DIFFER(X,Y) : The number of different features between X and Y

This decision usually requires a specific threshold to state that the features of both cases are same or not.

2) Tversky Method[15]

Case similarity with the original Tversky method can be formulated as:

$$Sim(X, Y) = \frac{\alpha|X \cap Y|}{\alpha|X \cap Y| + \beta|X \oplus Y|} \tag{3}$$

Where:

$$|X \oplus Y| = |X \cup Y| - |X \cap Y| \tag{4}$$

which also means :

$$Sim(X, Y) = \frac{\alpha(\#SAME(X, Y))}{\alpha(\#SAME(X, Y)) + \beta(\#DIFFER(X, Y))} \tag{5}$$

The value of  $\alpha$  dan  $\beta$  is parallel with their respective significance of value, usually formulated by experts or produced using machine based learning techniques.

3) Modified-Tversky Method

The Modified-Tversky Method is a modification of the Tversky Method formulated as:

$$Sim(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|} \tag{6}$$

By adding weight of Tversky-Method, the equation can be written as :

$$Sim(X, Y) = \frac{2\alpha(\#SAME(X, Y))}{2\alpha(\#SAME(X, Y)) + \beta(\#DIFFER(X, Y))} \tag{7}$$

VI. METHODOLOGY

The implementation of Case-based Reasoning method to diagnose mood [affective] disorder consists of three parts: construction of case base, input text processing and case retrieval process, as shown in Figure-1:

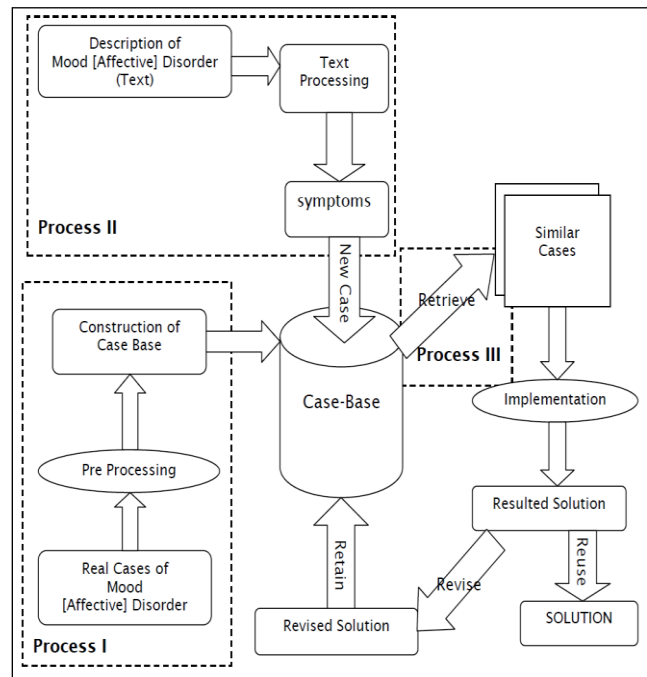


Fig. 1. The proposed CBR implementation method

Construction of case base is done by real life cases from various sources including the internet. The case base is then acquired as shown in Table-1:

TABLE I. CASE BASIS

No	Codes	Name of Mood Disorder	Symptoms Codes
1	F30.0	Hipomania	G13, G14, G15, G16, G25, G26, G27, G28
2	F30.0	Hipomania	G13, G14, G17, G19, G25, G26, G27, G28, G46
3	F32.1	Depression II	G36, G37, G38, G39, G40, G41, G42, G43, G44, G45, G46
4	F32.0	Depression I	G36, G37, G39, G44, G53
5	F32.0	Depression I	G36, G38, G41, G44, G53
6	F32.0	Depression I	G37, G39, G41, G42, G44, G45, G46
7	F32.0	Depression I	G35, G36, G37, G38, G39, G41, G42, G43, G44
8	F32.1	Depression II	G36, G37, G38, G40, G41, G42, G44, G45, G46, G47, G48
9	F32.1	Depression II	G35, G37, G38, G39, G40, G41, G43, G44, G46, G47, G50, G53
10	F31.1	Bipolar II	G13, G14, G15, G17, G21, G24, G25, G26, G27, G28, G36, G46, G49, G52
11	F31.1	Bipolar II	G13, G14, G15, G17, G21, G24, G25, G26, G27, G28, G36, G46, G48, G52
12	F31.0	Bipolar I	G13, G17, G25, G27, G37, G38, G39, G40, G41, G42, G44, G48, G52
13	F31.0	Bipolar I	G13, G14, G17, G26, G30, G36, G38, G39, G40, G44, G46, G52
14	F30.0	Hipomania	G13, G14, G15, G17, G18, G19, G21

The explanation meaning of the symptoms codes are shown in Table -2.

TABLE II. THE NAMES OF SYMPTOMS CODES

<b>Id</b>	<b>Name of symptom</b>
G13	increased mood
G14	increased energy and activity
G15	increased sociability, talkativeness and over familiarity
G16	increased sexual energy
G17	decreased need for sleep
G18	concentration and attention maybe impaired
G19	diminished the ability to settle down to work or to relaxation and leisure
G21	joviality
G23	easily distracted attention
G24	increased motoric activity
G25	talk too much
G26	self-esteem is inflated
G27	grandiose ideas
G28	over-optimistic ideas
G36	deminished energy and activity
G37	depressed mood
G38	loss of interest and enjoyment
G39	reduced concentration and attention
G40	reduced self-esteem and self-confidence
G41	idea of guilt and useless
G42	pessimistic
G43	idea of acts of self-harm
G44	disturbed sleep
G45	diminished appetite
G46	shows considerable distress
G47	loss of self-esteem
G48	attempted suicide
G52	decreased mood
G53	tiredness

Input text processing follow the steps above: tokenization, deletion of stop words, synonym search and replacement, N-gram configuration and symptoms diagnosis. Case retrieval is conducted by selecting a case from case base with the highest of case similarity degree. We implemented 3 methods to calculate the case similarity degree: Our Modified-Tversky, the original Tversky and Jaccard methods.

VII. RESULTS AND DISCUSSION

During text processing, testing is conducted through comparison of the number of symptoms used as the foundation of the input text configuration with those produced by text processing. The performance value is then formulated as:

$$P = \frac{NS}{NB} \times 100\% \tag{8}$$

Where:

P = Performance

NS = The number of same symptoms obtained from text processing

NB = The number of symptoms in the text (ground truth)

In the test conducted, text processing performance is shown in Table-3.

TABLE III. TEXT PROCESSING PERFORMANCE

<b>No</b>	<b>Symptoms in the text (ground truth)</b>	<b>N B</b>	<b>Symptoms obtained from text processing</b>	<b>NS</b>	<b>P(%)</b>
1	G36, G38, G39, G44, G53	5	G36, G38, G39, G44, G53	5	100
2	G36, G37, G39, G42, G53	5	G36, G37, G39, G42, G53	5	100
3	G13, G14, G16, G17, G18, G20	6	G13, G14, G16, G17, G18, G20	6	100
4	G36, G37, G38, G44	4	G36, G37, G38, G44	4	100
5	G13, G14, G15, G17, G18, G19, G20, G25	8	G13, G14, G15, G17, G18, G19, G20	7	87,5
6	G13, G14, G26, G27, G28, G36, G37, G38, G42, G43, G52	11	G13, G14, G27, G36, G37, G38, G42, G43, G48, G52	9	81,8
7	G13, G14, G16, G17, G18, G36, G46, G47, G48, G52, G53	11	G13, G14, G16, G17, G36, G43, G46, G48, G52, G53	9	81,8
8	G36, G37, G38, G39, G40, G41, G42, G46, G47, G48, G53	11	G37, G38, G39, G41, G43 G46, G47, G48, G53	8	72,7
9	G37, G38, G39, G42, G44, G53	6	G37, G38, G39, G44, G53	5	83,3
10	G36, G37, G38, G39, G43, G44, G53	7	G36, G37, G38, G39, G44, G53	6	85,7
Average Performance of Text Processing					<b>89,3</b>

Based on Table-2, the average performance of text processing of the test data is 89.3%. In several tests show that the performance of text processing does not reach 100 %, it indicates that the text processing algorithms in this study cannot always produce the correct result due to different interpretation of words, therefore the algorithm can be developed in order to improve performance.

During case retrieval, experiments are conducted by evaluating the correspondence of the diagnosis given by experts with the one produced by the CBR system and the case similarities degree generated. The entire results of the case retrieval experiment of 10 new problems are shown in Table-4:

TABLE IV. CASE RETRIEVAL TEST RESULTS

<b>No.</b>	<b>Expert Diagnose</b>	<b>CBR Diagnose</b>	<b>Case Similarity degree (%)</b>		
			<b>Modified-Tversky</b>	<b>Jaccard</b>	<b>Tversky</b>
1	F32.0	F32.0	88,89	66,67	80,00
2	F32.0	F32.0	88,89	66,67	80,00
3	F30.0	F30.0	44,44	44,44	28,57
4	F32.0	F32.0	36,44	36,44	22,22
5	F30.0	F30.0	83,33	55,56	71,43
6	F31.0	F31.0	43,75	43,75	28,75
7	F31.1	F31.1	41,18	41,18	25,93
8	F32.1	F32.1	86,49	61,54	76,19
9	F32.0	F32.0	88,89	66,67	80,00
10	F32.0	F32.0	95,24	83,33	90,91

The table shows that out of the three methods used in calculating of case similarity degree, the Modified-Tversky Method produces the best result. The statement is based on the all results of the test, as the suggested Modified-Tversky Method consistently generates better or at least equal outcome compared to other methods. This could be used as a reference in determining the method to calculate case similarity degree in future CBR researches. Based on Tabel-4 is also known that the diagnosis is generated by the CBR system in accordance with the diagnosis given by experts.

### VIII. CONCLUSION

In compliance with the test results and analysis carried out in this research, conclusions can be drawn as such: The modified-Tversky method outperforms the Jaccars and Tversky methods. In all the tests conducted, results of case similarity measures using the Modified-Tversky method is greater than or equal to the calculations performed using the Jaccard dan Tversky methods. The test results also provide an average level of performance in processing text input is 89.3%. However, in future work can be developed text processing algorithms in order to improve the performance of text processing. Cases on the case-base should be added so that the system can handle the problems of mood disorders. Therefore, it takes a lot of supporting data relating to the case of mood disorders.

### REFERENCES

- [1] A. Aamold, "Case-based reasoning: Foundation issues". *AICOM* 7, pp 39-59, 1994
- [2] A. Aamold and E. Plaza, "Case-based Reasoning: foundation issues, methodological variation and System approach", *AI Communication* 7(1), pp. 39-59, 1994
- [3] A. Nasir and A. Muhith, "The Fundamental of Psychiatric Nursing: Introduction and Theory", Salemba Medika, Jakarta, 2010
- [4] M. Denis, and A. Jasmin, "Applying Case-based reasoning for mobile support in diagnosing infective diseases". *International Conference on Signal Processing Systems*, IEEE Computer Society, pp 779-783, 2009
- [5] B. Chakraborty, I. Srinivas, P. Sood, V. Nabhi, D. Ghosh, "Case Based Reasoning Methodology for Diagnosis of Swine Flu", *IEEE GCC Conference and Exhibition*, 19-22 Februari 2011, Dubai, United Arab Emirates, pp. 132-135, 2011
- [6] P. P. S. Tomar, R. Singh, P. K. Saxena, and J. Sharma, "Case Based Medical Diagnosis of Occupational Chronic Lung Diseases From Their Symptoms and Signs", *International Journal of Biometrics and Bioinformatics (IJBB)*, Volume (5) : Issue (4), pp. 216-224, 2011
- [7] S. Pant, and S. R. Joshi, "Case-Based Reasoning In Neurological Domain", *IEEE* 978-1-4673-2590-5/12, 2012
- [8] K. Ghorai, S. Saha, A. Bakshi, A. Mahanti, and P. Ray, "An mHealth Recommender for Smoking Cessation using Case Based Reasoning", *46th Hawaii International Conference on System Sciences*, pp. 2695-2704, 2013
- [9] B. J. Sadock and V. A. Sadock, "Mood disorder" In : *Kaplan & Sadock's Synopsis of Psychiatry: Behavioral Sciences/Clinical Psychiatry, 10th Edition*. Wolters Kluwer/Philadelphia. Lippincott Williams & Wilkins, 2007
- [10] E. Dalami, "Nursing Clients with Mental Disorder", *Trans Info Media*, Jakarta, 2009
- [11] E. Vieta, "Managing Bipolar Disorder in Clinical Practice", second edition, London UK. Current Medicine Group Ltd., 2009
- [12] S. Ferilli, F. Esposito, and D. Grieco, "Automatic Learning of Linguistic Resources for Stopword Removal and Stemming from Text", *Procedia Computer Science*, 38, 116-123, 2014
- [13] O. Serban, G. Castellano, A. Pauchet, A. Rogozan, and J. P. Pecuchet, "Fusion of Smile, Valence and N-Gram Features for Automatic Affect Detection", *Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013
- [14] G. Kowalski, "Information Retrieval Architecture and Algorithm", Springer, New York, USA. E-Book (e-ISBN 978-1-4419-7716-8), 2011
- [15] S. K. Pal, and S. C. K. Shiu, "Foundation of Soft Case Based Reasoning", John Wiley & Sons, New Jersey, 2004