

Data Similarity Filtering of Wartegg Personality Test Result using Cosine-Similarity

<https://doi.org/10.3991/ijes.v6i3.9413>

Rosihan Ari Yuana^(✉), Dewanto Harjunowibowo,
Nugraha Arif Karyanta, Cucuk Wawan Budiyanto
Universitas Sebelas Maret, Surakarta, Indonesia
rosihanari@staff.uns.ac.id

Abstract—Wartegg test is a widely adopted personality evaluation instrument known for its drawing completion technique. Employee personality data, for instance, can be sorted by the closest similarity with the expected characters. Whereas, Wartegg test plays a significant role in data similarity filtering. Despite the potential contribution of personal characters identification technique, practical guidance is rarely found in the literature. This paper demonstrates the usage of cosine-similarity method for data similarity filtering on Wartegg personality test. The method used in this study is a case study, in which will be selected several Wartegg test subjects. By using the value of each character aspect derived from the Wartegg test, the cosine-similarity value will be calculated against the expected/ideal aspect character. Based on this value, the Wartegg test subjects will be filtered based on similarity to the expected/ideal character aspects. A technical procedure to perform the method is also presented in this paper. In order to find out the effectiveness, sample data scores of each character aspect from five test subjects, and also the ideal scores of the expected characters are given. By using FWAT, a graphical representation of the test subjects' characters to the ideal characters is generated. Then, this graph was compared to the results obtained from the cosine-similarity method. Drawn from the results, the cosine-similarity is effectively applied for Wartegg test data similarity filtering.

Keywords—Wartegg test, personality test, psychological application, similarity tool

1 Introduction

The Wartegg Zeichen Test (or WZT for short) is a widely adopted personality evaluation instrument known for its drawing completion technique [1, 2]. WZT suggests that personality could be projected through the way a person constructs the graphic elements from semi-structured signs [3]. Besides, the Wartegg test assumes that the content and the qualitative aspects of the signs imply the personality of the person who draws [4].

The Wartegg test has been widely implemented in many areas of study around the globe. In the field of education, for instance, it is used to predict the passing rate of

students [5]. Additionally, in the field of health, paramedics to help the healing process of cancer patients [6], examine psychopathology use the Wartegg test and personality in a group of patients affected by a certain disease [7], and can aid psychological evaluation during patient hospitalisation and psychotherapy [8]. Furthermore, the Wartegg test is also used by many companies in the selection process of prospective employees based on the job applicants' personalities [9].

Notwithstanding the test has been widely applied, the time taken for the analysis process delays the acquisition of the test outcome. Obtaining the graphic presentation from numerical test output is prolonged by the manual transfer of the scores obtained from the test [10]. Another problem faced by psychologists is how to quickly distinguish numeric test outcomes from several people. Some software such as E-Psychology [11], Presentation, and AcqKnowledge [12] have been developed to help psychologists to analyse the test result data. The expert system-based software has also successfully developed to assist psychologists in psychotherapy processes [13]. Specifically related to the Wartegg test, Campos [14], was created to analyse test results and also to provide a person's personality classification as its output. However, the expert system-based software was lacking in the interpretation of the subjects' characters to the predefined ideal criteria. For example, a company want to select employees that have personality close to the expected characters. Furthermore, employee personality data can be sorted by the closest similarity with the expected characters. This can be considered as a similarity data filtering problem. Whereas, this problem will be significant for users expecting to select subjects bearing personality with the most ideal characters.

A method that can be used for data similarity filtering is the cosine-similarity. This method has been widely used for measuring how close between two vectors, for example to detect errors in complex networks [15], as well as for filtering high-dimensional datasets [16]. The cosine similarity method gives the best value of proximity or similarity among other algorithms [17, 18].

Therefore, this paper contributes the usage of cosine-similarity method for data similarity filtering on Wartegg personality test. In addition, a Wartegg data scoring system is also presented to simplify the computation, so that it can be implemented to the computer application easily. In order to determine the effectiveness, the proposed methods will be implemented into the Fast Wartegg Analyzer Tool (FWAT), a web-based application to assist the acquisition process of Wartegg test data outcomes and discriminates individual subjects according to a set of predefined characters [19].

2 Method

The method used in this study is a case study, in which will be selected several Wartegg test subjects. By using the value of each character aspect derived from the Wartegg test, the cosine-similarity value will be calculated against the expected aspect character. Based on this value, the Wartegg test subjects will be filtered based on similarity with the aspect of the expected character.

In order to facilitate the computation of Wartegg test results, the input scoring system is needed. The scoring system taken in this study adopted the FWAT input scoring system. FWAT input is a number of 0, 0.5, 1, 1.5, 2, 2.5, or 3 as scores. Score 0 is given if that particular characters did not appear in the picture drawn by the test subject; Score 1 if a characters appears in the image, but is not overpowering; Score 2 if a characters appears in the picture powerfully; and Score 3 if a characters appears in a picture very predominantly. The score 0.5 is used if a characters appears between score 0 and 1, and similarly to scores 1.5 and 2.5, which are used to represent the middle scores between 1 and 2, and between 2 and 3, respectively. The input process is performed for each test subject with each picture (eight pictures) (see Figure 1).

Furthermore, scores that have been inputted will be calculated to obtain the total score for each characters (S_c) using the formula $S_c = \left[\sum_{i=1}^8 s_{ci} \right]$, where s_{ci} is a score for each characters c in the i -th picture, where $i = 1, 2, 3, \dots, 8$. Once s_{ci} is obtained, a description of the subject's characters can be achieved through graphics output generated by FWAT [19].

Criteria		Drawings							
		1	2	3	4	5	6	7	8
Nature	Animate	0	0	0	0	0	0	0	0
	Physiognomy	0.5	0	0	0	0	0	0	0
	Inanimate	1	0	0	0	0	0	0	0
	Atmosphere	1.5	0	0	0	0	0	0	0
Objects	Utility	2.5	0	0	0	0	0	0	0
	Ornaments	3	0	0	0	0	0	0	0

Fig. 1. FWAT score input form

Related to the data similarity computation, to identify similarities between a test subject's character aspects total score to the ideal characters score desired by a user, the cosine-similarity method is used. There are eight character aspects in Wartegg test, i.e. outgoing emotion; seclusive emotion; combinative imagination; creative imagination; practical intellect; speculative intellect; controlled activity; and dynamic activity. If each character aspect total score of the subject is presented as a vector $M = (m_1, m_2, \dots, m_8)$ and the ideal or expected character aspect score is $N = (n_1, n_2, \dots, n_8)$, then the cosine-similarity is

$$\cos \theta = \frac{\mathbf{M} \cdot \mathbf{N}}{\|\mathbf{M}\| \|\mathbf{N}\|} = \frac{\sum m_i n_i}{\sqrt{\sum m_i^2} \sqrt{\sum n_i^2}}$$

The closer the value obtained from the cosine-similarity is to 1, the closer in similarity the test subject's character aspects are to the ideal character aspects. Inversely, if the cosine value is not close to 1, then the subject's character aspects are increasingly incompatible with the ideal character aspects.

3 Result

Generating data similarity filtering outcomes requires a set of procedures as indicated in Figure 2. The procedure comprises the manual input of Wartegg test data and automatic selection of subjects and generation of graphical outcome. The data filtering process of the Wartegg test can be elucidated as follows. Based on the process in Figure 2, the user inputs Wartegg test result data of some test subjects manually using the scoring system as in Figure 1. Furthermore, this input data will be processed so that the total score for each aspect of character S_c is obtained, in this case S_c scores are presented in the form of vector M .

In the second step, the user sets the value for each expected ideal characters (N) as well as the similarity tolerance value used as the threshold for the filtering process. The similarity tolerance value is chosen in such a way that it approaches to 1. If the value is closer to 1, then the test subject with the closer to the ideal character will be obtained.

The next step is to compute the cosine-similarity value. Based on this value, if the similarity value of an individual test subject is more than or equal to the similarity tolerance value, then the subject will be nominated as the test output (filtering result). In contrast, if the subject's similarity value is less than the similarity tolerance value, then the result would not be shown. The user will be prompted to specify the criteria based on the character aspects he wants to view. In this case, the user may specify some combination of character aspects (at least two combined character aspects). For example, only the aspects of the outgoing and seclusive emotion are composed, or three-characters combinations on the aspects of outgoing, seclusive emotion and combinative imagination, etc. If the expected output is not obtained, it means that the value of the similarity of all subjects are smaller than the similarity tolerance value. Thus, in this case what needs to be done is to reduce the similarity tolerance value.

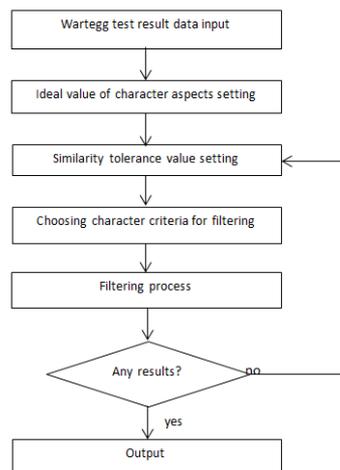


Fig. 2. Data similarity filtering process of Wartegg test results Discussion

In order to examine the effectiveness of the cosine-similarity method for data similarity filtering on Wartegg test, data samples taken from five test subjects. The effectiveness of this method will be seen by comparing the graphic distribution of ideal character aspect and test subjects to the filtering results with cosine-similarity. To obtain the graph, FWAT is used in this research. While to get the similarity filtering result, we have added this module to the FWAT.

Given the ideal scores and character aspect scores of each test subject, as presented in Table 1, a three-stage test was designed. In each stage, the selection of several different character aspect combinations was conducted with the focus on the identification of the subject with the closest value to the predefined ideal characters. In contrast, the furthest from the ideal subject was also identified to construct a comparison graph to represent the accuracy of the filtering process.

Filtering of outgoing aspects of emotion and creative imagination was performed in the first stage of test. The order of subjects' closeness to the ideal value was obtained subsequent to the calculation of similarity values of each subject using FWAT. The result is depicted in Figure 3.

Table 1. Score data sample of five test subjects and the ideal score

Character Aspects	Ideal Score	Test Subjects				
		1	2	3	4	5
Outgoing Emotion	36	26	19	22	31	14
Seclusive Emotion	45	27	22	24	48	42
Combinative Imagination	15	25	26	38	40	23
Creative Imagination	21	19	20	20	31	23
Practical Intellect	20	16	10	21	36	18
Speculative Intellect	38	19	11	15	8	14
Dynamic Activity	24	60	64	45	45	34
Controlled Activity	22	39	53	36	41	28

According to Figure 5, Subject 1 has outgoing emotion and creative imagination aspects scores that are closest to the expected ideal value. By using a 0.97 similarity tolerance value, it is recognised that, of the five subjects, there are only two subjects having a similarity value more than 0.97 – namely, Subject 1 and Subject 3 – while the rest of the participating subjects are not close enough to the ideal score (similarity value less than 0.97). In this experiment, Subject 5 indicates the lowest similarity value, or in another word, Subject 5 is the furthest from the expected ideal characters in the outgoing emotion and creative imagination aspects.

Filtering Result	
Character aspects: outgoing emotion, creative imagination	
Test Subjects	Similarity Values
Subject 1	0.99469968012721
Subject 3	0.97808495388173
Subject 4	0.96707453726265
Subject 2	0.96023346131666
Subject 5	0.87952510453374
Similarity tolerance value : 0.97	

Fig. 3. Filtering result for outgoing emotion and creative imagination aspects

Figure 3 shows the comparison of the Subject 1 and Subject 5 characters values to the ideal value. It is concluded from the graph that in the open emotion (outgoing emotion) aspect value, Subject 1 is closer to the ideal value than Subject 5. In this case, the open emotion aspect value of Subject 5 is much lower than Subject 1 against its ideal value in this aspect. While, in the aspect of creative imagination, the difference between Subject 1 and Subject 5 compared to the ideal value is not so significant. Thus, generally for a combination of both aspects, Subject 1 is closer to the ideal value than Subject 5.

Meanwhile, in the second experiment, a filtering test was conducted on three combination aspects. In this case, the selected aspects were seclusive emotion, practical intellect, and dynamic activity, using the similarity tolerance 0.90. From the filtering process, there were only two subjects with a similarity score over 0.90; Subject 5, which had a similarity score 0.99154426658371, was very close to the ideal, and also Subject 4 with score 0.96492060428270. The three other subjects were further from the ideal value (similarity value under 0.90). Meanwhile, Subject 2 had the lowest similarity value. Overall results of this experiment are shown in Figure 5.

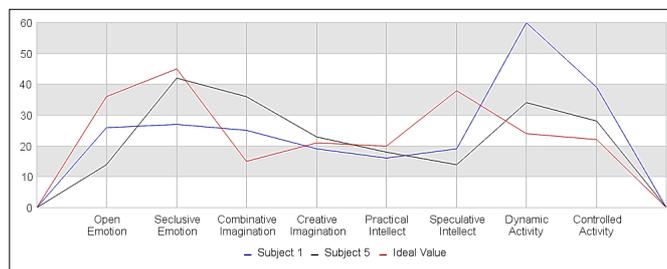


Fig. 4. Comparison graph of the character values of Subject 1 and Subject 5 against the ideal values

Filtering Result	
Character aspects: seclusive emotion, practical intellect, dynamic activity aspect	
Test Subjects	Similarity Values
Subject 5	0.99154426658371
Subject 4	0.96492060428270
Subject 1	0.88629909258042
Subject 3	0.88359843152965
Subject 2	0.74991405911192
Similarity tolerance value : 0.90	

Fig. 5. Filtering result for seclusive emotion, practical intellect, and dynamic activity aspects

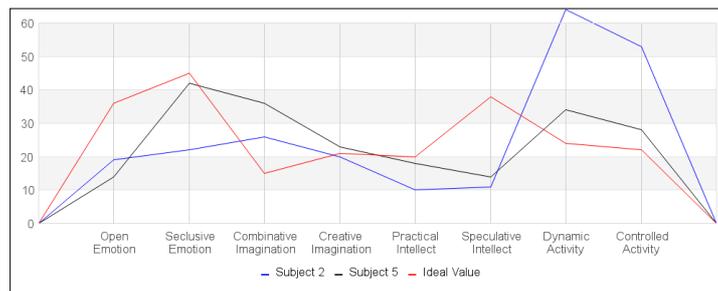


Fig. 6. Comparison graph of the character values of Subject 2 and Subject 5 against the ideal values

The graph presented in Figure 6 shows the comparison between the character values of Subject 2 and Subject 5 against the ideal values. From the graph, it appears that in the seclusive emotion aspect, Subject 5 is very close to its ideal value. Instead, Subject 2 is far below its ideal value. Meanwhile, the data for the practical intellect aspect shows the same result. As for the aspect of dynamic activity, the graph indicates that Subject 2 has a value that is much farther above the ideal value than Subject 5.

Furthermore, in the third stage of testing, filtering was done on four aspects: a combination of outgoing emotion, combinative imagination, practical intellect, and controlled activity. The result of this test, with the value of similarity of each aspect of his character, as shown in Figure 7, was obtained.

Filtering Result	
Character aspects: outgoing emotion, seclusive emotion, controlled activity, dynamic activity aspect	
Test Subjects	Similarity Values
Subject 4	0.95145097131087
Subject 5	0.92560573238540
Subject 3	0.85459940652819
Subject 1	0.83205540179866
Subject 2	0.75099982444564
Similarity tolerance value : 0.90	

Fig. 7. Filtering result of outgoing emotion, seclusive emotion, controlled activity, and dynamic activity aspects

By using a similarity tolerance value of 0.90, based on Figure 7, it can be seen that two subjects meet the level of proximity to the ideal value on the four selected aspects, namely Subject 4 and Subject 5. Subject 4 is the subject that has the closest character aspect to the ideal characters value, while Subject 2 has the most distant. Graphics comparison of characters values between Subject 2 and 4 compared to ideal values can be seen in Figure 8. Particularly in the outgoing emotion aspect, Subject 4 has a very close value to the ideal value compared to Subject 5. This difference appears very significant. While the seclusive emotion aspect also has the same characters, Subject 4 is very close to its ideal value with a significant difference compared to Subject 2. The same result also occurs in the 'controlled activity' and 'dynamic activity' aspects, in which Subject 4 is closer to the ideal value. Especially in these two aspects, it appears that the difference value of Subject 2 to the ideal value is almost two times greater than Subject 4.

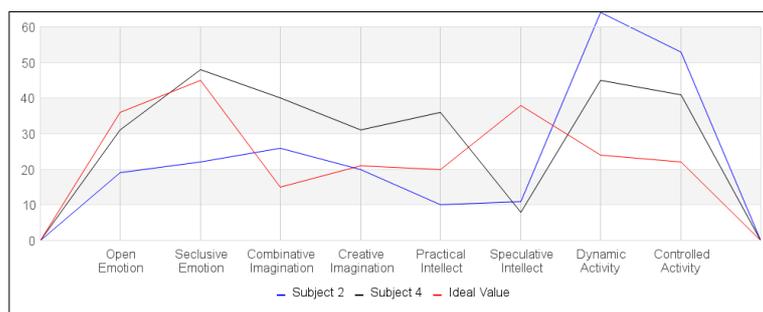


Fig. 8. Comparison graph of the characters values of Subject 2 and Subject 4 against the ideal values

4 Conclusion

Based on the experiment involving five test subjects, the use of cosine-similarity method for data similarity filtering on Wartegg test result to determine the subject having the character value close to the expected character value can be done. This conclusion is obtained by comparing the graph that represents the character aspect value of each test subject and the graph of the expected ideal aspect value of the character, with its cosine-similarity value. From this comparison, it can be seen that the cosine-similarity method can be used effectively for data similarity filtering. Based on conducted experiment, it also can be concluded that the cosine-similarity is effectively implemented for particular or even all character aspects of Wartegg test. Furthermore, this method can be developed by combining the expert system based on pattern recognition and fuzzy logic in order to interpret numerical data Wartegg test results automatically, and can also be used to search a data for a subject who has certain character using fuzzy query.

5 References

- [1] Roivainen, E., A Brief History of the Wartegg Drawing Test, *Gestalt Theory Journal*, 2009, 31:55–71.
- [2] Soilevuo, G. J., Grønnerød, C. The Wartegg Zeichen Test: A literature overview and a meta-analysis of reliability and validity, *Psychological Assessment*, 2012; 24:476–489. <https://doi.org/10.1037/a0026100>
- [3] Crisi, A., Dentale, F., The Wartegg Drawing Test: Inter-rater Agreement and Criterion Validity of Three New Scoring Categories, *International Journal of Psychology and Psychological Therapy*, 2016, 16:83–90.
- [4] Kinget, G. M., *The Drawing Completion Test – A Projective Technique for the Investigation of Personality*, Grune & Stratton Inc, New York. 1952.
- [5] Mellberg, K., The Wartegg Drawing Completion Test as a Predictor of Adjustment and Success in Industrial School, *Scandinavian Journal of Psychology*, 1972, 13:34–38. <https://doi.org/10.1111/j.1467-9450.1972.tb00046.x>
- [6] La Spina, C. M., Murasecchi, C., Stinco, S., Cassano, A., Strippoli, A., Trigila, N., Barone, C., The Psychological Projective Test Wartegg-Zeichentest: A contribution to The Knowledge of Patients with Cancer, *ASCO Annual Meeting Proceedings*, 2010, 28:110–117.
- [7] La Spina, C. M., Murasecchi, C., Stinco, S., Cassano, A., Strippoli, A., Trigila, N., Barone, C., The Psychological Projective Test Wartegg-Zeichentest: A contribution to The Knowledge of Patients with Cancer, *ASCO Annual Meeting Proceedings*, 2010, 28:110–117.
- [8] Rizzo, A., Villa, L. D., Crisi, A., Can the Problematic Internet Use evolve in a pre-psychotic state? A single case study with the Wartegg, *Computers in Human Behavior*, 2015, 51:532–538. <https://doi.org/10.1016/j.chb.2015.04.063>
- [9] Gregory, and Robert, J., *Psychological Testing: History, Principles, and Applications*. Third Edition. Boston : Allyn & Bacon, Inc, 2000.

- [10] Aulia, I., Kustimah, Frederick, D. P., Hubungan Antara Hasil Tes Rorschach dan Wartegg Zeichen Tes Dalam Menggali Aspek Emosi Dari Kepribadian, Fakultas Psikologi Universitas Padjajaran, 2007.
- [11] Drigas, A., Koukianakis, L., Papagerasimou, Y., Towards an ICT-based psychology: E-psychology, Computers in Human Behavior, 2011, 27:1416–1423. <https://doi.org/10.1016/j.chb.2010.07.045>
- [12] Blanch, A., Balada, F., Aluja, A., Presentation and Acknowledge: An application of software to study human emotions and individual differences, Computer Methods and Programs in Biomedicine, 2013, 110:89–98. <https://doi.org/10.1016/j.cmpb.2012.10.013>
- [13] Kim, S., Ryu, H., Hwang, J., Kim, M., An expert system approach to art psychotherapy, The Arts in Psychotherapy, 2006, 33:59–75. <https://doi.org/10.1016/j.aip.2005.07.004>
- [14] Lizarazo, J. F. S., Rincon, J. L. S., Higuera, B. M. R., Wartegg Automatic Test Qualifier, Bogota, 2008.
- [15] Geng, Z., Hu, H., and Han, Y., A Complex Network Fault Detection Method Based on Cosine Similarity and Its Application, Beijing Huagong Daxue Xuebao (Ziran Kexueban)/Journal of Beijing University of Chemical Technology (Natural Science Edition), 2017, 44(2): 87–94.
- [16] Dubey, Kumar, V., and Saxena, A. K., Cosine Similarity Based Filter Technique for Feature Selection, In ICCCCM 2016 - 2nd IEEE International Conference on Control Computing Communication and Materials, 2017.
- [17] Arora, Gaurav, and Mazumder. P., Cosine Similarity as Machine Reading Technique, In CEUR Workshop Proceedings, 2011.
- [18] Zahrotun, Lisna, Comparison Jaccard Similarity, Cosine Similarity and Combined Both of the Data Clustering With Shared Nearest Neighbor Method, Computer Engineering and Applications 5(11), 2016, 2252–4274.
- [19] Yuana, R. A., Harjunowibowo, D., Karyanta, N. A., FWAT (Fast Wartegg Analyzer Tool) for Personal Identification, Advanced Science Letters, 2015, 21:3114–3117. <https://doi.org/10.1166/asl.2015.6519>

6 Authors

Rosihan Ari Yuaa, Dewanto Harjunowibowo, Nugraha Arif Karyanta, and Cucuk Wawan Budiyanto are with Universitas Sebelas Maret, Surakarta, Indonesia. (rosihanari@staff.uns.ac.id).

Article submitted 22 August 2018. Final acceptance 06 October 2018. Final version published as submitted by the authors.