International Journal of Information System & Technology
Akreditasi No. 36/E/KPT/2019 | Vol. 5, No. 1, (2021), pp. 31-37

Detection Coronavirus using Cased-Based Reasoning with Extended Jaccard Coefficient

Murien Nugraheni

Information System and Technology, State University of Jakarta, Jakarta, Indonesia

E-mail: muriennugraheni@unj.ac.id

Abstract

Coronavirus Disease 2019 or known as COVID-19 is a new disease that can cause respiratory problems and pneumonia. This disease is caused by infection with Severe Acute Respiratory Syndrome Me Coronavirus 2 (SARS-CoV-2). Some of the clinical symptoms that appear vary, ranging from symptoms such as influenza, cough, cold, throat pain, muscle aches, headaches to those with serious complications such as pneumonia or sepsis. This research to build case-based reasoning for early detection of COVID-19 by looking at the characteristics of clinical symptoms seen in a person using the Extended Jaccard Coefficient method. The results show case-based reasoning for early detection of COVID-19 using the Extended Jaccard Coefficient method can model the level of similarity of a new case to an old case.

Keywords: *Coronavirus, Case-Based Reasoning, Extended Jaccard Coefficient*

1. Introduction

Coronaviruses are a large family of viruses that cause illness ranging from mild to severe symptoms. There are at least two types of Coronavirus that are known to cause illnesses that can cause severe symptoms such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). Coronavirus Disease 2019 (Covid-19) is a new type of disease that has never been previously identified in humans. The virus that causes COVID-19 is called SArs-CoV-2. Corona viruses are zoonotic (transmitted between animals and humans). Research states that SARS is transmitted from civet cats to humans and MERS from camels to humans. Meanwhile, the animal that is the source of COVI D-19 transmission is still unknown [1]. Lessons learned from the 2002 SARS outbreak have guided the development of COVID-19 identification and detection. Transmission electron microscopy was used to identify the morphology of the virus, genome sequencing was used to confirm the identification of the virus, and sequence data were used to help design Polymerase Chain Reaction (PCR) primers and probes [2]. Common signs and symptoms of COVID-19 infection include acute respiratory symptoms such as fever, cough and shortness of breath. The average incubation period is 5-6 days with the longest incubation period of 14 days. Severe cases of COVID-19 can because pneumonia, acute respiratory syndrome, kidney failure, and even death. Clinical signs and symptoms of having difficulty breathing, and X-rays show a large pneumonia infiltrate in both lungs [1].

From various studies, the main method of transmission of this disease is thought to be through respiratory droplets and close contact with sufferers. Droplets are small particles from a patient's mouth that can contain viral diseases, which are produced when coughing, sneezing, or speaking. The droplet can pass up to a certain distance (usually 1 meter). Droplets can stick to clothes or objects around the sufferer when coughing or sneezing. However, the droplet particles are large that not stay or settle in the air for very long. However, the community is required to wear a cloth mask that covers the nose and mouth to prevent the spread of droplets. Besides being able to infect through the nose and

ISSN : 2580-7250

mouth, this virus can infect through the eyes. So that people need to use anti-*Coronavirus* glasses. Besides that COVID-19 affects the financial life of the community and companies [3].

Research CoroNet about deep convolutional neural network model to automatically detect COVID-19 infection from chest X-ray images. The proposed model is based on exception architecture pre-trained on imagenet dataset and trained end to end on a dataset prepared by collecting COVID-19 and other chest pneumonia X-ray image from two different publically available databases [4]. Research about detection COVID-19 using artificial intelligence from total of 30 X-rays samples were selected randomly from each group (COVID-19 and healthy) for the test set, following the rule that this person (owner of this X-Ray) had not been included in the previous training or validation stages [5]. Artificial intelligence has several parts, including expert systems and case-based reasoning. Expert systems was presents for helping a group of doctors and medical specializations in diagnosing patients with different possible coronavirus disease symptoms. Doctors and medical specializations can get the diagnosis faster and more accurate than the traditional diagnosis [6]. And there research about expert system for diagnosing coronavirus using SL5 object language [7]. Case-based reasoning is computer reasoning that uses old knowledge to solve new problems. Case-based reasoning provides solutions to new cases by looking at old cases that have the closest resemblance to new cases. It would be very useful because it can eliminate the need to extract the model as required by the rules-based system. In addition, case-based reasoning can also be started from a small amount of knowledge because case-based reasoning knowledge can gradually increase when a case is added.

The underlying ideas of CBR can be applied consistently across application domains, the specific implementation of the CBR methods in particular retrieval and similarity functions is highly customized to the application at hand [8]. The process of looking for closeness or similarity between new cases and old cases to obtain solutions to new cases in CBR can use a variety of methods, where this method affect the success of CBR. One method that can be used in finding the closeness or similarity of a new case to an old case is Extended Jaccard Coefficient. This method correlates two continuous variable objects that have a linear relationship between the attributes of an object. The experiment result illustrate that Feature Graph Fusion (FGF) is robust and effective to face and object data sets in robot applications using Extended Jaccard Graph and word embedding method [9].

2. Research Methodology

2.1. Case-Based Reasoning

Case-based reasoning (CBR) is a problem-solving technique, which adopts solutions to previous problems that are similar to the new problems faced, to get a solution. CBR emphasizes problem solving based on the knowledge of previous cases, when there is a new case, it is stored in a knowledge base so that the system perform learning and knowledge possessed by the system increase. CBR can be represented as a process cycle which is divided into four sub processes [10]:

a) Retrieve which is looking for previous cases that are most similar to new case.
b) Reuse is reusing the most similar cases to get a solution for a new case.
c) Revise is making adjustments to the solution of previous cases so that they can be used as solutions for new cases.
d) Retain is to use a new solution as part of a new case, then the new case is updated on a case basis.

In Figure 1, it is explained about the CBR process stages, namely new cases are matched with existing cases in the case storage database and find one or more similar cases (retrieve). The suggested solution through case matching is then reused for a similar case, the proposed solution may be changed and adopted (revise). If new cases do

not match in the case storage database. CBR store the new cases (retain) in the knowledge database.

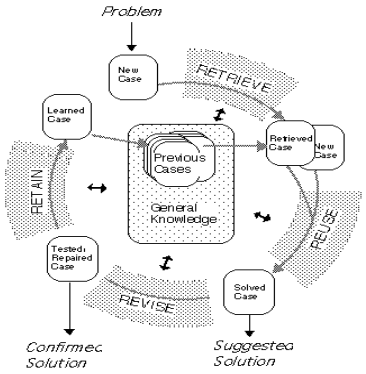

Figure 1. Case-based reasoning cycle [10]

2.2. Extended Jaccard Coefficient

Extended Jaccard Coefficient is one of the similarity methods used to calculate the similarity level of two objects. The Extended Jaccard Coefficient can be formulated as follows:

$$EJ(x,y) = \frac{x \,.\, y}{||x||^2 + ||y||^2 - x \,.\, y} \quad (1)$$

Where x is the old case and y is the new case and "|| ||" shows the norm in each vector. [11].

2.3. Coronavirus Disease 2019 (COVID-19)

Coronavirus is a positive, encapsulated, and unsegmented single-strain RNA virus. Coronavirus belongs to the order Nidovirales, family Coronaviridae. Coronaviridae are divided into two sub-families differentiated by serotype and genomic characteristics. There are four general, namely alpha coronavirus, betacoronavirus, deltacoronavirus, and gamma coronavirus [12]. Knowledge about COVID-19 infection in relation to pregnancy and the fetus is still limited and there are no specific recommendations for handling pregnant women with COVID-19. Based on this limited data and several examples of cases in the handling of the previous Coronavirus (SARS-CoV and MERS-CoV) and several cases of COVID-19, it is believed that pregnant women have a higher risk of severe disease, morbidity, and morality than the population [13].

COVID-19 infection can cause mild, moderate, or severe symptoms. The main clinical symptoms that appear are fever (temperature > 38 ° C), cough and difficulty breathing. In addition, it can be accompanied by severe shortness of breath, fatigue, myalgia, gastrointestinal symptoms such as diarrhea and other respiratory symptoms. Half of the patients developed shortness of breath within one week [12].

Table 1. Symptoms Patient Under Surveillance and People in Monitoring

Findings	**Patient Under Surveillance/ Suspect**				**People in Monitoring**	
	Criterion					
	C1	**C2**	**C3**	**C4**		
1. Fever/ history of fever	Yes	Yes	Yes	-	For either of these two points, for no other	Yes
2. Cough/ cold/ sore throat/ hard to breathe (one of choice)	Yes	-	Yes	-		

Findings	Patient Under Surveillance/ Suspect				People in Monitoring	
	Criterion					
	C1	C2	C3	C4		
					reason is clear	
3. Travel to a pandemic country area (the last 14 days)	Yes	-	-	Yes	Yes	
4. Contact with COVID-19 confirmation case	-	Yes	Yes	-		Yes
5. Pneumonia/ ISPA without cause	-	-	-	Yes		

(Source : Jurnal Penyakit Dalam Indonesia, 2020)

3. Result and Discussion

3.1. Case Representation

The cases be entered into the case base are taken from expert knowledge and literature related to diseases caused by Coronavirus. The cases that have been collected be represented in tabular form. As shown in Table 2. Contains the relationship between the name of the disease and the symptoms of the cause. So with this representation a case model can be made for the CBR system where the problem space is the symptoms of the disease and the solution space is the name of the disease.

Table 1. Shows the relationship between symptoms, classes, and solutions. In the table there are 21 cases, namely K1 to K21, with 9 symptoms, namely G1 to G9, each of which has a value of no (0) and yes (1). In addition, there are also classes categorized by KL1 is patient under surveillance, KL2 is people in monitoring, and KL3 is people without symptoms.

Table 2. Case Representation

Cases	G1	G2	G3	G4	G5	G6	G7	G8	G9	Class
K1	1	1	0	0	0	1	0	0	0	KL1
K2	1	0	1	0	0	1	0	0	0	KL1
K3	1	0	0	1	0	1	0	0	0	KL1
K4	1	0	0	0	1	1	0	0	0	KL1
K5	1	0	0	0	0	0	1	0	0	KL1
K6	1	1	0	0	0	0	1	0	0	KL1
K7	1	0	1	0	0	0	1	0	0	KL1
K8	1	0	0	1	0	0	1	0	0	KL1
K9	1	0	0	0	1	0	1	0	0	KL1
K10	0	0	0	0	0	1	0	1	0	KL1
K11	0	0	0	0	0	1	0	0	1	KL1
K12	1	0	0	0	0	1	0	0	0	KL2
K13	0	1	0	0	0	1	0	0	0	KL2
K14	0	0	1	0	0	1	0	0	0	KL2
K15	0	0	0	1	0	1	0	0	0	KL2
K16	0	0	0	0	1	1	0	0	0	KL2
K17	0	1	0	0	0	0	1	0	0	KL2

Cases	G1	G2	G3	G4	G5	G6	G7	G8	G9	Class
K18	0	0	1	0	0	0	1	0	0	KL2
K19	0	0	0	1	0	0	1	0	0	KL2
K20	0	0	0	0	1	0	1	0	0	KL2
K21	0	0	0	0	0	0	0	0	0	KL3

3.2. Acquiring New Cases Stage

Users who acts as the users of the system makes the target case or a new case is by way in experienced symptoms. Then calculation to find the value of similarity between the new case with the source case or cases in the case base. The comparison process produce a value that can determine whether or not there is a similarity between the target case and the source case, if there is a similarity, the solution of the source case be given to the user.

A case revision be carried out if there is no similarity between the target case and the source case after first confirming with the expert. After the expert confirms the revised case, the case is then stored in the case base to become new knowledge (retain). The retrieval technique used in this study is to find the closeness between new cases and old cases stored in the case base using the Extended Jaccard Coefficient as a similarity method.

3.3. New Case Diagnosis Results

The initial process in making a diagnosis is to include the symptoms feel, then will be results diagnosis appear looking similarity value with Extended Jaccard Similarity. The testing process for the early detection system for COVID-19 was carried out using a sample of case data. Then the value of this symptom is entered into a database containing all the symptoms with value: [1, 0, 0, 1, 1, 0, 0, 0, 0].

The similarity value is calculated using the Extended Jaccard Coefficient equation, where test cases or new cases are calculated the similarity value one by one from the old cases that have been stored in the case basis, can be seen in Table 3.

Table 3. The Similarity Value of Test Cases

Test	Case	Class	The Closest Old Case	Similarity Value
1	K22	PDP	K3	0,28571
2	K23	PDP	K1	0,28571
3	K24	PDP	K2	0,28571
4	K25	PDP	K3	0,28571
5	K26	PDP	K10	0,3333
6	K27	PDP	K10	0,5
7	K28	PDP	K05	0,3333
8	K29	PDP	K05	0,2
9	K30	ODP	K12	0,5
10	K31	ODP	K13	0,5

Of the 10 test cases, the similarity value for each new case was obtained by showing the old case which was the closest seen from the inputted symptoms. Although there are several that have the same similarity value, the results of the old cases stored in the case base have different results, because the system looks for the symptoms stored in the old cases that are closest to those of the test case.

4. Conclusion

The results of research Detection Coronavirus using Case-Based Reasoning *with* Extended Jaccard Coefficient, it can be concluded that:

a) Case-based reasoning can be applied to early detection of Covid-19 cases.
b) The Extended Jaccard Coefficient method can be used to find the similarity value in new cases by looking at the most similar old cases stored on a case basis.

It is hoped that in the development of Detection Coronavirus using Case-Based Reasoning *with* Extended Jaccard Coefficient, research can be carried out on other similarity methods which can produce a better similarity value.

References

[1] Kemenkes RI., *Pedoman Pencegahan dan Pengendalian Coronavirus Disease (Covid-19)*, Kementerian Kesehatan RI : Direktorat Jenderal Pencegahan dan Pengendalian Penyakit (P2P), (2020), Jakarta.

[2] Udugama, B., Kadhiresan, P., Kozlowski, H.N., Malekjahani, A., Osborne, M., Li, V.Y.C., Chen, H., Mubareka, S., Gubbay, J.B., and Chan, W.C.W., Diagnosing COVID-19 : The Disease and Tools for Detection, *American Chemical Society Nano 14*, (2020), pp.3822-3855.

[3] Rashwan, A.R.M.S and Alhelou, E.M.S., The Effect of the Covid-19 Pandemic on the Company's Suistainability under International Accounting Standards, *International Journal of Academic Accounting, Finance & Management Research*, Vol.4 No.7, (2020), pp.34-51.

[4] Khan, A.I., Shah, J.L., and Bhat, M.M., CoroNet : A Deep Neural Network for Detection and Diagnosis of COVID-19 from Chest X-Ray Images, *Elsevier Computer Methods and Programs in Biomedicine 196*, (2020), pp.1-9.

[5] Salman, F.M., Naser, S.S.A., Alajrami, E., Nasser, B.S.A. and Ashqar, B.A.M., COVID-19 Detection using Artificial Intelligence, *International Journal of Academic Engineering Research*, Vol.4 No.3, (2020), pp.18-25.

[6] Salman, F.M. and Naser, S.S.A., Expert System for COVID-19 Diagnosis, *International Journal of Academic Engineering Research*, Vol.4 No.3, (2020), pp.1-13.

[7] Almadhoun, H.R. and Naser, S.S.A., An Expert System for Diagnosing Coronavirus (COVID-19) using SL5, *International Journal of Academic Engineering Research*, Vol.4 No.4, (2020), pp.1-9.

[8] Agudo, B.D., Calero, P.A.G., Garcia, J.A.R., and Granados, A.A.S.R., Building CBR Systems with jCOLIBRI, *Elsevier Science of Computer Programming 69*, (2007), pp.68-75.

[9] Liu, S., Sun, M., Huang, X., Wang, W., and Wang, F., Featuring Fusion using Extended Jaccard Graph and Word Embedding for Robot, *Assembly Automation*, Vol.37 No.3, (2017), pp.278-284.

[10] Aamodt, A. and Plaza, E., Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approches. *AICom-Artificial Intelligence Communications, IOS Press*, Vol. 7: 1, (1994), pp.39-59.

[11] Tan, P.N., Steinbach, M., Kumar, V., *Data Mining*, Pearson Education, Inc, (2006).

[12] Perhimpunan Dokter Paru Indonesia (PDPI), *Pneumonia Covid-19 : Diagnosis dan Penatalaksanaan di Indonesia*, Perhimpunan Dokter Paru Indonesia, (2020), Jakarta.

[13] Pokja Infeksi Saluran Reproduksi Perkumpulan Obstetri Ginekologi Indonesia (POKJA ISR PP POGI), *Rekomendasi Penanganan Infeksi Virus Corona (Covid-19) pada Maternal (Hamil, Bersalin, dan Nifas)*, Pokja Infeksi Saluran Reproduksi : Perkumpulan Obstetri dan Ginekologi Indonesia, (2020), Jakarta.

[14] Susilo, A., Rumende, C.M., Pitoyo, C.W., Santoso, W.D., Yulianti, M., Herikurniawan, Sinto, R., Singh, G., Nainggolan, L., Nelwan, E.J., Chen, L.K., Widhani, A., Wijaya, E., Wicaksana, B., Maksum, M., Annisa, F., Jasirwan, C.O., and Yunihastuti, E., Coronavirus Disease 2019 : Tinjauuan Literatur Terkini, *Junal Penyakit Dalam Indonesia*, Vol.1 No.2, (2020), pp.45-67.

Author

Murien Nugraheni. Born in Yogyakarta on October 11th, 1987. Author has graduated with a Master Computer Science in Gadjah Mada University of Yogyakarta. Author is a lecture at the State University of Jakarta.