

Article

A Triangular Similarity Measure for Case Retrieval in CBR and Its Application to an Agricultural Decision Support System

Zhaoyu Zhai *, José-Fernán Martínez Ortega, Pedro Castillejo  and Victoria Beltran

Departamento de Ingeniería Telemática y Electrónica (DTE), Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación (ETSIST), Universidad Politécnica de Madrid (UPM), C/Nikola Tesla, s/n, 28031 Madrid, Spain; jf.martinez@upm.es (J.-F.M.O.); pedro.castillejo@upm.es (P.C.); mv.beltran@upm.es (V.B.)

* Correspondence: zhaoyu.zhai@upm.es

Received: 26 August 2019; Accepted: 22 October 2019; Published: 23 October 2019



Abstract: Case-based reasoning has been a widely-used approach to assist humans in making decisions through four steps: retrieve, reuse, revise, and retain. Among these steps, case retrieval plays a significant role because the rest of processes cannot proceed without successfully identifying the most similar past case beforehand. Some popular methods such as angle-based and distance-based similarity measures have been well explored for case retrieval. However, these methods may match inaccurate cases under certain extreme circumstances. Thus, a triangular similarity measure is proposed to identify commonalities between cases, overcoming the drawbacks of angle-based and distance-based measures. For verifying the effectiveness and performance of the proposed measure, case-based reasoning was applied to an agricultural decision support system for pest management and 300 new cases were used for testing purposes. Once a new pest problem is reported, its attributes are compared with historical data by the proposed triangular similarity measure. Farmers can obtain quick decision support on managing pest problems by learning from the retrieved solution of the most similar past case. The experimental result shows that the proposed measure can retrieve the most similar case with an average accuracy of 91.99% and it outperforms the other measures in the aspects of accuracy and robustness.

Keywords: triangular similarity measure; case retrieval; case-based reasoning; decision support system; sustainable agriculture

1. Introduction

Case-based reasoning (CBR) is defined as the process of solving new problems by matching and adapting cases that have been successfully managed before. The CBR approach mimics how humans would perform reasoning and learning. Thus, it seems to be a more psychologically plausible model of human reasoning [1]. This unique characteristic makes CBR a promising approach for building intelligent systems [2]. Due to its effectiveness and powerful reasoning capability, CBR has been applied to various fields such as healthcare [3–5], fault diagnosis [6–8], emergency response [9–11], and agricultural management [12–14].

The case-based reasoning approach can be formalized as four main steps: retrieve, reuse, revise, and retain [15]. A generic workflow of CBR is presented in Figure 1.

In Figure 1, the problem statements are transformed into a new case after data pre-processing. The retrieval process compares the new case with historical ones for the purpose of identifying the most similar past case. The solution of this retrieved case is reused to solve the new case. In order to perfectly fit the new case, a revision process is required to update the proposed solution. After

applying the updated solution, the solved case will be stored in the case base for further comparisons. Among these four steps, it is worth noting that the case retrieval process plays a key role because the rest of processes cannot proceed without successfully matching the most similar past case.

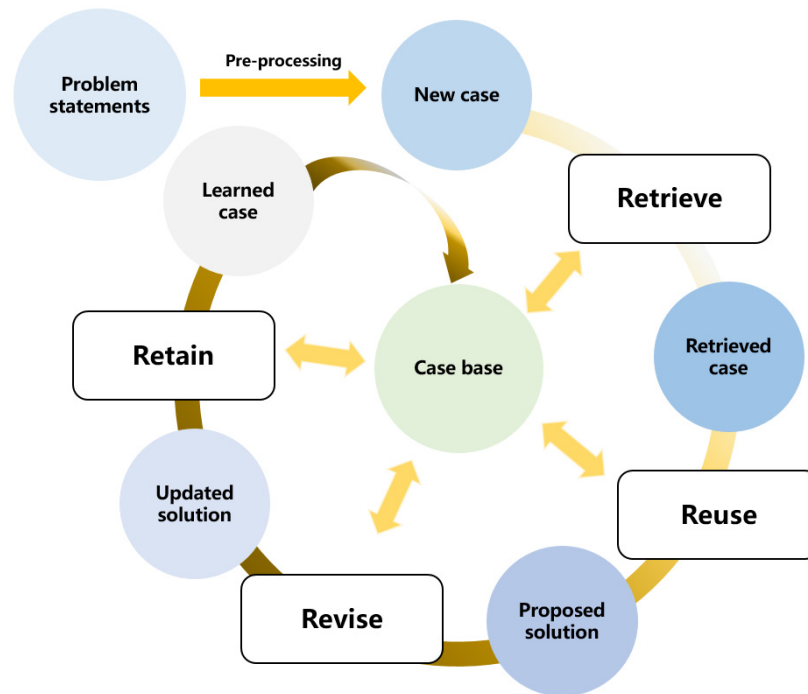


Figure 1. A generic workflow of case-based reasoning.

For case retrieval, many similarity measuring methods [16] have been used to evaluate the commonalities between cases, including angle-based and distance-based measures [17,18]. On the one hand, as one of the representative angle-based measures, cosine similarity measure [19] compares two non-zero vectors and calculates the cosine angle between them. It judges the similarity through their orientation, not magnitudes. A smaller cosine angle indicates that compared vectors are more similar. On the other hand, researchers have made tremendous contributions to distance-based similarity measures, especially the Euclidean distance [20], the Manhattan distance [21], the Chebyshev distance [22], and the Hamming distance [23]. Generally, the smaller the distance is, the more commonalities the compared vectors will have.

However, angle-based and distance-based measures may retrieve imprecise cases under certain extreme circumstances [24,25]. Thus, a novel triangular similarity measure (TSM) is presented in this paper to evaluate the commonalities between cases. The proposed measure not only considers the magnitude of vectors, but also their distances and magnitude differences for generating more accurate results. A case study on a CBR-based agricultural decision support system (ADSS) for pest management is explored to prove the effectiveness and robustness of the proposed measure. The triangular similarity measure is employed to retrieve similar past cases for providing farmers with quick decision support on agricultural activities.

The rest of this paper is organized as follows. In Section 2, related works on CBR-based ADSSs, angle-based and distance-based similarity measures are reviewed. According to detected drawbacks of current measures, the triangular similarity measure is explained in Section 3. A case study on applying the proposed measure to a CBR-based agricultural decision support system is explored in Section 4. Experimental results are presented in Section 5. Lastly, conclusions are drawn in Section 6.

2. Related Work

In this section, research works on applying the case-based reasoning approach to agricultural decision support systems are reviewed. It is demonstrated that how ADSSs could benefit from employing the CBR approach. Meanwhile, current states of angle-based and distance-based similarity measures are investigated for the purpose of detecting their drawbacks.

2.1. CBR Applications for Agricultural Decision Support Systems

An agricultural decision support system [26,27] can be defined as a human–computer system with the ability to utilize data from various resources such as climate, land use, and human labor. It aims at smoothening the decision-making processes and providing farmers with useful suggestions on agricultural activities, such as fertilization, irrigation, and pest management. With the help of sensing technologies and reasoning approaches, ADSSs are able to gather adequate data (e.g., sensor measurements, meteorological information, infrared images) and generate evidence-based knowledge.

As one of the most powerful reasoning approaches, CBR has been adopted by researchers in the aspect of agriculture for decades. Padma et al. [28] presented an intelligent decision support system for managing pest problems in an apple orchard. A hybrid algorithm was developed to optimize pest and disease protection decision-making processes by integrating case-based reasoning and database technologies. The nearest neighbor algorithm was used to determine the similarity between attributes of cases. Experimental results show that the accuracy of decision-making processes achieved 90.20%, thus the proposed ADSS can provide significant decision support for farmers in decision-making towards eco-friendly pest management practices. Evans et al. [14] mentioned that CBR is able to capture both experiential and expert knowledge to provide farmers with suggestions for potential action on agricultural practices. The advantage of adopting CBR in ADSSs is that domain experts do not need to ask a software engineer to update the knowledge base because step of retention in CBR can maintain the knowledge base on its own. Le Ber et al. [29] introduced a prototype of CBR and applied it to forecast miscanthus allocations. For case retrieval, a distance-based measure is used to compare the common attributes between the source and target problems. Then, the process of transformational adaptation [30] is conducted, starting from the solution of the selected source case and modifying it with respect to the differences between source and target cases. Sharaf-Eldeen et al. [31] developed a new case-based reasoning approach and improved the accuracy of case retrieval by range adaptation rules. The similarity between cases is calculated by a distance formula. The dissimilarity between attributes are used to generate the antecedent part of transformational adaptation rules while the differences between solutions of the compared cases become the consequent part of rules. The proposed approach was verified by a plant classification example. Over one hundred cases were considered in the case base and experimental results showed that the developed prototype achieved great accuracy in plant classification.

From the above research works, we conclude that the case-based reasoning approach has great potential in agricultural decision support systems. It is worth noting that case retrieval plays a key role in CBR. However, the similarity evaluation in the step of case retrieval remains to be further improved.

2.2. Angle-based Measures for Case Retrieval

Cosine similarity calculates the cosine angle between two non-zero vectors. It is widely used in case retrieval for case-based reasoning. The similarity of images, documents, and numeric values can be measured by the cosine similarity because all these data can be represented by vectors. The simplicity and effectiveness of cosine similarity measure has been acknowledged by the research community [32]. The cosine similarity of two vectors can be formularized in Equation (1) [33].

$$\text{Sim}(x_i, x'_i) = \cos(\theta) = \frac{\sum_{i=1}^n x_i * x'_i}{\sqrt{\sum_{i=1}^n x_i^2} * \sqrt{\sum_{i=1}^n x'^2_i}} \quad (1)$$

where θ is the angle between vectors x and x' . x_i and x'_i represents the i th attribute in the vectors x and x' , respectively. The similarity between these two vectors will increase as $\cos(\theta)$ increases.

Hassanien et al. [34] presented an automatic CBR based system for assessing water quality according to microscopic images of fish gills and livers. The cosine-based measure was used to find out a small number of cases from the case base with the highest similarity to the query. Compared with other similarity measures such as the Euclidean distance, the Canberra distance, and the squared chord distance, the proposed system achieved water quality prediction accuracy of 97.9% when using the cosine-based measure. More applications of the Cosine similarity measure are found in literature [35,36].

2.3. Distance-Based Measures for Case Retrieval

Distance-based measures include the Euclidean distance, the Manhattan distance, the Chebyshev distance, and the Hamming distance. These similarity measures have various applications, such as information retrieval, classification, clustering and so on.

The Euclidean distance, also known as the Euclidean metric, calculates the straight-line distance between two points in Euclidean space. It has been employed to measure the similarity of numeric, interval, fuzzy data, etc. The formula of Euclidean distance is defined in Equation (2) as follows.

$$Dist(x, x') = Dist(x', x) = \sqrt{\sum_{i=1}^n (x_i - x'_i)^2} \quad (2)$$

where $Dist()$ is the distance between compared two vectors, while x_i and x'_i represent the i th attribute in the vector x and x' , respectively. The similarity of these two vectors will increase as $Dist()$ decreases. The Euclidean distance similarity measure is adopted in literature [37,38].

The Manhattan distance between two vectors in N-dimensional space is the sum of the lengths of the line segments between two points onto the coordinate axes. Similar to the Euclidean distance, the Manhattan distance is also applicable for measuring the similarity of numeric, interval, fuzzy data, etc. The formula of Manhattan distance is defined in Equation (3) as follows.

$$Dist(x, x') = Dist(x', x) = \sum_{i=1}^n |x_i - x'_i| \quad (3)$$

where $Dist()$ is the distance between compared two vectors, while x_i and x'_i represent the i th attribute in the vectors x and x' , respectively. The similarity of these two vectors will increase as $Dist()$ decreases. The Manhattan distance similarity measure is employed in literature [39,40].

The Chebyshev distance is defined as the greatest difference between two vectors among any coordinate dimensions. Its application is found in measuring the similarity of images, fuzzy sets, interval data, and so on. The formula of Chebyshev distance is defined in Equation (4) as follows.

$$Dist(x, x') = Dist(x', x) = \max(|x_i - x'_i|) \quad (4)$$

where $Dist()$ is the distance between compared two vectors, while x_i and x'_i represent the i th attribute in the vectors x and x' , respectively. The similarity of these two vectors will increase as $Dist()$ decreases. Rashid [41] and Mousa and Yusof [42] used the Chebyshev distance for similarity measurement.

The Hamming distance is usually used to measure the similarity of strings. The length of strings should be equal. Hamming distance counts the minimum number of dis-matches between two strings. The formula of Hamming distance is defined in Equation (5) as follows.

$$Dist(x, x') = Dist(x', x) = \sum_{i=1}^n (x_i \oplus x'_i) \quad (5)$$

where $Dist()$ is the distance between compared two vectors, while x_i and x'_i represent the i th attribute in the vectors x and x' , respectively. The symbol \oplus represents the XOR operator. The similarity of

these two vectors will increase as $Dist()$ decreases. Applications of the Hamming distance similarity measure is detected in literature [43,44].

2.4. Drawbacks of Angle-Based and Distance-Based Measures

Though angle-based and distance-based measures are efficient and effective, their drawbacks cannot be ignored [45–47]. These measures may generate inaccurate results under certain extreme circumstances. For example, cosine similarity does not take the magnitudes of vectors into consideration, thus it may have difficulties when meeting the following situation shown in Figure 2.

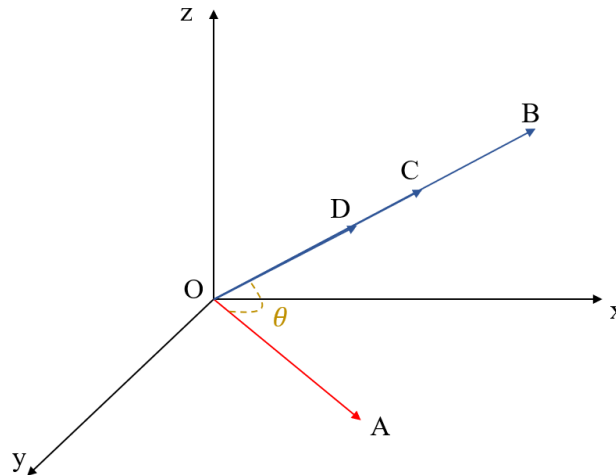


Figure 2. An extreme situation when measuring similarity with cosine similarity.

In Figure 2, the target vector is \vec{OA} and it is compared with vectors \vec{OB} , \vec{OC} , and \vec{OD} . On the one hand, though the angle between them is identical, there exists huge differences between their magnitudes. However, due to the same cosine similarity, vector \vec{OA} is considered similar to vectors \vec{OB} , \vec{OC} , and \vec{OD} . On the other hand, cosine similarity between \vec{OB} , \vec{OC} , and \vec{OD} all equals to one, meaning that these vectors are completely similar with each other. But the facts tell a different story. It is obvious that these vectors are not the same due to their magnitude differences. Thus, it is concluded that cosine similarity may match inaccurate cases in the step of case retrieval in CBR.

In terms of distance-based measures, the Euclidean distance also has certain drawbacks [48–50]. It may have difficulties when the distances between several vectors are the same. This situation is shown in Figure 3.

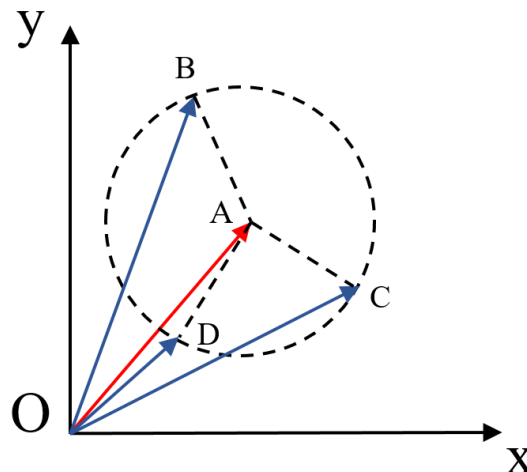


Figure 3. An extreme situation when measuring similarity with Euclidean distance.

In Figure 3, the target vector is \vec{OA} and it is compared with vectors \vec{OB} , \vec{OC} , and \vec{OD} . Though all these vectors have different magnitudes and orientations, the distance between them is identical, indicating that vectors \vec{OB} , \vec{OC} , and \vec{OD} are all similar to vector \vec{OA} . Meanwhile, it is acknowledged that each attribute in an individual vector may offer different contributions. However, the Euclidean distance treats all involved attributes equally. Thus, we can conclude that the Euclidean distance can be further improved to retrieve more accurate cases in CBR.

For the Manhattan distance and the Chebyshev distance, they both have the same drawbacks as the Euclidean distance [51–56]. When the distances between vectors are the same, these measures can hardly tell which vector is more similar to the target.

In conclusion, current similarity measures may be insufficient in the step of case retrieval for case-based reasoning. It is urgent to develop new similarity measures to improve retrieval accuracy and efficiency. Thus, a triangular similarity measure is proposed in this paper and explained in the next section.

3. Triangular Similarity Measure

After having a look in Figures 2 and 3, a triangular similarity measure is proposed to overcome the drawbacks of angle-based and distance-based measures. The proposed measure takes the angle between two vectors and their magnitudes. When evaluating the similarity, a triangle is formed by two compared vectors. The area of this triangle is considered as a similarity metric. Meanwhile, for enhancing the robustness of TSM, a coefficient is designed in the formula. It considers the magnitude differences between two compared vectors. The smaller the triangular area is, the more similar the two vectors will be. The formula of TSM is explained in the next sub-sections in detail.

3.1. Formular of Triangular Similarity Measure

For two vectors in three-dimensional space, $\vec{OA} = (a_1, a_2, a_3)$ and $\vec{OB} = (b_1, b_2, b_3)$, they can form a triangle as follows in Figure 4.

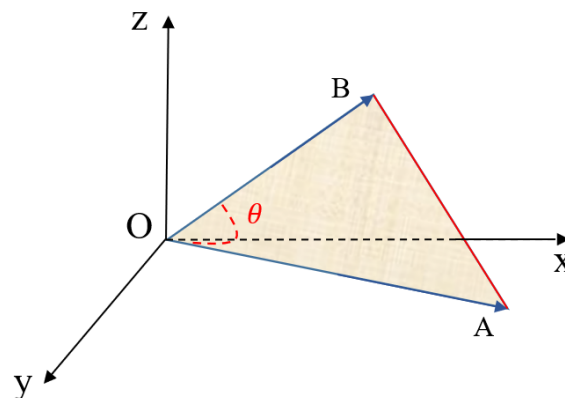


Figure 4. Two vectors \vec{OA} and \vec{OB} in three-dimensional space.

In Figure 4, a triangle $\triangle AOB$ is formed by two vectors \vec{OA} and \vec{OB} . The area of this triangle can be calculated by two sides and the included angle (SAS formula). Firstly, the magnitudes of the two vectors are calculated as follows.

$$|\vec{OA}| = \sqrt{a_1^2 + a_2^2 + a_3^2} \quad (6)$$

$$|\vec{OB}| = \sqrt{b_1^2 + b_2^2 + b_3^2} \quad (7)$$

The cosine value of these two vectors is calculated by Equation (1). For obtaining the included angle, the trigonometric inverse function $\cos^{-1} \theta$ can be employed. Then, the area of this triangle is calculated by the SAS formula as follows.

$$S_{\Delta AOB} = \frac{|\vec{OA}| * |\vec{OB}| * \sin(\theta)}{2} \quad (8)$$

By employing TSM, we can obtain that the area of the triangle ΔAOD is much smaller than the triangle ΔAOC in Figure 3. Under this circumstance, it is concluded that vectors \vec{OA} and \vec{OD} have more commonalities than vectors \vec{OA} and \vec{OC} , which is the expected result from common sense. Compared with distance-based measures, TSM is able to identify the most similar vector if when the distance between several vectors is the same. Thus, the proposed triangular similarity measure overcomes the drawbacks of distance-based measures.

3.2. Coefficient Design in Triangular Similarity Measure

The robustness of TSM can be further enhanced for overcoming the drawbacks of angle-based similarity measures. In order to deal with the situation shown in Figure 2, a coefficient is designed to improve the similarity precision. We call it K_{mag} because this coefficient mainly considers the magnitude difference between compared vectors. This coefficient is also formed by a triangle and its value is decided by calculating the area of the triangle. The coefficient for vectors \vec{OA} and \vec{OB} is shown in Figure 5.

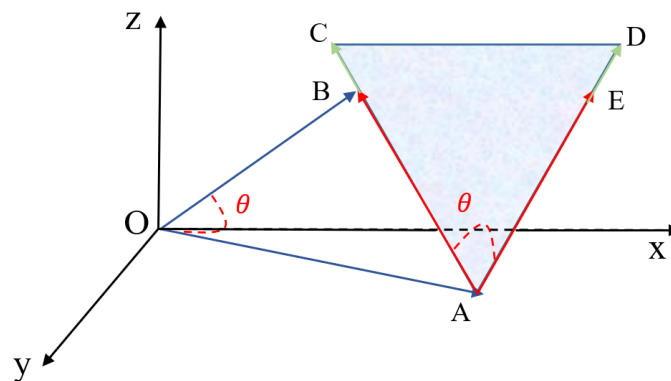


Figure 5. The coefficient triangle for vectors \vec{OA} and \vec{OB} .

As shown in Figure 5, the triangle ΔCAD is formed and its area represents the value of the coefficient for vectors \vec{OA} and \vec{OB} . The angle ($\angle CAD$) between \vec{AC} and \vec{AD} is the same as the angle $\angle AOB$. The magnitude of $|\vec{AB}|$ and $|\vec{AE}|$ equals to the Euclidean distance between vectors \vec{OA} and \vec{OB} , while the magnitude of $|\vec{BC}|$ and $|\vec{ED}|$ equals to the magnitude difference between vectors \vec{OA} and \vec{OB} . The magnitude of $|\vec{AC}|$ and $|\vec{AD}|$ is the sum of the Euclidean distance and the magnitude difference. The formula of $|\vec{AC}|$ and $|\vec{AD}|$ can be obtained by the following equation.

$$|\vec{AC}| = |\vec{AD}| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2} + \left| \|\vec{OA}\| - \|\vec{OB}\| \right| \quad (9)$$

The first part of Equation (9) is the Euclidean distance between vectors \vec{OA} and \vec{OB} , while the remaining part represents the magnitude difference between these two vectors. We detect that the Euclidean distance and the magnitude difference have a close relationship. When the magnitude

difference decreases, so does the Euclidean distance. Thus, both factors are taken into consideration when designing the coefficient. The area of the triangle $\triangle CAD$ is also calculated by the SAS formula.

$$K_{mag} = S_{\triangle CAD} = \frac{|\vec{AC}| * |\vec{AD}| * \sin(\theta)}{2} \quad (10)$$

With obtained areas of triangles $\triangle AOB$ and $\triangle CAD$, the similarity between vectors \vec{OA} and \vec{OB} can be measured by the following equation.

$$\text{Sim}(\vec{OA}, \vec{OB}) = K_{mag} * S_{\triangle AOB} \quad (11)$$

After multiplying the designed coefficient, TSM is complete and it is able to overcome the drawbacks of both angle-based and distance-based similarity measures.

For two vectors in N-dimensional space, $\vec{OA} = (a_1, a_2, \dots, a_n)$ and $\vec{OB} = (b_1, b_2, \dots, b_n)$, their similarity can be measured by the TSM formula defined as follows.

$$\text{Sim}(\vec{OA}, \vec{OB}) = \frac{\left(\sqrt{\sum_{i=1}^n (a_i - b_i)^2} + \sqrt{\sum_{i=1}^n a_i^2} - \sqrt{\sum_{i=1}^n b_i^2} \right)^2 * [\sin^2(\theta) + 0.001] * \sqrt{\sum_{i=1}^n a_i^2} * \sqrt{\sum_{i=1}^n b_i^2}}{4} \quad (12)$$

In Equation (12), the $\sin(\theta)$ is added by 0.001 for the following reason. When the compared two vectors are overlapped with each other, $\text{Sim}(\vec{OA}, \vec{OB})$ will equal to zero. As a consequence, the final result will be zero without this adjustment.

However, it may be difficult for human to understand the similarity by directly reading the triangle areas. It is better to adopt the inverse exponential function [57] to convert the area values into percentile values, ranging from 0 to 100%. Because the value presented by percentage can fit human notions better. This function is defined as follows.

$$\text{Sim}(\vec{OA}, \vec{OB})_{[0,100\%]} = \exp(-\text{Sim}(\vec{OA}, \vec{OB})) \quad (13)$$

A large value of $\text{Sim}(\vec{OA}, \vec{OB})_{[0,100\%]}$ means that vectors \vec{OA} and \vec{OB} are more similar. Only if vectors \vec{OA} and \vec{OB} are absolutely identical, the value of $\text{Sim}(\vec{OA}, \vec{OB})_{[0,100\%]}$ equals to one hundred percent. Because the value of $\text{Sim}(\vec{OA}, \vec{OB})$ equals to zero, indicating that the Euclidean distance between vectors and their magnitude differences are both zero.

4. Case Study: A CBR-Based Agricultural Decision Support System

In the previous section, the triangular similarity measure is explained in detail and it can be used for case retrieval in a CBR-based system. For verifying the effectiveness of the proposed TSM, we conduct a case study on applying the case-based reasoning approach to an agricultural decision support system. This ADSS aims at assisting farmers in managing pest problems under an eco-friendly manner. With the reduced chemical usages, soil fields can be greatly preserved, while applied chemicals can be precisely treated to infected and diseased crops.

The case study covers the following three parts: (1) problem statements for pest managements, (2) ADSS architecture, and (3) the process of retrieving similar past cases by the proposed TSM.

4.1. Problem Statements for Pest Management

The occurrence of pests has a close relationship with several factors, such as crop status, soil conditions, and meteorological information [58]. When managing pest problems, it is necessary to take pest, crop, and environment data into account. Thus, pest problems are defined by these three categories of attributes.

Firstly, different pest problems usually require specific chemical treatments. For example, rice growers may apply VIRTAKO or Prevathon for removing the eggs of stem borers [59]. Meanwhile, pest quantity has a great influence on applied dosages of chemicals. Therefore, it is necessary to obtain pest details beforehand.

Secondly, crop data are also essential. For example, the planting density of crops may affect the dilution concentration of chemicals [60]. Meanwhile, crops are sensitive to applied chemicals due to their growth stages. Toxic chemicals may be too strong, even lethal to those crops which are at the seed stage. Therefore, crops details are also important for the process of decision-making.

Lastly, environment variables play a significant role in pest management. The performance of applying chemical treatments may be easily affected by environmental changes. For example, the performance will decrease when it is raining. Moreover, it is not possible to use unmanned aerial vehicles (UAVs) for spraying the chemicals when the wind is too strong [61].

Conclusively, pest, crop, and environment variables are all considered in pest management in Table 1.

Table 1. List of attributes considered in pest management [62].

Category	Attribute Name	Attribute Unit	Content
Pest	Pest type	/	Name of pests
	Pest quantity	Unit	Number of pests
	Pest stage	/	Life cycle of pests
	Infected area	m ²	Area of infected districts
Crop	Crop name	/	Name of crops
	Growth stage	/	Life cycle of crops
	Planting density	Seeds/m ²	Seeds density in an area
	Temperature	Celsius degree	Minimum and maximum temperature
Environment	Humidity	/	Amount of water vapor in air
	Rainfall	/	Possibility of rainfall
	Sunlight	Lux	Daylight during the daytime
	Wind speed	Degree	Wind flow velocity

Since the case-based reasoning approach is applied to the agricultural decision support system, problem statements of pest management should be represented by cases. In this paper, we adopt the feature-value (attribute-value) pair representation to define each case [63]. Meanwhile, each past case has its own solution. An example of the pair representation is shown in Figure 6 as follows.

```

<Case>
  <Case_ID>1</Case_ID>

  <Features>
    <Pest_type datatype="int">1</Pest_type>
    <Pest_quantity datatype="int">200</Pest_quantity>
    <Pest_stage datatype="int">2</Pest_stage>
    <Infected_area datatype="int">50</Infected_area>
    <Crop_name datatype="int">2</Crop_name>
    ...
    ...
  </Features>

  <Solution_ID>1</Solution_ID>
  <Solution>S</Solution>

</Case>

```

Figure 6. An example of the pair representation in an agricultural decision support system (ADSS).

4.2. Agricultural Decision Support System Overview

This agricultural decision support system aims at assisting farmers in managing pest problems. After a new pest problem is reported by farmers, the ADSS tries to identify the most similar past case from the case base and reuse the retrieved solution for resolving the new case. The adopted framework of ADSS [64] is shown in Figure 7.

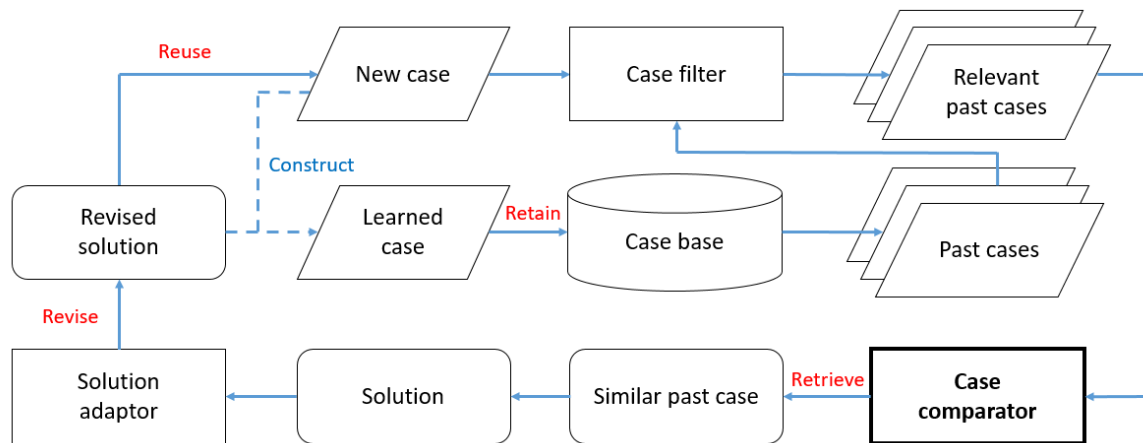


Figure 7. Framework of case-based reasoning (CBR)-based ADSS.

In Figure 7, the ADSS consists of three main components: a case filter, a case comparator, and a solution adaptor. Firstly, the case filter is used to remove irrelevant past cases. As stated in Section 4.1, each past case contains information about pests and crops. The pest type and crop name are selected as identifiers. For example, a new case is in regards to managing a rice planthopper problem for rice. Under this circumstance, the pest type is “rice planthopper” and the crop name is “rice”. The case filter component selects past cases from the case base which match these two identifiers. The process of case filtering enables the ADSS to improve the efficiency of similarity measurements because the number of past cases selected for comparisons is reduced. Secondly, the new case and relevant past cases are treated as input to the case comparator, which is the core of the ADSS, because steps of revision, reuse, and retention cannot be further processed without successful case retrieval. The case comparator component employs the proposed triangular similarity measure to match the most similar past case. Thirdly, the solution adaptor component is used to revise the retrieved solution for fitting the situation of the new case better. Lastly, learned cases are constructed by problem statements of the new case and the revised solution. These data are retained in the case base for further comparisons.

4.3. Case Retrieval in ADSS by TSM

The process of case retrieval in the CBR-based ADSS is performed by the case comparator component. After case filtering, relevant cases are obtained. Since feature-value pair representations are employed, each case can be considered as a vector in N-dimensional space. The new case is compared with past cases one by one, using Equation (12). After computing areas of all formed triangles, these area values are standardized and converted into percentile values by Equation (13). The past case with the highest percentile value is chosen as the most similar one and its solution is retrieved for further revision. It is worth mentioning that each attribute of cases is equally weighted [65], indicating that each attribute has the same priority.

5. Experiments and Discussion

After introducing the problem statements of pest management and the ADSS overview, the proposed triangular similarity measure is verified in this section. Firstly, the experiment settings are presented. In total, one thousand past cases were stored in the case base and 300 new cases were

prepared based on three categories of attributes defined in Section 4.1 for testing purposes. Then, we analysed the experimental results to verify whether the proposed TSM could retrieve the most similar past case or not. Lastly, the accuracy of TSM was compared with typical Euclidean distance and cosine similarity measures for demonstrating its effectiveness and superiority.

5.1. Experiment Settings

As mentioned in the previous section, this case study emphasises on applying the case-based reasoning approach to an agricultural decision support system for pest management. Two pests are considered in the case study: the rice planthopper (RP) [66] and the Chilo suppressalis (CS) [67]. In terms of crops, we focus mainly on managing pest problems for rice. In total, one thousand past cases were stored in the case base, half for RP and half for CS. The case base can be found at the following link: <https://github.com/ZhaoyuZHAI/Case-base/blob/master/past%20case>. Meanwhile, 300 new cases were designed for test purposes, 150 for RP and 150 for CS. Some data of new cases are given in Table 2. Each new case is compared with past cases, respectively. For now, we play with simulated data. The pest, crop, and environment data are randomly generated within a reasonable interval. For example, the attribute of crop planting density is generated from 180–525 seeds/m² [68]. Since TSM is developed within a European project, entitled Aggregate Farming in the Cloud (AFarCloud), we expect to receive data from real fields as soon as the sensor deployment is complete.

Table 2. Some data of new cases for test purposes.

New Case ID.	1	2	3	4	5	151	152	153	154	155
Pest type	RP	RP	RP	RP	RP	CS	CS	CS	CS	CS
Pest quantity	335	320	384	390	323	309	356	334	458	345
Pest stage	Pupae	Adult	Larvae	Larvae	Pupae	Larvae	Adult	Egg	Larvae	Egg
Infected area	77	56	87	109	76	75	146	168	95	157
Crop name	Rice	Rice	Rice	Rice	Rice	Rice	Rice	Rice	Rice	Rice
Growth stage	Repro	Ripen	Vege	Ripen	Vege	Repro	Ripen	Repro	Repro	Repro
Planting density	473	272	479	367	345	358	285	274	217	355
Temperature	[21,30]	[16,30]	[15,29]	[21,32]	[16,32]	[16,29]	[20,25]	[18,32]	[22,32]	[17,24]
Humidity	57.49%	60.27%	87.06%	35.22%	25.73%	36.04%	65.68%	43.35%	31.02%	46.05%
Rainfall	72.15%	37.39%	17.34%	82.90%	51.26%	12.50%	61.93%	25.79%	80.42%	14.48%
Sunlight	4830	831	3338	3273	1888	4146	3972	2680	995	4742
Wind speed	2	4	3	1	2	7	1	8	3	5

5.2. Experimental Results

The desired output from the case comparator component is the ID number of the most similar past case and its corresponding similarity compared with the new case. The experimental result is shown in Table 3.

Table 3. Some retrieved similar past cases and their corresponding similarity values.

New Case ID	1	2	3	4	5	151	152	153	154	155
Past case ID	148	412	378	387	38	611	700	918	580	672
Similarity	88.66%	91.82%	91.43%	89.92%	93.68%	89.97%	88.73%	89.72%	91.40%	94.01%

As shown in Table 3, new cases are matched with the most similar past cases, respectively. The attribute visualization of compared new and retrieved past cases in Table 3 is given in Appendix A at the end of this manuscript. From the visualization, it is determined that the attributes of retrieved past cases have great commonalities with the new cases correspondingly. Thus, the retrieved cases are considered similar to the new cases. The retrieved solutions of past cases can be used to resolve the new cases.

As 300 new cases are tested, 300 past cases are retrieved correspondingly. The mean similarity value of retrieved cases achieves 91.99%. The maximum and minimum similarity value of retrieved cases is 99.67% and 79.20%, respectively, as shown in Figure 8.

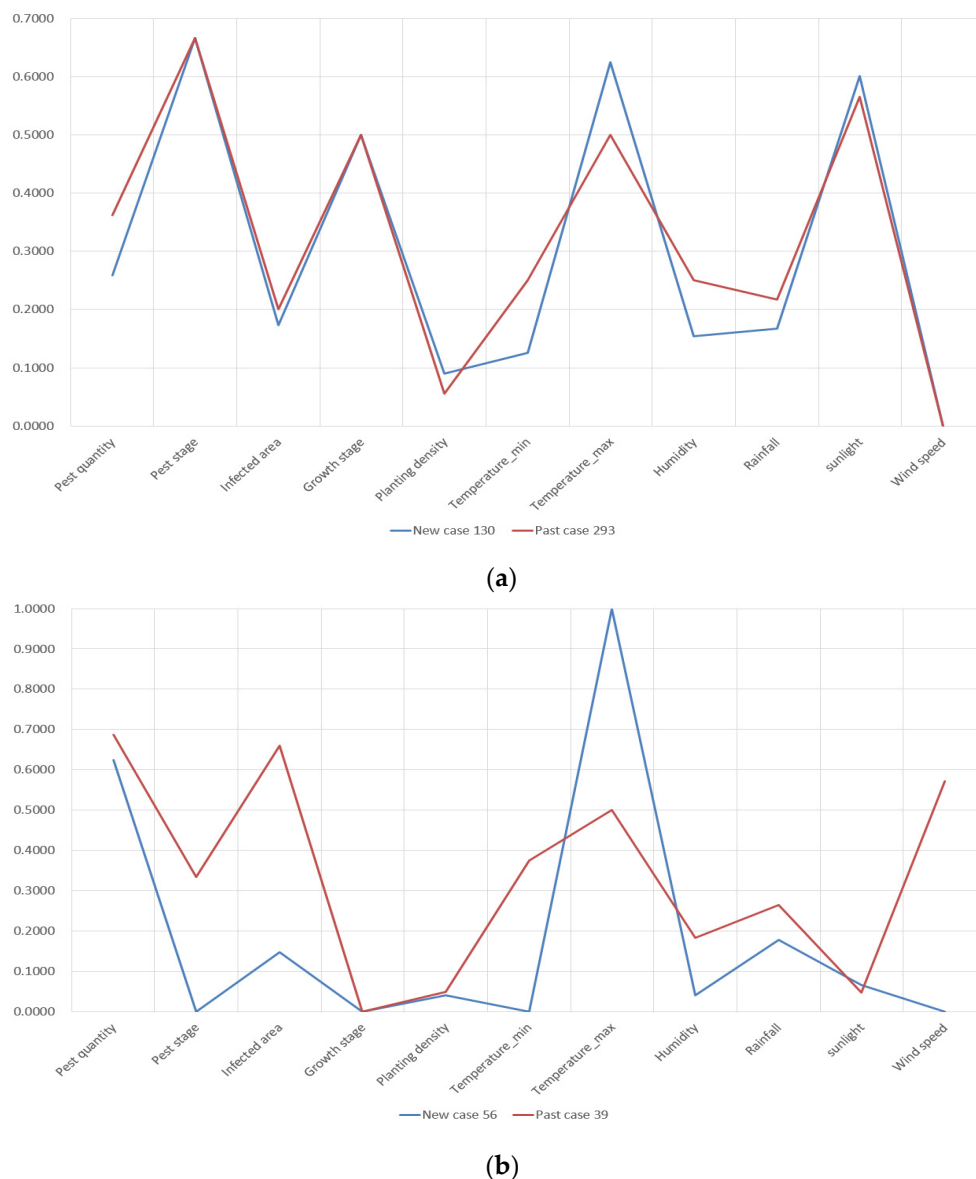


Figure 8. The similarity of retrieved cases: (a) the best situation; (b) the worst situation.

In Figure 8a, we can see that each attribute of New Case 130 and Past Case 293 matches each other with minor deviations. Thus, these two cases are considered highly similar. The retrieved solution of Past Case 293 has great potential in resolving the situation of New Case 130. In Figure 8b, the retrieved Past Case 39 does not match New Case 56 very well. The reason for this poor performance is that the case base contains only 1000 cases and it does not cover a situation like New Case 56. However, it is promising for the CBR-based ADSS to achieve an improvement because the case base will include more cases with time. As a consequence, future cases can have a greater chance for identifying similar ones in the case base and retrieving an effective solution.

By employing TSM, the most similar past case can be precisely retrieved. The pest, crop, and environment data of the retrieved past case have great commonalities with the new case. The successful case retrieval establishes a good foundation for the rest of processes in case-based reasoning. Meanwhile, the retrieved solutions of past cases are indeed useful guidance for farmers to manage

pest problems. Farmers can eliminate pests by referring to solutions of past cases which have been successfully managed before.

The main focus of this paper lies on proposing a novel triangular similarity measure for case retrieval in case-based reasoning. Therefore, the rest of processes like revision is not considered in this paper.

5.3. Comparisons with Typical Similarity Measures

In this section, the proposed triangular similarity measure is compared with two typical similarity measures: the Euclidean distance (ED) and cosine similarity. The experiment settings remain the same. However, instead of using Equation (12), the similarity between compared vectors is measured by Equation (1) and (2), respectively. Each attribute is also equally weighted. The inverse exponential function is still used in these two measures for presenting the final results. We select retrieval similarity as the main evaluation criteria for this comparison. The result of all 300 retrieval tasks is shown in Figure 9, while the comparative result of selected cases is shown in Table 4 as follows.

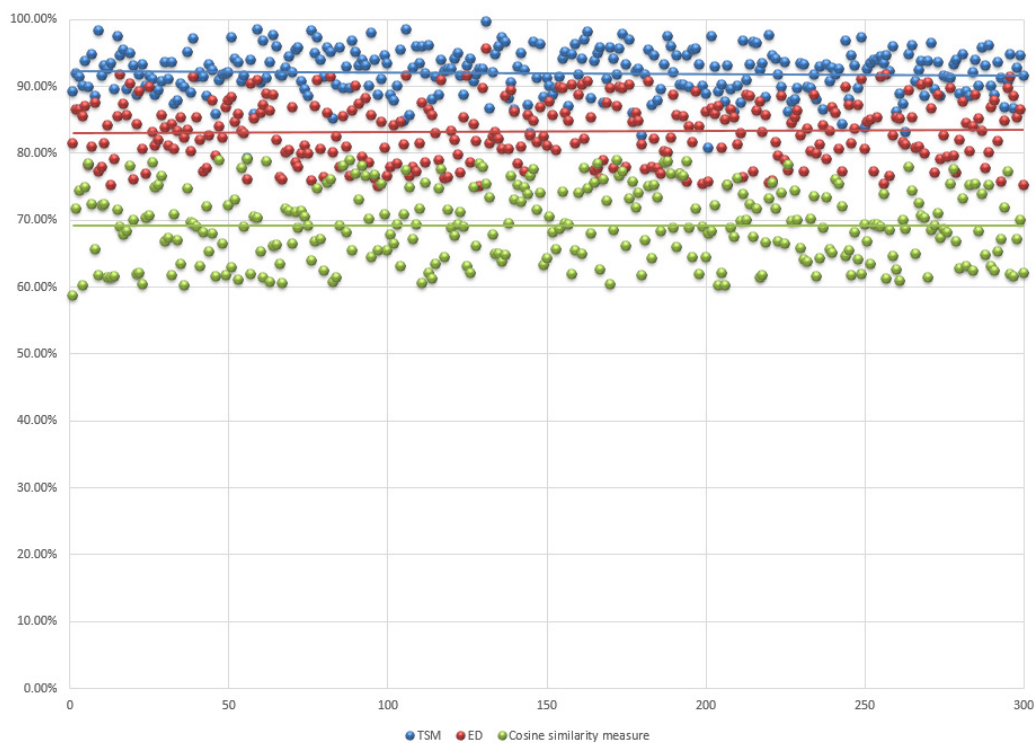


Figure 9. The result of all 300 retrieval tasks by TSM, ED, and cosine similarity measure.

Table 4. Comparative result of three similarity measures in selected cases.

New Case ID	TSM	Euclidean Distance	Cosine Similarity
1	148 (88.66%)	148 (81.84%)	237 (58.74%)
2	412 (91.82%)	412 (86.48%)	89 (71.57%)
3	378 (91.43%)	378 (86.41%)	253 (74.34%)
4	387 (89.92%)	387 (85.43%)	106 (60.16%)
5	38 (93.68%)	38 (86.89%)	183 (74.76%)
151	611 (89.97%)	983 (82.74%)	849 (70.53%)
152	700 (88.73%)	700 (85.63%)	732 (68.09%)
153	918 (89.72%)	918 (84.44%)	755 (65.62%)
154	580 (91.40%)	568 (84.48%)	772 (68.65%)
155	672 (94.01%)	672 (85.38%)	421 (74.23%)

In Figure 9, blue, red, and green points represent the similarity value of past cases which are retrieved by TSM, ED, and cosine similarity measures, respectively. The average similarity value of these three measures is 91.99%, 83.20%, and 69.21%, respectively. From this point of view, we can conclude that TSM can retrieve the most similar case with a greater accuracy. For further demonstrating the accuracy of retrieved past case, a particular new case will be presented later which all three measures retrieved different past cases.

In Table 4, the result is presented by “Past case ID (similarity value)”. Under the column of TSM, it shows that TSM is able to retrieve the most similar past cases with the highest accuracy. In terms of the Euclidean distance measure, it fails to match the most similar cases for New Cases 151 and 154. With regards to the cosine similarity measure, it obviously fails in all cases. Conclusively, the retrieval similarity of TSM outperforms both typical similarity measures. TSM achieves the highest performance in case retrieval for all ten test cases. Though the Euclidean distance measure is able to achieve the same similar past cases as TSM, its similarity accuracy is worse than TSM. For the cosine similarity measure, it is not suitable for such case retrieval tasks. Moreover, in Section 3, we have proved that the robustness of TSM is better than the Euclidean distance and cosine similarity measures. TSM can always provide correct results even in extreme situations.

For demonstrating the similarity accuracy of these measures, we select New Case 151 as an example. The standardized attributes of pest quantity, pest stage, infected area, growth stage of crops, planting density, temperature, humidity, rainfall, sunlight, and wind speed of the new case and retrieved past cases are presented by line charts in Figure 10.

In Figure 10, the past case retrieved by TSM is obviously more similar with the compared case than those retrieved by the ED and cosine similarity measures. Because the attribute deviation is smaller and the data trending between New Case 151 and Past Case 611 is more compliant. In terms of Past Case 983, its attributes of infected area, the minimum temperature, and humidity do not match New Case 151 perfectly. With regards to Past Case 849, retrieved by the cosine similarity measure, its attributes have a major difference from the target case.

Furthermore, we have compared the average precision of all these three similarity measures by Equation (14) [69].

$$\text{Average precision} = \frac{TP}{TP + FP} (\%) \quad (14)$$

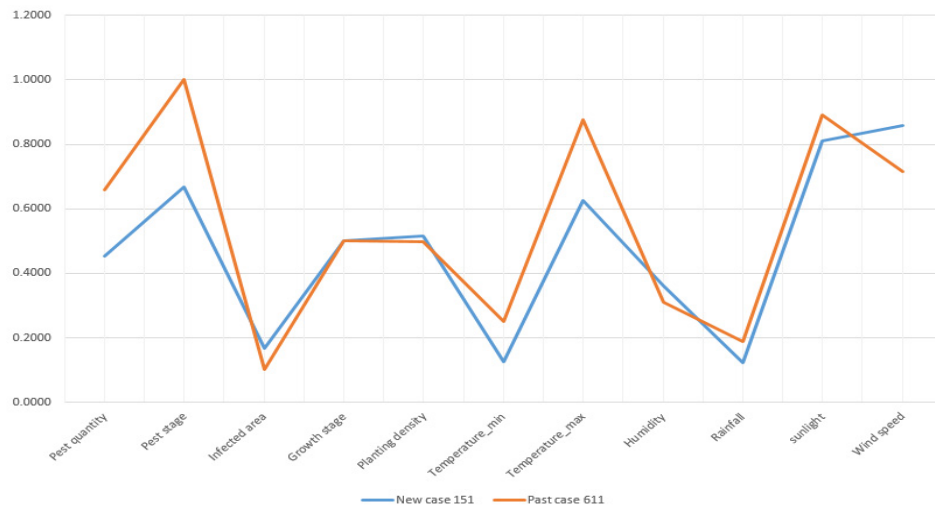
where TP means true positive and FP stands for false positive.

The average precision of all these three similarity measures are presented in Table 5 as follows. This criteria indicates the correctness of case retrieval.

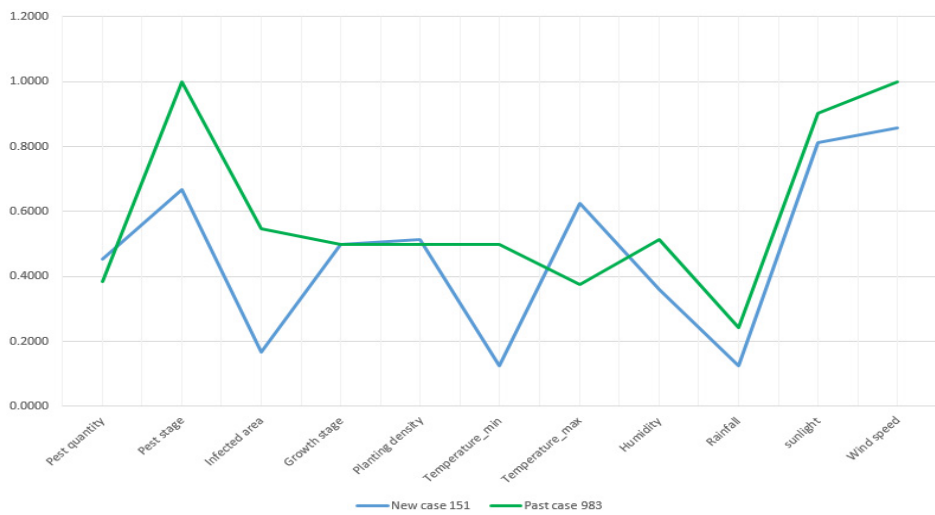
Table 5. The average precision of TSM, ED, and cosine similarity measures.

Similarity Measure	TSM	ED	Cosine Similarity Measure
Average precision	96.33%	94.00%	42.67%

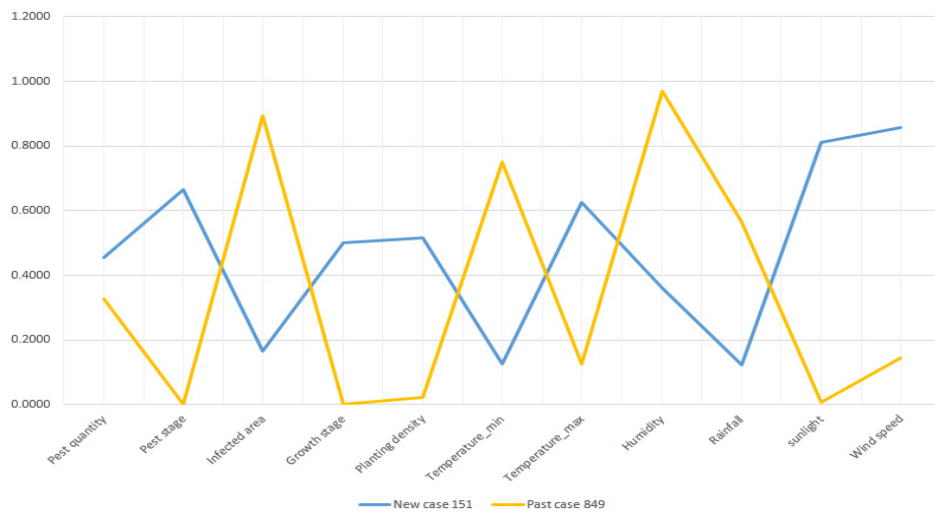
In Table 5, the average precision of TSM achieves the highest value at 96.33%, while the ED similarity measure ranks second at 94.00%, following by the cosine similarity measure in third place (42.67%). Thus, the effectiveness and accuracy of TSM has been proved through comparative experiments.



(a)



(b)



(c)

Figure 10. Comparative results of similarity accuracy: (a) past case retrieved by TSM; (b) case retrieved by ED; and (c) past case retrieved by cosine similarity measure.

6. Conclusions and Future Work

This paper focuses on proposing a triangular similarity measure for case retrieval in case-based reasoning systems. By reviewing current research works, we detected some drawbacks of angle-based and distance-based similarity measures. Typical measures may have difficulties in case retrieval under certain extreme situations. Though several cases sometimes have the same angle value or distance value, huge differences may exist among these cases, especially for their magnitude differences. Thus, a triangular similarity measure is presented to overcome these issues and provide users with a more accurate retrieval result. Meanwhile, a dynamic coefficient is designed to enhance the robustness of TSM. For verifying the proposed TSM, a case study on a CBR-based agricultural decision support system for pest management is conducted. TSM is adopted in the case comparator component for measuring the similarity between cases. The experimental result showed that TSM is able to retrieve the most similar past case with an average accuracy of 91.99%. By revising and reusing the retrieved solution, the ADSS can provide farmers with quick and accurate decision-making suggestions on managing pest problems for rice. Lastly, TSM is compared with typical Euclidean distance and cosine similarity measures. The effectiveness and accuracy of TSM is proved through the comparative experiments. Under the same experiment settings, the past case retrieved by TSM was more similar than those retrieved by the Euclidean distance and cosine similarity measures. By learning from an accurate past case, the rest of the processes, such as solution adaptation, may be easier and more efficient because data of retrieved past case have greater commonalities with the new case than those retrieved by typical similarity measures. Besides applying the proposed measure in CBR-based ADSS, it is also promising to employ this measure in any CBR-based systems like CBR-based clinical systems for disease diagnosis, CBR-based mechanical systems for fault detection, CBR-based emergency systems for emergency response, and so on.

Though the proposed triangular similarity measure has great potential in case retrieval, further developments should be continued for the purpose of improving its performance. Firstly, equal weights may not fully reflect the characteristics of cases. Naturally, weights address the relative importance of each attribute. Those important attributes should be assigned with larger weights during the process of decision-making. Secondly, for improving the efficiency of case retrieval, similar past cases in the case base may be associated with each other. Thus, the structured case representation is especially helpful to connect similar cases. Once a past case is compared with the new case, the similar ones will be preferentially extracted from the case base for comparisons. Thus, the number of visited cases can be reduced. Thirdly, there exists semantic relations within cases. Ontology and semantic reasoning techniques could be included in future work for improving the performance of case retrieval. Lastly, it is promising to verify the proposed TSM and CBR-based ADSS with collected data in real fields.

Author Contributions: Data curation, Z.Z.; Formal analysis, Z.Z.; Funding acquisition, J.-F.M.O.; Investigation, Z.Z.; Methodology, Z.Z.; Project administration, J.-F.M.O.; Software, Z.Z.; Supervision, Z.Z. and J.-F.M.O.; Validation, Z.Z.; Visualization, Z.Z.; Writing—original draft, Z.Z.; Writing—review & editing, P.C. and V.B.

Funding: The research leading to the presented results has been undertaken within the AFARCLOUD European Project (Aggregate Farming in the Cloud), under Grant Agreement No. 783221-AFarCloud-H2020- ECSEL-2017-2, and supported in part by the ECSEL JU and in part by the Spanish Ministry of Science, Innovation and Universities under Grant PCI2018-092965. This work has been also partially supported by the China Scholarship Council (CSC).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

This appendix contains the attribute visualization of compared cases in Table 3.

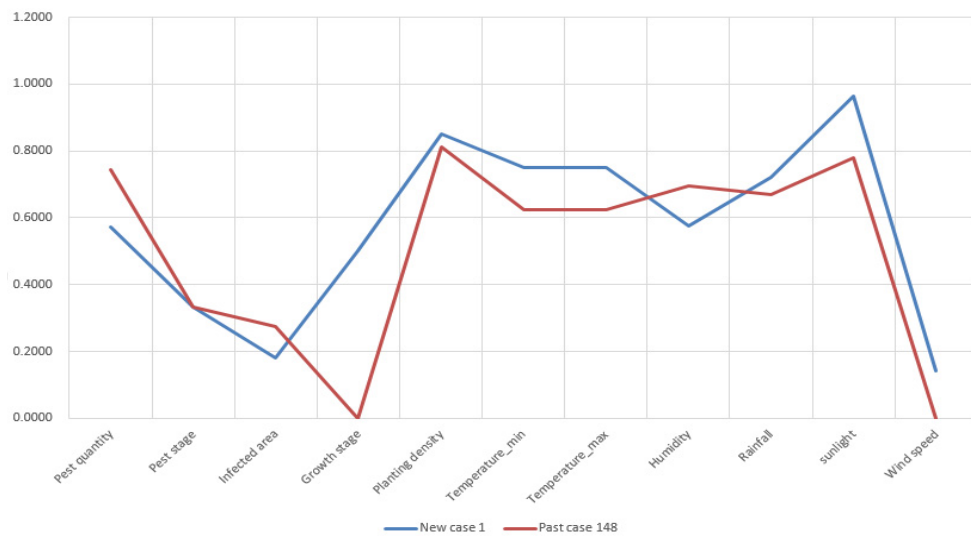


Figure A1. Attribute visualization of new case 1 and retrieved past case 148.

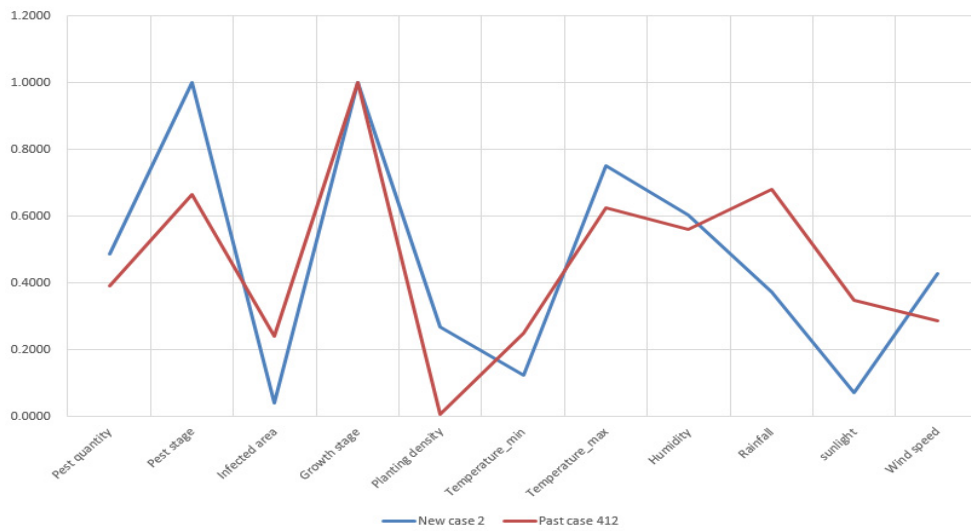


Figure A2. Attribute visualization of new case 2 and retrieved past case 412.

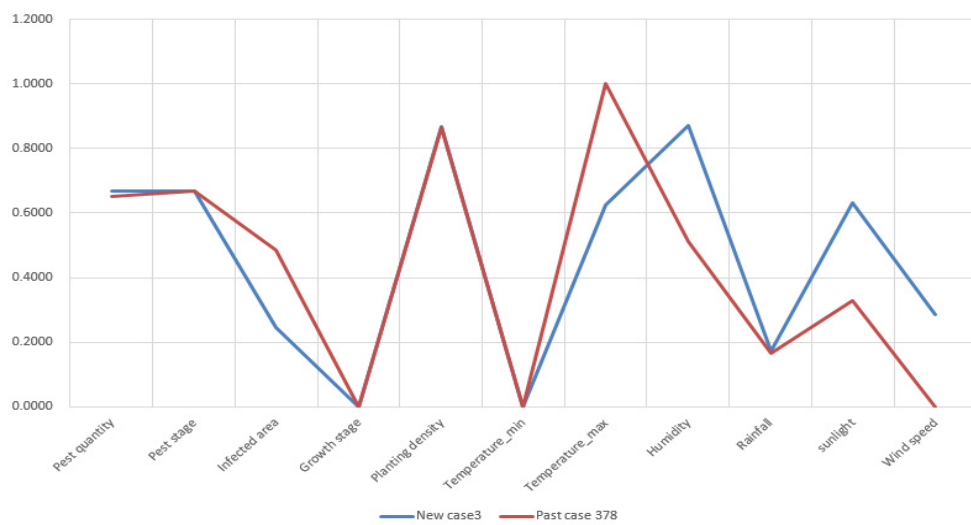


Figure A3. Attribute visualization of new case 3 and retrieved past case 378.

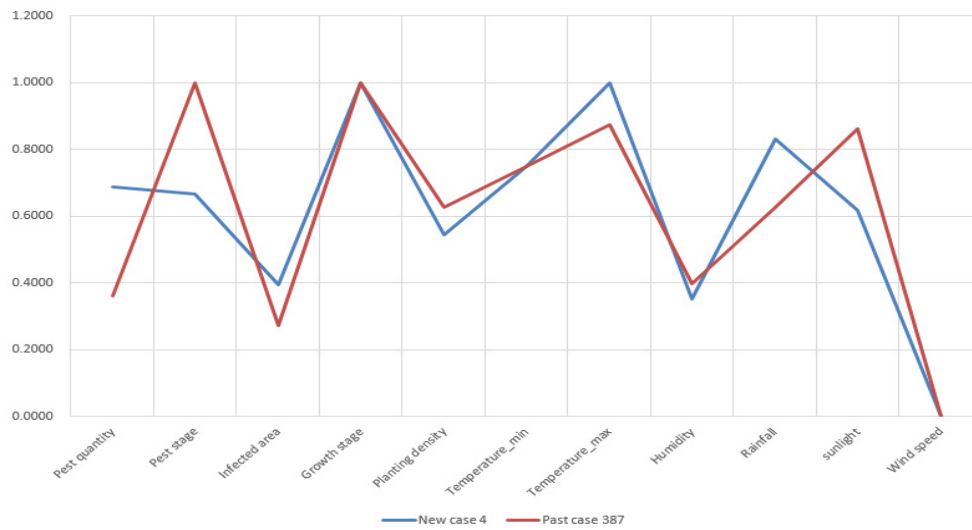


Figure A4. Attribute visualization of new case 4 and retrieved past case 387.

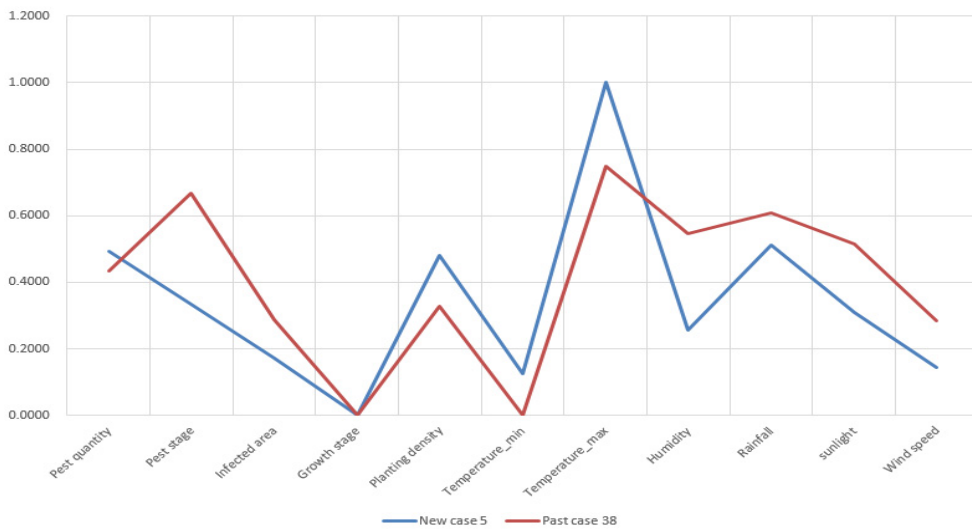


Figure A5. Attribute visualization of new case 5 and retrieved past case 38.

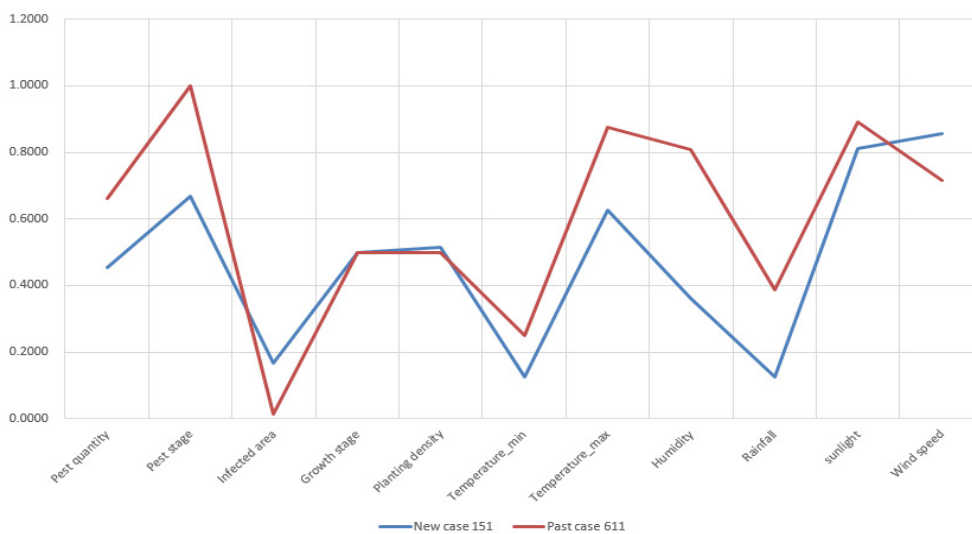


Figure A6. Attribute visualization of new case 151 and retrieved past case 611.

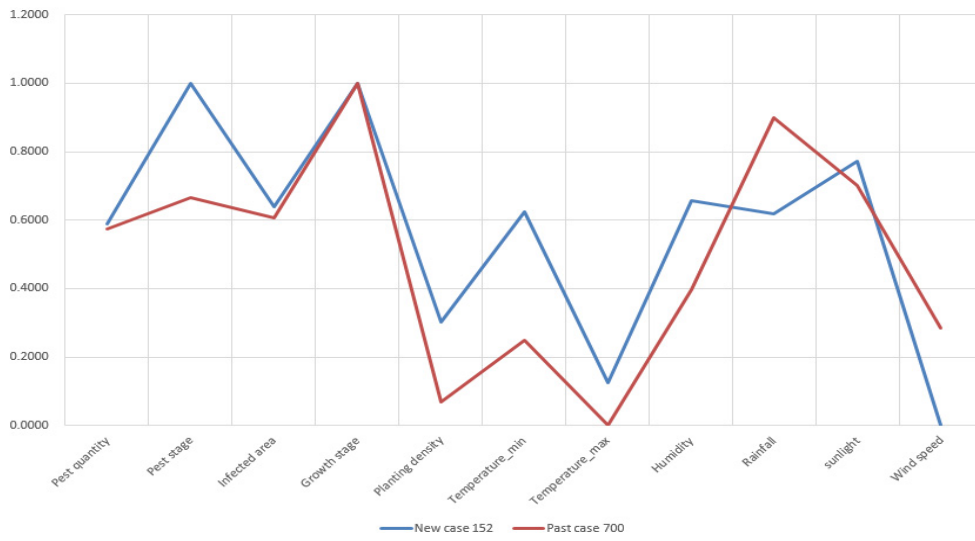


Figure A7. Attribute visualization of new case 152 and retrieved past case 700.

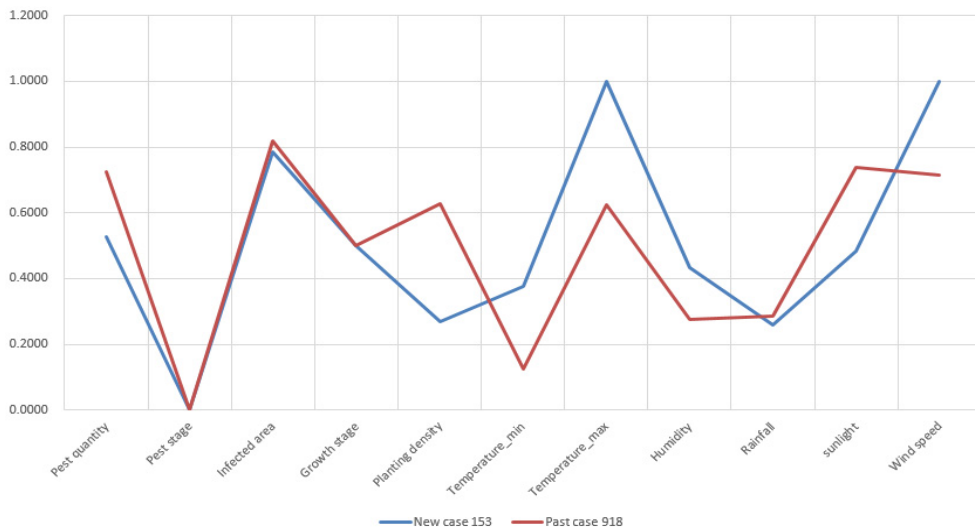


Figure A8. Attribute visualization of new case 153 and retrieved past case 918.

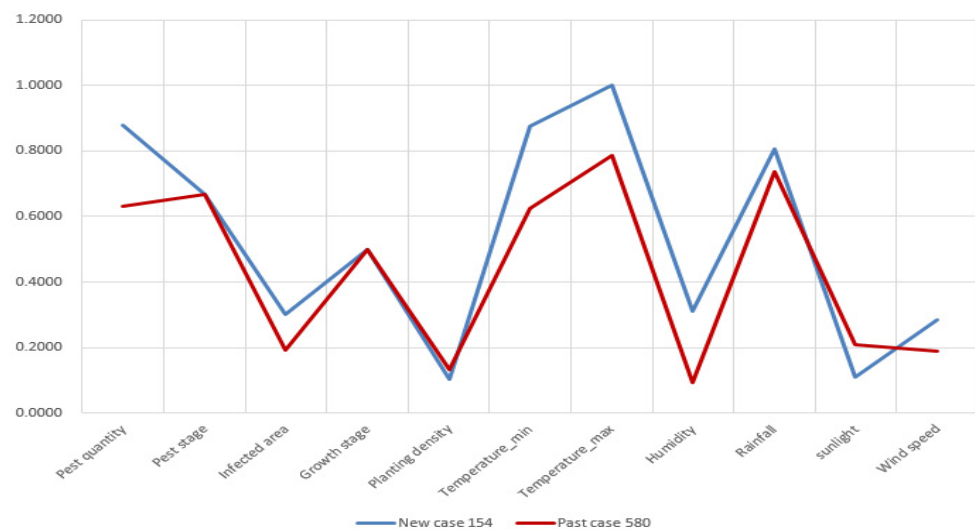


Figure A9. Attribute visualization of new case 154 and retrieved case 580.

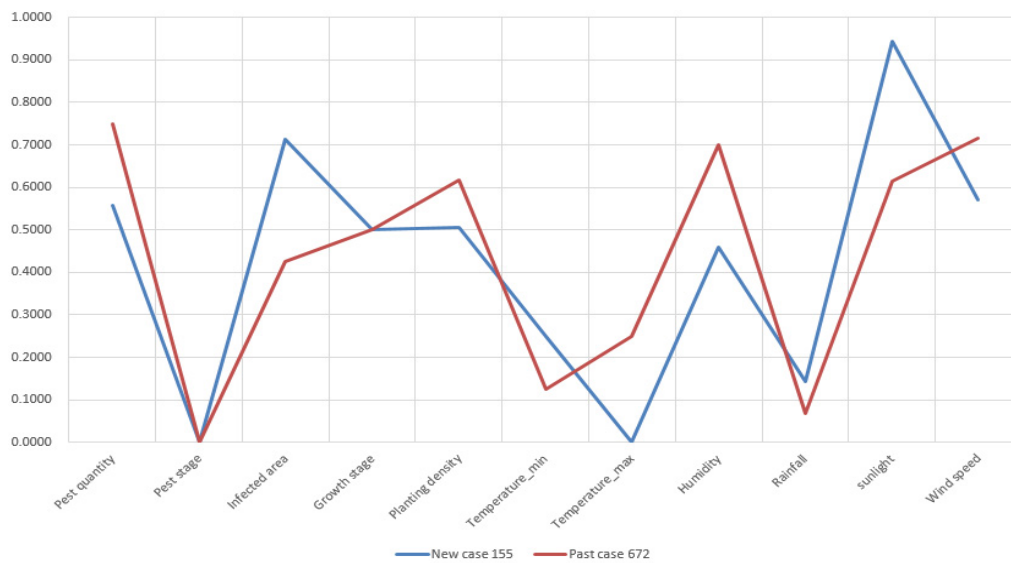


Figure A10. Attribute visualization of new case 155 and retrieved past case 672.

References

1. Watson, I.; Marir, F. Case-based reasoning—A review. *Knowl. Eng. Rev.* **1994**, *9*, 327–354. [[CrossRef](#)]
2. Dutta, S.; Wierenga, B.; Dalebout, A. Case-based reasoning: From automation to decision-aiding and simulation. *IEEE Trans. Knowl. Data Eng.* **1997**, *9*, 911–922. [[CrossRef](#)]
3. Begum, S.; Barua, S.; Filla, R.; Ahmed, M.U. Classification of physiological signals for wheel loader operators using multi-scale entropy analysis and case-based reasoning. *Expert Syst. Appl.* **2014**, *41*, 295–306. [[CrossRef](#)]
4. Begum, S.; Barua, S.; Ahmed, M.U. Physiological sensor signals classification for healthcare using sensor data fusion and case-based reasoning. *Sensors* **2014**, *14*, 11770–11785. [[CrossRef](#)]
5. Janssen, R.; Spronck, P.; Arntz, A. Case-based reasoning for predicting the success of therapy. *Expert Syst.* **2015**, *32*, 165–177. [[CrossRef](#)]
6. Mohammed, M.A.; Ghani, M.K.A.; Arunkumar, N.; Obaid, O.I.; Mostafa, S.A.; Jaber, M.M.; Burhanuddin, M.A.; Matar, B.M.; Abdullatif, S.K.; Ibrahim, D.A. Genetic case-based reasoning for improved mobile phone faults diagnosis. *Comput. Electr. Eng.* **2018**, *71*, 212–222. [[CrossRef](#)]
7. Zhong, Z.W.; Xu, T.H.; Wang, F.; Tang, T. Text case-based reasoning framework for fault diagnosis and prediction by cloud computing. *Math. Probl. Eng.* **2018**. [[CrossRef](#)]
8. Zhao, H.; Liu, J.W.; Dong, W.; Sun, X.Y.; Ji, Y.D. An improved case-based reasoning method and its application on fault diagnosis of Tennessee Eastman process. *Neurocomputing* **2017**, *249*, 266–276. [[CrossRef](#)]
9. Amaief, K.; Lu, J. Ontology-based case-based reasoning approach for intelligent m-Government emergency response services. *Decis. Support. Syst.* **2013**, *55*, 79–97. [[CrossRef](#)]
10. Yu, F.; Li, X.Y. Improving emergency response to cascading disasters: Applying case-based reasoning towards urban critical infrastructure. *Int. J. Disaster Risk Reduct.* **2018**, *30*, 244–256.
11. Yu, F.; Li, X.Y.; Han, X.S. Risk response for urban water supply network using case-based reasoning during a natural disaster. *Saf. Sci.* **2018**, *106*, 121–139. [[CrossRef](#)]
12. Le Ber, F.; Napoli, A.; Metzger, J.L.; Lardon, S. Modeling and comparing farm maps using graphs and case-based reasoning. *J. Univers. Comput. Sci.* **2003**, *9*, 1073–1095.
13. Shih, M.L.; Huang, B.W.; Chiu, N.H.; Chiu, C.; Hu, W.Y. Farm price prediction using case-based reasoning approach—A case of broiler industry in Taiwan. *Comput. Electron. Agric.* **2009**, *66*, 70–75. [[CrossRef](#)]
14. Evans, J.; Terhorst, A.; Kang, B.H. From data to decisions: Helping crop producers build their actionable knowledge. *Crit. Rev. Plant. Sci.* **2017**, *36*, 71–88. [[CrossRef](#)]
15. Aamodt, A.; Plaza, E. Case-based reasoning—Foundational issues, methodological variations, and system approaches. *AI Commun.* **1994**, *7*, 39–59.
16. Liao, T.W.; Zhang, Z.M.; Mount, C.R. Similarity measures for retrieval in case-based reasoning systems. *Appl. Artif. Intell.* **1998**, *12*, 267–288. [[CrossRef](#)]

17. Ye, J. Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Math. Comput. Model.* **2011**, *53*, 91–97. [[CrossRef](#)]
18. Tang, H.C.; Yang, S.T. Counterintuitive test problems for distance-based similarity measures between intuitionistic fuzzy sets. *Mathematics* **2019**, *7*, 437. [[CrossRef](#)]
19. Rachkovskij, D.A. Binary vectors for fast distance and similarity estimation. *Cybern. Syst. Anal.* **2017**, *53*, 138–156. [[CrossRef](#)]
20. Elmore, K.L.; Richman, M.B. Euclidean distance as a similarity metric for principal component analysis. *Mon. Weather Rev.* **2001**, *129*, 540–549. [[CrossRef](#)]
21. Jiang, W.; Wang, M.J.; Deng, X.Y.; Gou, L.F. Fault diagnosis based on TOPSIS method with Manhattan distance. *Adv. Mech. Eng.* **2019**, *11*. [[CrossRef](#)]
22. Esmaeili-Najafabadi, H.; Ataei, M.; Sabahi, M.F. Designing sequence with minimum PSL using Chebyshev distance and its application for chaotic MIMO radar waveform design. *IEEE Trans. Signal Process.* **2017**, *65*, 690–704. [[CrossRef](#)]
23. Mustafa, A.A.Y. A modified Hamming distance measure for quick detection of dissimilar binary images. In Proceedings of the International Conference on Computer Vision and Image Analysis Applications (ICCVIA), Sousse, Tunisia, 18–20 January 2015.
24. Cheng, M.Y.; Tsai, H.C.; Chiu, Y.H. Fuzzy case-based reasoning for coping with construction disputes. *Expert Syst. Appl.* **2009**, *36*, 4106–4113. [[CrossRef](#)]
25. Rezvan, M.T.; Hamadani, A.Z.; Shalbahzadeh, A. Case-based reasoning for classification in the mixed data sets employing the compound distance methods. *Eng. Appl. Artif. Intell.* **2013**, *26*, 2001–2009. [[CrossRef](#)]
26. Jakku, E.; Thorburn, P.J. A conceptual framework for guiding the participatory development of agricultural decision support systems. *Agric. Syst.* **2010**, *103*, 675–682. [[CrossRef](#)]
27. Lindblom, J.; Lundstrom, C.; Ljung, M.; Jonsson, A. Promoting sustainable intensification in precision agriculture: Review of decision support systems development and strategies. *Precis. Agric.* **2017**, *18*, 309–331. [[CrossRef](#)]
28. Padma, T.; Mir, S.A.; Shantharajah, S.P. Intelligent decision support systems for sustainable computing: Paradigms and applications. In *Studies in Computational Intelligence*, 1st ed.; Sangaiah, A.K., Abraham, A., Siarry, P., Sheng, M., Eds.; Springer: Cham, Switzerland, 2017; pp. 225–245.
29. Le Ber, F.; Dolques, X.; Martin, L.; Mille, A.; Benoit, M. A reasoning model based on perennial crop allocation cases and rules. In Proceedings of the 25th International Conference on Case-Based Reasoning (ICCBR), Trondheim, Norway, 26–28 June 2017.
30. Yurin, A.Y. Group decision-making methods for adapting solutions derived from case-based reasoning. *Sci. Tech. Inf. Process.* **2015**, *42*, 375–381. [[CrossRef](#)]
31. Sharaf-Eldeen, D.A.; Moawad, I.F.; El Bahnasy, K.; Khalifa, M.E. Learning and applying range adaptation rules in case-based reasoning systems. In Proceedings of the 1st International Conference on Advanced Machine Learning Technologies and Applications (AMLTA), Cairo, Egypt, 8–10 December 2012.
32. Thompson, V.U.; Panchev, C.; Oakes, M. Performance evaluation of similarity measures on similar and dissimilar text retrieval. In Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), Lisbon, Portugal, 12–14 November 2015.
33. Xia, P.P.; Zhang, L.; Li, F.Z. Learning similarity with cosine similarity ensemble. *Inf. Sci.* **2015**, *307*, 39–52. [[CrossRef](#)]
34. Hassanien, A.E.; El-Bendary, N.; Sweidan, A.H.; Mohamed, A.; Hegazy, O.M. Hybrid-biomarker case-based reasoning system for water pollution assessment in Abou Hammad Sharkia, Egypt. *Appl. Soft Comput.* **2016**, *46*, 1043–1055. [[CrossRef](#)]
35. Senanayake, S.M.N.A.; Malik, O.A.; Iskandar, P.M.; Zaheer, D. A hybrid intelligent system for recovery and performance evaluation after anterior cruciate ligament injury. In Proceedings of the 12th International Conference on Intelligent Systems Design and Applications, Kochi, India, 27–28 November 2012.
36. Mabotuwana, T.; Lee, M.C.; Cohen-Solal, E.V. An ontology-based similarity measure for biomedical data—Application to radiology reports. *J. Biomed. Inform.* **2013**, *46*, 857–868. [[CrossRef](#)]
37. Rahman, M.M.; Desai, B.C.; Bhattacharya, P. Image retrieval-based decision support system for dermatoscopic images. In Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems, Salt Lake City, UT, USA, 22–23 June 2006.

38. Kwon, N.; Lee, J.; Park, M.; Yoon, I.; Ahn, Y. Performance evaluation of distance measurement methods for construction noise prediction using case-based reasoning. *Sustainability* **2019**, *11*, 871. [[CrossRef](#)]
39. Li, H.; Sun, J. Predicting business failure using multiple case-based reasoning combined with support vector machine. *Expert Syst. Appl.* **2009**, *36*, 10085–10096. [[CrossRef](#)]
40. Ferreira, J.R.; Oliveira, M.C. Evaluating margin sharpness analysis on similarity pulmonary nodule retrieval. In Proceedings of the 28th IEEE International Symposium on Computer-Based Medical Systems (CBMS), Sao Paulo, Brazil, 22–25 June 2015.
41. Rashid, E. Construction cost prediction on the basis of multiple parameters using case-based reasoning method. *Int. J. Ser. Tech. Manag.* **2017**, *23*, 255–261. [[CrossRef](#)]
42. Mousa, A.; Yusof, Y. An improved Chebyshev distance metric for clustering medical images. In Proceedings of the 2nd Innovation and Analytics Conference and Exhibition (IACE), Alor Setar, Malaysia, 29 September–1 October 2015.
43. Mustafa, A.A.Y. Probabilistic binary similarity distance for quick binary image matching. *IET Image Proces.* **2018**, *12*, 1844–1856. [[CrossRef](#)]
44. Zhang, L.; Zhang, Y.D.; Tang, J.H.; Lu, K.; Tian, Q. Binary code ranking with weighted Hamming distance. In Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013.
45. Ye, J. Improved cosine similarity measures of simplified neutrosophic sets for medical diagnoses. *Artif. Intell. Med.* **2015**, *63*, 171–179. [[CrossRef](#)] [[PubMed](#)]
46. Liu, H.F.; Hu, Z.; Mian, A.; Tian, H.; Zhu, X.Z. A new user similarity model to improve the accuracy of collaborative filtering. *Knowl. Based Syst.* **2014**, *56*, 156–166. [[CrossRef](#)]
47. Liu, C.J. Discriminant analysis and similarity measure. *Pattern Recognit.* **2014**, *47*, 359–367. [[CrossRef](#)]
48. Ling, H.B.; Jacobs, D.W. Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 286–299. [[CrossRef](#)]
49. Zheng, Y.H.; Jeon, B.; Xu, D.H.; Wu, Q.M.J.; Zhang, H. Image segmentation by generalized hierarchical fuzzy C-means algorithm. *J. Intell. Fuzzy Syst.* **2015**, *28*, 961–973.
50. Wu, J.X.; Rehg, J.M. Beyond the Euclidean distance: Creating effective visual codebooks using the histogram intersection kernel. In Proceedings of the 12th IEEE International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009.
51. Liao, Q.; Yang, F.; Zhao, J.M. An improved parallel K-means clustering algorithm with MapReduce. In Proceedings of the 15th IEEE International Conference on Communication Technology (ICCT), Guilin, China, 17–19 November 2013.
52. Li, C.G.; Qiu, Z.Y.; Liu, C.T. An improved weighted K-nearest neighbour algorithm for indoor positioning. *Wirel. Pers. Commun.* **2017**, *96*, 2239–2251. [[CrossRef](#)]
53. Wang, W.; Zhang, G.Q.; Lu, J. Collaborative filtering with entropy-driven user similarity in recommender systems. *Int. J. Intell. Syst.* **2015**, *30*, 854–870. [[CrossRef](#)]
54. Kobayashi, E.; Fushimi, T.; Saito, K.; Ikeda, T. Similarity search by generating pivots based on Manhattan distance. In Proceedings of the 13th Pacific Rim International Conference on Artificial Intelligence (PRICAI), Gold Coast, Australia, 1–5 December 2014.
55. Li, H.Y.; Shi, R.J.; Chen, W.B.; Shen, I.F. Image tangent space for image retrieval. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR), Hong Kong, China, 20–24 August 2006.
56. Demirci, S.; Erer, I.; Ersoy, O. Weighted Chebyshev distance classification method for hyperspectral imaging. In Proceedings of the Conference on Next-Generation Spectroscopic Technologies VIII, Baltimore, MD, USA, 20–22 April 2015.
57. Billot, A.; Gilboa, I.; Schmeidler, D. Axiomatization of an exponential similarity function. *Math. Soc. Sci.* **2008**, *55*, 107–115. [[CrossRef](#)]
58. Meisner, M.H.; Rosenheim, J.A.; Tagkopoulos, I. A data-driven, machine learning framework for optimal pest management in cotton. *Ecosphere* **2016**, *7*, e01263. [[CrossRef](#)]
59. Ho, G.T.T.; Le, C.V.; Nguyen, T.H.; Ueno, T.; Nguyen, D.V. Incidence of yellow rice stem borer scirpophaga incertulas walker in Haiphong, Vietnam and control efficiency of egg mass removal and insecticides. *J. Fac. Agric. Kyushu Univ.* **2013**, *58*, 301–306.

60. Leistra, M.; Zweers, A.J.; Warinton, J.S.; Crum, S.J.H.; Hand, L.H.; Beltman, W.H.J.; Maund, S.J. Fate of the insecticide lambda-cyhalothrin in ditch enclosures differing in vegetation density. *Pest. Manag. Sci.* **2014**, *60*, 75–84. [[CrossRef](#)] [[PubMed](#)]
61. Wang, J.; Lan, Y.B.; Zhang, H.H.; Zhang, Y.L.; Wen, S.; Yao, W.X.; Deng, J.J. Drift and deposition of pesticide applied by UAV on pineapple plants under different meteorological conditions. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 5–12. [[CrossRef](#)]
62. Decision Support System Data for Farmer Decision Making. Available online: <https://pdfs.semanticscholar.org/6304/fb7c0884183d00629af93bf1e05de5e3d0da.pdf> (accessed on 16 September 2019).
63. El-Sappagh, S.H.; Elmogy, M. Case based reasoning: Case representation methodologies. *Int. J. Adv. Comput. Sci. Appl.* **2015**, *6*, 192–208.
64. Mir, S.A.; Padma, T. Generic multiple-criteria framework for the development of agricultural DSS. *Decis. Syst.* **2017**, *26*, 341–367. [[CrossRef](#)]
65. Jahromi, M.Z.; Parvinnia, E.; John, R. A method of learning weighted similarity function to improve the performance of nearest neighbour. *Inf. Sci.* **2009**, *179*, 2964–2973. [[CrossRef](#)]
66. Rice Planthopper Problems and Relevant Causes in China. Available online: <http://ag.udel.edu/delphi/5789.pdf> (accessed on 12 July 2019).
67. Cheng, X.A.; Chang, C.; Dai, S.M. Responses of striped stem borer, chilo suppressalis (Lepidoptera: Pyralidae), from Taiwan to a range of insecticides. *Pest. Manag.* **2010**, *66*, 762–766. [[CrossRef](#)]
68. Sonderskov, M.; Fritzsche, R.; del Mol, F.; Gerowitt, B.; Goltermann, S.; Kierzek, R.; Krawczyk, R.; Bojer, O.M.; Rydahl, P. DSSHerbicide: Weed control in winter wheat with a decision support system in three South Baltic regions—Field experimental results. *Crop. Prot.* **2015**, *76*, 15–23. [[CrossRef](#)]
69. Garcia-Balboa, J.L.; Alba-Fernandez, M.V.; Ariza-Lopez, F.J.; Rodriguez-Avi, J. Analysis of thematic similarity using confusion matrices. *ISPRS Int. J. Geoinf.* **2018**, *7*, 233. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).