Journal of Computer Science Research
https://journals.bilpubgroup.com/index.php/jcsr

ARTICLE

Similarity Intelligence: Similarity Based Reasoning, Computing, and Analytics

Zhaohao Sun

Department of Business Studies, PNG University of Technology, Private Mail Bag, Lae 411, Morobe, Papua New Guinea

ABSTRACT

Similarity has been playing an important role in computer science, artificial intelligence (AI) and data science. However, similarity intelligence has been ignored in these disciplines. Similarity intelligence is a process of discovering intelligence through similarity. This article will explore similarity intelligence, similarity-based reasoning, similarity computing and analytics. More specifically, this article looks at the similarity as an intelligence and its impact on a few areas in the real world. It explores similarity intelligence accompanying experience-based intelligence, knowledge-based intelligence, and data-based intelligence to play an important role in computer science, AI, and data science. This article explores similarity-based reasoning (SBR) and proposes three similarity-based inference rules. It then examines similarity computing and analytics, and a multiagent SBR system. The main contributions of this article are: 1) Similarity intelligence is discovered from experience-based intelligence consisting of data-based intelligence and knowledge-based intelligence. 2) Similarity-based reasoning, computing and analytics can be used to create similarity intelligence. The proposed approach will facilitate research and development of similarity intelligence, similarity computing and analytics, machine learning and case-based reasoning.

Keywords: Similarity intelligence; Similarity computing; Similarity analytics; Similarity-based reasoning; Big data analytics; Artificial intelligence; Intelligent agents

*CORRESPONDING AUTHOR:
Zhaohao Sun, Department of Business Studies, PNG University of Technology, Private Mail Bag, Lae 411, Morobe, Papua New Guinea; Email: zhaohao.sun@gmail.com

ARTICLE INFO
Received: 19 March 2023 | Revised: 20 May 2023 | Accepted: 26 May 2023 | Published Online: 9 June 2023
DOI: https://doi.org/10.30564/jcsr.v5i3.5575

CITATION
Sun, Zh.H., 2023. Similarity Intelligence: Similarity Based Reasoning, Computing, and Analytics. Journal of Computer Science Research. 5(3): 1-14. DOI: https://doi.org/10.30564/jcsr.v5i3.5575

1. Introduction

Similarity, similarity relations, and similarity metrics have been playing an important role in computer science, artificial intelligence (AI), and data science [1-6]. Similarity has also played an important role in machine learning and case-based reasoning (CBR) [7,8]. Machine learning including deep learning has had important impacts on our life and work [7,9]. CBR as an AI technique has also played a significant role in experience-based reasoning and experience management [8-10]. From a meta viewpoint, what are the relationships between machine learning and CBR?

Intelligence has been also playing an important role in AI, business intelligence, machine learning, and CBR [3,8,11,12]. Intelligence can be defined as the collection, analysis, interpretation, visualization, and dissemination of strategic data, information, and knowledge for discovering and using the knowledge patterns and insights at the right time in the decision-making process [13]. Are machine learning and CBR related to similarity intelligence? This implies that similarity intelligence has been ignored in these disciplines. More specifically, research issues in this direction are:

1) Why is similarity intelligence important?

2) What are the relationships between similarity intelligence and experience intelligence, knowledge-based intelligence and data-based intelligence?

3) What are similarity computing and analytics and their impacts on similarity intelligence?

This article will explore similarity intelligence, similarity-based reasoning, similarity computing and analytics, and their relationships. To address the first question, this article looks at the similarity of intelligence and its impact on a few areas in the real world. To address the second question, it explores similarity intelligence that has been accompanying experience-based intelligence, knowledge-based intelligence and data-based intelligence to play an important role in computer science, AI, and data science. After reviewing the fundamentals of similarity, this article explores similarity-based reasoning (SBR) and proposes three similarity-based inference rules. This article then examines similarity computing and analytics, and a multiagent SBR system. The main contributions of this article are: 1) Similarity intelligence is discovered from experience-based intelligence consisting of data-based intelligence and knowledge-based intelligence. 2) Similarity-based reasoning, computing and analytics can be used to create similarity intelligence.

The rest of this article is organized as follows: Section 2 looks at why similarity intelligence is important. Section 3 examines machine learning and CBR as experience based intelligence. Section 4 examines the fundamentals of similarity. Section 5 explores similarity-based reasoning and proposes similarity-based inference rules for conducting SBR. Section 6 examines similarity computing and analytics. Section 7 proposes a multiagent architecture for an SBR system, and Section 8 ends this article with some concluding remarks.

2. Why is similarity intelligence important?

This section highlights why similarity intelligence is important.

Similarity has been playing an important role in mathematics, computer science, AI, and data science. Similarity has also played a significant role in fuzzy logic [2] and big data [5]. However, similarity intelligence has been ignored in these disciplines.

Similarity intelligence is a process for discovering intelligence from two or more objects or cases using similarity algorithms and techniques. The Turing test [14] has already mentioned that intelligence computing machinery is similar to that of human beings. This is a kind of similarity intelligence. Similarity intelligence includes similar relationships consisting of patterns and insights between machines, human beings, and software apps [14,5]. In other words, similarity intelligence is not only from human beings, but also from machines or software or apps.

Similarity also plays an important role in ChatGPT, because similarity is crucial in natural language understanding and processing. Based on the research analyzing 1000 texts produced by ChatGPT, it found that on average, the similarity varies between 70%

and 75% [15]. Therefore, one of the important tasks of ChatGPT (http://wwww.openAi.com) is to discover similarity intelligence from two or more objects, texts, and cases.

Similarity intelligence is important because it enables us to identify similarities and patterns in data sets, which can be used to make more informed decisions and predictions. By identifying similarities between different sets of data, objects, and cases [8], we can better understand relationships and draw insights that might not be immediately apparent, at least similarity intelligence can allow us to select one from a similarity class as a representative and then we can analyze it as a characteristic of the similarity class [16]. For example, in the field of customer relationship management [12], intelligence can be used to identify patterns and preferences in consumer behaviors through similarity metrics. The patterns and preferences can then be used to develop targeted advertising and product recommendations that are more likely to appeal to specific groups of consumers.

Therefore, similarity intelligence is important not only for computer science, AI, big data, and data science, but also for businesses and organizations in a wide range of industries, enabling decision makers to obtain more informed decisions in an intelligent experience-based, knowledge-driven, and data-driven world.

3. Experience-based intelligence

Experience-based intelligence is a process of discovering intelligence from experiences, based on experience-based reasoning [17]. Experience-based intelligence consists of data-based intelligence and knowledge-based intelligence. This section looks at similarity intelligence from experience-based intelligence using two examples, machine learning and CBR. Machine learning is data-based intelligence. CBR is knowledge-based intelligence.

3.1 Machine learning

Similarity has always been important in pattern recognition, graphical pattern recognition, machine learning [7,18], because as soon as we have created patterns, and we have to use similarity to match what was input to the systems and compare it with our patterns.

Machine learning is about how to build computers and apps that improve automatically through experience [19], that is, machine learning is a process of discovering intelligence from experience using computers and software. Therefore, machine learning is an experience-based Intelligence.

Machine learning is about how a computer can use a model and algorithm to observe some data about the world, and adapt to new circumstances and detect and extrapolate patterns [11]. Therefore, machine learning is a process of discovering intelligence from data, that is, machine learning is data-based intelligence, a process of discovering intelligence from data, because it is a process of using probabilistic models and algorithms on data to create intelligence through data [11].

One of the unsupervised machine learning is clustering [7]. How we calculate the similarity between two clusters or two objects is important for clustering [4,7,18]. There are a few methodologies that are utilized to calculate the similarity: For example, Min, Max, the distance between centroids and other similarity matrices mentioned in Section 4.4. Therefore, machine learning is similarity intelligence, a process for creating intelligence through similarity.

Overall, machine learning is an experience-based Intelligence, a process of discovering Intelligence through experience [4]. Machine learning is data-based intelligence, a process of discovering intelligence from data. Machine learning is also similarity intelligence, a process for creating intelligence through similarity.

3.2 Case-based reasoning

CBR is a process of discovering similarity intelligence from a case base, just as data mining is a process of discovering data intelligence from a large DB [12]. Similarity intelligence includes the exact case that has been used in the past for solving the problem encountered recently.

CBR is a reasoning paradigm based on previous experiences or cases [12,8]. CBR is based on two principles about the nature of the world [8]: The types of problems an agent encounters tend to recur. Hence, future problems are likely to be similar to current problems. The world is regular: similar problems have similar solutions or similar causes bring similar effects [8]. Consequently, solutions to similar prior problems are a useful starting point for new problem solving. The first principle implies that CBR is a kind of experience-based reasoning (EBR), while the second principle is the guiding principle underlying most approaches to similarity-based reasoning (SBR) [8]. "Two cars with similar quality features have similar prices" is one application of the above-mentioned second principle, and also a popular experience principle summarizing many individual experiences of buying cars. It is a kind of SBR. In other words, SBR is a concrete realization of CBR. The CBR system (CBRS) is an intelligent system based on CBR, which can be modelled as [8]:

$$\text{CBRS} = \text{Case Base} + \text{CBRE} \tag{1}$$

where the case base (CB) is a set of cases, each of which consists of the previously encountered problem and its solution. CBRE is a CBR engine, which is the inference mechanism for performing CBR, in particular for performing SBR. The SBR can be formalized as:

$$\frac{\text{P}', \text{P}' \sim \text{P}\ \text{P}, \rightarrow \text{Q}}{\therefore Q'} \tag{2}$$

where *P*, P', Q' and Q' represent compound propositions, P' ∼ P means that if P' and *P* are similar (in terms of similarity relations, metrics and measures, see Section 4) and then Q and Q' are also similar. (2) is called *generalized modus ponens*, that is, (2) is one of the inference rules for performing modus ponens based on SBR. Typical reasoning in CBR, known as the CBR cycle, consists of (case) Repartition, Retrieve, Reuse, Revise and Retain [8]. Each of these five stages is a complex process. SBR dominates all these five stages [16]. Therefore, CBR is a process for discovering intelligence through SBR, because Similar problems have similar solutions.

One significant contribution of CBR research and development is that it points out the importance of experience and similarity [9,16]. CBR is experience-based intelligence, a process for discovering intelligence based on experience. Because case base is a kind of knowledge base [10,8], so that, CBR is also a knowledge-based intelligence [11].

Overall, similarity intelligence accompanies experience-based intelligence [10,8], data-based intelligence [4] and knowledge-based intelligence [11] to provide constructive insights and decision supports for businesses and organizations.

4. Fundamentals of similarity

The similarity is a fundamental concept for many fields in mathematics, mathematical logic, computer science, AI, data science, and other sciences [16,9,20,21]. This section first briefly looks at similarity and then focuses on similarity relations, fuzzy similarity relations, and similarity metrics.

4.1 Introduction

The concept of similarity has been studied by numerous researchers from different disciplines such as in mathematics [20], big data [5], computer science [22,23], AI and fuzzy logic [1,2,21], to name a few. For example, Klawonn and Castro [24] examined similarity in fuzzy reasoning and showed that similarity is inherent to fuzzy sets. Fontana and Formato [25] extended the resolution rule as the core of a logic programming language based on similarity and discussed similarity in deductive databases. The concepts of similarity and similarity relations play a fundamental role in many fields of pure and applied science [26,20]. The notion of a metric or distance between objects has long been used in many contexts as a measure of similarity or dissimilarity between elements of a set [27,22,18]. Thus, there exist a wide variety of techniques for dealing with problems involving similarity, similarity relations, similarity measures, and similarity metrics [21,23]. For example, fuzzy logic [1,2], databases [5], data mining [18] and CBR [8] provides a number of concepts and techniques for dealing with similarity relations, similarity measures, and similarity metrics.

In what follows, we briefly introduce similarity relations, fuzzy similarity relations, and similarity metrics.

4.2 Similarity relations

The concept of a similarity relation is a natural generalization of similarity between two triangles and two matrices in mathematics [16,18]. More precisely:

Definition 1. A binary relation S on a non-empty set X is called a similarity relation if it satisfies:

(1) $\forall x$, xSx,

(2) If xSy, then ySx,

(3) If xSy, ySz then xSz

The conditions (1), (2), and (3) are the reflexive, symmetric, and transitive laws. If xSy we say that x and y are similar [20,16].

Example 1. Matrices B and C in $M_{n,n}$ are similar if $C = PBP^{-1}$ for an invertible P, in which case we write B~C. It is easy to prove that is ~ is a similarity relation in $M_{n,n}$ [20].

This example implies that the concept of a similarity relation here is a generalization of the similarity between matrices in $M_{n,n}$.

Similarity relations can be used for classification through partition [16] and clustering [18].

4.3 Fuzzy similarity relations

As an extension of similarity relations, fuzzy similarity relations were introduced by Zadeh in 1971 [2] and have attracted much attention since then [1,27,21]. For example, fuzzy similarity relations have been used in CBR [16]. For the sake of brevity, we use standard fuzzy set theory notation for operations min and max, although there are many alternative choices for these operations available in fuzzy set theory (Zimmermann, 1996). S is still used to denote a fuzzy similarity relation if there is not any confusion arising.

Definition 2. A fuzzy binary relation S on a non-empty set is a fuzzy similarity relation in X if it is reflexive, symmetric, and transitive [28,2], that is:

$$S(x, x) = 1 \tag{3}$$

$$S(x, y) = S(y, x) \tag{4}$$

$$S \geq S \cdot S \tag{5}$$

where · is the composition operation of fuzzy binary relations based on min and max operations. A more explicit form of Equation (5) is

$$S(x,z) \geq \vee_y (S(x,y) \wedge S(y,z)) \tag{6}$$

Equation (6) is called max-min transitivity [2]. The revised form of this definition was given by Ovchinnikov in 1991 [28]. The main difference between the definition of Zadeh and that of Ovchinnikov lies in that instead of Equation (6), Ovchinnikov viewed the following model as max-min transitivity.

$$S(x,z) \geq S(x,y) \wedge S(y,z) \tag{7}$$

4.4 Similarity metrics

Generally speaking, similarity in mathematics is considered as a relation, while similarity in CBR is considered both a relation and a measure, a function, and a metric [16,8].

Definition 3. A relation, denoted by S_m, on non-empty X, is a similarity metric if it satisfies [16]:

1) S_m is a similarity relation on X;

2) 1- S_m is a metric on X; that is, it is a function from $X \times X$ to [0,1], provided that:

- For any $x \in X$, $S_m(x,x) = 1$
- For all $x, y \in X$, $S_m(x,y) = S_m(y,x)$ (8)
- For all $x, y, z \in X$, $S_m(x,z) \geq S_m(x,y) \wedge S_m(y,z)$

where ∧ is min operator. Equation (8) in this definition is called the similarity inequality. It should be noted that the similarity metric here, S_m, can not directly satisfy the triangle inequality [16]. Equation (8) is based on Ovchinnikov's concept of fuzzy similarity relations [28].

In comparison with the definition of fuzzy similarity relations, we emphasize that the similarity metric here is first a traditional similarity relation, and also just a metric, maybe to some extent, because the similarity between two objects is the necessary condition to further discuss how similar they are in the context [16].

5. Similarity-based reasoning and inference rules

This section highlights similarity-based reasoning and its three inference rules.

5.1 Similarity-based reasoning

Similarity-based reasoning (SBR) has been studied by many researchers from different fields. For example, Sun [9] examined integration of rule-based and SBR from an AI viewpoint. He considered SBR as a reasoning-based similarity matching. Bogacz and Giraud-Carrier considered SBR as "reasons from similarity" from a neural network viewpoint [29]. The relationship between CBR and SBR has drawn some attention [8,16]. However, what is similarity-based reasoning? There is still no definition of it, to our knowledge. In fact, many methods of SBR seem to lack a sound theoretical or logical basis [30]. We need a relatively precise definition of SBR, in order to investigate similarity-based approaches to SBR.

Definition 4. Let P, P', Q, and Q' represent compound propositions, $P \to Q$ is a production rule, denoting if P then Q. A proposition can be inferred from propositions P and $P \to Q$, provided that P, and P' are similar ($P' \sim P$), and then Q and Q' are also similar; that is:

$$\frac{P', P' \sim P, P \to Q}{\therefore Q'} \tag{9}$$

Then, this reasoning paradigm is called similarity-based reasoning (SBR).

More generally, a proposition Q' can be similarity-based inferred from propositions P_1' $P_2', \ldots, P_n'$ provided P_1 P_2, ..., $P_n \to Q$, and $P_i' \sim P_i$ $(i = \{1, 2, ..., n\})$, Q and Q' are also similar; that is:

$$\frac{\begin{array}{c} P_1', P_2', \cdots, p_n' \\ P_1, P_2, \cdots, P_n \to Q \end{array}}{\therefore Q'} \tag{10}$$

It should be noted that the definition is based on modus ponens [31]. Therefore, the reasoning defined above can be considered as a kind of SBR with respect to modus ponens (which will be examined further in the next subsection). It can be also considered as a composite reasoning paradigm. Furthermore, the above definition is general, and its generality lies in that we have not assigned any special meaning or semantics to the similarity used in the definition.

Example 4. Google Chrome as a search engine is based on similarity-based reasoning, that is, "similarity-based reasoning" which is searched by https://www.google.com.au/ and found 50,000 results (on 02 March 2023). However, not every one of the found results is related to "similarity-based reasoning" but to "similarity" or "reasoning". This is the reason why we call Google Chrome as similarity-based reasoning. It does not really result in complying with the inference rule of modus ponens (see latter). This is also the reason why some hope to get exact results rather than 50,000 or millions of found results searched by Chrome and many other search engines.

In order to perform SBR, it is necessary to note the following three points:

1) What we examined in the previous subsections: Similarity relations, fuzzy similarity relations, and similarity metrics are concrete forms of similarity used in the above definition. In other words, each of them can lead to a class of SBR. We can, therefore, examine SBR from a viewpoint of either similarity relations or fuzzy similarity relations or similarity metrics. If so, we will make our investigation very complex, although the corresponding research results are of significance in applications.

2) SBR is treated in a more general way in this article; that is, two different forms, $\sim$ and $\approx$, of similarity (e.g., similarity relations, fuzzy similarity relations, and similarity metrics) are used in the context. The first $\sim$ is associated with the similarity between P and P', while the second $\approx$ is associated with the similarity between Q and Q'. In the context of CBR, these two similarity relations (or fuzzy relations or metrics) are in different worlds [8]; that is, the first $\sim$ is associated with the possible world of problems, while the second $\approx$ is associated with the possible world of solutions.

3) The readers can consider $\sim$ and $\approx$, from now on in this article, as either similarity relations or fuzzy

similarity relations or similarity metrics in a consistent way in a real-world application. In other words, for a real-world problem, if one considers them similarity metrics, then she/he should use them consistently.

In order to perform SBR using Equation (9), we define a degree of similarity between propositions P and P' [9]) as

$$S(P, P') = 2\,\frac{|F_P \cap F_{P'}|}{|F_P| + |F_{P'}|} \quad (11)$$

where F_x is the set of the features of proposition x, and $|F_x|$ is the size of the set of features of F_x. The degree of similarity $S(P, P')$ has the following properties [18]:

1) $0 \le S(P, P') \le 1$.

Note that a similarity function is a special similarity relation.

2) If $P = P'$, that is, P and P' are the same propositions, then $S(P, P') = 1$.

5.2 Similarity-based inference rules

This section will look at similarity-based reasoning (SBR) and its three inference rules.

Similarity-based modus ponens

In the previous section, we defined SBR concerning modus ponens. In order to emphasize the importance of similarity between P and P', Q and Q' and show the difference of the inference rule of SBR from the generalized modus ponens of fuzzy reasoning [1], we call Equation (9) similarity-based modus ponens (SMP), and its form will be replaced with the following:

$$\frac{P',\ P' \sim P,\ P \to Q,\ Q \approx Q'}{\therefore Q'} \quad (12)$$

The reasoning paradigm of the similarity-based deductive system and similarity-based agent [32] is based on Equation (12), because similarity-based resolution [32] is an alternative form of Equation (12). Equation (12) is also a logical foundation for CBR [8]. SMP is one of the most important reasoning paradigms in many other disciplines, because it is the basic form of any similarity-based deductive reasoning paradigm [29,30,9].

In a CBR customer support system [8], P' is the problem description of the customer, $P' \sim P$ means that P' and P are similar, $P \to Q$ is the case retrieved from the case base C based on a similarity-based retrieval algorithm. $Q \approx Q'$ means that Q and Q' are similar, and Q' is the satisfactory solution to the requirement of the customer.

In the context of fuzzy similarity relations and similarity metrics [22], we assume that P' corresponds to $\widetilde{P_0}$, $P' \sim P$ corresponds to $\widetilde{F_{01}}$, $P \to Q$ corresponds to $\widetilde{F_{11}}$, $Q \approx Q'$ corresponds to $\widetilde{F_{10}}$, and Q' corresponds to $\widetilde{Q_1}$. Then, using the compositional rule of inference [1,33], we obtain:

$$\widetilde{Q_1} = \widetilde{P_0} \circ \widetilde{F_{01} \circ F_{11}} \circ \widetilde{F_{10}} \quad (13)$$

where $\widetilde{P_0}$ is a fuzzy set in W_P. $\widetilde{F_{01}}$, $\widetilde{F_{11}}$, and $\widetilde{F_{10}}$ are a similarity metric, a fuzzy rule and a fuzzy similarity metric in $W_P \times W_Q$ respectively, and $\widetilde{Q_1}$ is a fuzzy set on W_Q. This is a computational foundation for similarity-based modus ponens [8]. In the case of $Q \approx Q'$, $\widetilde{F_{10}}$ is a unit metric, and Equation (13) is then simplified into:

$$\widetilde{Q_1} = \widetilde{P_0} \circ \widetilde{F_{01} \circ F_{11}} \quad (14)$$

When $\widetilde{P_0}$, $\widetilde{F_{01}}$, $\widetilde{F_{11}}$, and $\widetilde{Q_1}$ are only a numerical similarity measure respectively, (14) essentially degenerates into the computational form.

In fact, many other reasoning paradigms also follow, to some sense, Equation (14), for example, analogical reasoning [1], although they have different semantics and operational algorithms for performing their own reasoning based on different real-world scenarios.

While fuzzy reasoning is essentially computational reasoning, SBR can be considered as both symbolic reasoning and computational reasoning [8]. If we regard SBR as computational reasoning, then we can consider it as a special kind of fuzzy reasoning, to some extent, because the similarity between P and P', $P \sim P'$, and the similarity between Q and Q', $Q \approx Q'$, are replaced by the fuzziness between them in the context of fuzzy logic. This is the reason why we can use fuzzy reasoning to examine the similarity-based modus ponens in CBR [8].

Similarity-based modus tollens

Similarity-based modus tollens (SMT) is another inference rule for SBR. From a traditional viewpoint, we can consider SMT as an integration of SBR and modus tollens. The general form of SMT is as follows:

$$\frac{\neg Q', Q' \approx Q, Q \to P, P \to P'}{\therefore \neg P'} \tag{15}$$

Although fuzzy modus tollens have not been investigated in fuzzy logic [1], this is the first time that similarity-based modus tollens is discussed. With the increasing importance of similarity, SMT and its corresponding SBR will find their applications in business and mathematics.

Example 5. Similarity-based modus tollens. Let:

- *RAT*: The applicant has a good credit rating,
- *REP*: The applicant has a good financial reputation,
- The loan officer has an experienced rule, *RAT→REP*: If the applicant has a good credit rating, then the applicant has a good financial reputation.

In this case, the loan officer knows the information from applicant *A*, ¬*REP*: The applicant has an unsatisfactory financial reputation.

Because "a satisfactory financial reputation" is similar to "a good financial reputation", that is, *RAT→REP,* therefore, the loan officer uses the above SMT to make the decision and obtain ¬*REP*: The applicant has an unsatisfactory credit rating, since "a good credit rating" is similar to "a satisfactory credit rating".

In the context of fuzzy similarity relations or similarity metrics [22], using the compositional rule of inference [1] to the above Equation (15) we obtain:

$$\widetilde{P_0} = 1 - (1 - \widetilde{Q_1}) \circ \widetilde{F_{10}} \circ \widetilde{F_{11}} \circ \widetilde{F_{01}} \tag{16}$$

This is a computational foundation for similarity-based modus tollens. In the case $Q \approx Q'$, $\widetilde{F_{10}}$ is a unit metric, and Equation (16) is then simplified into

$$\widetilde{P_0} = 1 - (1 - \widetilde{Q_1}) \circ \widetilde{F_{11}} \circ \widetilde{F_{01}} \tag{17}$$

Similarity-based abduction

Abduction has been used in system diagnosis or medical diagnosis [8] and scientific discovery [34].

Abduction is an important reasoning paradigm in SBR. Similarity-based abductive reasoning (SAR) is a natural development of abductive reasoning [35], or an application of SBR in abductive reasoning. Its general form is as follows:

$$\frac{Q', Q' \approx Q, Q \to P, P \to P'}{\therefore P'} \tag{18}$$

Example 6. Similarity-based abductive reasoning. As in Example 5, let:

- *RAT*: The applicant has a good credit rating,
- *REP*: The applicant has a good financial reputation,
- The loan officer has an experienced rule, *RAT→REP*: If the applicant has a good credit rating, then the applicant has a good financial reputation.

In this case, the loan officer knows the information from applicant *A*, *REP′* : The applicant has a satisfactory financial reputation. Because "a satisfactory financial reputation" is similar to "a good financial reputation"; that is, *REP* ~ *REP′*, the loan officer uses the above similarity-based abductive reasoning to make the decision and obtain *REP′*: The applicant has a satisfactory credit rating, because "a good credit rating" is similar to "a satisfactory credit rating". It is obvious that "The applicant has a satisfactory credit rating" is an explanation for "The applicant has a satisfactory financial reputation." Therefore, similarity-based abductive reasoning can be also used for generations of explanation, as abductive reasoning does scientific discovery [34,36].

In the context of fuzzy similarity relations and similarity metrics, using the compositional rule of inference [1] to the above Equation (18), we obtain:

$$\widetilde{P_0} = \widetilde{Q_1} \circ \widetilde{F_{10}} \circ \widetilde{F_{11}} \circ \widetilde{F_{01}} \tag{19}$$

This is a computational foundation for similarity-based abductive reasoning. In the case of $Q \approx Q'$, $\widetilde{F_{10}}$ is a unit metric, and Equation (19) is then simplified into:

$$\widetilde{P_0} = \widetilde{Q_1} \circ \widetilde{F_{11}} \circ \widetilde{F_{01}} \quad (20)$$

5.3 Summary

Table 1 summarizes the well-known inference rules: Modus ponens, modus tollens, abduction, and proposes three inference rules with respect to SBR, corresponding to the traditional forms: Modus ponens, modus tollens, and abduction [37,31]. So far, we have examined three different inference rules for SBR (see **Table 1**) in a unified viewpoint, each of them has been thoroughly used in computer science, mathematics, mathematical logic [38], and other sciences [30,39,34]. However, they are all the abstractions and summaries of SBR, natural reasoning, and ordinary reasoning in the real world. Furthermore, CBR has been only based on either modus ponens or modus tollens or abduction [33,8], whereas SBR is based on the mentioned three inference rules. It should be noted that reasoning paradigms can be classified into simple (atomic or first level) reasoning paradigms and composite (second level) reasoning paradigms [40], just as propositions can be divided into simple (atomic) propositions and compound propositions [39]. The simplest reasoning paradigm is an inference rule, which is the basis for any reasoning paradigm.

A composite reasoning paradigm consists of more than one inference rule. For example, fuzzy modus ponens [2] is a composite reasoning paradigm that integrates modus ponens and fuzzy rules. Any process model of a reasoning paradigm in AI is a method for obtaining composite reasoning paradigms. For example, the simplest rule-based expert system (RBES) can mainly consist of the knowledge base (KB) and an inference engine (IE), where IE is an inference mechanism for performing modus ponens or modus tollens or abduction. However, in order to manipulate the knowledge in the KB, the RBES must deal with knowledge representation, knowledge explanation, and knowledge utility which are the main components of the process model [11,8]. Therefore, the reasoning involved in RBES can be considered as a composite reasoning paradigm. In this way, we can differentiate reasoning paradigms in mathematical logic and AI. What we have examined in this article are simple or atomic inference rules for SBR. In future work, we will examine composite reasoning paradigms for SBR, which constitute a "reasoning chain" [3], "reasoning network" or "reasoning tree" with some depth, and correspond to natural reasoning in human professional activities.

It should be noted that the above-mentioned abductive reasoning and its SBR are unsound reasoning paradigms from a logical viewpoint [31]. However, like nonmonotonic reasoning, which is also unsound reasoning [8], this inference rule and its similarity-based abduction is the summarization of SBR used by people in the real-world situations.

6. Similarity computing and analytics

Similarity computing and analytics are science, technology, system and tools used in data, information, and knowledge analysis to measure and compare the similarity between different data, information, and knowledge sets. They are used in various fields such as AI including machine learning, data science, natural language understanding and processing, image recognition, and information retrieval. This section will examine similarity computing and

Table 1. Three inference rules for similarity-based reasoning.

	Modus ponens	Modus tollens	Abduction
Traditional form	$\dfrac{P, P \to Q}{\therefore Q}$	$\dfrac{\neg Q, P \to Q}{\therefore \neg P}$	$\dfrac{Q, P \to Q}{\therefore P}$
similarity-based form	$\dfrac{P', P' \sim P, \; P \to Q, Q \approx Q'}{\therefore Q'}$	$\dfrac{\neg Q', Q' \approx Q, \; Q \to P, P \sim P'}{\therefore \neg P'}$	$\dfrac{Q', Q' \approx Q, \; Q \to P, P \sim P'}{\therefore P'}$

analytics in some detail.

Similarity computing is a science, technology, system, and tool for determining the degree of similarity or dissimilarity between two or more objects to create intelligence. That is, based on the research of Sun [41],

$$\begin{aligned}\text{Similarity computing} = {} & \text{Similarity science}\\ & + \text{Similarity engineering}\\ & + \text{Similarity technology}\\ & + \text{Similarity system}\\ & + \text{Similarity tools} \quad (21)\end{aligned}$$

Similarity relations, fuzzy similarity relations [2] and similarity metrics [1] such as Cosine similarity, Jaccard similarity, Euclidean distance, and Pearson correlation coefficient, among others are fundamentals for realizing similarity or dissimilarity between two or more objects for similarity computing [18].

Analytics is science, technology, system and tools for mining data, information, knowledge to discover meaningful intelligence, insights, patterns, and knowledge from big data in a database or data warehouse or knowledge in a knowledge base using similarity [42]. This can be achieved using database and data warehouse techniques, statistical techniques, knowledge base techniques, data visualization techniques, machine learning algorithms, and other data and knowledge processing tools [4,43]. Similarity analytics can be represented below [41],

$$\begin{aligned}\text{Similarity Analytics} = {} & \text{Similarity science}\\ & + \text{Similarity engineering}\\ & + \text{Similarity technology}\\ & + \text{Similarity system}\\ & + \text{Similarity tools} \quad (22)\end{aligned}$$

Basically, similarity analytics is a part of similarity computing, just as analytics is a part of computing [41]. Both aim to discover similarity intelligence in the domain. Even so, not only similarity computing but also similarity analytics can enable the analysis of large datasets, information sets and knowledge sets to identify and discover intelligence, patterns, knowledge and insights, and prediction of outcomes or cases. For example, in machine learning, similarity computing is used to find similarities between different data points, and analytics is used to train models that can make predictions based on those similarities [4,18].

Although similarity science has not been proposed in academia, similarity engineering, similarity technology, similarity systems (see the next section) and similarity tools based on similarity models, methods, and algorithms are well-known in the market [1,7,44].

Overall, similarity computing and analytics are science, technology, and system in modern data, information, and knowledge analysis to enable researchers and practitioners to gain similarity intelligence, knowledge and insights, and make predictions in various fields.

7. A multiagent SBR systems

In AI, a reasoning paradigm usually corresponds to an intelligent system. This section proposes a multiagent SBR system as an example, which constitutes an important basis for developing any multiagent SBR systems (MSBRS).

7.1 A general architecture of an SBR system

Similarity case base (SCB) is similar to a case base in a case base system [8] illustrated in **Figure 1**. The SCB is a text case base in natural language processing systems [4] and an insight base in data mining system and data analytics systems [41]. SCB consists of all the cases that the SBR System collects periodically. A user interface is used to interact with the SCB and MIE in the SBR System (see Section 7.3). The MIE is a multi-inference engine that consists of the mechanism for implementing three reasoning paradigms based on the above-mentioned three similarity-based inference rules and their algorithms for SBR with manipulating the SCB to infer similarity-based problem solving and decision making requested by the user. The remarkable difference between the mentioned SBRS and the traditional CBR system (CBRS) lies in that the latter's inference engine is based on a unique reasoning paradigm (or inference rule), while the MIE is based on many different reasoning paradigms. This implies that a CBRS is only a subsystem of the SBRS. Therefore, this SBR System is the extension of CBRS and similarity-based reasoning [8,31].

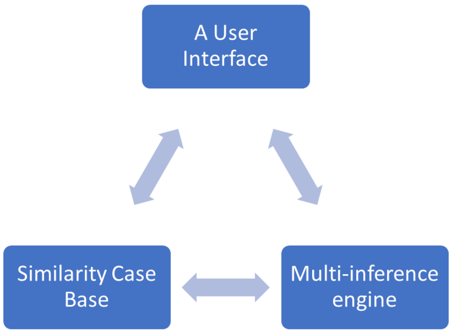

Figure 1. A general architecture for a SBR system.

7.2 MEBIE: A multiagent framework for similarity based inference engine

As mentioned in the previous subsection, the MIE is a multi-inference engine for SBRS [45]. MIE could automatically adapt itself to the changing situation and perform one of the mentioned similarity-based inference rules for SBR (**Figure 2**). However, any existing intelligent system has not reached such a high level [46]. The alternative strategy is to use multiagent technology to implement the MIE. Based on this idea, we propose a multi-agent framework for a similarity-based inference engine (for short MABIE), which is a core part of a multiagent SBR system (MSBRS), as shown in **Figure 1**. In this framework, three rational agents (from SMP agent to SAR agent) are semiautonomous [8]. These three agents are mainly responsible for performing SBR corresponding to three similarity-based inference rules in the SBRS respectively. In what follows, we discuss each of them in some detail.

1) The SMP agent in the MABIE is responsible for manipulating the SCB based on similarity-based modus ponens and its algorithm (also see Section 5.1) to infer the similarity-based problems and solutions requested by the user. This agent can be considered as an agentization of an inference engine in a traditional CBR system. The function of the SMP agent can be extended to infer the cases in the SCB based on fuzzy modus ponens [46,23].

2) The SMT agent manipulates the SCB to infer the case requested by the user based on similarity-based modus tollens and its algorithms (see Section 5.2).

3) The SAR agent is responsible for manipulating the SCB to infer the case requested by the user based on similarity-based abductive reasoning and its algorithm (see Section 5.3). This agent can generate the explanation for the experience-based reasoning inferred by the MEBIE. This agent can be considered as an agentization of an inference engine in an abductive CBR system [33].

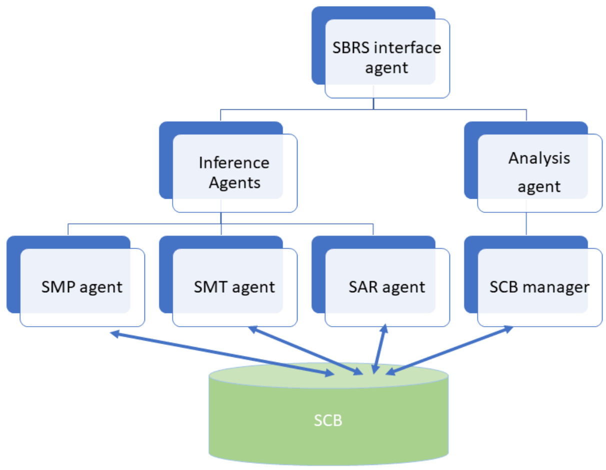

Figure 2. MIE and other agents in a MSBRS.

7.3 Some other agents in MSBRS

For the proposed MSBRS, there are some other intelligent agents, shown in **Figure 2**. These are an interface agent, an analysis assistant and a SCB manager. In what follows, we will look at them in some depth [45].

The SBRS interface agent is an advisor to help the MSBRS user to know which reasoning agent she/he should ask for help. Otherwise, the SBRS interface agent will forward the problem of the user to all agents in the MIE for further processing.

The output provided by the MIE can be considered as a sub output. The final output as the solutions to the similarity-based problem of the user will be processed with the help of the analysis agent. Since different agents in the MIE use different inference rules, and then produce different, conflicting results with knowledge inconsistency. How to resolve such knowledge inconsistency is a critical issue for the

MSBRS. This issue will be resolved by the Analysis assistant of the MSBRS. The analysis assistant will:

- Rank the degree of importance of the sub outputs from the MAMIE taking into account the knowledge inconsistency,
- Give an explanation for each of the outputs from the MIE and how the different results are conflicting,
- Combine or vote to establish the best solutions,
- Forward them to the SBRS interface agent who then forwards them to the user.

The SCB manager is responsible for administering the SCB. Its main tasks are SCB creation and maintenance, similarity case base evaluation, reuse, revision, and retention. Therefore, the roles of the SCB manager are an extended form of the functions of a CBR system [8], because case base creation, case retrieval, reuse, revision and retention are the main tasks of the CBR system [16].

7.4 Workflows of agents in MSBRS

Now let us have a look at how the MSBRS works. The user, U, asks the SBRS interface agent to solve the problem, *p*. The SBRS interface agent asks U whether a special reasoning agent is needed [45]. U does not know. Thus, the SBRS interface agent forwards p (after formalizing it) to all agents in the MIE for further processing. The agent in the MIE manipulates the case in the SCB based on p, and the corresponding reasoning mechanism, and then obtains the solution, which is forwarded to the Analysis assistant. After the Analysis assistant receives all solutions to p, it will rank the degree of importance of the solutions, give an explanation for each of the solutions and how the results are conflicting or inconsistent, and then forward them (with p) to the SBRS interface agent who would then forward them to U. If U accepts one of the solutions to the problem, then the MSBRS completes this mission. In this case, the SCB manager will look at whether this case is a new one. If yes, then it will add it to the SCB. Otherwise, it will keep some routine records to update the SCB. If U does not accept the solution provided, the SBRS interface agent will ask U to adjust some aspects of the problem p, which is changed into *p*', then the SBRS interface agent will once again forward the revised problem *p*' to the MIE for further processing.

8. Conclusions

Artificial intelligence (AI) has addressed experience-based intelligence and knowledge-based intelligence at their early stage. Big data has been experiencing significant progress in the past 10 years, AI has been developing machine learning and deep learning to address data-based Intelligence. In fact, similarity intelligence has been accompanying experience-based intelligence, knowledge-based intelligence, and data-based Intelligence to play an important role in computer science, AI, and data science in general and similarity computing and analytics in particular. The main contributions of this article are:

1) It explored similarity intelligence, based on the similarity discovered from experience-based intelligence in machine learning and CBR. Similarity intelligence will be developed and created by many systems and algorithms in AI, computer science, and data science.

2) It explored similarity-based reasoning and proposed its three different rules, which constitute the fundamentals for all SBR paradigms.

3) It highlighted similarity-based reasoning, computing, and analytics to create similarity intelligence. As an example, the article also proposed a multiagent architecture for an SBR system (MSBRS).

Overall, similarity intelligence is discovered from big data, information, and knowledge using similarity relations, fuzzy similarity relations and metrics, SBR, similarity computing and semantics.

Furthermore, the similarity-based approach to similarity intelligence, SBR, similarity computing and analytics proposed in the article opens a new way to integrate machine learning (e.g. machine learning algorithms such as instance-based learning and k-Nearest Neighbor (kNN) classifier and experience-based reasoning based on SBR, which will be examined in future work. Knowledge management and experience management have drawn increasing attention in business, e-commerce, and computer science. Their correspondence to intelligent systems is similarity-based systems such as CBR systems and machine learning. How to apply similarity intelligence in Knowledge management, experience management, and similarity-based systems will be also examined in future work.

Measurement of intelligence is based on the ability to solve difficult problems. How to define the measurement of similarity intelligence is still a weakness of this article. In future work, we will explore the measurement of similarity intelligence.

Conflict of Interest

There is no conflict of interest.

References

[1] Zimmermann, H.J., 2011. Fuzzy set theory—and its applications. Springer Science & Business Media: Berlin.

[2] Zadeh, L.A., 1971. Similarity relations and fuzzy orderings. Information Sciences. 3(2), 177-200.

[3] Minsky, M., 1988. Society of mind. Simon and Schuster: New York.

[4] Aroraa, C., Chitra, L., Munish, J., 2022. Data analytics: Principles, tools, and practices. BPB Publications: New Dalhi.

[5] Sun, Z., 2022. A mathematical theory of big data. Journal of Computer Science Research. 4(2), 13-23.

[6] Zhang, D.G., Ni, C.H., Zhang, J., et al., 2022. A novel edge computing architecture based on adaptive stratified sampling. Computer Communications. 183, 121-135.

[7] Milošević, P., Petrović, B., Jeremić, V., 2017. IFS-IBA similarity measure in machine learning algorithms. Expert Systems with Applications. 89, 296-305.

[8] Finnie, G., Sun, Z., 2004. Intelligent techniques in E-commerce: A case based reasoning perspective. Springer-Verlag: Berlin.

[9] Sun, R., 1995. Robust reasoning: Integrating rule-based and similarity-based reasoning. Artificial Intelligence. 75(2), 241-295.

[10] Bergmann, R., 2002. Experience management: Foundations, development methodology, and internet-based applications. Springer: Berlin.

[11] Russell, S., Norvig, P., 2020. Artificial intelligence: A modern approach (4th Edition). Prentice Hall: Upper Saddle River.

[12] Laudon, K.G., Laudon, K.C., 2020. Management information systems: Managing the digital firm (16th Edition). Pearson: Harlow.

[13] López-Robles, J.R., Otegi-Olaso, J.R., Gómez, I.P., et al., 2019. 30 years of intelligence models in management and business: A bibliometric review. International Journal of Information Management. 48, 22-38.

[14] Turing, A., 1950. Computing machinery and intelligence. Mind. 49, 433-460.

[15] Schwab, P.N., 2023. ChatGPT: 1000 Texts Analyzed and up to 75,3% Similarity [Internet] [cited 2023 Mar 17]. Available from: https://www.intotheminds.com/blog/en/chatgpt-similarity-with-plan/

[16] Sun, Z., Finnie, G., Weber, K., 2004. Case base building with similarity relations. Information Sciences. 165(1-2), 21-43.

[17] Finnie, G., Sun, Z., 2003. R5 model for case-based reasoning. Knowledge-Based Systems. 16(1), 59-65.

[18] Kantardzic, M., 2011. Data mining: Concepts, models, methods, and algorithms. John Wiley & Sons: Hoboken.

[19] Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. Science. 349(6245), 255-260.

[20] Epp, S.S., 2010. Discrete mathematics with applications. Cengage Learning: Boston.

[21] Zhang, D.G., Ni, C.H., Zhang, J., et al., 2022. New method of vehicle cooperative communication based on fuzzy logic and signaling game strategy. Future Generation Computer Systems. 142, 131-149.

[22] Finnie, G., Sun, Z., 2002. Similarity and metrics

in case-based reasoning. International Journal of Intelligent Systems. 17(3), 273-287.

[23] Zhang, D., Wang, W., Zhang, J., et al., 2023. Novel edge caching approach based on multi-agent deep reinforcement learning for Internet of vehicles. IEEE Transactions on Intelligent Transportation Systems. 24(6), 1-16.

[24] Klawonn, F., Castro Peña, J.L., 1995. Similarity in fuzzy reasoning. Mathware & Soft Computing. 2(3), 197-228.

[25] Fontana, F.A., Formato, F., 2002. A similarity-based resolution rule. International Journal of Intelligent Systems. 17(9), 853-872.

[26] Biacino, L., Gerla, G., Ying, M., 2000. Approximate reasoning based on similarity. Mathematical Logic Quarterly. 46(1), 77-86.

[27] Kundu, S., 2000. Similarity relations, fuzzy linear orders, and fuzzy partial orders. Fuzzy Sets and Systems. 109(3), 419-428.

[28] Ovchinnikov, S., 1991. Similarity relations, fuzzy partitions, and fuzzy orderings. Fuzzy Sets and Systems. 40(1), 107-126.

[29] Bogacz, R., Giraud-Carrier, C., 2000. A novel modular neural architecture for rule-based and similarity-based reasoning. Hybrid neural systems. Springer: Berlin. pp. 63-77.

[30] Hüllermeier, E., 2001. Similarity-based inference as evidential reasoning. International Journal of Approximate Reasoning. 26(2), 67-100.

[31] Sun, Z., 2017. A logical approach to experience-based reasoning. New Mathematics and Natural Computation. 13(1), 21-40.

[32] Loia, V., Senatore, S., Sessa, M.I., 2004. Combining agent technology and similarity-based reasoning for targeted E-mail services. Fuzzy Sets and Systems. 145(1), 29-56.

[33] Sun, Z., Finnie, G., Weber, K., 2005. Abductive case-based reasoning. International Journal of Intelligent Systems. 20(9), 957-983.

[34] Magnani, L., 2011. Abduction, reason and science: Processes of discovery and explanation. Springer Science & Business Media: Berlin.

[35] Baral, C., 2000. Abductive reasoning through filtering. Artificial Intelligence. 120(1), 1-28.

[36] Console, L., Dupré, D.T., Torasso, P., 1991. On the relationship between abduction and deduction. Journal of Logic and Computation. 1(5), 661-690.

[37] Sun, Z., Finnie, G., Sun, J. (editors), 2005. Four new fuzzy inference rules for experience based reasoning. Fuzzy Logic, Soft Computing and Computational Intelligence (IFSA2005); 2005 May 30; Beijing. p. 188-193.

[38] Reeves, S., Clarke, M., 1990. Logic for computer science. Addison-Wesley: Wokingham.

[39] Hurley, P.J., 2000. A concise introduction to logic. Thomson Learning: Wadsworth.

[40] Dosen, K., 1993. A historical introduction to substructrual logics. Substructrual logics. Clarendon Press: Oxford. pp. 1-30.

[41] Sun, Z., 2022. Problem-based Computing and Analytics. International Journal of Future Computer and Communication.11(3), 52-60.

[42] Sun, Z., Stranieri, A., 2021. The nature of intelligent analytics. Intelligent analytics with advanced multi-industry applications. IGI-Global: Hershey. pp. 1-22.

[43] Sun, Z., Pambel, F., Wu, Z., 2022. The elements of intelligent business analytics: Principles, techniques, and tools. Handbook of research on foundations and applications of intelligent business analytics. IGI-Global: Hershey. pp. 1-20.

[44] Iantovics, L.B., Kountchev, R., Crişan, G.C., 2019. ExtrIntDetect—A new universal method for the identification of intelligent cooperative multiagent systems with extreme intelligence. Symmetry. 11(9), 1123.

[45] Sun, Z., Finnie, G., 2016. A Similarity Based Approach to Experience Based Reasoning (Prepprint). Available: https://ro.uow.edu.au/

[46] Sun, Z., Finnie, G., 2005. MEBRS: A multiagent architecture for an experience based reasoning system. Knowledge-based intelligent information and engineering systems. Springer: Berlin. pp. 972-978.