

Paper ID: **Case_Based_Reasoning-7**

1. Write a short (2-3) sentence about the paper.

The paper presents the general CBR category where solutions to previous cases are applied in order to solve new problems. The focus is on this overarching topic in every part of the paper. Similarity and its importance for CBR is presented in part 3, and the section for current applications discusses a generalized CBR cycle with focus on similarity, based on the additional course material paper of Stahl (2005).

2. Relevance: Does the paper fit in the overall topic?

The paper fits well with the overall topic of CBR, but this follows from the fact that it mainly focuses on the CBR category in general. Despite that the concept of similarity is explored, the paper fails the task to select and focus on a specific method within the category of CBR. Similarity can be a part of this method, but in that case it should in my opinion be contextualized with a specific implementation of the retrieval step and presented in part 2. The lack of a specific method within the category of CBR is a problem with the foundation of the paper, which naturally leads to other shortcomings. Part 3 and 4 are not answering the given tasks in my opinion, as they both focus on new (general) theory on CBR instead of presenting improvements to a method / alternative methods (part 3) and current research (part 4).

Relevance:	1 (poor)	2	3 (good)	4	5 (excellent)
-------------------	----------	---	-----------------	---	---------------

3. Technical Soundness

The theoretical foundations of the general CBR method is presented in a good way, and the theory on similarity is also satisfactory. The paper is however not very technically comprehensive, and I think this follows from the lack of focus on a specific method within CBR. As mentioned throughout this review, I think it is a good idea to choose such a method in order to apply more of the curriculum and show that you are able to go in depth on more specific/advanced theory from the field of Machine Learning.

In part 2 there are no formulas at all, but the CBR-cycle is presented systematically using general examples and theory from the mandatory course material. In part 3 when addressing similarity, the paper applies some formulas in order to describe the objects that are to be compared and how the sum of different (weighted) aspects of similarity accounts for the total similarity. It is good that you explain the variables!

The theory is quite simple and it is easy to follow the explanation, however there are some gaps. How to actually calculate similarity is only described in words, and I think this is where there is room for going more in depth and discussing one of the many specific methods from ML, using even more formulas and theoretical justification. As my list of suggestions below imply, I think this is a perfect opportunity to write about a specific implementation of the retrieval step with focus on a supervised learning method in order to calculate similarity, e.g. K-NN to retrieve the k most similar cases to the object problem. It is important to address this in part 2, since you are supposed to write about improvements or alternative methods in part 3.

In my opinion, the lack of specific theoretical focus results in below average technical soundness, and should be improved/elaborated in the final paper.

Technical Soundness:	1 (poor)	2	3 (good)	4	5 (excellent)
-----------------------------	----------	----------	----------	---	---------------

4. Clarity and Presentation

The presentation of the paper is quite clear. The arguments are logical and build on each other to explain general CBR. The written language is solid with few grammatical errors. There are however some things to note. It is supposed to be a scientific paper, and the words "we", "us" and "our" occur way too many times in my opinion. These words should be used to reflect on own work, but you actively use them a lot throughout the entire paper. This, in addition to the overuse of adjectives (e.g. in the first paragraph in part 2) makes the text a bit informal. I would also advise against your current use of footnotes and extremely direct addressing of the reader (e.g. in the last paragraph, page 2), since this also contradicts the formality required by a scientific paper. I think the length of your paragraphs are great, but I miss an active use of subsections. If you implement at least some subheadings, it would make the paper better structured and it is a lot easier for the reader to follow your presentation of theory and arguments.

Clarity:	1 (poor)	2	3 (good)	4	5 (excellent)
-----------------	----------	---	-----------------	---	---------------

5. Does the paper point out current (successful) use of the method?

In my opinion the paper fails to point out current (successful) use of the method. The only research that is referred to is the additional course material of Stahl from 2005 (!!!). This is applied in order to present an even more general approach to CBR, with the justification of shortcomings to the classical CBR-cycle. This does not fulfill the given task of presenting research from the last 5 years, and the paper does not investigate other research than provided in the course. As mentioned, I think some of the reason for this problem with the paper follows from the fact that you haven't focused on a specific implementation of a method within CBR. I have elaborated this throughout the review, and in main suggestions for improvement (3.) below, I provide concrete advice on how to improve this section of your paper.

Current applications:	1 (poor)	2	3 (good)	4	5 (excellent)
------------------------------	-----------------	---	----------	---	---------------

6. List 3 good points of the paper and make 3 suggestions how to improve the paper.

Good:

1. Good presentation and discussion of the general CBR method.
2. Good understanding and explanation of similarity
3. Good written language with few grammatical errors + great length of paragraphs.

Main suggestions for improvement:

1. **Whole paper / part 2:** Focus a larger proportion of the paper (part 2) on one specific element within the CBR-cycle and explain an implementation of this in depth using relevant theory, formulas and figures. In part 3 you focus on similarity – this is something you can use as the main focus of the paper, preferably in combination with an implementation of a specific ML method (See 7. *List of suggestions (a.)* below for examples). It is in part 2 you should explain the foundations for this. I.e. in part 2, commence with presenting the general CBR and quickly shift focus over to the specific implementation of an element in the cycle. In part 1 it is a good idea to contextualize the chosen method within the taxonomy of ML. E.g. if you choose K-NN & similarity, this should be related to supervised learning. (You already talk about distance metrics in part 3, and it is possible to incorporate this with a K-NN approach to the paper)

2. **Whole paper / part 3:** I think you have misunderstood part 3. As of now, you present new theory on similarity, a concept in CBR, while you by definition of the task should either present improvements to the method or alternative methods. This can involve similarity, and e.g. if you choose retrieval with K-NN, I recommend you to compare this with some of the other options I present in 7. *List of suggestions (a.)*.
 3. **Part 4:** In the first paragraph of part 4 you state that CBR can make use of a broad range of topics from ML, more specific supervised learning and k-means clustering (unsupervised). You also state that it is hard to "nail down one specific paper or application" because of the many options to choose from. Instead of using 2,5 pages to present more general theory on CBR, I highly recommend that you choose one or two specific implementations/current research relevant to the element in the CBR-cycle you choose as the main area of focus in your paper. ICCBR is a yearly conference on CBR and present the latest research from the field of study. If you google "ICCBR <year> papers" you will get access to these. After a quick look I found several relevant papers using supervised learning methods for retrieval/similarity in CBR, including some papers discussing K-NN. PS, here is a link to the 2020 papers (not published yet) <https://drive.google.com/drive/folders/1VC2xRnioGPij2H105nZQ452puSI-3t29>
7. List of suggestions / recommendations
- a. The specific element you choose to examine more in depth can be either one in the CBR-cycle, but I highly recommend **Retrieval**. This aligns well with your current work on similarity! There are several different approaches for retrieval/similarity and a lot of good literature on the alternatives, some of which are:
 - i. K-NN as a Similarity measure. (<https://youtu.be/09mb78oiPkA>). This is probably the simplest way to implement CBR. Two possible metrics to investigate are Hamming- and Euclidean distance.
 - ii. MAC/FAC for Similarity based retrieval. [Potential source: MAC/FAC: A Model of Similarity-based Retrieval. Forbus, Genter & Law (1994), Cognitive Science (p. 141-205)]
 - iii. Bayesian analysis, similarity/retrieval. Source: <https://folk.idi.ntnu.no/kerstinb/paper/2018-ICCBR-Nikpour-et-al.pdf>
 - iv. KD-Trees.
 - b. You currently have approximately 1.5 pages of available space – make use of this.
 - c. In part 2 you use bold font both when addressing the elements of the CBR-cycle and for highlighting "similarity" and "adapting". This gets a bit confusing, and you could for example use italics when referring to the elements in the cycle.
 - d. Even though the thought is good, I would reconsider using the last sentence in the first paragraph of the introduction, since I find the metaphor unreasonable.
 - e. On page 5, I think the bank loan example using the table needs some more work. You state that it is tricky to implement a simple comparison between two variables into your function (and therefore need to introduce a new, virtual variable), but I do not understand why this is the case?
 - f. OBS: Sthal vs. Stahl
 - g. "Further work" in the conclusion is very thin.

I have tried to provide constructive feedback continuously throughout this the review, and I hope that you find the time to look into my suggestions. I wish you the best of luck!