User Evaluation to Measure the Perception of Similarity Measures in Artworks

Belén Díaz-Agudo[0000−0003−2818−027X], Guillermo Jimenez-Diaz[0000−0003−4070−825X], and Jose Luis Jorro-Aragoneses[0000−0002−2917−3912]

Department of Software Engineering and Artificial Intelligence
Instituto de Tecnologías del Conocimiento
Universidad Complutense de Madrid, Spain
{belend,gjimenez,jljorro}@ucm.es

Abstract. Similarity measures do not typically capture subjective elements of perception of similarity. Our research contributes an experimental methodology for validating and learning similarity computation algorithms against human perceptions of similarity in subjective domains like Art and emotions. In this paper, we explain the first experiment to check our hypotheses and methodology. We have obtained promising results that explain the differences between users profiles and their perception of similarity between artworks and how to combine local similarity functions to be able to compute similarity measures reflecting users' perception.

Keywords: Similarity measures, Arts, Similarity perception

1 Introduction

The well-known main assumption in case-based reasoning (CBR) relies on the hypothesis that similar problems should have similar solutions. No need to say that similarity is a core concept for different processes of the CBR cycle. Similarity between the query and the cases is typically computed using the description features represented using attribute-value pairs. These features can be simple, textual or, in some applications, it may be necessary to use derived features obtained by inference based on domain knowledge. In yet other applications, cases are represented by complex structures (such as graphs or first-order terms) and retrieval requires an assessment of their structural similarity. As might be expected, the use of deep features or structural similarity is computationally expensive; however, the advantage is that relevant cases are more likely to be retrieved [14]. Our research group has previous works on semantic and structured semantic similarity with ontological an taxonomic knowledge[20, 21, 9].

In this paper we highlight that *similarity measures* on the structures that represent the cases (either attribute-value or graphs) does not typically capture *subjective* elements of perception of similarity. For example, when comparing two menus that are similar for me, the similarity can be due to the fact that both include my favourite meals and not because of the similarity between their ingredients. There are other examples, like recommender systems for songs or movies, where retrieval performance is affected when similarity between items depends on subjective criteria for different users. There are many different application domains where human subjectivity is an issue. However, perception of similarity is difficult to measure, even if the item's descriptions is structured and includes semantic domain knowledge. In this paper we present a case study in the Art domain where similarity perception is clearly a subjective criteria.

The context of the research conducted in this paper is the SPICE project[1]. The overall aim of SPICE is to develop tools and methods to support Citizen Curation [7], in which citizens actively engage in curatorial activities in order to learn more about themselves and develop a better understanding of, and empathy for, other communities.

One challenge is to be able to identify communities of citizens that allow the reflection processes inside and between communities. This is a two way process. On one side, community detection relies heavily on the definition and use of semantic similarity measures over complex graph structures representing citizens, opinions, artworks, contributions, reflections and emotions [2]. On the other side, communities of users can be seen as useful resources to identify common profiles from the similarity perception point of view.

Our research contributes an experimental method for validating similarity computation algorithms against human perceptions of similarity. Such validation enables researchers to ground their similarity methods in context of intended use instead of relying on assumptions of fit. In addition to the methodology, this paper presents the results of experimentation using real data with artworks from the Prado Museum. We also present some analysis of potential causes of differences between the compared cases in which this model matches human perceptions of similarity. This method will allow to personalise a similarity measure to compare items using subjective perception criteria adapted to the user who compares the items. She is the user that retrieves the case in CBR systems, or the user that get the recommendation in a recommender system or for any other applications relying on similarity computation where similarity perception is an issue.

The paper runs as follows. Section 2 reviews some related work about similarity. Section 3 describes out methodology for capturing and learning knowledge reflecting human perceptions of similarity. Section 4 describes an experiment associated to the step 1 of our methodology. In the experiment with Art data from the Prado Museum we define local similarity measures for comparing attributes of artworks and validate them regarding user perception of similarity. Section 5 concludes the paper and review some lines of future work.

---

[1] Social cohesion, Participation, and Inclusion through Cultural Engagement - Horizon 2020 programme https://spice-h2020.eu/

[2] SPICE relies in Linked Data technologies that include a huge mass of interlinked knowledge

2 Related work about Similarity

CBR relies strongly in similarity computation. However, Similarity measures are considered also essential tools to solve problems in a broad range of AI domains and applications, specially when semantic matters. For example semantic web and linked data [29], recommender systems [11], Natural Language Processing [22], Information Retrieval, Knowledge Engineering [3],and many others. There are several similarity measures that have been used in CBR systems, and some comparison studies and frameworks exist [18, 14]. The results obtained in these studies show that the different similarity measures have a performance strongly related to the type of attributes representing the case and to the importance of each attribute. Thus, it is very different to deal with only continuous data, with ordered discrete data or non-ordered discrete data. In [15] authors distinguish between case similarity measures that are learnt from data and those that are typically modelled by experts with the relevant domain knowledge together with CBR experts, who know how to encode this domain knowledge into the similarity measures by selecting what are the properties of the case descriptions have more impact in the similarity of the solutions. For example, in a cars for sale application the amount of miles driven has a greater importance than the color of the car [6].

When dealing with conceptual background domain models, like graphs, networks or taxonomies, another possibility is the representational approach that assigns similarity meaning to the path joining two individuals. In general, a graph-based semantic similarity measure is a mathematical tool used to estimate the strength of the semantic interaction between entities (concepts or instances) based on the analysis of ontologies[12]. Similarity is computed for a given pair of individuals. An individual is defined in terms of the concepts of which is an instance and the properties asserted for it, which are represented as relations connecting the individual to other individuals or primitive values (fillers). In [9] we have described a similarity framework where we distinguished between the *structural similarity* that will be computed based on the composition relations (part-of, has-part), the *semantics similarity* is due to all the concepts and relations describing the meaning of the case, and the *contextual similarity* that depends on the case context relations and the *adaptation similarity* that will use the adaptation related knowledge (also used in previous approaches like [23]). Note that the application of this measure is strongly dependent on the availability of an ontology or conceptual model that represents the application domain. The work in [17] classifies the distance and similarity functions on graph-based representations in four types: (1) graph matching, (2) based on edit distances, (3) based on the types of relationships and refinement operators and (4) based on kernels.

Regarding perception of similarity related work exists in the field of human psychology, where similarity is defined a relationship that holds between two perceptual or conceptual objects and serves to classify objects, form concepts and make generalizations [26]. As it is noted in [5], similarity between objects is not solely dependent on the characteristics of those objects. It is also affected by the context, and by other present and immediately past stimuli, as well as long-term experience with related objects. A well-known example is that humans have the effect of experience on similarity among phonemes. Native English speakers find spoken "L" and "R" quite distinct, whereas to native speakers of Japanese they sound extremely similar.

Perception of similarity is also relevant when dealing with textual representations. For example, in [25] authors deal with the problem of navigation in large text collections (blogs, forums, idea management systems, online deliberation platforms...), and analyse how the algorithmic similarity measures being used match up with human perceptions. They found out that in favourable conditions human similarity judgements and algorithmic similarity measurement often (75%) agree. However, that agreement is not so good (66%) when documents are selected more generally. Other previous studies have also examined the match between typical algorithmic similarity based approaches (such LDA, or cosine similarity) and human perceptions of text document similarity. For example, in [27] authors compare the relevance (according to human judges) of the results of the retrieval task on an abstracted document collection given an information-need query. Also related is the work of [10] that refers to the individual word level. They compared the computed cosine similarity between feature vectors that incorporated information from lexicons and large corpora, against benchmark datasets containing pairs of English words that had been assigned similarity ratings by humans, finding out discrepancies between the perceived and the computed similarity.

In [4] authors describe a CBR system that helps the users make online privacy decisions by identifying similar situations from the past. They calculate the similarity between privacy policies and provide results from a focus group study on the perceived similarity of data items and data handling purposes from a privacy point of view. Particular attention has been also placed by the similarity perceived by experts on the use of analogical reasoning [13, 28].

3 Methodology for learning similarity measures reflecting human perceptions

In this section, we propose a methodology for the construction of similarity measures that reflect the perception of similarity. The challenge is that similarity perception is different for different people, so it can not be computed with a common similarity measure that is shared for all the users.

Our proposal aims at configuring different similarity measures for different users and being able to generalise these measures for users of the same *profile*, supposing that users who belong to the same profile have similar perception of similarity. Our methodology relies on the following hypothesis:

1. **Hypothesis 1**. Different users have different perceptions of similarity and consider differently the attributes describing the items. In this paper, we study how local similarity measures on the individual attributes relate with the perceived similarity in each one of these attributes.

2. **Hypothesis 2**. Users can be grouped together using profiles and users from the same profile have similar perceptions of similarity. That means that a similarity measure can be learnt for each profile group.
3. **Hypothesis 3**. Profiles can be learnt from the common properties of users of the same community. Our proposal consist on applying community detection algorithms and use the communities as the profiles to construct a similarity measure between items.

These hypothesis needs to be proven by cross-validation and it is very dependent on how the profiles are defined.

3.1 Methodology for learning perception aware similarity measures

Our methodology aims to build improved measures to compare both similarity between items and between users reflecting perception:

1. SIM'_xItem: similarity measure **between items** that is a computable model that is adjusted either to a particular user u (SIM'_uItem) or to the users of the same profile p (SIM'_pItem).
2. $SIM'User$: similarity measure **between users** that reflects shared perceptions regarding items. Similar users will be those that have similar emotions regarding similar items.

The process starts with the following **input requirements**:

- Set of $Items$ defined by the set of descriptive attributes (atr_j)
- Set of $Users$ defined by the set of descriptive attributes ($UserAtr_j$)
- Basic similarity measure between users ($SimUsers$) defined as a linear combination of the user descriptive attributes $UserAtr_j$.
- Set of Profiles to classify users. To simplify we assume from now that each user belong to exactly one profile (see section 3.2).

We propose a methodology organised in three steps:

1. **Step 1**. Definition and validation of the Local Similarity measures for each individual attribute atr_j. We define local similarity measures associated to each attribute describing the items, so $SimAtr_k(I_{ik}, I_{jk})$ is the local similarity between the value of attribute k in items i and j. We assume that, in this way, we can define weighted similarity measures $SIM(I_i, I_j) = \sum_k w_k \cdot SimAtr_k(I_{ik}, I_{jk})$, where I_i and I_j are two items; w_k is the weight or importance assigned to attribute k. Local similarity measures can be complex (graph based) or simple depending on the domain background knowledge. It is necessary to study how each local measure affect the perception of global similarity and study the correlation within the different user profiles.
2. **Step 2**. Construction of SIM'_xItem as a computable model to calculate the similarity between items. As this measure should reflect perception it will reflect either perception of one specific user u, or more interestingly perception of a group of users p sharing a common profile. Because this measure

is a weighted similarity measures $SIM(I_i, I_j)$, we address the fundamental problem of learning a weight model for features. i.e, it is necessary to give a greater similarity contribution to an important attribute than to other less important ones regarding perception of similarity:
   - SIM'_uItem with $u \in Users$. The weight for each attribute is adjusted to reflect the perception of the specific user u using the results of the experiment for *user u*.
   - SIM'_pItem The weight for each attribute is adjusted to reflect the perception of the users of the *profile p*. $p \in Profiles$.

   We plan to use an approach similar to the one described in [24] to learn weights of the local similarity measures through a genetic algorithm.
3. **Step 3** Construct a similarity measure $Sim'User$ that combines $SimUser$ with the polarity of the compared users with items. The similarity measure $Sim'User$ should reflect that users with similar emotions on similar items are similar (using the perceived measure SIM'_pItems).

One advantage of this approach is that it is scalable. New users and new items can be included in the system. New users benefit of personalised similarity measures reflecting perceptions of similar users. A key aspect of this methodology is the definition of user profiles that is described next.

3.2 Profile definition

As our methodology depends on the existence of profiles we consider two options:

- Manual definition of simple profiles at-hand reflecting the knowledge about the domain. This option is used in this paper using the knowledge in the Art domain, where there is a dependency between the level of expertise with the perception of similarity between artworks (see section 4).
- Use community detection algorithms [1] and define the *profile* as the common features for the users in this community. Again, community detection algorithms rely on similarity measures between users. As future work we will explore an iterative process to improve Community Detection processes by improving the similarity measure between users, as follows:
  1. Initial Community detection using basic $SimUser$
  2. Use each community in the communities set $c \in C$ as profiles for steps 1,2 and 3 of the methodology to learn $Sim'Users$ and Sim'_cItems.
  3. Recalculate Communities using the improved similarity measure $Sim'Users$ and study community model adequacy.

4 Experiment on the perception of similarity for Artworks

In this paper we propose to validate similarity computation algorithms against human perceptions of similarity. This section describes an experiment associated to the Step 1 of our methodology (described in Section 3): defining local similarity measures for comparing attributes of items and validating them regarding user perception of similarity.

Our experiment uses an artwork dataset. The set of Items is a set of Artworks from the Prado Museum described by four attributes atr_j: the dominant colour, the motion evoked to the users, the content depicted in the artwork and the domain knowledge that the user has about the artwork (like the painter or its art movement). Our first experimental goal is to validate the general acceptance of these aspects by real users and the difference of criteria in the perception of similarity between artworks in different user profiles.

The set of Items (artworks) employed in the experiments come from a dataset created as an excerpt from Wikiart Emotion Dataset [16]. We use this dataset because it includes data about the emotions that the artworks evoked to different users, so they will be employed to compute the local similarity concerning to evoked emotions. This dataset contains 30 artworks from Prado Museum and 1760 annotations of emotions from 171 different users. We limited the number of emotions to the set of emotions described by Plutchik Emotion Theory [19] (anger, anticipation, joy, trust, fear, surprise, sadness, and disgust), so the dataset is reduced to 1040 annotations from 168 unique users. The original artwork dataset has also been enriched with the Wikidata URLs of the paintings and artists, as long as the entity identifier in Wikidata[3] in order to compute the local similarity measure concerning the content depicted in the artwork.

According to our methodology, we first define local similarity measures associated to each attribute describing the items. This will be employed to exemplify the validation of our similarity measures against the user perceptions. Additionally, we will check if the combination of these local similarities can enhance the precision of a similarity measure according to the perceived similarity by users.

The experiment is divided into two steps: the implementation of local similarity measures and the gathering of user data about perceived similarity. These steps will be described in following subsections.

4.1 Definition of Local Similarity measures

The four attributes selected as the ones that support the similarity between two artworks I_i and I_j has been converted into four local similarity measures ($SimAtr_k(I_{ik}, I_{jk})$):

1. **Colour similarity**: This measure uses the weighted euclidean distance between the dominant colour of each painting in HSV space [8]. The dominant colour is the center of the biggest cluster when applying k-means on the artwork image pixels in RGB space.

$$SimAttr_{col}(I_i, I_j) = 1 - dist(hsv_i, hsv_j)$$

$$dist(hsv_i, hsv_j) = \sqrt{(v_i - v_j)^2 + {s_i}^2 + {s_j}^2 + 2s_i s_j(h_i - h_j)}$$

[3] Wikidata: https://www.wikidata.org

where hsv_i is the dominant colour or artwork I_i in HSV space.

2. **Content similarity**: This measure employs the knowledge about the elements depicted in an artwork stored in Wikidata. For each artwork, we created a list of contents collecting the values for the "depict" property in Wikidata. A first test over these lists highlighted that common contents were not frequent so we enlarge the list of contents using the concept hierarchy defined in Wikidata with the properties "instance of" and "subclass". The final list is computed traversing these hierarchy up to 2 levels. Finally, the similarity measure is computed using Jaccard over the list of contents.

$$SimAttr_{con}(I_i, I_j) = Jaccard(C_i, C_j) = \frac{C_i \cap C_j}{C_i \cup C_j}$$

   where C_i is the list of contents in artwork a_i.

3. **Emotion similarity**: This similarity uses the annotations in Wikiart Emotion Dataset about the emotions evoked by the artworks in different users. It is computed using the 3 most popular emotions and calculating the distance between emotions according to the Plutchik wheel of emotions –that places similar emotions close together and opposites 180 degrees apart, like complementary colours (see Figure 1).

$$SimAttr_{emo}(I_i, I_j) = 1 - \frac{1}{3}\sum_{k=1}^{3} dist(e_{ik}, e_{jk})$$

$$dist(e_i, e_j) = min_{dist}(e_i, e_j)/4$$

   where e_{ik} is the k-th most popular emotion in artwork a_i.

4. **Knowledge similarity**: This similarity uses the information about the artist and the art movement that the artworks belong to. These information is extracted from the WikiArt Emotion Dataset.

$$SimAttr_{kno}(I_i, I_j) = \begin{cases} \alpha & \text{if } author(a_i) = author(a_j) \\ \beta & \text{if } artMov(a_i) = artMov(a_j) \\ 0 & \text{otherwise} \end{cases}$$

   where $author(a_i)$ is the artist who painted a_i; $artMov(a_i)$ is the art movement that a_i belongs to and α and β are constants in $[0, 1]$ and $\alpha > \beta$.

4.2 Data gathering of perceived similarity

We have collected user perceptions of similarity between different artworks through an online questionnaire[4]. This questionnaire also collects user information in order to sketch some initial profiles that will be employed to study how the perception relates with the different user profiles and learn similarity measures that reflect the common perceptions (see Step 2 in Section 3).

[4] The questionnaire is available at https://tinyurl.com/2dn7wey4

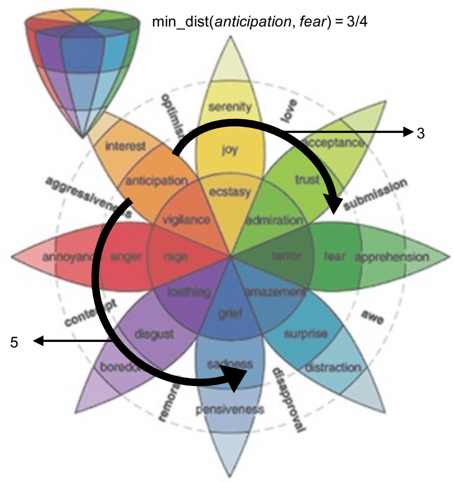

Fig. 1. Emotion wheel conceptualised by Plutchik [19]

Figure 2 (left) shows the first part of the questionnaire. In this part, we collect the information used to create user profiles. A tentative profile for this experiment is based on demographic aspects (age and gender), the user expertise or knowledge about art (professional, amateur, a fan of an artist or not interested in art), and the user habits on how often they visit museums (rarely, sporadic or often). This information allowed us to manually define different user profiles and evaluate the perception of similarity among them.

The next step of the questionnaire aims to gather the perception of similarity between different artworks. In this step, the application shows two different artworks (Figure 2, right) and users should select a value of similarity between 1 –artworks are very different– and 5 –artworks are very similar. The artworks are extracted from the Wikiart Emotion Dataset, described above. Although the pair of artworks is randomly chosen, the experiment is designed in a way that the randomisation process tries to balance the number of times that every pair is presented in the questionnaire.

To understand the reasons behind similarity perception, we include another question where users choose which criteria they applied to rate this similarity between both artworks. The criteria categories are the colour, the content represented, the user knowledge about these artworks (like the author, style, etc.) and evoked emotions by these artworks. Users can select more than one criteria category and they can add any other criteria not included in the questionnaire.

4.3 Experimental results

During the experiment, 92 unique users filled the questionnaire for assessing the perceived similarity. Figure 3 shows the users distribution based on each profile variable described in Section 4.

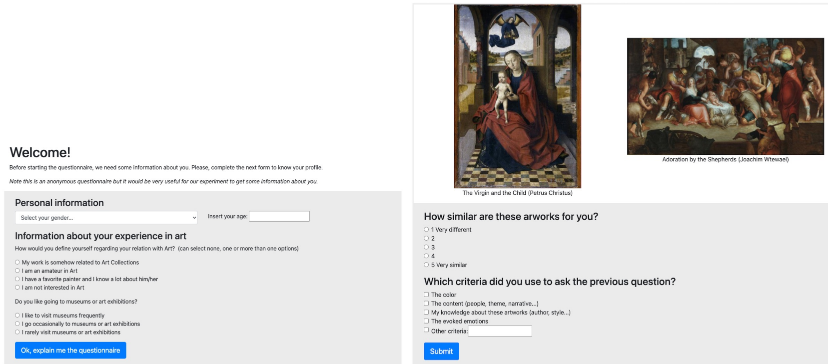

Fig. 2. Web application for collecting perceived similarity. On the left, questions about user profile information. On the right, the interface for assessing the perceived similarity between two artworks for choosing the criteria applied to explain the similarity perceived

These users generated a total of 1792 answers about the similarity perceived for a pair of artworks. All artworks received at least 2 answers, and most of them received 3 or 4 answers. The left graph included in Figure 4 shows the distribution of the perceived similarity values provided by users. It is worth noting that most of the answers correspond to perceiving dissimilarity (value equal to 1) between the artworks shown.

Next, we analysed the criteria employed by users to explain the similarity value provided. The right graph in Figure 4 shows the frequency of each criterion employed by users. It is important to remember that users can choose several criteria to explain the perceived similarity value. This graph shows that content is the most criteria employed, followed by the colour and the emotion evoked by artworks. In addition, we obtained 95 answers that considered other criteria out of the initial categories provided (i.e. colour, content, knowledge and evoked emotion). After a revision of these answers, we added 3 additional criteria to the previous categories: composition (it refers to the artwork layout, perspective, point of view, etc.), light (how the light is used in the artwork, contrast between foreground and background) and preference (user likes or dislike both artworks). Although these new criteria are not included in the rest of the analysis, they represent an important conclusion of our experiment and our future work.

When we correlate these criteria with the different user profiles, we can see some dependencies. In Table 1, we see that content criterion is more employed by users categorised in a lower knowledge level profile (fan of an artist/no interest) than amateur and professional profiles. Professional and amateur profiles use more Colour and Emotion criteria, as long as the composition and light criteria (other criteria) discovered during the questionnaire data analysis. This fact supports our Hypothesis 2 that different profiles are affected by different aspects in the perception of similarity between artworks.

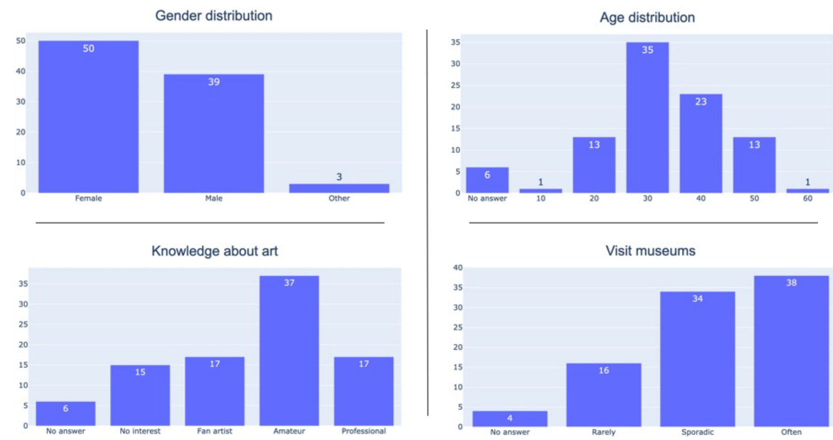

Fig. 3. Distribution of users by age, gender, their knowledge about art and how often they visit a museum.

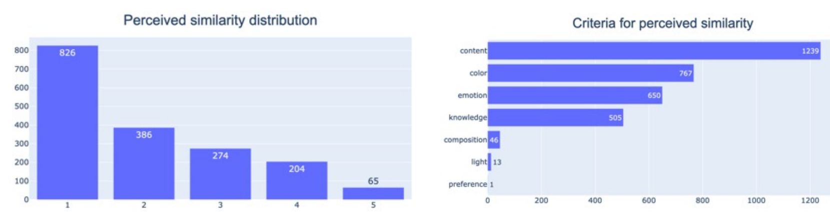

Fig. 4. Distribution of perceived similarity by users (left) and criteria frequency (right) to explain the perceived similarity

Although a single criterion is, in general, widely used to explain the perceived similarity, users also have employed combinations of these criteria. Table 2 summarises the top selected combinations of criteria. The combinations of colour, content and emotion are the most popular criteria to explain the perceived similarity.

The next step in our analysis is to determine if the two first steps of our methodology can be applied to this problem. In Step-1, we define and validate local similarity measure for each attribute. To do that, we calculated the Mean Absolute Error (MAE) and the Main Squared Error (MSE) between the local similarity measures explained in Section 3 and each corresponding values of humans perceptions of similarity. Table 3 shows the results, and we can observe that the *emotion similarity measure* is the least accurate comparing with the users' perceptions. On the other hand, the *knowledge similarity function* has the most accurate results regarding similarity perceptions.

We have additionally calculated the corresponding error values combining the top selected criteria. To do that, we compare the perception of users that combines these criteria with a simple weighted similarity function that uses the

	Colour	**Content**	**Knowledge**	**Emotion**	**Other**
Professional	28,03%	31,68%	18,83%	20,00%	1,46%
Amateur	24,54%	36,53%	14,72%	22,20%	2,01%
Fan artist	9,09%	59,09%	22,08%	6,49%	3,25%
No interest	16,37%	50,44%	15,49%	16,81%	0,88%
No answer	33,49%	33,97%	4,31%	27,27%	0,96%

Table 1. Distribution of criteria used for explaining perceived similarity according to user knowledge

Criteria combination	Frequency
content	439
colour-content-knowledge-emotion	177
colour-content	172
colour	162
colour-content-emotion	136
content-emotion	115
knowledge	81
emotion	81

Table 2. Most frequent criteria combinations to explain the perceived similarity

average of the local similarity measure based on these criteria. Table 4 presents the result of this analysis. Although, in the current state of this work, we have applied the same weight for each local similarity value, results show that the combination of similarity functions increases the accuracy. These are preliminary promising results and, in the next step (Step-2) of our methodology, we will apply learning algorithms to better adjust these weights to users' perceptions. In summary, in the experiment conducted in this paper we have validated Step 1 (correlation between local similarity measures and perception of similarity), we have observed that our Hypotheses 1 and 2 work on this experiment, and we have obtained promising informal results for Step 2 (learning and validating similarity measure $SIM'Item$ reflecting perception).

5 Conclusions and Future Work

Defining similarity measures is a requirement for some AI methods including CBR. Typically most of the approaches capture and define similarity measures analytically. However, research about automatically learning similarity measures has also been an active area of research in CBR. In this paper we have considered the problem of similarity measures definition for tasks where subjectivity in the perception of similarity is an issue. We have proposed a methodology for the definition of similarity measures that reflects the perception of similarity and applied it to the domain of Art.

In the experiment described in this paper we have have captured datasets from the similarity perception between artworks. This dataset contains the knowl-

	Colour	*Content*	*Knowledge*	*Emotion*
N	366	401	314	336
MAE	0.271	0.246	0.236	0.337
MSE	0.105	0.087	0.091	0.153

Table 3. Results of Mean Absolute Error (MAE) and Mean Squared Error (MSE) between similarity measures and user similarity perception

	Col-Con-K-E	*Col-Con*	*Col-Con-E*	*Con-E*
N	247	362	305	333
MAE	0.140	0.148	0.166	0.155
MSE	0.031	0.037	0.040	0.037

Table 4. Results of Mean Absolute Error (MAE) and Mean Squared Error (MSE) between combination of local similarity measures and user similarity perceptions

edge to construct or learn such a similarity measure. We have proposed a methodology, acquired the dataset, and validated Hypothesis 1 –different users have different perceptions of similarity and consider differently the attributes describing the items; and Hypothesis 2 –users can be grouped together using profiles and users from the same profile have similar perceptions of similarity. This means that a similarity measure can be learnt for each profile group. This will be done as future work in the Step 2 of the methodology that aims at the construction of a computable model to calculate the similarity between items reflecting the perception data acquired during step 1. We will automate the construction of similarity measures using machine learning from the acquired data. We address the fundamental problem of weight model for features. i.e, it is necessary to give a greater similarity contribution to an important attribute than to other less important ones regarding perception of similarity. We will explore different approaches [15] to automate the construction of similarity measures using machine learning algorithms. We also need to deal with the heterogeneity problem that arises when different attributes are used to describe different cases. We plan to use an approach similar to the one described in [24] to learn weights of the local similarity measures through an evolutionary algorithm (EA) and apply different solutions in the metric learning research area [2].

Also as future work in Step 3 we will construct a similarity measure to compare users ($Sim'User$) by combining a simple measure $SimUser$ with the polarity and emotions of the compared users regarding the domain items. The similarity measure $Sim'User$ should reflect that users with similar emotions on similar items are similar regarding the measure SIM'_pItems. This methodology will be validated and applied in the SPICE project. Our proposal is applying community detection algorithms and use the communities as the profiles to learn improved similarity measures.

Acknowledgements The research leading to this publication has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement SPICE No 870811

References

1. Arinik, N., Labatut, V., Figueiredo, R.: Characterizing and comparing external measures for the assessment of cluster analysis and community detection. IEEE Access **9**, 20255–20276 (2021). https://doi.org/10.1109/ACCESS.2021.3054621
2. Bellet, A., Habrard, A., Sebban, M.: A Survey on Metric Learning for Feature Vectors and Structured Data. Research report, Laboratoire Hubert Curien UMR 5516 (2013), https://hal.inria.fr/hal-01666935
3. Bergmann, R., Müller, G.: Similarity-based retrieval and automatic adaptation of semantic workflows. In: Nalepa, G.J., Baumeister, J. (eds.) Synergies Between Knowledge Engineering and Software Engineering, Advances in Intelligent Systems and Computing, vol. 626, pp. 31–54. Springer (2018). https://doi.org/10.1007/978-3-319-64161-4_2
4. Bernsmed, K., Tøndel, I.A., Nyre, Å.A.: Design and implementation of a cbr-based privacy agent. In: 2012 Seventh International Conference on Availability, Reliability and Security. pp. 317–326. IEEE (2012)
5. Blough, D.S.: The Perception of Similarity, Edited and Published by Dr. Robert G. Cook Department of Psychology Tufts University In cooperation with Comparative Cognition Press, vol. Avian Visual Recognition. Department of Psychology, Brown University (2001), http://www.pigeon.psy.tufts.edu/avc/dblough/
6. Cunningham, P.: A taxonomy of similarity mechanisms for case-based reasoning. IEEE Transactions on Knowledge and Data Engineering 21(11):1532–1543 (2009)
7. Daga, E., Marilena, D., Diaz-Kommonen, L., Gangemi, A., Kuflik, T., Mguire, M., Mulholland, P., Peroni, S., Pescari, S., Weckler, A.: Enabling multiple voices in the museum: Challenges and approaches. Digital Culture & Society (2020)
8. Dasari, H., Bhagvati, C., Jain, R.K.: Distance measures in RGB and HSV color spaces. In: Hu, G. (ed.) 20th International Conference on Computers and Their Applications, CATA 2005, Proceedings. pp. 333–338. ISCA (2005)
9. Díaz-Agudo, B., González-Calero, P.A.: A declarative similarity framework for knowledge intensive CBR. In: Aha, D.W., Watson, I.D. (eds.) Case-Based Reasoning Research and Development, ICCBR 2001, Proceedings. Lecture Notes in Computer Science, vol. 2080, pp. 158–172. Springer (2001)
10. Faruqui, M., Dodge, J., Jauhar, S.K., Dyer, C., Hovy, E.H., Smith, N.A.: Retrofitting word vectors to semantic lexicons. In: Mihalcea, R., Chai, J.Y., Sarkar, A. (eds.) The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 1606–1615 (2015). https://doi.org/10.3115/v1/n15-1184
11. Gazdar, A., Hidri, L.: A new similarity measure for collaborative filtering based recommender systems. Knowl. Based Syst. **188** (2020). https://doi.org/10.1016/j.knosys.2019.105058
12. Harispe, S., Ranwez, S., Janaqi, S., Montmain, J.: Semantic similarity from natural language and ontology analysis. CoRR **abs/1704.05295** (2017), http://arxiv.org/abs/1704.05295
13. Holyoak, K.J., Koh, K.: Surface and structural similarity in analogical transfer. Memory & cognition **15**(4), 332–340 (1987). https://doi.org/https://doi.org/10.3758/BF03197035

14. de Mántaras, R.L., McSherry, D., Bridge, D.G., Leake, D.B., Smyth, B., Craw, S., Faltings, B., Maher, M.L., Cox, M.T., Forbus, K.D., Keane, M.T., Aamodt, A., Watson, I.D.: Retrieval, reuse, revision and retention in case-based reasoning. Knowl. Eng. Rev. **20**(3), 215–240 (2005)
15. Mathisen, B.M., Aamodt, A., Bach, K., Langseth, H.: Learning similarity measures from data. Progress in Artificial Intelligence **9**(2), 129–143 (2020)
16. Mohammad, S., Kiritchenko, S.: WikiArt Emotions: An Annotated Dataset of Emotions Evoked by Art. In: Procs. of the 11th International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA) (2018)
17. Ontañón, S.: An overview of distance and similarity functions for structured data. Artificial Intelligence Review **53**(7), 5309–5351 (2020)
18. Osborne, H., Bridge, D.G.: A case base similarity framework. In: Smith, I.F.C., Faltings, B. (eds.) Advances in Case-Based Reasoning, Third European Workshop, EWCBR-96, Lausanne, Switzerland, November 14-16, 1996, Proceedings. Lecture Notes in Computer Science, vol. 1168, pp. 309–323. Springer (1996). https://doi.org/10.1007/BFb0020619, https://doi.org/10.1007/BFb0020619
19. Plutchik, R.: The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. American Scientist **89**(4), 344–350 (2001)
20. Puga, G.F., Díaz-Agudo, B., González-Calero, P.A.: Similarity measures in hierarchical behaviours from a structural point of view. In: Guesgen, H.W., Murray, R.C. (eds.) Proceedings of the Twenty-Third International Florida Artificial Intelligence Research Society Conference. AAAI Press (2010)
21. Sánchez-Ruiz, A.A., Ontañón, S.: Structural plan similarity based on refinements in the space of partial plans. Comput. Intell. **33**(4), 926–947 (2017)
22. Shanavas, N., Wang, H., Lin, Z., Hawe, G.I.: Knowledge-driven graph similarity for text classification. Int. J. Mach. Learn. Cybern. **12**(4), 1067–1081 (2021). https://doi.org/10.1007/s13042-020-01221-4
23. Smyth, B., T.Keane, M.: Adaptation-guided retrieval: questioning the similarity assumption in reasoning. https://doi.org/10.1016/S0004-3702(98)00059-9 (1998)
24. Stahl, A., Gabel, T.: Using evolution programs to learn local similarity measures. In: Ashley, K.D., Bridge, D.G. (eds.) 5th International Conference on Case-Based Reasoning, ICCBR 2003, Trondheim, Norway, June 23-26, 2003, Proceedings. Lecture Notes in Computer Science, vol. 2689, pp. 537–551. Springer (2003). https://doi.org/10.1007/3-540-45006-8_41
25. Towne, W.B., Rosé, C.P., Herbsleb, J.D.: Measuring similarity similarly: LDA and human perception. ACM Trans. Intell. Syst. Technol. **8**(1), 7:1–7:28 (2016). https://doi.org/10.1145/2890510
26. Tversky, A.: Features of similarity. Psychological Review **84**(4), 327–352 (1977). https://doi.org/10.1037/0033-295X.84.4.327
27. Voorhees, E.M.: TREC: continuing information retrieval's tradition of experimentation. Commun. ACM **50**(11), 51–54 (2007). https://doi.org/10.1145/1297797.1297822
28. Voskoglou, M.G., Salem, A.M.: Analogy-based and case-based reasoning: Two sides of the same coin. CoRR **abs/1405.7567** (2014), http://arxiv.org/abs/1405.7567
29. Wenige, L., Ruhland, J.: Similarity-based knowledge graph queries for recommendation retrieval. Semantic Web **10**(6), 1007–1037 (2019). https://doi.org/10.3233/SW-190353