

# Ontology-based Semantic Similarity in Generating Context-aware Collaborator Recommendations

Siying LI

*Sorbonne universités, Université de  
technologie de Compiègne  
CNRS UMR 7253, HEUDIASYC  
60203 Compiègne, France  
siying.li@utc.fr*

Marie-Hélène ABEL

*Sorbonne universités, Université de  
technologie de Compiègne  
CNRS UMR 7253, HEUDIASYC  
60203 Compiègne, France  
marie-helene.abel@utc.fr*

Elsa NEGRE

*Paris-Dauphine University, PSL  
Research University  
CNRS UMR 7243, LAMSADE  
75016 Paris, France  
elsa.negre@lamsade.dauphine.fr*

**Abstract**—Nowadays, the number of collaborative tools has increased significantly. This makes it difficult for users to find the collaborators that are most relevant to their needs among these tools. Besides, their needs can also be influenced by the context of collaboration (e.g., workplace, tools, and resources). This raises an issue: how to help users find their collaborators within the collaboration context. In our research, we propose an ontology-based semantic similarity and employ it in a collaboration context ontology to generate context-aware collaborator recommendations for users. In this paper, we present how to calculate and apply the semantic similarity in context-aware recommendation algorithms.

**Keywords**—Ontology-based semantic similarity, Context-aware recommendations, Collaborator, Collaboration context.

## I. INTRODUCTION

With the advancement of information technology, collaborative tools have been rapidly developed (e.g. instant messaging and resource management tools), leading to a significant increase in their amount. This offers users with convenience during collaborations [1] [2]. However, it also causes difficulties for users in finding the most relevant collaborators to their needs among these tools [3] [4]. Besides, due to the pandemic, users collaborate remotely under different contexts (e.g., workplace, tools, and resources) affecting users' needs during collaborations. This requires us to take the context of the collaboration into account when searching for collaborators. Therefore, one of the current issues is how to help users find their collaborators within the collaboration context.

To bridge the gap, collaborator recommendations are generated and intended to aid users in seeking out and choosing collaborators [3] [5]. However, due to the missing consideration of the context, these recommendations are sometimes not consistent with users' needs. Thus, we intend to incorporate the collaboration context into recommendation generation processes for producing more relevant recommendations, context-aware collaborator recommendations [6]. Among the different methods of generating such recommendations [6], we decide to employ contextual pre-filtering and post-filtering due to their advantages in terms of cost and computational complexity.

Specifically, the collaboration context is defined in an ontology, where all information is specified through semantic 3-uples  $\langle \text{Subject, Predicate, Object} \rangle$  [7] [8]. While dealing

with such information, ontology-based semantic similarities can be applied [9]–[15]. It allows us to measure the closeness between two collections of semantic 3-uples, which represent two collaborations and their contexts. Hence, we develop and utilize a new ontology-based semantic similarity to process information in the collaboration context ontology and thus generate context-aware collaborator recommendations for users.

The remainder of this paper is constructed as follows. Section II studies context-aware recommendations and ontology-based semantic similarities. Section III presents our contributions in (i) developing a new ontology-based semantic similarity, (ii) calculating it in a collaboration context ontology, (iii) employing it in two methods to generate context-aware collaborator recommendations for users. We then discuss the strengths and weaknesses of the ontology-based semantic similarity in Section IV. Finally, some conclusions and future work are shown in Section V.

## II. RELATED WORK

This section introduces what are context-aware recommendations and how they can be generated. We also investigate existing ontology-based semantic similarities and analyze their characteristics.

### A. Context-aware recommendations

When generating recommendations for users, two dimensions of data are usually applied: users and items<sup>1</sup> [16]. Here, items are objects that are recommended to users [17]. In 2D collaborator recommendations, a collaborator is an item. These items are sorted by their ratings, indicating how a particular user liked a specific item [17].

While producing context-aware recommendations for users, the context is incorporated into recommendation generation processes [6]. Such recommendations are generated by dealing with at least three dimensions of data: User, Item and Context<sup>2</sup>. Sometimes, it can even handle multi-dimensions. For example, [16] utilized 5 dimensions: User, Item, Place, Time and Companion for context-aware movie recommendations. Thus, with

<sup>1</sup>Such recommendations are mentioned as 2D recommendations in the rest.

<sup>2</sup>Here, the context is any information that can be used to characterize the situation of users, items, or interactions between users and items [18].

n-dimensional data and an initial set of ratings, users' unknown ratings for items can be predicted by (1) [16]:

$$R : D_1 \times D_2 \times \dots \times D_n \rightarrow Rating(n \geq 3, n \in \mathbb{N}^*) \quad (1)$$

where  $D_1, D_2, \dots, D_n$  represent the n dimensions of data:  $D_1 = User$ ;  $D_2 = Item$ ;  $D_3, \dots, D_n = Context$  (e.g., place, time).

Particularly, [6] proposed three methods to incorporate the context in different phases of recommendation processes (see Fig. 1)<sup>3</sup>, as follows:

- **Contextual pre-filtering (PreF)** starts by applying  $D_3, \dots, D_n$  to filter ratings that are irrelevant to specific contexts [16]. Afterwards,  $D_1, D_2$ , and approaches of 2D recommendations are used to predict unknown ratings and thus generate context-aware recommendations for users.
- **Contextual post-filtering (PoF)** first utilizes  $D_1, D_2$ , and approaches of 2D recommendations to produce 2D recommendations. Then  $D_3, \dots, D_n$  is employed to filter out irrelevant 2D recommendations or adjust their orders, resulting in context-aware recommendations [6].
- **Contextual modeling (CM)** directly processes  $D_1, D_2, D_3, \dots, D_n$  inside each phase of the recommendation generation process, which gives rise to truly multidimensional recommendations.

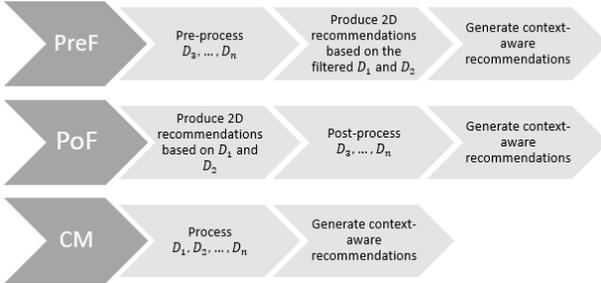


Fig. 1. Main phases of PreF, PoF, and CM [6].

Among these methods, CM is too costly when data volume and complexity are heavy. It can not handle the same amount of users and items as other methods (i.e., PreF and PoF). In each phase of the recommendation generation process, CM needs to process  $n$  dimensions of data, whereas PreF and PoF only focus on partial dimensions (either  $D_1 \times D_2$  or  $D_3 \times \dots \times D_n$ ). Besides, as  $n$  increases, the computational complexity of CM becomes higher. However, for PreF and PoF, only the phase of processing  $D_3 \times \dots \times D_n$  becomes complicated. Thus, we concentrate on applying PreF and PoF to generate context-aware collaborator recommendations for users.

## B. Ontology-based semantic similarity

An ontology is a formal and explicit specification of a shared, agreed and detailed conceptualization [19] [20]. It applies semantic 3-uples <Subject, Predicate, Object> to describe and represent information. In ontologies, the closeness between two

terms<sup>4</sup> is computed by ontology-based semantic similarities, which can be classified into different types based on the information contained in terms and/or the calculation method.

First, three types of semantic similarities are summarized, depending on the information contained in terms [13]–[15]: 1) Path-based similarities are calculated based on the shortest path between two terms [14]. The longer the shortest path, the more semantically different the two terms are [13]; 2) Feature-based similarities consider a term as a set of ontological features [13] [14]. The more common features and the less non-common features two terms have, the more similar they are [21]. To compare features, several coefficients on the sets are applicable, such as Jaccard index  $J(O^x, O^y) = \frac{|O^x \cap O^y|}{|O^x \cup O^y|}$ <sup>5</sup> [22], Dice coefficient  $D(O^x, O^y) = \frac{2|O^x \cap O^y|}{|O^x| + |O^y|}$ <sup>6</sup> [23], and Tversky index  $T(O^x, O^y) = \frac{|O^x \cap O^y| + \alpha|O^x - O^y| + \beta|O^y - O^x|}{|O^x \cap O^y| + \alpha|O^x - O^y| + \beta|O^y - O^x|}$ <sup>7</sup> [24]; 3) Information content-based similarities measure the amount of information provided by a common ancestor of two terms in an ontology [13] [14]. Particularly for a term  $x$ ,  $IC(x) = -\log p(x)$ <sup>8</sup> is utilized to identify the amount of  $x$ 's provided information.

Particularly, when the compared terms are collections of semantic 3-uples, semantic similarities can be divided into pairwise and groupwise similarities, based on the calculation method [9]. Pairwise similarity is a combination of similarities between individual subjects, predicates, or objects in the two collections, which can consider either every pair or only the best-matching pair [10]. Secondly, groupwise similarities are directly calculate by one of three collection presentations: set, graph, or vector, such as the similarities between two graphs [9]–[12].

These ontology-based semantic similarities have their own advantages and disadvantages. For example, path-based similarities do not necessitate the detailed information of each term, which is an advantage. However, they have a strong dependence on the degree of completeness and coverage of predicates in an ontology [25]. As for feature-based similarities, they can be employed in cross ontologies (i.e. when the two terms belong to different ontologies) [26], but require detailed information about the features of each term. Besides, using information content-based similarities, infrequent terms are more informative than the frequent ones [13]. But they request the recursive computation of all subjects', objects' and predicates' appearances in an ontology. If any subject, object, or predicate changes, recalculations are mandatory. Pairwise similarities can apply all the first types of similarities (i.e., path-based, feature-based, and information content-based similarities) for each pair,

<sup>4</sup>A term may be a subject, a predicate, an object or a collection of semantic 3-uples.

<sup>5</sup>Here,  $x$  and  $y$  are two concepts (instances) in ontologies;  $O^x$  and  $O^y$  refer to their sets of ontological features;  $|O^x \cap O^y|$  the number of common features in the sets  $O^x$  and  $O^y$ ;  $|O^x|$  denotes the number of features in the set  $O^x$ . The range of  $J(O^x, O^y)$  is  $[0, 1]$ .

<sup>6</sup>The range of  $D(O^x, O^y)$  is  $[0, 1]$ .

<sup>7</sup>Here,  $O^x - O^y$  denotes the relative complement of  $O^y$  in  $O^x$ .  $\alpha, \beta \geq 0$  are parameters of the Tversky index.

<sup>8</sup>Here,  $p(x)$  is the probability of  $x$ 's appearance in the ontology.

<sup>3</sup>Fig. 1 also illustrates how  $D_1, D_2, \dots, D_n$  are utilized in different methods.

while groupwise similarities are limited due to the collection presentation.

Nevertheless, a common disadvantage is that they utilize only partial information in ontologies. For instance, path-based similarities employ information related to the shortest path, while feature-based similarities only focus on features of each term. This makes it hard to compare two terms with all their detailed information. Therefore, inspired by feature-based, information content-based, and groupwise similarities, a new ontology-based semantic similarity is built and employed in our research. It can not only compare the common and non-common features of subjects (objects) in two collections of semantic 3-uples, but also assess the amount of information provided by different involved predicates.

### III. CONTRIBUTIONS

This section explains how to generate context-aware collaborator recommendations by computing an ontology-based semantic similarity in a collaboration context ontology and implementing it in two methods (i.e., PreF and PoF). In doing so, two corresponding context-aware recommendation algorithms are constructed, separately following PreF and PoF. Finally, these algorithms are presented.

#### A. A collaboration context ontology

The collaboration context ontology MEMORAe-Collaboration-Context (MCC) [7] employs a subject and its involved semantic 3-uples to define the collaboration context. Specifically, the subject itself indicates a collaboration. Surrounding it, eight dimensions of the collaboration context are represented by different predicates and objects, including: Goal, Collaborator, Activity, Resource, Time, Location, Relation, and Satisfaction [7]. Thus, a collaboration  $c$  and its context are formulated as a collection of semantic 3-uples in MCC:  $\{ \langle c, p_g^{c,t}, o_g^{c,t} \rangle \mid g \leq G_c, t \leq T, g, t \in \mathbb{N}^+ \}$ .

Here,  $p_g^{c,t}$  is a predicate that the collaboration  $c$  contains, belonging to the  $t^{\text{th}}$  dimension;  $P^{c,t} (= \{ p_g^{c,t} \})$  is a set of predicates that the collaboration  $c$  contains in the  $t^{\text{th}}$  dimension;  $P^c (= \{ p_g^c \})$  is the set of all predicates that the collaboration  $c$  contains;  $o_g^{c,t} (\in O^{c,t})$  is an object of the predicate  $p_g^{c,t}$ ;  $O^{c,t} (= \{ o_g^{c,t} \})$  is a set of objects that the collaboration  $c$  contains through predicates in  $P^{c,t}$ ;  $O^c (= \{ o_g^c \})$  is the set of all objects that the collaboration  $c$  contains through predicates in  $P^c$ ;  $G_c$  is the number of predicates that the collaboration  $c$  contains;  $T$  represents the maximum number of dimensions that the collaboration  $c$  can contain.

Within such collections of semantic 3-uples in MCC, we are able not only to compare the common and non-common objects in two collections, but also to assess the amount of information provided by a specific predicate in two collections. Thus, we develop a new groupwise semantic similarity to measure the closeness between two collections of semantic 3-uples.

#### B. An ontology-based semantic similarity

The ontology-based groupwise semantic similarity aims to compare two collections of semantic 3-uples  $\langle \text{Subject, Predicate, Object} \rangle$ , which respectively stand for two collaborations

(represented by two subjects  $c$  and  $d$ ) and their contexts. Inspired by the existing ontology-based semantic similarities (cf. Section II-B), it is calculated from two aspects: **Object** (built on feature-based semantic similarities) and **Predicate** (built on information content-based semantic similarities).

a) *Object*: This aspect measures the common and non-common objects that are related to the same predicates in two collections of semantic 3-uples. Particularly, both qualitative and quantitative objects are considered.

Qualitative objects contain non-numerical and descriptive information. For example, when comparing whether the two collections include the same collaborators, the collaborators are qualitative objects. These objects can be associated with different predicates in the two collections. When they are linked to a same predicate  $p_g^i (i = 1, 2, \dots, T)$ , they constitute separately two sets of objects:  $O_{p_g^i}^c$  and  $O_{p_g^i}^d$ , which are measured by (2)<sup>9</sup>.

$$S_1^i(c, d) = \sum_{g=1}^{|P^{cd,i}|} \frac{|O_{p_g^i}^c \cap O_{p_g^i}^d| \times \frac{IC(P^{c,i})}{\sum_{t=1}^T IC(P^{c,t})}}{|O_{p_g^i}^c \cap O_{p_g^i}^d| + \alpha |O_{p_g^i}^c - O_{p_g^i}^d| + \beta |O_{p_g^i}^d - O_{p_g^i}^c|} \quad (2)$$

where  $p_g^i$  is a predicate that both  $c$  and  $d$  contain, belonging to the  $i^{\text{th}}$  dimension;  $O_{p_g^i}^d$  indicates a set of qualitative objects that  $d$  relates through the predicate  $p_g^i$ ;  $|O_{p_g^i}^c - O_{p_g^i}^d|$  denotes the number of non-common objects in the relative complement of  $O_{p_g^i}^d$  in  $O_{p_g^i}^c$ ;  $|O_{p_g^i}^c \cap O_{p_g^i}^d|$  represents the number of common objects associated to  $c$  and  $d$  through the predicate  $p_g^i$ ;  $|P^{cd,i}|$  represents the number of predicates belonging to the  $i^{\text{th}}$  dimension that both  $c$  and  $d$  contain;  $IC(P^{c,i})$  expresses the amount of information provided by the predicates that  $c$  contains, belonging to the  $i^{\text{th}}$  dimension<sup>10</sup>;  $\sum_{t=1}^T IC(P^{c,t})$  is the sum amount of information provided by the predicates that  $c$  contains<sup>11</sup>;  $\alpha, \beta (\geq 0)$  indicates the weights of  $c$  and  $d$ . The smaller the semantic similarity  $S_1^i(c, d)$  is, the more differences there are between  $c$  and  $d$  in the  $i^{\text{th}}$  dimension of the collaboration context.

Quantitative objects include numerical information, such as start time and end time of a collaboration [7]. To compare two such objects, the absolute difference  $|o_{p_g^j}^c - o_{p_g^j}^d| (j = 1, 2, \dots, T)$  is utilized. The smaller the absolute difference is, the greater the similarity between  $o_{p_g^j}^c$  and  $o_{p_g^j}^d$ . Equation (3)<sup>12</sup> is thus applied.<sup>13</sup>

$$S_2^j(c, d) = \sum_{g=1}^{|P^{cd,j}|} \frac{1}{|o_{p_g^j}^c - o_{p_g^j}^d| + 1} \times \frac{IC(P^{c,j})}{\sum_{t=1}^T IC(P^{c,t})} \quad (3)$$

where  $o_{p_g^j}^c$  indicates a quantitative object that  $c$  contains through  $p_g^j$ ;  $|o_{p_g^j}^c - o_{p_g^j}^d|$  denotes the absolute differences between two

<sup>9</sup>The range of  $S_1^{cd,i}$  is  $[0, \frac{IC(P^{c,i})}{\sum_{t=1}^T IC(P^{c,t})}]$ .

<sup>10</sup>The calculation of  $IC(P^{c,t})$  will be presented later.

<sup>11</sup>Here,  $\frac{IC(P^{c,i})}{\sum_{t=1}^T IC(P^{c,t})}$  serves to normalize values of  $S_1^{cd,i}$ .

<sup>12</sup>The range of  $S_2^j(c, d)$  is  $[0, \frac{IC(P^{c,j})}{\sum_{t=1}^T IC(P^{c,t})}]$ .

<sup>13</sup>Equation (3) intends to convert  $|o_{p_g^j}^c - o_{p_g^j}^d|$ , which is inspired from (2) and the discussion on the site <https://stats.stackexchange.com/questions/158279/how-i-can-convert-distance-euclidean-to-similarity-score>.

qualitative objects that  $c$  and  $d$  contains through the predicate  $p_g^j$ . Notably, (3) is only suitable for  $o_{p_g^j}$  whose range of is  $[0, +\infty)$ . Particularly, if the range of  $o_{p_g^j}$  is  $[0, L]^{14}$ , then (3) is transformed into:

$$S_2^j(c, d) = \sum_{g=1}^{|P^{c,d,j}|} \left( \frac{\frac{L+1}{L}}{|o_{p_g^j}^c - o_{p_g^j}^d| + 1} - \frac{1}{L} \right) \times \frac{IC(P^{c,j})}{\sum_{t=1}^T IC(P^{c,t})} \quad (4)$$

b) *Predicate*: This aspect computes the amount of information provided by a specific predicate, reflecting the importance of this predicate. Particularly, the appearance frequency and contribution of a predicate in one collection of semantic 3-uples matter in this computation. Thus, we decide to employ TF-IDF [27]. Specifically, we replace “term, document, and corpus” in TF-IDF as following for “predicate (factor), a collaboration, and a collection of collaborations” (see Fig. 2).

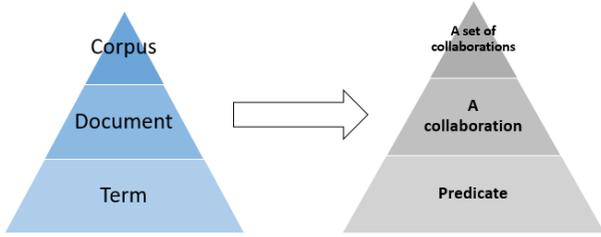


Fig. 2. Apply TF-IDF to measure the amount of information provided by predicates.

For a predicate  $p$  in a collaboration  $c$  ( $c \in X$ ,  $X$  represents a set of collaborations) and the total number of collaborations in the set is  $|X|$ , we have<sup>15</sup>

$$\begin{aligned} tf(p, c) &= \frac{f(p, P^c)}{\sum_{p' \in P^c} f(p', P^c)} \\ idf(p, X) &= \log \frac{|X|}{1 + |c \in X: p \in P^c|} + 1 \\ tf \cdot idf(p, c, X) &= tf(p, c) \cdot idf(p, X) \end{aligned} \quad (5)$$

where  $f(p, P^c)$  represents the appearance frequency of the predicate  $p$  in  $c$ ;  $\sum_{p' \in P^c} f(p', P^c)$  denotes the appearance frequencies of all predicates in  $c$ .

Then the amount of information provided by the predicate  $p$  in the collaboration  $c$  is  $IC(p) = tf \cdot idf(p, c, X)$ . As for the amount of information provided by the predicates belonging to the  $t^{th}$  dimension:  $IC(P^{c,t})(t = 1, 2, \dots, T)$ , it is calculated through (6)<sup>16</sup>:

$$IC(P^{c,t}) = \sum_{q=1}^{|P^{c,t}|} tf \cdot idf(p_q^{c,t}, c, X) \quad (6)$$

where  $|P^{c,t}|$  is the number of predicates that the collaboration  $c$  contains, belonging to the  $t^{th}$  dimension;  $p_q^{c,t} \in P^{c,t}$ .

<sup>14</sup>Here,  $L$  represents the maximum value in the range of  $o_{p_g^j}$ .

<sup>15</sup>Particularly, the range of  $tf(p, c)$  is  $[0, 1]$  and the range of  $idf(p, X)$  is  $[\log \frac{|X|}{1 + |c \in X: p \in P^c|} + 1, \log |X| + 1]$ . Thus, the range of  $tf \cdot idf(p, c, X)$  is  $[0, \log |X| + 1]$ .

<sup>16</sup>The range of  $IC(P^{c,t})$  is  $[0, 2(\log |X| + 1)]$ .

c) *Calculation*: Combining eq. (2), 3, and 4, the semantic similarity  $S(c, d)$  between  $c$  and  $d$  can be computed as follows. Its range is  $[0, 1]$ .

$$S(c, d) = \sum_{i=1}^I S_1^i(c, d) + \sum_{j=1}^J S_2^j(c, d) \quad (7)$$

where  $I$  represents the number of dimensions that contain qualitative objects;  $J$  represents the number of dimensions that contain quantitative objects;  $I + J = T$ . In MCC, we have  $T = 8$ ,  $I = 6$  (i.e., Goal, Collaborator, Activity, Resource, Location, and Relation) and  $J = 2$  (i.e., Time and Satisfaction) [7]. Thus, the ontology-based semantic similarity  $S(c, d)$  in MCC is calculated as shown in (8)<sup>17</sup>.

$$S(c, d) = \sum_{i=1}^6 S_1^i(c, d) + \sum_{j=1}^2 S_2^j(c, d) \quad (8)$$

Particularly,  $\sum_{j=1}^2 S_2^j(c, d) = \sum_{g_2=1}^{|P^{xy, T^i}|} \frac{1}{|o_{p_{g_2}^{T^i}}^c - o_{p_{g_2}^{T^i}}^d| + 1} \times \frac{IC(P^{c, T^i})}{\sum_{t=1}^8 IC(P^{c, t})} + \left( \frac{\frac{L_{S_a} + 1}{L_{S_a}}}{|o_{p_{S_a}^c} - o_{p_{S_a}^d}| + 1} - \frac{1}{L_{S_a}} \right) \times \frac{IC(P^{c, S_a})}{\sum_{t=1}^8 IC(P^{c, t})}$ , where  $L_{S_a}$  is the maximum value in the range of individuals' satisfactions to collaborations<sup>18</sup>;  $\overline{o_{p_{S_a}^c}}$  denotes the average of all given satisfactions to the collaboration  $c$ .

Within this ontology-based semantic similarity  $S(c, d)$ , the closeness between two collections of semantic 3-uples can be measured. Each collection contains information relevant to a collaboration and its context. Therefore, through  $S(c, d)$ , the collaboration context in MCC can be processed to generate context-aware collaborator recommendations for users.

### C. Context-aware collaborator recommendation algorithms

Within the ontology-based semantic similarity  $S(c, d)$  in MCC, the context-aware collaborator recommendation problem is formulated as:

Given a rating matrix  $R$  and a user  $u$  in a collaboration  $c$  ( $u \in O^{c, Col}$ ), the top  $K$  collaborators  $v$  ( $v \neq u, v \notin O^{c, Col}$ ) that can facilitate  $u$ 's collaboration  $c$  with the highest probabilities will be recommended to  $u$ . Here, the context indicates the context of  $u$ 's collaboration  $c$ , which is represented by a collection of semantic 3-uples  $\{ \langle c, p_g^{c,t}, o_g^{c,t} \rangle | g \leq G_c, t \leq 8, g, t \in \mathbb{N}^+ \}$ .

To address this problem, two methods (i.e., PreF and PoF) are applied to generate context-aware collaborator recommendations. Particularly, their main phases (see Fig. 1) can be resumed below, regardless of their order:

- Employ  $D_1 (= User)$ ,  $D_2 (= Item)$ , and approaches of 2D recommendations to produce 2D recommendations
- Process  $D_3, \dots, D_n (= Context)$  (pre-process in PreF and post-process in PoF)
- Produce context-aware recommendations based on the previous results

Based on these phases, two corresponding algorithms are constructed (see Algorithms 1 and 2).

<sup>17</sup>Its range is  $[0, 1]$ .

<sup>18</sup>For example, if individuals' satisfactions to a collaboration belong to the interval  $[0, 5]$ , then  $L_{S_a} = 5$ .

---

**Algorithm 1:** PreF algorithm.

---

**Input:** The rating matrix:  $R$ ,  
the user:  $u$ ,  
the collaboration:  $c$ ,  
the set of members in the collaboration  $c$ :  $O_{Col}^c$ ,  
the number of recommendations:  $K$ ,  
the number of known collaborations:  $|X|$ ,  
the number of users:  $m$ .  
**Initialize:** Two lists with zeros:  $SC$  (of length  $|X|$ ) and  $VU$  (of length  $m$ ).

- 1 **for**  $d \in \{1, 2, \dots, |X|\}$  **do**
- 2     **if**  $d == c$  **then**
- 3          $S(d, c) \leftarrow 0$  ;
- 4     **else**
- 5          $S(d, c) \leftarrow \sum_{i=1}^6 S_1^i(d, c) + \sum_{j=1}^2 S_2^j(d, c)$ ;
- 6      $SC.insert(S(d, c), d)$ ;
- 7 Rank  $SC$  in decreasing order based on the value of ontology-based semantic similarity and copy  $K$  collaborations with  $K$  highest semantic similarities to a new list  $SC'$ ;
- 8 **for**  $k \in \{1, 2, \dots, m\}$  **do**
- 9     **if**  $R_{uk}$  is unknown **then**
- 10          $VU(k) \leftarrow 1$ ;
- 11     **else**
- 12         **for**  $d \in \{1, 2, \dots, length(SC')\}$  **do**
- 13             **if**  $k \in O_{Col}^d$  **then**
- 14                  $VU(k) \leftarrow 1$ ;
- 15 Filter  $R$  of  $m \times m$  to  $R'$  of  $m \times (m - length(SC'))$  by deleting all ratings of irrelevant collaborators  $v'$  ( $VU(v') = 0$ );
- 16 **for**  $k \in \{1, 2, \dots, m\}$  **do**
- 17     **if**  $R'_{uk}$  is unknown **then**
- 18         apply approaches of 2D recommendations to predict users' ratings  $r^*(u, k)$ ;
- 19      $R'_{uk} \leftarrow VU(k) \times r^*(u, k)$ ;
- 20 Rank  $R'_{u*}$  in decreasing order and get  $K$  highest ratings;

**Output:** User  $u$ 's ratings  $R'_{u*}$ ,  $K$  collaborators with  $K$  highest ratings.

---

Specifically, Algorithm 1 first calculates ontology-based semantic similarities between the collaboration  $c$  and other collaborations  $d$  ( $d \in X, d \neq c$ ) (line 1-6). Then, it filters out irrelevant collaborators  $v'$  that did not participate in similar collaborations with  $c$  (line 7-15). Next, approaches of 2D recommendations are applied to predict users' ratings for relevant collaborators (line 16-19). Finally, the top  $K$  collaborators with higher ratings will be recommended to the user  $u$  (line 20). Briefly, Algorithm 1 first processes the collaboration context by means of the ontology-based semantic similarity, then applies approaches of 2D recommendations, and finally generates

context-aware collaborator recommendations for users.

---

**Algorithm 2:** PoF algorithm.

---

**Input:** The rating matrix:  $R$ ,  
the user:  $u$ ,  
the collaboration:  $c$ ,  
the set of members in the collaboration  $c$ :  $O_{Col}^c$ ,  
the number of recommendations:  $K$ ,  
the number of known collaborations:  $|X|$ ,  
the number of users:  $m$ .  
**Initialize:** Two lists with zeros:  $SC$  (of length  $|X|$ ) and  $VU$  (of length  $m$ ).

- 1 **for**  $h \in \{1, 2, \dots, m\}$  AND  $k \in \{1, 2, \dots, m\}$  **do**
- 2     **if**  $R_{hk}$  is unknown **then**
- 3         apply approaches of 2D recommendations to predict users' ratings  $R_{hk} \leftarrow r^*(h, k)$ ;
- 4         **if**  $h == u$  **then**
- 5              $VU(k) \leftarrow 1$ ;
- 6 **for**  $d \in \{1, 2, \dots, |X|\}$  **do**
- 7     **if**  $d == c$  **then**
- 8          $S(d, c) \leftarrow 0$  ;
- 9     **else**
- 10          $S(d, c) \leftarrow \sum_{i=1}^6 S_1^i(d, c) + \sum_{j=1}^2 S_2^j(d, c)$ ;
- 11      $SC.insert(S(d, c), d)$ ;
- 12 Rank  $SC$  in decreasing order based on the value of ontology-based semantic similarity and copy  $K$  collaborations with  $K$  highest semantic similarities to a new list  $SC'$ ;
- 13 **for**  $k \in \{1, 2, \dots, m\}$  **do**
- 14     **for**  $d \in \{1, 2, \dots, length(SC')\}$  **do**
- 15         **if**  $k \in O_{Col}^d$  **then**
- 16              $VU(k) \leftarrow 1$ ;
- 17 **for**  $k \in \{1, 2, \dots, m\}$  **do**
- 18      $R_{uk} \leftarrow R_{uk} \times VU(k)$
- 19 Rank  $R_{u*}$  in decreasing order and get  $K$  highest ratings;

**Output:** User  $u$ 's ratings  $R_{u*}$ ,  $K$  collaborators with  $K$  highest ratings.

---

In Algorithm 2, we first employ approaches of 2D recommendations to predict users' unknown ratings for collaborators (line 1-5). Then, it computes ontology-based semantic similarities between the collaboration  $c$  and other collaborations  $d$  ( $d \in X, d \neq c$ ), and find similar collaborations with the  $K$  highest semantic similarity (line 6-12). Next, irrelevant collaborators  $v'$  that did not participate in these similar collaborators are filtered out (line 13-18). Finally, the top  $K$  collaborators with higher ratings will be recommended to the user  $u$  (line 19). To summarize, Algorithm 2 starts by applying approaches of 2D recommendations, then calculates the ontology-based semantic similarity, and eventually generates context-aware collaborator recommendations for users.

Using the above algorithms of PreF and PoF methods (Algorithms 1 and 2), context-aware collaborator recommendations can be generated for users. Particularly, both PreF and PoF methods employ the ontology-based semantic similarities in the recommendation generating processes, but in different orders.

#### IV. DISCUSSION

The new ontology-based semantic similarity, which is calculated by (7), measures the closeness between two collections of semantic 3-uples representing two collaborations and their contexts. This semantic similarity not only measures the objects but also considers the importance of their related predicates, which are reflected by the amount of information provided by these predicates.

Besides, applying this semantic similarity in PreF and PoF methods permit us to generate more accurate collaborator recommendations, context-aware collaborator recommendations<sup>19</sup>. This also enables us to identify users' ratings for collaborators within the whole context of collaboration, instead of the partial context of user or item (collaborator). For instance, [28] only took into account some of item context in their algorithm following PreF method. Furthermore, since both users and items belong to a collaboration, our algorithms collectively consider users and items within the collaboration context, rather than considering separately the context of user or item.

#### V. CONCLUSION AND FUTURE WORK

In this paper, we focus on how to recommend collaborators for users within the collaboration context. To solve this issue, we develop a new ontology-based semantic similarity, calculate it in a collaboration context ontology, and utilize it in two context-aware collaborator recommendation algorithms. Based on the literature review, we explain context-aware recommendations and compare existing ontology-based semantic similarities. We also investigate how to compute the new ontology-based semantic similarity in a collaboration context ontology and employ it in two algorithms to generate context-aware collaborator recommendations. Our future work is to finalize and improve the experiments about the two algorithms.

#### REFERENCES

- [1] H. Patel, M. Pettitt, and J. R. Wilson, "Factors of collaborative working: A framework for a collaboration model," *Applied ergonomics*, vol. 43, no. 1, pp. 1–26, 2012.
- [2] P. W. Mattessich and B. R. Monsey, *Collaboration: what makes it work. A review of research literature on factors influencing successful collaboration*. ERIC, 1992.
- [3] W. Wang, J. Liu, Z. Yang, X. Kong, and F. Xia, "Sustainable collaborator recommendation based on conference closure," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 2, pp. 311–322, 2019.
- [4] C. Yang, T. Liu, L. Liu, and X. Chen, "A nearest neighbor based personal rank algorithm for collaborator recommendation," in *2018 15th International Conference on Service Systems and Service Management (ICSSSM)*. IEEE, 2018, pp. 1–5.

- [5] C. Yang, J. Sun, J. Ma, S. Zhang, G. Wang, and Z. Hua, "Scientific collaborator recommendation in heterogeneous bibliographic networks," in *2015 48th Hawaii International Conference on System Sciences*. IEEE, 2015, pp. 552–561.
- [6] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," in *Recommender systems handbook*. Springer, 2011, pp. 217–253.
- [7] S. Li, M. Abel, and E. Negre, "Towards a collaboration context ontology," in *2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2019, pp. 93–98.
- [8] L. Siying, A. Marie-Hélène, and E. NEGRE, "Managing and recommending resources in web-based collaborative working environments," in *2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 2020, pp. 287–290.
- [9] Q. Zhang and D. Haglin, "Semantic similarity between ontologies at different scales," *IEEE/CAA Journal of Automatica Sinica*, vol. 3, no. 2, pp. 132–140, 2016.
- [10] C. Pesquita, D. Faria, A. O. Falcao, P. Lord, and F. M. Couto, "Semantic similarity in biomedical ontologies," *PLoS comput biol*, vol. 5, no. 7, p. e1000443, 2009.
- [11] X. Hu, Z. Feng, S. Chen, K. Huang, J. Li, and M. Zhou, "Accurate identification of ontology alignments at different granularity levels," *IEEE Access*, vol. 5, pp. 105–120, 2016.
- [12] M. Wang, Z. Ding, G. Liu, C. Jiang, and M. Zhou, "Measurement and computation of profile similarity of workflow nets based on behavioral relation matrix," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 10, pp. 3628–3645, 2018.
- [13] D. Sánchez, M. Batet, D. Isern, and A. Valls, "Ontology-based semantic similarity: A new feature-based approach," *Expert systems with applications*, vol. 39, no. 9, pp. 7718–7728, 2012.
- [14] X. Zhang, L. Jing, X. Hu, M. Ng, J. X. Jiangxi, and X. Zhou, "Medical document clustering using ontology-based term similarity measures," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 4, no. 1, pp. 62–73, 2008.
- [15] Q. Zhao, C. Wang, P. Wang, M. Zhou, and C. Jiang, "A novel method on information recommendation via hybrid similarity," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, pp. 448–459, 2016.
- [16] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin, "Incorporating contextual information in recommender systems using a multidimensional approach," *ACM Transactions on Information Systems (TOIS)*, vol. 23, no. 1, pp. 103–145, 2005.
- [17] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender systems handbook*, 2011, pp. 1–35.
- [18] A. K. Dey, "Understanding and using context," *Personal and ubiquitous computing*, vol. 5, no. 1, pp. 4–7, 2001.
- [19] J. Peckham and F. Maryanski, "Semantic data models," *ACM Computing Surveys (CSUR)*, vol. 20, no. 3, pp. 153–189, 1988.
- [20] Y. Du, S. Ranwez, N. Sutton-Charani, and V. Ranwez, "Apports des ontologies aux systèmes de recommandation : état de l'art et perspectives," in *30es Journées Francophones d'Ingénierie des Connaissances, IC 2019*, Toulouse, France, Jul. 2019, pp. 64–77.
- [21] G. Varelas, E. Voutsakis, P. Raftopoulou, E. G. Petrakis, and E. E. Milios, "Semantic similarity methods in wordnet and their application to information retrieval on the web," in *Proceedings of the 7th annual ACM international workshop on Web information and data management*, 2005, pp. 10–16.
- [22] P. Jaccard, "Étude comparative de la distribution florale dans une portion des alpes et des jura," *Bull Soc Vaudoise Sci Nat*, pp. 547–579, 1901.
- [23] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [24] A. Tversky, "Features of similarity," *Psychological review*, vol. 84, no. 4, p. 327, 1977.
- [25] P. Cimiano, *Ontology Learning and Population from Text — Algorithms, Evaluation and Applications*, 01 2006.
- [26] E. G. Petrakis, G. Varelas, A. Hliaoutakis, and P. Raftopoulou, "X-similarity: Computing semantic similarity between concepts from different ontologies," *Journal of Digital Information Management*, vol. 4, no. 4, 2006.
- [27] A. Rajaraman and J. D. Ullman, *Mining of massive datasets*. Cambridge University Press, 2011.
- [28] I. M. Al Jawarneh, P. Bellavista, A. Corradi, L. Foschini, R. Montanari, J. Berrocal, and J. M. Murillo, "A pre-filtering approach for incorporating contextual information into deep learning based recommender systems," *IEEE Access*, vol. 8, pp. 40 485–40 498, 2020.

<sup>19</sup>Due to limited space, we can't show the results of our experiments about the two algorithms (i.e., Algorithms 1 and 2). But, most of the results signify that with our ontology-based semantic similarity, the algorithms can enhance their performances in terms of accuracy.