Available online at http://scik.org
J. Math. Comput. Sci. 2022, 12:X
https://doi.org/10.28919/jmcs/7406
ISSN: 1927-5307

CONCEPT SIMILARITY IN FORMAL CONCEPT ANALYSIS

ANNA FORMICA

Istituto di Analisi dei Sistemi ed Informatica (IASI), National Research Council,
Via dei Taurini 19, I-00185, Rome, Italy

Abstract. The identification of syntactically different concepts that are semantically similar, also referred to as *Similarity Reasoning*, is fundamental in several research areas such as Artificial Intelligence, Software Engineering, Cognitive Science and, in particular, in Semantic Web. Formal Concept Analysis (FCA) is a mathematical framework which is revealing very interesting in supporting fundamental activities for the development of Semantic Web. In order to model uncertainty information, FCA with *many-valued* contexts is addressed and, in particular, FCA with *Ordinal scaling* (OFCA), and FCA with *Interordinal scaling* (IFCA). Concept similarity in IFCA, i.e., in many-valued contexts where attribute values are intervals, is a problem that has been marginally investigated, although the increasing interest in the literature in this topic.

Keywords: Formal Concept Analysis; similarity reasoning; many-valued contexts; FCA with *Ordinal scaling*; FCA with *Interordinal scaling*.

2010 AMS Subject Classification: 68U35.

1. INTRODUCTION

Formal Concept Analysis (FCA) is a formal framework commonly used for data analysis which is based on lattice theory [11, 15]. In the so-called *one-valued* contexts, FCA attributes are crisp, i.e., any object either has or does not have an attribute of that context. However,

*Corresponding author
E-mail address: anna.formica@iasi.cnr.it
Received April 01, 2022

in real life most of attributes are fuzzy, i.e., "it is a matter of degree to which an object has a (fuzzy) attribute" [1]. In other words, an object may have different attributes with different values, and an attribute may apply to different objects with different values. This is the case of *many-valued* contexts [11]. *Fuzzy Formal Concept Analysis* (FFCA) is a generalization of FCA where contexts are many-valued, and the attribute values are real numbers in the range [0,1] or intervals. This kind of FCA is referred to as OFCA, i.e., FCA with *Ordinal scaling* [11], or IFCA, i.e., FCA with *Interordinal scaling* [7].

Similarity Reasoning, i.e., the identification of syntactically different concepts that are semantically close, is fundamental in several research areas such as Artificial Intelligence, Software Engineering, Cognitive Science, and Semantic Web [10, 12], and in different applications, such as for instance in GIS [9]. Concept similarity in the framework of IFCA, i.e., in many-valued contexts where attribute values are intervals, is a problem that has been marginally investigated in the literature, although the increasing interest in this topic.

A concept similarity measure in IFCA has been defined in [7, 8], and combines the *Interval Type-2 Fuzzy Sets* (IT2 FSs) framework [19], with regard to concept extents, and the *information content* approach [13], with regard to concept intents. The latter has been extensively investigated and experimented in the literature, and has a higher correlation with human judgment with respect to the traditional approaches.

The paper is organized as follows. In the next section, the Related Work is given, and in Section 3 first FCA, OFCA and IFCA are informally presented and, then, evaluating concept similarity in IFCA is addressed. Finally Section 4 concludes.

2. Related Work

FCA concept similarity has been addressed in [4], by relying on human domain expertise, and in [5, 17], according to the information content approach, but in both cases within one-valued contexts. In particular, in [17], a method for measuring the similarity of FCA concepts has been proposed, and the Pearson and Spearman correlation coefficients with human judgment have been provided for some of the existing approaches, which is one of the open challenge of this research topic.

Many-valued contexts have been addressed in [6], but in the case of FCA with *Ordinal scaling* (OFCA). With regard to IFCA, a formal framework, referred to as *L-Fuzzy concept theory*, has been defined in [2] which is probably the first research paper providing a theoretical foundation about it. Successively, some interesting works have been defined in the literature which have investigated and deepened the mathematics underlying specific aspects of IFCA, as for instance [3].

In [16] the need for IT2 fuzzy analytical systems for the development of Semantic Web is emphasized, and a similarity measure for IFCA is proposed. It is based on the similarity measure for IT2 FSs defined in [18], the approach presented in [5], and relies on the experimental results given in [6].

3. FCA WITH ONE AND MANY-VALUED CONTEXTS

In order to intuitively recall FCA, the context named *Sardinia Hotels* presented in [7] is used.

In FCA, a *one-valued context* (*context* for short) is a triple (O,A,R), where O is a set of *objects*, A is a set of *attributes*, and R is a binary relation between O and A. In the *Sardinia Hotels* context recalled below, the set O is defined by the following six objects representing six different hotels:

$O = \{H1, H2, H3, H4, H5, H6\}$,

and the set A is defined by the three following attributes:

$A = \{SwPool, Sea, Meal\}$

where *SwPool* stands for swimming pool. Furthermore, the relation R among hotels and attributes is defined by Table 1.

A concept of the *Sardinia Hotels* context is, for instance, the pair (E, I) where E is a set of objects, referred to as concept *extent*, and I is a set of attributes, referred to as concept *intent*, defined as follows:

$((H1, H3, H5),(Sea, Meal))$

since the objects $H1$, $H3$, and $H5$ have both the attributes *Sea* and *Meal*, and vice versa, both these attributes apply to the objects $H1$, $H3$, and $H5$.

	SwPool	**Sea**	**Meal**
H1		×	×
H2	×		×
H3		×	×
H4	×	×	
H5		×	×
H6	×	×	

TABLE 1. The FCA *Sardinia Hotels* context

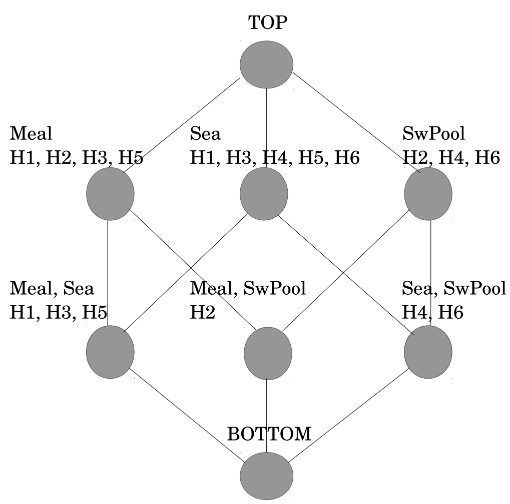

FIGURE 1. Concept Lattice of the *Sardinia Hotels* context

Given a context (O,A,R), consider the set of all the concepts of this context, indicated as $\mathscr{L}(O,A,R)$. Then:

$(\mathscr{L}(O,A,R),\leq)$

is a complete lattice called *Formal Concept Lattice* (*Concept Lattice* for short), i.e., for each subset of concepts, the greatest lower bound (the greatest common subconcept) and the least upper bound (the least common superconcept) exist. For instance, the Concept Lattice constructed from the context of Table 1 is shown in Figure 1.

In a one-valued context an attribute is a property that an object may have or may not have. For instance, according to the one-valued context *Sardinia Hotels* above, each of the attributes

SwPool, *Sea*, and *Meal* applies or does not apply to each of the hotel objects. However, in real world, an attribute may apply to different objects with different values, i.e., it can be many-valued.

Analogously to one-valued contexts, many-valued contexts can be represented by tables, where rows are labeled by objects and columns are labeled by attributes. Many-valued contexts can be transformed into one-valued contexts according to a *conceptual scaling* process [11]. In particular, in this process, each attribute of a many-valued context is interpreted by means of a context, referred to as *conceptual scale* [11]. Typical conceptual scales are *Nominal*, *Ordinal*, and *Interordinal* scales. Nominal scales are used for attribute values which mutually exclude each other, for instance in the case of the attribute values {*human, animal, plant*}. Ordinal scales are suitable when attribute values are ordered, and each value implies the weaker ones, e.g., {*extremely active, very active, active*}. *Interordinal scales* are used for attributes which have a range of possible values (intervals), e.g., {*fully, very much, very few, not at all*}.

In many-valued contexts attributes do not describe objects in a uniform way, i.e., a given attribute applies to different objects in different ways. For instance, in the *Sardinia Hotels* context above, consider the attribute *Meal*. In general, when reserving an hotel, we would like to know whether the hotel provides both lunch and dinner, or half-board. Without the introduction of fuzzy information, we have no way to specify how appropriate is an attribute to a given object.

Consider the many-valued context *Sardinia Hotels* which is specified by the fuzzy relation given in Table 2. Note that crosses in Table 1 have been replaced by grades of membership, from 0 to 1, each allowing us to quantify "how much" an object has, or is described by, an attribute and vice versa an attribute applies to an object.

In the table, the presence of attributes with grade of membership equal to 1.0, such us for instance the attributes *Sea* or *Meal* of the object $H1$, means that the attribute fully applies to the object and vice versa the object is properly described by the attribute. This does not hold for lower membership grades. For example, consider the attribute *Meal* of the object $H2$ which has membership value equal to 0.5. This means that the attribute *Meal* partially applies to the hotel $H2$, for instance because the hotel just provides half board.

	SwPool	Sea	Meal
H1		1.0	1.0
H2	1.0		0.5
H3		0.7	0.5
H4	1.0	1.0	
H5		0.3	1.0
H6	1.0	0.8	

TABLE 2. The many-valued OFCA *Sardinia Hotels* context

Consider now the many-valued context *Sardinia Hotels* which is specified by the fuzzy relation given in Table 3, where crosses in Table 1 have been replaced by *words*, each allowing us to specify "how much" an object has, or is described by, an attribute and vice versa an attribute applies to an object. For instance the hotel $H2$ in Table 3 has the attribute *SwPool* with grade of membership *Fully*, which means that such it fully applies to the hotel $H2$ (and vice versa the hotel $H2$ can be properly described by the attribute *SwPool*). Instead, the object $H2$ has the attribute *Meal* with a membership value *Very*, which means that such an attribute partially applies to this hotel (for instance it could provide meals just for lunch).

	SwPool	Sea	Meal
H1		Fully	Fully
H2	Fully		Very
H3		Very much	Very
H4	Fully	Fully	
H5		Very Few	Fully
H6	Fully	Very much	

TABLE 3. The IFCA *Sardinia Hotels* context, by using words

In order to elaborate such grades of membership, words are replaced by intervals (IT2 FS grades of membership). The association of words with intervals is a problem which has been extensively investigated in the literature and is still attracting a lot of attention [14]. Suppose that words in Table 3 are associated with intervals, as defined in the IFCA context of in Table 4.

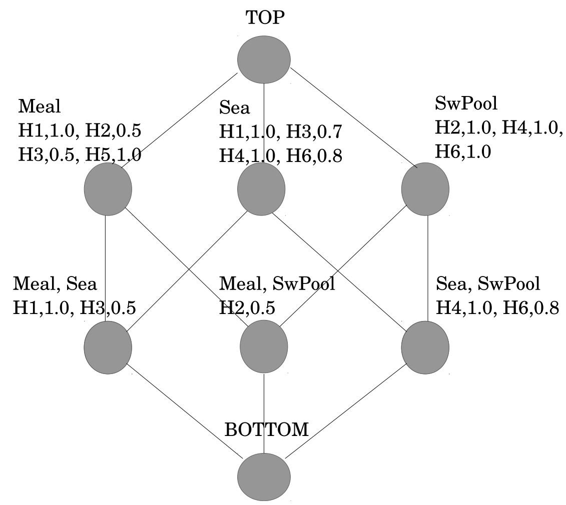

FIGURE 2. Concept Lattice of the OFCA *Sardinia Hotels* context

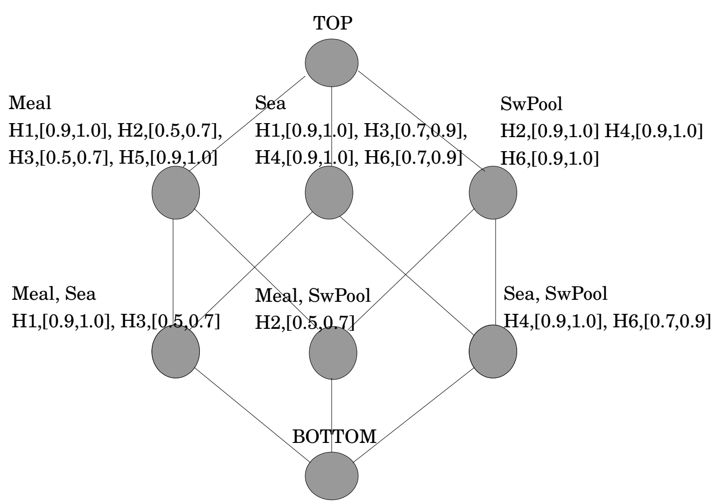

FIGURE 3. Concept Lattice of the IFCA *Sardinia Hotels* context

The OFCA and IFCA Concept Lattices constructed from the contexts of Tables 2 and 4 are shown in Figures 2 and 3, respectively. Note that in the case two or more attributes apply to an object with different grades of membership, e.g., different intervals, the object is associated with the interval having, as lower bound and upper bound, the minimum between the lower bounds and the upper bounds, respectively. The IFCA concept similarity measure proposed in [7, 8] combines the similarity of the concept extents, i.e., the *Interval Type-2 Fuzzy Sets* (IT2 FSs) of objects [18], and the similarity of concept intents, i.e., the sets of attributes. In particular, concept extents are evaluated according to the widely accepted crisp similarity measure for IT2 FSs defined in [18]. Such a notion is used in most applications of general Type-2 Fuzzy Sets due to the simpler underlying mathematics, and allows a relevant simplification about the definition of similarity between fuzzy sets. Concept intents are evaluated according to the *information content* approach [13], which has been extensively experimented in the literature and has a higher correlation with human judgment. Currently, to our knowledge, there are no other proposals for evaluating IFCA concept similarity. The impact about the use of the information content approach within IFCA has been experimented in [6]. In the mentioned paper, the experimental results show that the correlation with human judgment has an average increment of about 0.3, with respect to the compared proposals. Besides the use of the information content approach, this significant increment is due to the combination of the concept extent and the concept intent similarities.

	SwPool	**Sea**	**Meal**
H1		[0.9,1.0]	[0.9,1.0]
H2	[0.9,1.0]		[0.5,0.7]
H3		[0.7,0.9]	[0.5,0.7]
H4	[0.9,1.0]	[0.9,1.0]	
H5		[0.1,0.3]	[0.9,1.0]
H6	[0.9,1.0]	[0.7,0.9]	

TABLE 4. The IFCA *Sardinia Hotels* context

4. Conclusion

In this paper evaluating IFCA concept similarity has been addressed, and the related literature has been recalled. According to [7, 8], it concerns the combination of the similarity of concept extents, that are IT2 FSs, and the similarity of concept intents, that are sets of concept nouns. In particular, concept extents are compared according to the widely accepted crisp similarity measure for IT2 FSs, that allows a relevant simplification about the definition of similarity between general T2 FSs. Concept intents are evaluated according to the *information content* approach, which has been extensively experimented in the literature and has a higher correlation with human judgment.

Although the interest in this research topic is increasing, unfortunately in the literature there are no further significant proposals in this direction that can be compared with the mentioned similarity measure.

Conflict of Interests

The author declares that there is no conflict of interests.

References

[1] R. Belohlavek, What is a fuzzy concept lattice? II, in: S.O. Kuznetsov, D. Slezak, D.H. Hepting, B.G. Mirkin (Eds.), Rough sets, fuzzy sets, data mining and granular computing, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011: pp. 19–26.

[2] A. Burusco, R. Fuentes-González, The study of the interval-valued contexts, Fuzzy Sets Syst. 121 (2001), 439-452.

[3] Y. Djouadi, H. Prade, Interval-valued fuzzy formal concept analysis. In: Rauch, J. et al. (Eds.): Foundations of intelligent systems, International Symposium on Methodologies for Intelligent Systems (ISMIS 2009), Lecture Notes in Computer Science (LNCS) 5722, 2009, pp. 592-601.

[4] A. Formica, Ontology-based concept similarity in formal concept analysis, Inform. Sci. 176 (2006), 2624–2641.

[5] A. Formica, Concept similarity in formal concept analysis: An information content approach, Knowl.-Based Syst. 21 (2008), 80–87.

[6] A. Formica, Similarity reasoning for the semantic web based on fuzzy concept lattices: An informal approach, Inform. Syst. Front. 15 (2013), 511–520.

[7] A. Formica, Similarity reasoning in formal concept analysis: from one- to many-valued contexts, Knowl. Inform. Syst. 60 (2019), 715–739.

[8] A. Formica, Concept similarity in formal concept analysis with many-valued contexts, Comput. Inform. 40 (2021), 469–488.

[9] A. Formica, E. Pourabbas, Content based similarity of geographic classes organized as partition hierarchies, Knowl. Inform. Syst. 20 (2009), 221–241.

[10] A. Formica, F. Taglino, An enriched information-theoretic definition of semantic similarity in a taxonomy, IEEE Access. 9 (2021), 100583–100593.

[11] B. Ganter, R. Wille, Formal concept analysis: Mathematical foundations. Springer (1999).

[12] P. Hitzler, M. Krötzsch, S. Rudolph, Foundations of semantic web technologies. S. Chapman & Hall/CRC, Taylor & Francis Group (2009).

[13] D. Lin, An information-theoretic definition of similarity, Proc. of the Int. Conference on Machine Learning, Madison, Wisconsin, USA, Morgan Kaufmann (1998) 296-304.

[14] J.M. Mendel, Computing with words and its relationships with fuzzistics, Inform. Sci. 177 (2007), 988–1006.

[15] C.M. Rocco, E. Hernandez-Perdomo, J. Mun, Introduction to formal concept analysis and its applications in reliability engineering, Reliab. Eng. Syst. Safe. 202 (2020), 107002.

[16] H. Safaeipour, M.H.F. Zarandi, I.B. Turksen, Developing type-2 fuzzy FCA for similarity reasoning in the semantic web. Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), IEEE (2013) 1477-1482.

[17] F. Wang, N. Wang, S. Cai, W. Zhang, A similarity measure in formal concept analysis containing general semantic information and domain information, IEEE Access. 8 (2020), 75303–75312.

[18] D. Wu, J.M. Mendel, A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets, Inform. Sci. 179 (2009), 1169–1192.

[19] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning—I, Inform. Sci. 8 (1975), 199–249.