

Participatory Learning in Linked Open Data

Marek Z. Reformat¹ and Ronald R. Yager²

¹University of Alberta, Edmonton, Canada

²Iona Collage, New York, USA

Abstract

The introduction of Resource Description Framework (RDF) as a fundamental data representation format of Semantic Web is changing a way how data is stored on the Internet. The intrinsic features of RDF data, i.e., its interconnections and simplicity of expressing information as triples: two entities connected by a property, provide new possibilities of analyzing and absorbing information. The extended format of participatory learning methodology based on propositions is an attractive way of integrating new knowledge. It mimics a human-like way of accepting new facts, expressed as propositions, that could be in contradiction with already known facts.

This paper proposes a method of application of participatory learning to integrate RDF triples collected on the web. The approach presented here includes the concept of conjunctive and disjunctive variables. The learning process is presented, and a simple case study is provided.

Keywords: participatory learning, RDF triples, conjunctive and disjunctive variables, information assimilation, Linked Open Data, Semantic Web

1. Introduction

The concept of Semantic Web [1,7] has been introduced as a new paradigm of storing and utilizing data on the web. The primary data representation format used by the Semantic Web is ontology [4]. Ontology is a set of concepts in a specific domain, together with their detailed definitions, ordered in a hierarchical way. A fundamental building block of any ontology-based representation is Resource Description Framework – RDF [15]. At the same time, RDF becomes a basic element of Linked Open Data (LOD) paradigm [2]. In LOD, all the information is represented as a vast network of interconnected RDF triples.

An RDF triple contains three elements: <subject-property-object>. A <subject- > is an element that a particular piece of information is about, an <-object> is an element that describes the <subject- >, and a <-property- > indicates relationship between the < subject- > and the <-object> of a given triple. There are no restrictions on the type of items that can be used as subjects, objects, and properties. In general, a set of triples with the same subject in each triple constitutes a de-

scription of an entity associated with the subject. This means that any entity represented by RDF triples can be perceived as a set of triples where each of them provides a piece of the description: <-property-object>.

Processes of gathering information and learning in the LOD (RDF environment) require collecting and merging RDF descriptions of entities. Variety of sources with RDF triples results in different quantity and quality of triples defining the same entity. Some of the sources could have extensive descriptions containing many triples for each entity, some sources could have new versions of triples (descriptions), yet some other sources could have different triples, as well as triples conveying different – even contradicting – facts. Therefore, it is important to have a mechanism mimicking a human way of determining if collected pieces of RDF information should be accepted or not, and if not – how many instances of different information should a system encounter before it “changes its attitude” and accepts new, different information.

This paper addresses this issue and proposes application of participatory learning [11-13] and the concept of conjunctive and disjunctive variables [8-10] to develop a mechanism of accumulating information encountered on the web. The contributions of the paper are:

- identification of two types of features: conjunctive and disjunctive, in feature-based RDF descriptions of entities: (Section 3.2);
- representing RDF triples as propositions and application of participatory learning, extended to include conjunctive and disjunctive based propositions, in a process of integrating RDF-based descriptions (Section 4);
- proposing, generally, a concept of human-like information accumulating processes where collecting and integrating information is performed in a common-sense way where accepting new facts found on the web occurs in reference to the current state of knowledge, with careful and prudent assimilation of new and contradicting pieces of information.

2. Background

2.1. RDF

A single RDF-triple <subject-property-object> can be perceived as a feature of an entity identified by the subject. In other words, each single triple is a feature of its subject. Multiple triples with the same subject consti-

tute a definition of a given entity. A simple illustration of this is shown in Fig. 1. It is a definition of the city Berkeley. If we “think graphically” about it, a definition of entity resembles a star.

Quite often a subject and object of one triple can be involved in multiple other triples, i.e., they can be objects or subjects of other triples. In such a case, multiple definitions – RDF-stars – can share features, or some features can be centers of another RDF-stars. Such interconnected triples constitute a network of interleaving definitions of entities.

A description of the same entity can exist in multiple places on the web. We can state that collecting such sets of RDFs can be perceived as a learning process. In such a case, accumulating descriptions of one and the same entity is equivalent to a repetitive process of acquiring information and eventually gaining confidence in gathered descriptions, i.e., RDF triples. This idea is a pivotal aspect of the learning approach described here. It enables a gradual learning, and leads to a strong belief in individual triples. As the result the information about an entity is composed of triples associated with different levels of confidence. And all this depends on a process of assimilating new information.

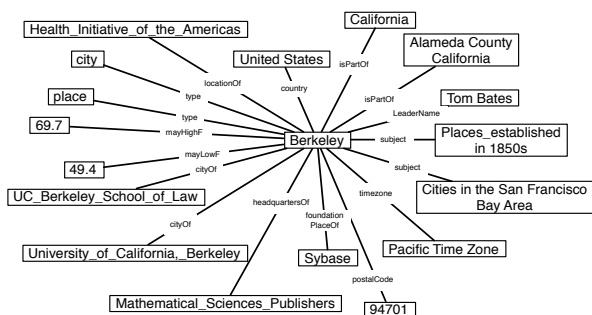


Fig. 1: RDF-based description of Berkeley

2.2. Participatory Learning of Propositional Knowledge

The participatory learning has been proposed as a learning paradigm suitable for accruing new data and information in the presence of existing knowledge base, i.e., in the presence of data and information that have been already learned and believed in [11,12]. It has been applied in multiple applications [5,6], and recently, it has been extended to the environment in which learnt information and knowledge is expressed in terms of declarative statements [13].

The participatory learning is based on the premise that learning is not context free. Every aspect of any learning process fully depends on already known information. The central idea is that new information – in order to be useful and contribute to the overall body of knowledge – has to have some degree of compatibility/consistency with already known information. Before it is accepted, new information is being assessed, i.e., its compatibility with the known information is determined. If a degree of compatibility is high, the new information is integrated, if its compatibility is low – the

information is not added to the existing knowledge base. However, the process of discarding new information is being monitored. A special mechanism called “arousal” is used to conclude if there is something wrong with the known information, and a “radical” change should occur. The arousal mechanism provides a way of accepting information that is incompatible with the current information and that should be discarded because it is very likely that it is incorrect. For details, see [11-12].

The same process is followed in a version of participatory learning for knowledge expressed in the form of propositions [13]. Let us assume the knowledge base – KB – contains a number of propositions:

$$P_i : V_i \text{ is } S_i \text{ and } KB = \{P_1, P_2, \dots, P_k\}$$

where S_i is a subset defined on V_i . A consistency of knowledge base is determined based on the conjunction of the propositions: $Con(KB) = P_1 \cap P_2 \cap \dots \cap P_k$. The base is consistent $Con(KB) \neq \emptyset$ if there exists some interpretation that satisfies all propositions.

New information is also represented in a similar way:

$$P_j^N : V_j \text{ is } S_j \text{ and } {}^N KB = \{P_1, P_2, \dots, P_m\}$$

and its consistency is determined based on the conjunction of propositions.

In general, propositions could be equipped with levels of certainty:

$$P_i : V_i \text{ is } S_i \text{ is } \alpha_i - \text{certain}$$

that represent levels of belief in their exactness.

According to Zadeh’s theory of approximate reasoning [14] we can transform such propositions into their equivalent non-qualified form:

$$P_i : V_i \text{ is } S_i \text{ is } \alpha - \text{certain} \Rightarrow V_i \text{ is } F_i$$

$$F_i(x) = \max(S_i(x), (1 - \alpha_i)) = S_i(x) \vee (1 - \alpha_i)$$

where x ’s are elements of the variable V_i defined on the domain X . We note that Dubois and Prade [3] codified many of these ideas and provided a weighted deductive logic that they called possibility logic.

An essential part of participatory learning is to determine a degree of compatibility ($Comp()$) between the known knowledge KB and the new knowledge ${}^N KB$.

$$Comp(KB, {}^N KB) = \rho = \frac{Con(KB \cup {}^N KB)}{Con(KB)}$$

As it can be seen, the compatibility depends on consistency ($Con()$, see Section 4.2 for its definition) of integrated knowledge bases. A level of compatibility influences the whole process of accepting the new knowledge ${}^N KB$. Besides a degree of compatibility ρ ,

there is another parameter called arousal δ . This parameter is adjusted at the time new information is evaluated, and it echoes a history of “differences” between the known KB and the new ${}^N KB$. A value closer to 1 indicates “persistent differences” and implies that new information should be accepted while the known information should be discarded. Both, ρ and δ , are updated when the new information is considered for integration. The arousal

$$\delta^* = \delta + (1 - \text{Con}({}^N KB)) \cdot \text{Con}({}^N KB) \cdot \beta \cdot (1 - \rho - \delta)$$

reflects a disagreement between the knowledge bases. It takes into account consistency of the new knowledge, its compatibility with the known base, and a learning rate β set up a priori and reflecting “quickness” of the whole learning process. The value of compatibility level is also adjusted:

$$\rho^+ = \rho \vee (\delta^* \wedge \text{Con}({}^N KB))$$

The assimilation process takes place according to the following formula:

$$KB^* = (KB, 1 - \delta^*) \cup ({}^N KB, \rho^+)$$

In other words, each knowledge base has its own “certainty level”: $1 - \delta^*$ for KB , and ρ^+ for ${}^N KB$. An integration process is performed once all propositions are “updated” with respective certainty levels. See [13] for a detailed explanation of the learning, and Section 4 for details related to its application to RDF data.

2.3. Disjunctive and Conjunctive Variables

A closer look at variables’ features and their behavior allows us to identify two types of variables: disjunctive and conjunctive [8-10].

A disjunctive variable is characterized by the fact that it can only assume one value in the base set. The set object associated with the value of a disjunctive variable is indicative of some uncertainty, lack of specificity, in our knowledge about the variable’s value. An example of a disjunctive variable would be a person’s weight, person’s age, or person’s father or mother.

A conjunctive variable on the other hand allows for multiple solutions from the base set. In this case, the set associated with the value of such a variable is indicative of this multiplicity of solution. An example of a conjunctive variable would be the friends of a person.

Let us assume a set $S = \{x_1, x_2, x_3, x_4\}$ and two variables V^D and V^C disjunctive and conjunctive, respectively (NOTE: we will use superscript D to identify disjunctive variables, and C for conjunctive variables). Then

$$V_i^D \text{ is } S \leftrightarrow V_i^D x_1 \text{ or } x_2 \text{ or } x_3 \text{ or } x_4$$

$$V_i^C \text{ are } S \leftrightarrow V_i^C x_1 \text{ and } x_2 \text{ and } x_3 \text{ and } x_4$$

The details of the theory of conjunctive variables and its reasoning schemas are in [8-10].

3. RDF-based Description of Items

3.1. RDF Features

A very important aspect of utilization of RDF triples for entity representation is related to a feature-based method of defining an entity. Each RDF triple is perceived as a single feature of the entity. It means that a feature is composed of two elements: property and object. In other words, a property “defines” a type of relation between the subject and the object, and the object identifies another entity or term (numerical value or string) that the subject is in relation with. A set of triples that share the same subject represents a set of features of this subject – a definition of entity.

An example of a set of triples is shown in Fig. 1. All triples share the same object – Berkeley – and all of them constitute a description of the entity Berkeley. Some of the features of this entity are: $\langle \text{-type-city} \rangle$ and $\langle \text{-type-place} \rangle$, $\langle \text{-country-United States} \rangle$, $\langle \text{-isPartOf-California} \rangle$, $\langle \text{-subject-places established in 1850s} \rangle$, and $\langle \text{-subject-cities in San Francisco Bay Area} \rangle$.

If we take a closer look at the RDF description of the entity Berkeley, Fig. 1, we notice that a number of features have the same property. For example, features $\langle \text{-type-city} \rangle$ and $\langle \text{-type-place} \rangle$ have the same property $\langle \text{-type-} \rangle$. Such a scenario is quite normal in the case of RDF descriptions of entities. Similarly, this can be observed for the properties $\langle \text{-subject-} \rangle$, $\langle \text{-cityOf-} \rangle$, and $\langle \text{-isPartOf-} \rangle$, as well as $\langle \text{-headquartersOf-} \rangle$, and $\langle \text{-foundationPlaceOf-} \rangle$. All these properties can be linked with multiple different objects/values. Other properties, like $\langle \text{-country-} \rangle$, $\langle \text{-postalCode-} \rangle$ can be linked only with a single object/value.

In general, an RDF description of an entity will have two types of properties: ones that can be linked with a single value only, and ones that can be linked with many values.

3.2. RDF: Disjunctive and Conjunctive Features

The feature-based RDF description of an entity, as explained in Section 3.1, provides us with an interesting and important notion of identifying two types of features in RDF descriptions: conjunctive features, and disjunctive features.

In Section 2.3, we have presented basic definitions and characteristics of both types of data. We can say that a feature-based RDF description has features that are “built” using the same property linked to different data – conjunctive features, and features that have to be linked only to a single value – disjunctive features.

The real world examples of RDF descriptions of variety of different entities (taken from dbpedia.org) confirm such categorization of features. Some examples of conjunctive and disjunctive features are presented in Figs 2 and 3, respectively.

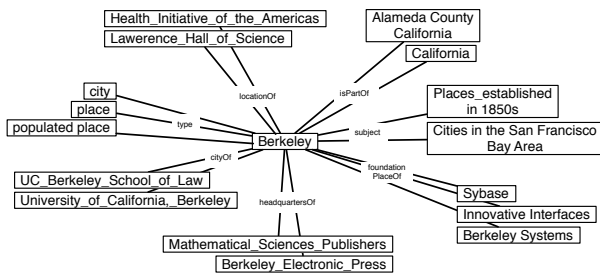


Fig. 2: RDF-based description – conjunctive features

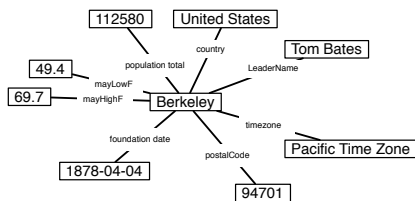


Fig. 3: RDF-based description – disjunctive features

4. RDF and Participatory Learning

4.1. RDF Triples as Propositions

The ability to perceive a single feature of an entity as an RDF triple leads to a very important observation that becomes a basic idea of the application of participatory learning to RDF based data. We state that each RDF triple is a single proposition defined on the domain of values that the triple’s property can assume. For example, a highlighted triple in Fig. 4

Berkeley-country-United States

can be expressed as a proposition with a fuzzy subset S_i

$$P_i : V_{BCountry} \text{ is } S_i : \left\{ \frac{1}{US}, \frac{0}{\text{other ValuesOf } V_{BCountry}} \right\}$$

where $V_{BCountry}$ is a variable, and US as its value is associated with I , while all other possible values of $V_{BCountry}$ have 0 . This means that the Berkeley’s country is US .

With such an approach, a set of RDF triples that defines a given entity can be treated as a knowledge base with propositions denoting features of the entity.

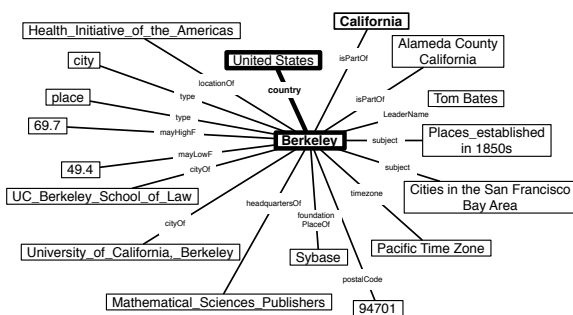


Fig. 4: RDF triples as propositions about Berkeley

Due to the fact that an RDF description of entity can have both disjunctive and conjunctive types of variables, we can distinguish two types of propositions de-

scribing the entity. The propositions built using disjunctive variables V^D that can have only one value, for example propositions built based on properties: <Leader-Name>, <Country>, <Timezone>, Fig. 2; and propositions with conjunctive variables V^C – the ones with properties: <subject>, <type>, <cityOf>, Fig 3. A few examples:

$$P_j^D : V_{BCountry}^D \text{ is } S_j^D = \left\{ \frac{1}{US}, \frac{0}{\text{other ValuesOf } V_{BCountry}^D} \right\}$$

where the variable $V_{BCountry}^D$ can assume a single value from a set of countries;

$$P_n^C : V_{BType}^C \text{ is } S_n^C = \left\{ \frac{1}{city}, \frac{1}{place}, \frac{0}{\text{other ValuesOf } V_{BType}^C} \right\}$$

where the variable V_{BType}^C can assume multiple values from a set of possible types.

4.2. Participatory Learning with RDF Data

As it has been stated earlier, the idea of application of participatory learning is based on a simple concept of treating a set of RDF triples – a description of an entity – as a knowledge base. Therefore, the RDF triples already known are the known knowledge base KB , while a set of triples about an entity that has been acquired from the web is perceived as new information, a new knowledge base NKB . Fig. 5 shows possible differences between already known and new descriptions of the same entity.

The participatory learning process applied to RDF data requires some modifications and adjustments due to existence of two types of variables. The most significant alternation of the original learning method is related to a process of determining consistency and compatibility of known and new knowledge bases. This process is done using only propositions built on disjunctive variables. Once consistency and compatibility values are calculated, the assimilation of new information is performed. It includes, of course, propositions with both types of variables.

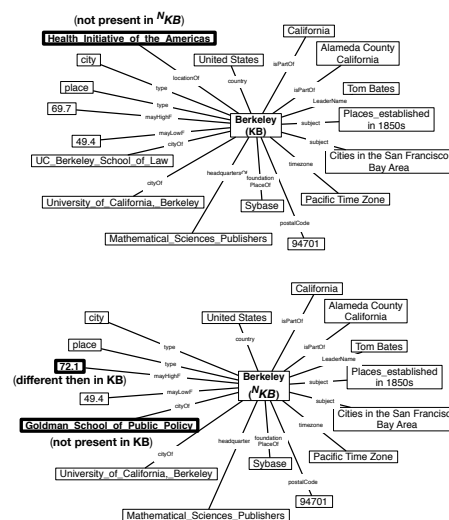


Fig. 5: Possible differences between descriptions of the same entity between KB and NKB

Let us take a look at a formal description of the participatory learning in the environment of RDF triples. This description allow us to show detailed processes of determining consistencies, compatibilities, and a mechanism of combining both knowledge bases, i.e., sets of RDF triples.

Let a known knowledge base, KB , be a set of two types of propositions built based on two types of variables: disjunctive variables (NOTE: we will use the subscript i for disjunctive based propositions)

$$P_i^D : V_i^D \text{ is } S_i \text{ is } \alpha_i - \text{certain}$$

and conjunctive based propositions (NOTE: we will use the subscript n for conjunctive based propositions)

$$P_n^C : V_n^C \text{ is } S_n \text{ is } \alpha_n - \text{certain}$$

They can be represented as equivalent propositions:

$$\begin{aligned} P_i^D : V_i^D \text{ is } [F_i(x_i^D) = \text{Max}(S_i(x_i^D), (1 - \alpha_i))] \\ P_n^C : V_n^C \text{ is } [F_n(x_n^C) = \text{Min}(S_n(x_n^C), \alpha_n)] \end{aligned} \quad (1)$$

where $x_i^D \in X_i^D$, and X_i^D is a domain of disjunctive variable V_i^D , while $x_n^C \in X_n^C$ is a variable from a domain of conjunctive variable V_n^C .

The consistency of the KB is determined only based on propositions built on disjunctive variables. So:

$$\begin{aligned} \text{Con}(KB) &= \text{Con}(KB^{D\text{-only}}) \\ &= \text{Min}\{\text{Max}_{x_1^D}[F_1(x_1^D)], \dots, \text{Max}_{x_i^D}[F_i(x_i^D)], \dots\} \end{aligned} \quad (2a)$$

The same approach is used for representing a new knowledge base and its consistency:

$$\begin{aligned} {}^N P_p^D : V_p^D \text{ is } {}^N F_p(x_p^D) = \text{Max}({}^N S_p(x_p^D), (1 - \alpha_p)) \\ {}^N P_q^C : V_q^C \text{ is } {}^N F_q(x_q^C) = \text{Min}({}^N S_q(x_q^C), \alpha_q) \end{aligned}$$

(NOTE: for the new knowledge base we will use the subscript p for disjunctive based propositions, and q for conjunctive based propositions)

$$\begin{aligned} \text{Con}({}^N KB) &= \text{Con}({}^N KB^{D\text{-only}}) \\ &= \text{Min}\{\text{Max}_{x_1^D}[{}^N F_1(x_1^D)], \dots, \text{Max}_{x_p^D}[{}^N F_p(x_p^D)], \dots\} \end{aligned} \quad (2b)$$

Right now, we can determine a compatibility level ρ between known and new knowledge based:

$$\text{Comp}(KB) = \frac{\text{Con}(KB^{D\text{-only}} \cup {}^N KB^{D\text{-only}})}{\text{Con}(KB^{D\text{-only}})} = \rho \quad (3)$$

where

$$\begin{aligned} \text{Con}(KB^{D\text{-only}} \cup {}^N KB^{D\text{-only}}) &= \\ &= \text{Min}\{ \\ &\quad \text{if } X_i^D = X_p^D : \\ &\quad \quad \text{Max}_{x_i^D}[F_i(x_i^D) \wedge {}^N F_p(x_p^D)] \\ &\quad \dots, \\ &\quad \text{otherwise:} \\ &\quad \quad \text{Max}_{x_i^D}[F_i(x_i^D)] \\ &\quad \quad \text{Max}_{x_p^D}[{}^N F_p(x_p^D)] \\ &\quad \dots\} \end{aligned} \quad (3a)$$

As we can see, the consistency of a combined sets of propositions, i.e., $KB^{D\text{-only}} \cup {}^N KB^{D\text{-only}}$, is determined by finding the maximum of intersection of fuzzy sets on the same domain. Now we are ready to modify two parameters: compatibility ratio ρ and arousal δ .

$$\begin{aligned} \delta^* &= \delta + (1 - \text{Con}({}^N KB^{D\text{-only}})) \cdot \text{Con}({}^N KB^{D\text{-only}}) \\ &\quad \cdot \beta \cdot ((1 - \rho) - \delta) \end{aligned} \quad (4)$$

where β is a learning rate set up for the whole learning process, and

$$\rho^+ = \rho \vee (\delta^* \wedge \text{Con}({}^N KB^{D\text{-only}})) \quad (5)$$

Once we determine the values of δ^* and ρ^+ we can combine the known and new knowledge bases. The new, update knowledge base is built using the following approach:

$$KB^* = (KB, (1 - \delta^*)) \cup ({}^N KB, \rho^+) \quad (6)$$

At the time both knowledge bases are integrated, propositions that use disjunctive variables and propositions with conjunctive variables are combined.

4.3. Integration of Knowledge Bases: Special Cases

An interesting aspect of the integration process is related to how the propositions are modified before the bases are merged. If we look at each side of the union (Eq. 6), propositions of each base are ‘‘augmented’’: KB by a value related to arousal δ^* , and ${}^N KB$ by the compatibility value ρ^+ . This process can be represented in the following way:

$$\begin{aligned} KB^* &= (KB, (1 - \delta^*)) \cup ({}^N KB, \rho^+) \\ &= (F_i(x) \vee \delta^*) \wedge ({}^N F_p(x) \vee (1 - \rho^+)) \text{ for disjunctive var.} \\ &= (F_n(x) \wedge (1 - \delta^*)) \vee ({}^N F_q(x) \wedge \rho^+) \text{ for conjunctive var.} \end{aligned}$$

Let us take a look at different cases, i.e., situations in which both bases are fully compatible, and not compatible, as well as different values of arousal. If the ba-

ses are fully compatible, i.e., $\rho^+ = 1$, and the arousal $\delta^* = 0$, then

$$\begin{aligned} KB^* &= (KB, (1 - \delta^*)) \cup ({}^N KB, \rho^+) \\ &= F_i(x) \wedge {}^N F_p(x) \text{ for disjunctive var.} \quad (6a) \\ &= F_n(x) \vee {}^N F_q(x) \text{ for conjunctive var.} \end{aligned}$$

As we can see, the situation of full compatibility of known and new knowledge bases and low value of arousal means even if we *and-ing* and *or-ing* pieces of information, the known base does not change. If we consider a situation with full compatibility but with the value of $\delta^* = 1$ then

$$\begin{aligned} KB^* &= (KB, (1 - \delta^*)) \cup ({}^N KB, \rho^+) \\ &= {}^N F_p(x) \text{ for disjunctive var.} \quad (6b) \\ &= {}^N F_q(x) \text{ for conjunctive var.} \end{aligned}$$

In such a case, the new knowledge base replaces what is know.

If the bases are not compatible, i.e., $\rho = 0$, so $\rho^+ = \delta^* \wedge \text{Con}({}^N KB^{D\text{-only}})$. In the case the arousal is $\delta^* = 0$, then

$$\begin{aligned} KB^* &= (KB, (1 - \delta^*)) \cup ({}^N KB, \rho^+) \\ &= F_i(x) \text{ for disjunctive var.} \quad (6c) \\ &= F_n(x) \text{ for conjunctive var.} \end{aligned}$$

This means that the new knowledge base is discarded – it is not compatible, and the arousal level does not allow for its assimilation, even if what is known and what has been found do not much – in other words, we are not convinced that the known knowledge should be replaced with a new information.

This situation changes if the value of arousal is $\delta^* = 1$. Now, the new information is accepted – however, it is altered via the level of its consistency.

$$\begin{aligned} KB^* &= (KB, (1 - \delta^*)) \cup ({}^N KB, \rho^+) \\ &= ({}^N F_p(x) \vee (1 - \text{Con}({}^N KB^{D\text{-only}}))) \text{ for disjunctive v.} \quad (6d) \\ &= ({}^N F_q(x) \wedge \text{Con}({}^N KB^{D\text{-only}})) \text{ for conjunctive var.} \end{aligned}$$

5. Case Study

In our case study we would like to illustrate the application of participatory learning in a simple scenario of a small set of RDF triples that represent features of the city Berkeley. The knowledge base KB that contains these triples and propositions built based on them is presented below:

P_1 : Berkeley **type** City

$$P_1 : V_{BType}^C \text{ is } \left\{ \frac{1}{V_{BType}^C = \text{City}}, \frac{0}{V_{BType}^C = \text{other}} \right\} \alpha_1 = 1.0$$

P_2 : Berkeley **type** Place

$$P_2 : V_{BType}^C \text{ is } \left\{ \frac{1}{V_{BType}^C = \text{Place}}, \frac{0}{V_{BType}^C = \text{other}} \right\} \alpha_2 = 1.0$$

P_3 : Berkeley **postalCode** 94701

$$P_3 : V_{BPCode}^D \text{ is } \left\{ \frac{1}{V_{BPCode}^D = 94701}, \frac{0}{V_{BPCode}^D = \text{other}} \right\} \alpha_3 = 1.0$$

P_4 : Berkeley **MayHighTemperature** 96.7

$$P_4 : V_{BHTemp}^D \text{ is } \left\{ \frac{1}{V_{BHTemp}^D = 96.7}, \frac{0}{V_{BHTemp}^D = \text{other}} \right\} \alpha_4 = 1.0$$

KB contains two propositions P_1 and P_2 that are built on the conjunctive variable V_{BType}^C , and two propositions P_3 and P_4 with disjunctive variables V_{BPCode}^D and V_{BHTemp}^D , respectively. In order to illustrate the process of integrating new information in relevance to the possible scenarios from Section 4.3, we assume that a level of certainty is 1.0 for each proposition, and the arousal δ is set up to 1.0. This mimics a scenario of high confidences in the known information, as well as high willingness to assimilate new information. The values of certainty levels α equal to 1.0 mean that the equivalent propositions are the same as the original propositions (see Eq. 1).

5.1. Scenario A: Assimilation of Information We Already Know

The first scenario is an example of an attempt to assimilate information that is already in the knowledge base KB . A set of new RDF triples, called hereafter KB_A , contains just three triples shown below with the propositions:

KB_A:

P_{A1} : Berkeley **type** City

$$P_{A1} : V_{BType}^C \text{ is } \left\{ \frac{1}{V_{BType}^C = \text{City}}, \frac{0}{V_{BType}^C = \text{other}} \right\} \alpha_{A1} = 0.5$$

P_{A2} : Berkeley **type** PopulatedPlace

$$P_{A2} : V_{BType}^C \text{ is } \left\{ \frac{1}{V_{BType}^C = \text{PPlace}}, \frac{0}{V_{BType}^C = \text{other}} \right\} \alpha_{A2} = 0.5$$

P_{A3} : Berkeley **postalCode** 94701

$$P_{A3} : V_{BPCode}^D \text{ is } \left\{ \frac{1}{V_{BPCode}^D = 94701}, \frac{0}{V_{BPCode}^D = \text{other}} \right\} \alpha_{A3} = 0.5$$

The assumption about the levels of certainty – the value of 0.5 – reflects a partial belief in the triples. In the first step we build equivalent propositions (Eq. 1):

$$P_{A1} : V_{BType}^C \text{ is } \left\{ \frac{0.5}{V_{BType}^C = \text{City}}, \frac{0.5}{V_{BType}^C = \text{other}} \right\}$$

$$P_{A2} : V_{BType}^C \text{ is } \left\{ \frac{0.5}{V_{BType}^C = PPlace}, \frac{0.5}{V_{BType}^C = other} \right\}$$

$$P_{A3} : V_{BPCode}^D \text{ is } \left\{ \frac{1}{V_{BPCode}^D = 94701}, \frac{0.5}{V_{BPCode}^D = other} \right\}$$

At this stage we calculate the consistency of KB_A based on the disjunctive proposition (Eq. 2b):

$$Con(KB_A) = Con(P_{A3}) = 1.0,$$

and its compatibility with the knowledge base KB (Eq. 3a):

$$Com(KB / KB_A) = \frac{Con(KB^{D-only} \cup KB_A^{D-only})}{Con(KB^{D-only})} = \rho = 1.0$$

This means that both bases contain the same information. The updated value of δ^* (Eq. 4) is:

$$\delta^* = \delta + (1 - Con(KB_A^{D-only})) \cdot Con(KB_A^{D-only}) \\ \cdot \beta \cdot ((1 - \rho) - \delta) = \delta = 1.0$$

where β is a learning rate set up to 1.0. The modified compatibility rate ρ^+ (Eq. 5) is:

$$\rho^+ = \rho \vee (\delta^* \wedge Con(KB_A^{D-only})) = \rho = 1.0.$$

With such values of the compatibility rate and arousal, the new KB^* composed of KB and KB_A is (Eq. 6b):

$$KB^* = (KB, (1 - \delta^*)) \cup (KB_A, \rho^+) \\ = (F_i(x) \vee \delta^*) \wedge ({}^N F_p(x) \vee (1 - \rho^+)) \text{ for disjunctive var.} \\ = (F_n(x) \wedge (1 - \delta^*)) \vee ({}^N F_q(x) \wedge \rho^+) \text{ for conjunctive var.} \\ \text{so} \\ = (F_i(x) \vee 1.0) \wedge (F_{A,p}(x) \vee 0.0) \text{ for disjunctive var.} \\ = (F_n(x) \wedge 0.0) \vee (F_{A,q}(x) \wedge 1.0) \text{ for conjunctive var.}$$

what means that:

$$KB^* = \\ = F_{A,p}(x) \text{ for disjunctive var.} \\ = F_{A,q}(x) \text{ for conjunctive var.}$$

Because ${}^N KB$ and KB are fully compatible, the new information is assimilated but it does not add “anything new” when the triple P_{A1} replaces P_1 , and P_{A3} replaces P_3 . The only change is addition of the conjunctive-based proposition P_{A2} not known before.

5.2. Scenario B: Assimilation of Information that is in Conflict with What We Know

This scenario involves the original knowledge base KB and two new RDF triples, named KB_B . One of them – P_{B2} – is in a conflict with the triple (proposition) P_4 from the original KB .

KB_B:

P_{B1} : Berkeley type PopulatedPlace

$$P_{B1} : V_{BType}^C \text{ is } \left\{ \frac{1}{V_{BType}^C = PPlace}, \frac{0}{V_{BType}^C = other} \right\} \alpha_{B1} = 0.5$$

P_{B2} : Berkeley MayHighTemperature 98.1

$$P_{B2} : V_{BHTemp}^D \text{ is } \left\{ \frac{1}{V_{BHTemp}^D = 98.1}, \frac{0}{V_{BHTemp}^D = other} \right\} \alpha_{B2} = 0.5$$

Once the equivalent propositions are created, we calculate the consistency of KB_B based on the disjunctive proposition (Eq. 2b):

$$Con(KB_B) = Con(P_{B2}) = 1.0,$$

The compatibility of KB_B with the knowledge base KB (Eq. 3a) is:

$$Comp(KB / KB_B) = \frac{Con(KB^{D-only} \cup KB_B^{D-only})}{Con(KB^{D-only})} = \rho = 0.0$$

and this reflects the fact that the information carried by the disjunctive proposition P_{B2} of KB_B is different when compared with P_4 of KB . At the same time, the values of δ^* (Eq. 4) and ρ^+ (Eq 5) are:

$$\delta^* = \delta + (1 - Con(KB_B^{D-only})) \cdot Con(KB_B^{D-only}) \\ \cdot \beta \cdot ((1 - \rho) - \delta) = \delta = 1.0$$

where $\beta = 1.0$

$$\rho^+ = \rho \vee (\delta^* \wedge Con(KB_B^{D-only})) = Con(KB_B^{D-only})$$

Overall, this leads to the following integration of both bases (Eq. 6d):

$$KB^* = (KB, (1 - \delta^*)) \cup (KB_B, \rho^+) \\ = (F_i(x) \vee \delta^*) \wedge ({}^N F_p(x) \vee (1 - \rho^+)) \text{ for disjunctive var.} \\ = (F_n(x) \wedge (1 - \delta^*)) \vee ({}^N F_q(x) \wedge \rho^+) \text{ for conjunctive var.} \\ \text{so} \\ = F_{B,p}(x) \vee (1 - Con(KB_B^{D-only})) \text{ for disjunctive var.} \\ = F_{B,q}(x) \wedge Con(KB_B^{D-only}) \text{ for conjunctive var.}$$

The fact that the consistency of a new knowledge base $Con(KB_B^{D-only}) = 1.0$, means that

$$KB^* =$$

$$= F_{B,p}(x) \text{ for disjunctive var.}$$

$$= F_{B,q}(x) \text{ for conjunctive var.}$$

Therefore, the new and different piece of information is accepted, i.e., the “conflicting” triple (propositions) P_{B2} replaces the triple P_4 .

6. Conclusion

The Resource Description Framework (RDF) perceived as a data representation format in Linked Open Data that automatically interconnects pieces of information and yet is simple, is a promising way of storing information. The simple translation of RDF triples into a propositional format creates an interesting opportunity to provide a new, more human-like method of accumulating information, reason about it, and eventually create knowledge.

This paper proposes an approach of combining participatory learning and conjunctive/disjunctive variables, and their application to integrating RDF-based propositions. The introduced method of aggregating propositions equipped with levels of belief in them, and a methodology of handling inconsistent pieces of information is illustrated here.

References

- [1] T. Berners-Lee, J. Hendler, and O. Lassila, The semantic web, *Scientific American* 284, 34-43, 2001.
- [2] C. Bizer, T. Heath, and T. Berners-Lee, Linked data-the story so far, *International Journal on Semantic Web and Information Systems* 4, 1-22, 2009.
- [3] D. Dubois, and H. Prade, Possibilistic logic: a retrospective and prospective view, *Fuzzy Sets and Systems* 144, 3-23, 2004.
- [4] T. R. Gruber, A translation approach to portable ontology specifications, *Knowledge Acquisition* 5, 199-220, 1993.
- [5] M. Hell, R. Ballini, P. Costa, and F. Gomide, Training Neurofuzzy Networks with Participatory Learning, *FUZZ-IEEE Conference*, 1-6, 2007.
- [6] L. Maciel, F. Gomide, and R. Ballini, Enhanced evolving participatory learning fuzzy modeling: an application for asset returns volatility forecasting, *Evolving Systems* 5, 75-88, 2014.
- [7] N. Shadbolt, W. Hall, and T. Berners-Lee, The semantic web revisited, *Intelligent Systems* 21, 96-101, 2006.
- [8] R.R. Yager, Set based representations of conjunctive and disjunctive knowledge, *Information Sciences* 41, 1-22, 1987.
- [9] R.R. Yager, Towards a theory of conjunctive variables, *Int. J. General Systems* 13, 203-227, 1987.
- [10] R.R. Yager, Reasoning with conjunctive knowledge, *Fuzzy Sets and Systems* 28, 69-83, 1988.
- [11] R.R. Yager, A model of participatory learning,

IEEE Transactions on Systems, Man and Cybernetics 20, 1229-1234, 1990.

- [12] R.R. Yager, Participatory learning with granular observations, *IEEE Transactions on Fuzzy Sets and Systems* 17, 1-13, 2009.
- [13] R.R. Yager, Participatory learning of propositional knowledge, *IEEE Transactions on Fuzzy Sets and Systems* 20, 715-727, 2012.
- [14] L.A. Zadeh, A theory of approximate reasoning, *Machine Intelligence* 9, 149-194, 1979.
- [15] RDF Primer, <http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/> (accessed January 31, 2015).