Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 323–330
ISBN 978-83-60810-51-4

Dominance-Based Rough Set Approach for Decision Systems over Ontological Graphs

Krzysztof Pancerz
Institute of Biomedical Informatics
University of Information Technology and Management in Rzeszów, Poland
Email: kpancerz@wsiz.rzeszow.pl

***Abstract*—In the paper, we continue the research in the field of decision systems over ontological graphs. Some relations defined over attribute values in such systems are partial order relations. Therefore, we try to incorporate Dominance-Based Rough Set Approach (DRSA) for decision systems over ontological graphs. Due to this approach, we can define decision rules, similar to that defined in DRSA, which give us a new look on data included in such systems.**

I. Introduction

IN [1], we have defined information systems (decision systems) over ontological graphs. In such systems, attribute values are considered in the ontological (semantic) space. Ontology is represented by means of graph structures. Information (decision) systems over the ontological graphs can be created in different ways:

1) Attribute values of a given information system are concepts from ontologies assigned to attributes - a simple information system over ontological graphs.
2) Attribute values of a given information system are local ontological graphs of ontologies assigned to attributes - a complex information system over ontological graphs.

In our current investigation, we focus on the first approach.

Comparing this approach to the classic one, simple sets of attribute values are replaced by ontological graphs which deliver us some new knowledge about meanings of attribute values. For this case, decision rules in decision systems can be seen from different perspectives, for example, taking into consideration synonymy, generality or some more sophisticated properties determining meanings of attribute values.

Incorporation of ontological graphs into a knowledge discovery process from information (decision) systems can give us some benefits. Such benefits serve as motivation, briefly described in Section II, for developing a new approach. In Section III, we recall basic notions defined for simple information systems (decision systems) over ontological graphs. Among others, relations over attribute values are mentioned. Some of them are partial order relations, for example, a generalization relation and a specialization relation. Such relations define preference-orders of attribute domains. The original rough set approach is not able to deal with preference ordered attribute domain. Therefore, some modification, called Dominance-Based Rough Set Approach (DRSA), has been proposed by Greco, Slowinski and Matarazzo (cf. [2]). In this paper, we try to incorporate that approach for decision systems over ontological graphs. Decision rules, similar to that defined in DRSA, give us a new look on data included in such systems when the knowledge about meanings of attribute values is used.

II. Motivation

Classic rough set theory is based on an indiscernibility relation defined over a set of objects, described by attribute value vectors, in a given information (decision) system (cf. [3]). Two objects are indiscernible if and only if for each attribute its value is the same for both objects. "The same" means that attribute values are expressed by the same symbol. In this case, we can say that we are interested in exact (literal, symbol-precise) meaning of attribute values. Especially, it is a problem in case of categorical values. Such an approach does not take into consideration a semantic meaning of attribute values. We can give two simple examples:

- Let us consider two concepts car and $automobile$ which can be two values of a given attribute. The semantic meaning of such concepts is the same (they are synonyms). However, from the point of view of the classic indiscernibility relation, they are treated as different values (concepts).
- Let us consider two concepts $house$ and $bungalow$ which can be two values of a given attribute. The semantic meaning of $house$ is wider, $bungalow$ is a house that is all on one level, without stairs. However, $bungalow$ is also $house$. It is an example of the classic "is-a" relation known in many computer science areas, e.g., in object oriented programming. From the point of view of the classic indiscernibility relation, they are also treated as different values (concepts).

To avoid such nuisances in the interpretation of meanings of attribute values, we propose to put attribute values in the ontological (semantic) space. Replacing attribute value sets by ontological graphs leads to generalization of different notions defined in classic rough set theory. Therefore, notions defined in the presented paper are based on classic notions.

III. Decision Systems over Ontological Graphs

In this section, we recall basic notions related to information systems (decision systems) over ontological graphs. According to definitions of ontology given by Neches et al. [4] and Kohler [5], ontology is constructed on the basis of a controlled vocabulary and the relationships of the concepts in the controlled vocabulary. Formally, the ontology can be represented by means of graph structures. In our approach, the graph representing the ontology $\mathcal{O}$ is called the ontological graph. In such a graph, each node represents one concept from $\mathcal{O}$, whereas each edge represents a relation between two concepts from $\mathcal{O}$.

Definition 3.1: Let $\mathcal{O}$ be a given ontology. An ontological graph is a quadruple $OG = (\mathcal{C}, E, \mathcal{R}, \rho)$, where

- $\mathcal{C}$ is a nonempty, finite set of nodes representing concepts in the ontology $\mathcal{O}$,
- $E \subseteq \mathcal{C} \times \mathcal{C}$ is a finite set of edges representing relations between concepts from $\mathcal{C}$,
- $\mathcal{R}$ is a family of semantic descriptions (in natural language) of types of relations (represented by edges) between concepts,
- $\rho : E \to \mathcal{R}$ is a function assigning a semantic description of the relation to each edge.

In the proposed approach, we take into consideration the following family of semantic descriptions of relations between concepts (cf. [1]):

$$\mathcal{R} = \{\text{"is synonymous with"}, \text{"is generalized by"}, \text{"is specialized by"}\}.$$

We will use the following notation: $R_\sim$ - "is synonymous with", $R_\lhd$ - "is generalized by", $R_\rhd$ - "is specialized by". Therefore, for simplicity $\mathcal{R} = \{R_\sim, R_\lhd, R_\rhd\}$.

Relations have the following properties:

- $R_\sim$ is reflexive, symmetric and transitive (it is an equivalence relation),
- $R_\lhd$ is reflexive and transitive,
- $R_\rhd$ is reflexive and transitive.

Remark 3.1: In the graphical representation of the ontological graph, for readability, we will omit reflexivity of relations. However, the above relations are reflexive, i.e., a given concept is synonymous with itself, a given concept is generalized by itself, a given concept is specialized by itself.

Definition 3.2: A simple decision system SDS^{OG} over ontological graphs is a tuple

$$SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d),$$

where:

- U is a nonempty, finite set of objects,
- C is a nonempty, finite set of condition attributes,
- D is a nonempty, finite set of decision attributes,
- $\{OG_a\}_{a \in C}$ is a family of ontological graphs associated with condition attributes from C,
- $V_d = \bigcup_{a \in D} V_a$, V_a is a set of values of the decision attribute $a \in D$,
- $c : U \times C \to \mathcal{C}$, where $\mathcal{C} = \bigcup_{a \in C} \mathcal{C}_a$, is an information function such that $f(u, a) \in \mathcal{C}_a$ for each $u \in U$ and $a \in C$, where $\mathcal{C}_a$ is a set of concepts from the graph OG_a,
- $d : U \times D \to V$ is a decision function such that $f(u, a) \in V_d$ for each $u \in U$ and $a \in D$.

Remark 3.2: It is not necessary for an information function to be a total function, i.e., $c : U \times C \to \mathcal{C}^* \subseteq \mathcal{C}$.

Example 3.3: Let us consider a simple decision system $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$, represented by a decision table (see Table I), over ontological graphs shown in Figures 1 and 2. For this system, we have:

- $U = \{u_1, u_2, ..., u_8\}$ is a set of eight persons described with respect to their material status,
- $C = \{Vehicle, Dwelling\}$ is a set of condition attributes describing selected possessions of persons,
- $D = \{Material\ Status\}$ is a set of decision attributes, D consists of one attribute evaluating material status of persons,
- $OG_{Vehicle} = (\mathcal{C}_V, E_V, \mathcal{R}, \rho_V)$ is an ontological graph associated with the attribute $Vehicle$, where
$$\mathcal{C}_V = \{Bicycle, Car, Lorry, Minivan, SUV, Vehicle\}$$
and the rest of elements can be obtained from Figure 1,
- $OG_{Dwelling} = (\mathcal{C}_D, E_D, \mathcal{R}, \rho_D)$ is an ontological graph associated with the attribute $Dwelling$, where
$$\mathcal{C}_D = \{Bungalow, Condominium, Cottage, Dwelling, House, Flat, Logging\}$$
and the rest of elements can be obtained from Figure 2,
- $\mathcal{R} = \{R_\sim, R_\lhd, R_\rhd\}$,
- $V_d = \{Low, Medium, High\}$,
- c is an information function and d is a decision function, both defined in the tabular form in Table I.

Obviously, ontological graphs used in this example have been simplified in comparison to ontological graphs expressing real-world relations between concepts.

IV. Relations over Sets of Attribute Values

In our approach presented in [1], we have proposed to consider some relations defined over sets of attribute values in simple information systems over ontological graphs:

- an exact meaning relation,
- a synonym meaning relation,
- a general meaning relation of at most k-th order,
- a proper generalization relation of at most k-th order,
- a generalization relation of at most k-th order,
- a proper specialization relation of at most k-th order,
- a specialization relation of at most k-th order.

In the defined relations, we use some additional knowledge about relationships between attribute values which is included in ontological graphs.

We will take into consideration two of the recalled relations: a generalization relation and a specialization relation. However, the presented approach is also valid for a proper generalization relation as well as a proper specialization relation, respectively. In the second case, we do not admit synonyms in generalization (specialization) paths.

TABLE I
A SIMPLE DECISION SYSTEM OVER ONTOLOGICAL GRAPHS

$U/C \cup D$	$Vehicle$	$Dwelling$	$Material\ Status$
u_1	Car	$Lodging$	$Medium$
u_2	$Minivan$	$House$	$High$
u_3	Car	$Flat$	$Medium$
u_4	$Bicycle$	$Lodging$	Low
u_5	SUV	$Bungalow$	$High$
u_6	Car	$Lodging$	Low
u_7	Car	$Condominium$	$Medium$
u_8	Car	$Cottage$	$Medium$

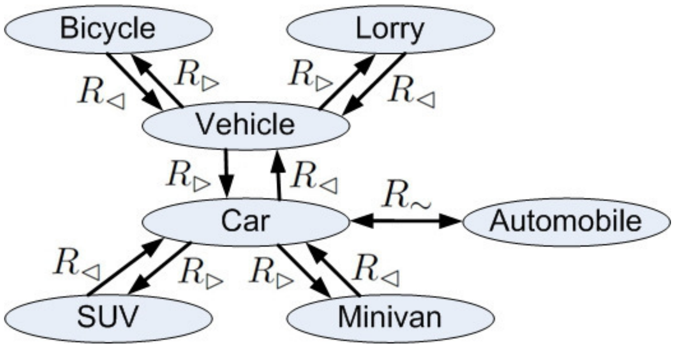

Fig. 1. An ontological graph $OG_{Vehicle}$ associated with the attribute $Vehicle$.

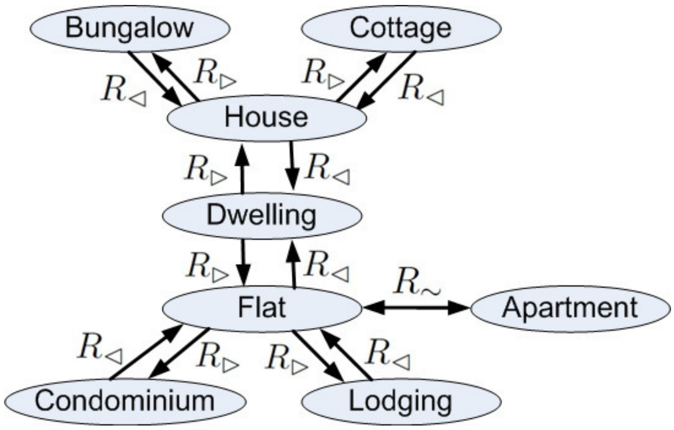

Fig. 2. An ontological graph $OG_{Dwelling}$ associated with the attribute $Dwelling$.

Let $OG = (\mathcal{C}, E, \mathcal{R}, \rho)$ be an ontological graph. We will use the following notation: $[c_i, c_j]$ is a simple path in OG between $c_i, c_j \in \mathcal{C}$, $\mathcal{E}([c_i, c_j])$ is a set of edges from E belonging to the simple path $[c_i, c_j]$, $\mathcal{P}(OG)$ is a set of all simple paths in OG. In the literature, there are different definitions for a simple path in the graph. In this paper, we follow the definition in which a path is simple if no node or edge is repeated, with the possible exception that the first node is the same as the last. Therefore, the path $[c_i, c_j]$, where $c_i, c_j \in \mathcal{C}$ and $c_i = c_j$ can also be a simple path in OG.

Definition 4.1: Let an ontological graph $OG_a = (\mathcal{C}_a, E_a, \mathcal{R}, \rho_a)$ be associated with the attribute a in a simple information system, where $\mathcal{R} = \{R_\sim, R_\lhd, R_\rhd\}$.

- A generalization relation $GR^k(a)$ of at most k-th order is a set of all pairs $(c_1, c_2) \in \mathcal{C}_a \times \mathcal{C}_a$ satisfying the following condition: there exists $[c_1, c_2] \in \mathcal{P}(OG_a)$ such that $\underset{e \in \mathcal{E}([c_1,c_2])}{\forall} \rho_a(e) \in \{R_\sim, R_\lhd\}$ and $card(\{e \in \mathcal{E}([c_1, c_2]) : \rho_a(e) = R_\lhd\}) \leq k$.
- A specialization relation $SR^k(a)$ of at most k-th order is a set of all pairs $(c_1, c_2) \in \mathcal{C}_a \times \mathcal{C}_a$ satisfying the following condition: there exists $[c_1, c_2] \in \mathcal{P}(OG_a)$ such that $\underset{e \in \mathcal{E}([c_1,c_2])}{\forall} \rho_a(e) \in \{R_\sim, R_\rhd\}$ and $card(\{e \in \mathcal{E}([c_1, c_2]) : \rho_a(e) = R_\rhd\}) \leq k$.

Moreover, let us omit the order of relations. Therefore, we have:

- A generalization relation $GR(a)$ being a set of all pairs $(c_1, c_2) \in \mathcal{C}_a \times \mathcal{C}_a$ satisfying the following condition: there exists $[c_1, c_2] \in \mathcal{P}(OG_a)$ such that $\underset{e \in \mathcal{E}([c_1,c_2])}{\forall} \rho_a(e) \in$

$\{R_\sim, R_\triangleleft\}$.

- A specialization relation $SR(a)$ being a set of all pairs $(c_1, c_2) \in \mathcal{C}_a \times \mathcal{C}_a$ satisfying the following condition: there exists $[c_1, c_2] \in \mathcal{P}(OG_a)$ such that $\forall_{e \in \mathcal{E}([c_1,c_2])} \rho_a(e) \in \{R_\sim, R_\triangleright\}$.

GR and SR are reflexive and transitive relations.

We can identify a generalization relation with a dominance relation whereas a specialization relation with an inverse dominance relation.

Definition 4.2: Let an ontological graph $OG_a = (\mathcal{C}_a, E_a, \mathcal{R}, \rho_a)$ be associated with the attribute a in a simple information system and $c_1, c_2 \in \mathcal{C}_a$. It is said that c_1 dominates c_2, written as $D^{\geq}(c_1, c_2)$, if $(c_2, c_1) \in SR(a)$, i.e., c_2 is specialized by c_1.

Definition 4.3: Let an ontological graph $OG_a = (\mathcal{C}_a, E_a, \mathcal{R}, \rho_a)$ be associated with the attribute a in a simple information system and $c_1, c_2 \in \mathcal{C}_a$. It is said that c_1 is dominated by c_2, written as $D^{\leq}(c_1, c_2)$, if $(c_2, c_1) \in GR(a)$, i.e., c_2 is generalized by c_1.

A generalization relation is colloquially called an "is-a" relation. In the Dominance-Based Rough Set Approach (DRSA) [2], an outranking relation S_q [6] corresponding to criterion q is used. In this case $S_q(x, y)$ means that "x is at least as good as y with respect to criterion q". In our approach, covering an "is-a" relation, the meaning will be quite similar, i.e., "x is at least y with respect to the ontological graph OG".

Definition 4.4: Let a simple decision system over ontological graphs $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be given, $a \in C$ and $u \in U$. An a-dominating set with respect to u is a set $D_a^+(u) = \{v \in U : D^{\geq}(a(v), a(u))\}$. An a-dominated set with respect to u is a set $D_a^-(u) = \{v \in U : D^{\leq}(a(u), a(v))\}$.

Remark 4.1: Let a simple decision system over ontological graphs $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be given, $a \in C$ and $v \in \mathcal{C}_a$ of OG_a. We use the following notation:

- By D_a^{+v} we denote the set $\{u \in U : D^{\geq}(a(u), v)\}$, i.e., the set of all objects u in U for which $a(u)$ dominates v.
- By D_a^{-v} we denote the set $\{u \in U : D^{\leq}(a(u), v)\}$, i.e., the set of all objects u in U for which $a(u)$ is dominated by v.

The a-dominating set with respect to u is a set of all objects from U having concepts assigned to them by the attribute a which specialize the concept assigned to u by the attribute a according to an ontological graph OG_a. Analogously, the a-dominated set with respect to u is a set of all objects from U having concepts assigned to them by the attribute a which generalize the concept assigned to u by the attribute a according to an ontological graph OG_a.

V. Decision Rules Consistent with Dominance Principle

In this section, we use definitions related to the Dominance-Based Rough Set Approach, given, among others, in [2], to provide notions for decision rules in a simple decision system over ontological graphs.

Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system over ontological graphs, $a_d \in D$ and $\mathbf{Cl}_{a_d} = \{Cl_t : t \in T\}$, where $T = \{1, \ldots, n\}$, be a set of classes of U determined by a_d, such that $u \in U$ belongs to one and only one class $Cl_t \in \mathbf{Cl}_{a_d}$. Moreover, suppose that we can define a complete preorder, i.e., a strongly complete and transitive binary relation S_{a_d}, for a decision attribute $a_d \in D$ in SDS^{OG}. For $r, s \in T$, $r > s$ means that each element of Cl_r is preferred (strictly or weakly) to each element of Cl_s. For $u, v \in U$, $(u, v) \in S_{a_d}$ means "u is at least as good as v".

For a family $\mathbf{Cl}_{a_d}$ of classes, we define:

- an upward union of classes: $Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s$,
- a downward union of classes: $Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s$,

where $Cl_t, Cl_s \in \mathbf{Cl}_{a_d}$.

Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system over ontological graphs, $a_c \in C$, $a_d \in D$ and $Cl_t^{\geq}$, $Cl_t^{\leq}$ be unions of classes determined by a_d, upward and downward, respectively. We define:

- the a_c-lower approximation of $Cl_t^{\geq}$:
$$\underline{a_c}(Cl_t^{\geq}) = \{u \in U : D_{a_c}^+(u) \subseteq Cl_t^{\geq}\},$$
- the a_c-upper approximation of $Cl_t^{\geq}$:
$$\overline{a_c}(Cl_t^{\geq}) = \bigcup_{u \in Cl_t^{\geq}} D_{a_c}^+(u),$$
- the a_c-lower approximation of $Cl_t^{\leq}$:
$$\underline{a_c}(Cl_t^{\leq}) = \{u \in U : D_{a_c}^-(u) \subseteq Cl_t^{\leq}\},$$
- the a_c-upper approximation of $Cl_t^{\leq}$:
$$\overline{a_c}(Cl_t^{\leq}) = \bigcup_{u \in Cl_t^{\leq}} D_{a_c}^-(u).$$

Let $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be a simple decision system over ontological graphs. Let $\mathcal{C} = \bigcup_{a \in C} \mathcal{C}_a$, where $\mathcal{C}_a$ is a set of concepts from the graph OG_a associated with a given $a \in C$. For the decision system SDS^{OG}, we define:

- condition descriptors which are expressions $(a, v)^{\geq}$ over C and $\mathcal{C}$, where $a \in C$ and $v \in \mathcal{C}$, read as "a is at least v" according to OG_a,
- decision descriptors which are expressions $(a, v)^{\geq}$ over D and V_d, where $a \in D$ and $v \in V_d$, read as "a is at least v" according to a complete preorder defined for a.

Analogously, we define:

- condition descriptors which are expressions $(a, v)^{\leq}$ over C and $\mathcal{C}$, where $a \in C$ and $v \in \mathcal{C}$, read as "a is at most v" according to OG_a,
- decision descriptors which are expressions $(a, v)^{\leq}$ over D and V_d, where $a \in D$ and $v \in V_d$, read as "a is at most v" according to a complete preorder defined for a.

In this paper, we consider elementary decision rules. Complex decision rules are subjected to further research. An elementary rule is said to be a rule with one condition descriptor.

A complex rule is said to be a rule with many condition descriptors linked with propositional connectives.

The following two types of elementary decision rules can be considered in a simple decision system over ontological graphs $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$:

1) a $D_{\geq}$-elementary decision rule:

$$(a_c, r_c)^{\geq} \Rightarrow (a_d, v_d)^{\geq},$$

2) a $D_{\leq}$-elementary decision rule:

$$(a_c, r_c)^{\leq} \Rightarrow (a_d, v_d)^{\leq},$$

where $a_c \in C$, $r_c \in \mathcal{C}_{a_c}$ of OG_{a_c}, $a_d \in D$, $v_d \in V_d$ and $\phi \Rightarrow \psi$ is read: "if ϕ, then ψ.

Decision rule (1) can be read in the following way: "if a_c is at least r_c, then a_d is at least v_d". Analogously, decision rule (2) can be read in the following way: "if a_c is at most r_c, then a_d is at most v_d".

Decision rule (1) is true (valid, certain) in SDS^{OG} if and only if:

$$D_{a_c}^{+r_c} \subseteq Cl_{v_d}^{\geq}$$

and

$$D_{a_c}^{+r_c} \neq \emptyset,$$

where Cl_{v_d} denotes the class of objects $u \in U$ such that $a_d(u) = v_d$. Analogously, decision rule (2) is true (valid, certain) in SDS^{OG} if and only if:

$$D_{a_c}^{-r_c} \subseteq Cl_{v_d}^{\leq}$$

and

$$D_{a_c}^{-r_c} \neq \emptyset,$$

where Cl_{v_d} denotes the class of objects $u \in U$ such that $a_d(u) = v_d$.

Example 5.1: Let us consider a simple decision system over ontological graphs $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ given in Example 3.3. In this example, we will examine $D_{\geq}$-elementary decision rules and $D_{\leq}$-elementary decision rules in SDS^{OG}.

1) Let us consider the following rule in SDS^{OG}:

$$(Dwelling, House)^{\geq} \Rightarrow (Material\ Status, Medium)^{\geq},$$

This rule means that if $Dwelling$ is at least $House$, then $Material\ Status$ is at least $Medium$. From $OG_{Dwelling}$, we obtain that $Bungalow$ dominates $House$, i.e., $D^{\geq}(Bungalow, House)$, as well as $Cottage$ dominates $House$, i.e., $D^{\geq}(Cottage, House)$. This rule is true in SDS^{OG} because:
- $D_{Dwelling}^{+House} = \{u_2, u_5, u_8\}$,
- $Cl_{Medium}^{\geq} = \{u_1, u_2, u_3, u_5, u_7, u_8\}$,
- hence $D_{Dwelling}^{+House} \subseteq Cl_{Medium}^{\geq}$.

2) Let us consider the following rule in SDS^{OG}:

$$(Vehicle, Car)^{\geq} \Rightarrow (Material\ Status, Medium)^{\geq},$$

This rule means that if $Vehicle$ is at least Car, then $Material\ Status$ is at least $Medium$. From $OG_{Vehicle}$, we obtain that SUV dominates Car, i.e., $D^{\geq}(SUV, Car)$, as well as $Minivan$ dominates Car, i.e., $D^{\geq}(Minivan, Car)$. This rule is not true in SDS^{OG} because:
- $D_{Vehicle}^{+Car} = \{u_1, u_2, u_3, u_5, u_6, u_7, u_8\}$,
- $Cl_{Medium}^{\geq} = \{u_1, u_2, u_3, u_5, u_7, u_8\}$,
- but $D_{Vehicle}^{+Car} \supset Cl_{Medium}^{\geq}$.

3) Let us consider the following rule in SDS^{OG}:

$$(Dwelling, Flat)^{\leq} \Rightarrow (Material\ Status, Medium)^{\leq},$$

This rule means that if $Dwelling$ is at most $Flat$, then $Material\ Status$ is at most $Medium$. This rule is true in SDS^{OG} because:
- $D_{Dwelling}^{-Flat} = \{u_3\}$,
- $Cl_{Medium}^{\leq} = \{u_1, u_3, u_4, u_6, u_7, u_8\}$,
- hence $D_{Dwelling}^{-Flat} \subseteq Cl_{Medium}^{\leq}$.

4) Let us consider the following rule in SDS^{OG}:

$$(Dwelling, Lodging)^{\leq} \Rightarrow (Material\ Status, Low)^{\leq},$$

This rule means that if $Dwelling$ is at most $Lodging$, then $Material\ Status$ is at most Low. This rule is not true in SDS^{OG} because:
- $D_{Dwelling}^{-Lodging} = \{u_1, u_3, u_4, u_6\}$,
- $Cl_{Low}^{\leq} = \{u_4, u_6\}$,
- hence $D_{Dwelling}^{-Lodging} \nsubseteq Cl_{Low}^{\leq}$.

Example 5.1 shows that, in simple decision systems over ontological graphs, the validity of a rule depends on the viewpoint from which attribute values (their meanings) are seen. For example, generality of meanings of attribute values changes a view on a decision system, and, among others, a view on the validity of rules in decision systems. This is a very important problem in data mining and data analysis. If we consider literally attribute values, then our solutions are sensitive to different mistakes of the data meaning, uncertainties, vagueness, ambiguities, etc.

From rough set theory, we know that lower approximations generate rules true in decision systems. In our case, each nonempty a_c-lower approximation of $Cl_t^{\geq}$ and a_c-lower approximation of $Cl_t^{\leq}$ generates $D_{\geq}$-elementary decision rules and $D_{\leq}$-elementary decision rules, respectively, true in a given decision system.

Complex decision rules will be considered in the future works. However, let us have a look at another example.

Example 5.2: Let us consider the following rules in a simple decision system over ontological graphs SDS^{OG} given in Example 3.3:

$$(Vehicle, Car) \wedge (Dwelling, House)^{\leq} \Rightarrow (Material\ Status, Medium)^{\leq},$$

and

$$(Vehicle, Minivan) \wedge (Dwelling, House)^{\leq} \Rightarrow (Material\ Status, High)^{\leq},$$

where $\wedge$ denotes the "and" propositional connective. We can check that both rules are true in SDS^{OG} and they are consistent to each other. The second rule does not contradict the first one, but it is more precise, i.e., it proposes a more precise decision. It means that if we specialize a concept in the first descriptor of the rules, i.e., instead of Car we put $Minivan$, then the decision "$Material\ Status$ is at least $High$" is more precise than "$Material\ Status$ is at least $Medium$". This example shows that to extract more precise knowledge from a decision system we can pinpoint data included within it, i.e., specialize the meaning of data.

VI. Generation of Elementary Decision Rules Consistent with Dominance Principle

A generalization relation is the standard subtype / supertype relationship that frequently appears in data modeling (cf. [7]). It is often expressed as A is-a B where A is referred to as the specific entity type and B the generic entity type. The well-known inheritance principle of "is-a" relationships [8] states that anything that is true about the generic entity type, B, must also be true of the specific entity type, A. It is worth noting that an opposing statement is not true, i.e., if something is not true about the generic entity type, B, then it does not mean that it is not true of the specific entity type, A. These properties seem to be important in case of rules generated from information (decision) systems.

Example 6.1: Let us consider again a simple decision system over ontological graphs given in Example 3.3. We can formulate informally the following rule "if $Dwelling$ is $House$, then $Material\ Status$ is at least $Medium$". This rule means that owing a house makes at least a medium material status. According to the inheritance principle another rule can be formulated, namely, "if $Dwelling$ is $Bungalow$, then $Material\ Status$ is at least $Medium$", because $Bungalow$ is a kind of $House$ ($Bungalow$ is generalized by $House$). Therefore, Rule (1) shown in Example 5.1 is valid. If we take into consideration another rule "if $Vehicle$ is Car, then $Material\ Status$ is at least $Medium$", then it is easy to see, on the basis of the decision system shown in Example 3.3, that this rule is not true. We can say, according to this rule and our decision system, that owing a car does not make at least a medium material status. However, the rule "if $Vehicle$ is SUV, then $Material\ Status$ is at least $Medium$" is valid in spite of the fact that SUV is a kind of Car (SUV is generalized by Car).

Example 6.1 shows that an important task is to find the most general rules with respect to their condition parts valid in simple decision systems over ontological graphs.

A generalization /specialization relation is hierarchical, i.e., it makes a partial order [8]. "Is-a" connections form a hierarchy (or, in some cases, a lattice) of the types (concepts) being connected. In our case, we consider only a generalization / specialization relation that is hierarchical. Therefore, for this relation, we obtain a generalization / specialization hierarchy which can be presented in the form of a tree.

Let $OG_a = (\mathcal{C}_a, E_a, \mathcal{R}, \rho_a)$ be an ontological graph associated with the attribute a of some simple decision system over ontological graphs. By $Children_{\rhd}(c)$, where $c \in \mathcal{C}_a$ we denote the set of all concepts $c' \in \mathcal{C}_a$ such that c' specializes the concept c according to OG_a, i.e., $\rho_a((c, c')) = R_{\rhd}$. By $Parent_{\lhd}(c)$, where $c \in \mathcal{C}_a$ we denote the concept $c' \in \mathcal{C}_a$ such that c' generalizes the concept c according to OG_a, i.e., $\rho_a((c, c')) = R_{\lhd}$.

Definition 6.1: Let a simple decision system over ontological graphs $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ be given. An elementary rule $(a_c, r_c)^{\geq} \Rightarrow (a_d, v_d)^{\geq}$, where $a_c \in C$, $r_c \in \mathcal{C}_{a_c}$ of OG_{a_c}, $a_d \in D$, $v_d \in V_d$, is said to be the most general rule with respect to its condition part for a fixed decision part $(a_d, v_d)^{\geq}$ if and only if the rule $(a_c, r_c)^{\geq} \Rightarrow (a_d, v_d)^{\geq}$ is true in SDS^{OG}, but the rule $(a_c, r'_c)^{\geq} \Rightarrow (a_d, v_d)^{\geq}$, where $r'_c = Parent_{\lhd}(r_c)$, is not true in SDS^{OG}.

Example 6.2: Let us consider two elementary rules in SDS^{OG}:

$$(Vehicle, Car)^{\geq} \Rightarrow$$
$$(Material\ Status, Medium)^{\geq}.$$

and

$$(Vehicle, SUV)^{\geq} \Rightarrow$$
$$(Material\ Status, Medium)^{\geq}.$$

For a fixed decision part, i.e.,

$$(Material\ Status, Medium)^{\geq},$$

the second rule is the most general one in SDS^{OG} with respect to its condition part, because this rule is true in SDS^{OG}, the first rule is not true, and $Car = Parent_{\lhd}(SUV)$ according to $OG_{Vehicle}$ (see Figure 1).

For mining the most general elementary rules, in a given simple decision system over ontological graphs, with respect to their condition parts for fixed decision parts, we can use an algorithm based on the depth-first search technique with pre-pruning. The pre-pruning enables us to terminate some of the branches prematurely as a tree structure generated by a generalization / specialization relation is searched. Input data for the algorithm are the following:

- $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ - a simple decision system over ontological graphs,
- $a_c \in C$ - a selected condition attribute,
- $(a_d, v_d)^{\geq}$, where $a_d \in D$, $v_d \in V_d$ - a fixed decision part of the most general elementary rules with respect to their condition parts.

As the output we obtain a set $MGER(a_c, a_d, v_d)$ of all most general elementary rules with respect to their condition parts including the attribute a_c for a fixed decision part $(a_d, v_d)^{\geq}$. In this algorithm, we can use a recursive procedure shown as Procedure 1. The procedure starts with the root concept of the hierarchy determined by the generalization / specialization relation.

Example 6.3: Let us consider a simple decision system over ontological graphs $SDS^{OG} = (U, C, D, \{OG_a\}_{a \in C}, V_d, c, d)$ given in Example 3.3. Suppose we are interested in all

Procedure MGER

Input : r_c - a concept belonging to $\in \mathcal{C}_{a_c}$ of OG_{a_c}.
if *the rule* $(a_c, r_c)^{\geq} \Rightarrow (a_d, v_d)^{\geq}$ *is true in* SDS^{OG} **then**
  $MGER(a_c, a_d, v_d) \longleftarrow$
  $MGER(a_c, a_d, v_d) \cup \{(a_c, r_c)^{\geq} \Rightarrow (a_d, v_d)^{\geq}\}$;
else
  for *each* $r'_c \in Children_{\triangleright}(r_c)$ **do**
    Call the procedure MGER for r'_c;
  end
end

most general elementary rules with respect to their condition parts designating $Material\ Status$ at least $Medium$. Tree structures searched by the proposed algorithm for attributes $Vehicle$ and $Dwelling$ are shown in Figures 3 and 4, respectively. Bordered nodes are termination nodes defining mined rules. A set of all most general elementary rules in SDS^{OG} with respect to their condition parts for the fixed decision part $(Material\ Status, Medium)^{\geq}$ includes the following rules:

- $(Vehicle, SUV)^{\geq} \Rightarrow$ $(Material\ Status, Medium)^{\geq}$,
- $(Vehicle, Minivan)^{\geq} \Rightarrow$ $(Material\ Status, Medium)^{\geq}$,
- $(Dwelling, Condominium)^{\geq} \Rightarrow$ $(Material\ Status, Medium)^{\geq}$,
- $(Dwelling, House)^{\geq} \Rightarrow$ $(Material\ Status, Medium)^{\geq}$.

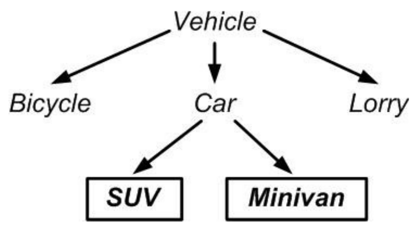

Fig. 3. The searched tree structure for the attribute $Vehicle$

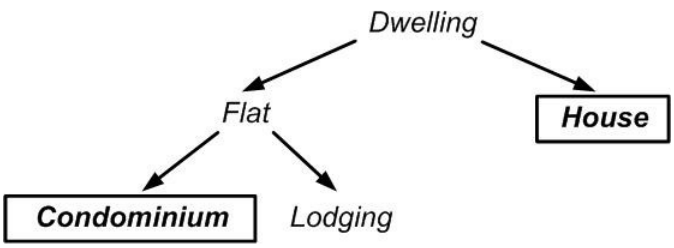

Fig. 4. The searched tree structure for the attribute $Dwelling$

VII. Possible Applications

The presented paper constitutes the first attempt to dealing with decision rules in simple decision systems over ontological graphs. Therefore, it has a rather rudimentary (introductory) character. In this section, we outline possible applications of the proposed approach.

Recent research in the area of data mining shows that, in many situations, data (i.e., attribute values describing objects) alone are not sufficient. There is a need to add some expert knowledge about relationships within data expressing the meaning of data. Such knowledge is included in ontologies. Ontologies play increasingly significant role in information systems and computer science [4], [9]. It is one of the areas of modern computer science that evolves dynamically. Occurrence of a huge amount of data is one of the causes of such popularity. Such data, treated only as values, do not enable us to use information and knowledge hidden in them in an efficient way. This problem was recognized years ago in different areas of computer science, especially, data bases and the Internet. Storage or transmission of a large amount of data without description of their meaning is not enough in modern intelligent computer tools. Therefore, there have been proposed semantic relationships between objects (concepts) stored in data bases or transmitted through computer networks. Semantic Web based on semantic networks [10] and the whole set of methodologies and tools for creating ontologies in computer science are one of the benefits of such workings. Covering data semantics is important, especially, in situations of data coming from different sources (the so-called information fusion [11]). Values semantically equivalent can be recorded in a literally different way, i.e., by means of different symbolical names. Moreover, different databases can store data at different levels of rigorousness (generality or specificity). In many cases of classic data mining algorithms, a variety of ambiguity, vagueness and uncertainty in data causes extracting incorrect information and knowledge. This, in turn, leads to errors in, for example, classification and reasoning processes. The approach presented in this paper, based on supplementing data with information about their semantics in the form of ontological graphs, should prevent (or, at least, curb) data mining from appearing such unwanted effects. The ontological graph can indicate, for example, values being synonyms, values being in a generalization/specialization relation, etc. Therefore, the problems of synonymic names or names at different levels of rigorousness will not influence data mining results and they will enable us to build tools sensitive to data semantics.

Proposed modifications will allow creating data mining tools sensitive to data semantics, which can be used in different areas, e.g., medicine, biology, economy, sociology, etc. In these areas, we often deal with categorical data. Covering data semantics seems to be especially important in medicine, because the data gathered are characterized relatively often by ambiguity, vagueness and uncertainty.

VIII. Conclusions

The approach presented in this paper generalizes a look on decision rules in decision systems if some new knowledge describing semantic meanings of attribute values is available. This new knowledge enables us to look on properties of decision systems and decision rules from different perspectives taking into consideration different relations between attribute values, for example, a level of generality in their semantic meaning. The paper is a new contribution to the research area devoted to embedding ontologies within information systems. We can distinguish, for further work, several directions well known in classic rough set theory. We will consider, among others:

- complex decision rules in simple decision systems over ontological graphs, i.e., rules with descriptors linked with propositional connectives,
- properties of simple decision systems over ontological graphs from the DRSA perspective.

REFERENCES

[1] K. Pancerz, "Toward information systems over ontological graphs," in *Rough Sets and Current Trends in Computing*, J. T. Yao *et al.*, Eds. Berlin Heidelberg: Springer-Verlag, 2012, to appear.

[2] S. Greco, B. Matarazzo, and R. Słowiński, "Rough sets theory for multicriteria decision analysis," *European Journal of Operational Research*, vol. 129, no. 1, pp. 1–47, 2001.

[3] Z. Pawlak, *Rough Sets. Theoretical Aspects of Reasoning about Data*. Dordrecht: Kluwer Academic Publishers, 1991.

[4] R. Neches, R. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator, and W. Swartout, "Enabling technology for knowledge sharing," *AI Magazine*, vol. 12, no. 3, pp. 36–56, 1991.

[5] J. Köhler, S. Philippi, M. Specht, and A. Rüegg, "Ontology based text indexing and querying for the semantic web," *Knowledge-Based Systems*, vol. 19, pp. 744–754, 2006.

[6] B. Roy, *Méthodologie Multicritère d'Aide à la Décision (in French)*. Paris: Economica, 1985.

[7] R. Hull and R. King, "Semantic database modeling: Survey, applications, and research issues," *ACM Computing Surveys*, vol. 19, no. 3.

[8] R. Brachman, "What IS-A is and isn't: An analysis of taxonomic links in semantic networks," *Computer*, vol. 16, no. 10, pp. 30–36, 1983.

[9] A. Gomez-Perez, M. Fernandez-Lopez, and O. Corcho, *Ontological Engineering*. London: Springer-Verlag, 2004.

[10] M. Quillian, "Word concepts: A theory and simulation of some basic semantic capabilities," *Behavioral Science*, vol. 12, pp. 410–430, 1967.

[11] V. Torra, *Information Fusion in Data Mining*. Berlin Heidelberg: Springer-Verlag, 2003.

[12] B. Hjorland, "Semantics and knowledge organization," *Annual Review of Information Science and Technology*, vol. 41, pp. 367–405, 2007.

[13] J. L. Milstead, "Standards for relationships between subject indexing terms," in *Relationships in the organization of knowledge*, C. A. Bean and R. Green, Eds. Dordrecht: Kluwer Academic Publishers, 2001, pp. 53–66.

[14] K. Pancerz and P. Grochowalski, "Matching ontological subgraphs to concepts: a preliminary rough set approach," in *Proceedings of the ISDA 2010*, A. Hassanien *et al.*, Eds., Cairo, Egypt, 2010, pp. 1394–1399.

[15] Z. Pawlak and A. Skowron, "Rudiments of rough sets," *Information Sciences*, vol. 177, pp. 3–27, 2007.