A Well-founded Graph-based Summarization Framework for DLs

Motivations

Graph summarization has received considerable attention in the literature for timely topics like:
- exploration and visualization of large graphs;
- optimization of graph data management systems.

- ABoxes can be seen as multigraphs with typed vertices and typed edges (Fig. 1).

Quotient operation for ABoxes

Definition 1 (Quotient ABox). Let \( \mathcal{A} \) be an ABox, \( \equiv \) be some equivalence relation between constants, and let \( a_1, \ldots, a_n \) denote a slight abuse of notation both the equivalence classes of the constants in \( \mathcal{A} \) w.r.t. \( \equiv \) and the names of these equivalence classes. The quotient ABox of \( \mathcal{A} \) w.r.t. \( \equiv \) is the ABox \( \mathcal{A}_\equiv \) such that:

\[- C(a_i) \in \mathcal{A}_\equiv \text{ iff there exists } a \in a_i \text{ such that } C(a) \in \mathcal{A}, \text{ for } 1 \leq i \leq n, \]

\[- R(a_i,a_j) \in \mathcal{A}_\equiv \text{ iff there exist } a \in a_i \text{ and } a' \in a_j \text{ such that } R(a,a') \in \mathcal{A}, \text{ for } 1 \leq i,j \leq n. \]

ABoxes are not just graphs because they also have a first-order semantics.

- Does there exist a semantic relationship between an ABox and a summary of it?

Characterization of ABox summaries

A summary is more specific than the summarized ABox.

Property 1. Let \( \mathcal{T} \) be a TBox, \( a_1, \ldots, a_m \) be the constants in an ABox \( \mathcal{A} \) and \( a'_1, \ldots, a'_n \) be the constants in some summary \( \mathcal{A}_\sigma \) of \( \mathcal{A} \). If we consider these constants as existential variables, then \( \exists a_1 \ldots \exists a_m \mathcal{A} \equiv \exists a'_1 \ldots \exists a'_n \mathcal{A}_\sigma \) holds.

- To which extent a summary is more specific than the ABox it summarizes?

For an ABox and a TBox, the most precise description of a set of constants is the conjunction of concepts these constants are all instances of, which is called their most specific concept (msc).

Theorem 1. Let \( \mathcal{T} \) be a TBox, \( \mathcal{A} \) be an ABox and \( \mathcal{A}_\sigma \) be the summary of \( \mathcal{A} \) w.r.t. \( \equiv \equiv \) equivalence relation. If \( a_1, \ldots, a_n \) are all the constants in \( \mathcal{A} \) that belong to the equivalence class \( a_\sigma \) according to \( \equiv \), then the following holds:

\[ msc^{\mathcal{A},\mathcal{T}}(a_\sigma) \subseteq \bigcap_{i=1}^{m} msc^{\mathcal{A}_\sigma}(a_i) \subseteq \bigcap_{i=1}^{n} msc^{\mathcal{A},\mathcal{T}}(a_i) \subseteq \bigcup_{i=1}^{n} msc^{\mathcal{A}_\sigma}(a_i, \ldots, a_n). \]

Application to ontology-based data management

Consistency-checking

Property 2. Let \( \mathcal{T} \) be a TBox, \( \mathcal{A} \) be an ABox and \( \mathcal{A}_\sigma \) be some summary of \( \mathcal{A} \). If \( \mathcal{A}_\sigma \) is consistent w.r.t. \( \mathcal{T} \), then \( \mathcal{A} \) is consistent w.r.t. \( \mathcal{T} \).

Query answering

Property 3. Let \( \mathcal{T} \) be a TBox, \( \mathcal{A} \) be an ABox\( \mathcal{A}_\sigma \) be some summary of \( \mathcal{A} \), and let \( q \) be a UCQ. If \( q^{\mathcal{T},\mathcal{A}} \neq \emptyset \) holds then \( q^{\mathcal{T},\mathcal{A}_\sigma} \neq \emptyset \) holds, with \( q_\sigma \), the query \( q \) in which every constant \( u \) is replaced by its image \( h(u) \) through the \( \mathcal{A}-\mathcal{A}_\sigma \) homomorphism \( h \).

Property 2 and the contraposition of Property 3 are of practical interest to check rapidly if, for sure, an ABox is consistent and a query has no answer resp., by using the typically (much) smaller ABox summary.

Perspectives

- Devise DL-specific equivalence relations between constants (i.e., which use both the TBox and the ABox) to explore, visualize or optimize reasoning on and the management of ABoxes.

Bibliography