Journal of Southeast University (English Edition) Vol. 25, No. 4, pp. 482 – 485 Dec. 2009 ISSN 1003—7985

Optimizing reasoning in EL^{++} ontologies by using boundary-based module

Fang Jun Guo Lei Yang Ning

(School of Automation, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: In order to optimize ontology reasoning, a novel boundary-based modular extraction method is introduced for ontologies in EL^{++} description logics. The proposed module extraction method is capable of identifying relevant axioms in an ontology based on the notion of boundaries of symbols, with respect to a given reasoning task. Exactness of the method is ensured by discovering all axioms in the original ontology that may be directly or indirectly relevant to boundaries of symbols used in the reasoning task. Compactness of the method is ensured by boundary partition and intersection operation performed in the process of module extraction. The theoretical foundation and a practical algorithm for computing the proposed axiom-based modules are presented. The proposed algorithm is implemented for the description logic EL^{++}. Experimental results on real-world ontologies show that, based on the proposed modularization method, the performance of ontology reasoning can be significantly improved.

Key words: boundary; module extraction; reasoning optimization; axiom-based module

Modularization is a promising technique to meet the scalability challenge in reasoning with very large ontologies. With the identification of the modular structure of an ontology, a reasoning task may be carried out more efficiently by pinpointing the relevant modules, or by dividing the overall reasoning task into smaller components that can be computed against modules. To achieve high accuracy and efficiency of inference, a modularization-based reasoning approach needs to meet two critical requirements: exactness and compactness. 1) Exactness means that the answer of a reasoning task performed in the modular way and the answer to the same task that is performed in the conventional way on the whole ontology are always identical; 2) Compactness means that a reasoning task should involve the minimal set of axioms that are necessary for performing the reasoning task.

Most relevant work to our study are ontology modularization with structural[1-3] or logic-based approaches[4-5]. However, modules extracted by using structural methods in general may not be both exact and compact in query answering. Logic-based approaches provide ways to extract modules with an exactness guarantee, e. g., sign-ature-based module extraction. In this method, given a set of symbols (i. e., a signature), a fragment can be extracted from an ontology that can answer any reasoning problems for that signature in an exact manner. However, the modules discovered by this approach are in general not compact.

In this work, we present a novel axiom-based approach for module extraction that behaves well for all requirements. Signature-based modularization is targeted at extracting a module that can preserve all knowledge about the signature in question from the entire ontology. This may result in axioms only being needed in rather different reasoning scenarios to be included in the generated module. In our approach, we adopt the axiom-based module extraction approach, so that only a small subset of axioms might be directly related to a reasoning task, which is included in the generated module. Thus, the compactness of our approach is significantly better than that of the signature-based approach.

The main idea of our approach is that an axiom may specify the "boundaries" for the interpretation of an expression of a symbol. Based on the notion of boundary, a set of axioms which are relevant to a given reasoning task can be extracted. Exactness of the method is ensured by discovering all the axioms in the original ontology that may be directly or indirectly relevant to boundaries of symbols used in the reasoning task.

The present paper focuses on the description logic EL^{++} which is a notable subset of OWL-DL. The advantage of EL^{++} is that it combines tractability with expressive power which is sufficient for many important applications of ontologies, especially for life science ontologies. Therefore, EL^{++} serves as the underlying logic of the OWL 2 EL profile.

1 Optimizing Reasoning by Using Boundary-Based Module

1.1 Theory of boundary

Definition 1 (boundary of symbols) Let O be an ontology and s be a symbol in O. The boundary of s is a pair (l_s^O, u_s^O). l_s^O and u_s^O are two functions. Then for a given I and $X \subseteq^{I}$, let I' be a new interpretation obtained from I by letting $s^{I'} = X$ and keep other symbols interpretations unchanged, and then $I' \in \text{model}(O)$ iff $l_s^O(I) \subseteq X \subseteq u_s^O(I)$. We call l_s^O and u_s^O the lower and upper boundaries of s in O, and we drop the superscript O if it is clear from the context.

Based on the above theory, the relationship between boundary and logical entailment is formalized as follows:

Lemma 1 (relationship between boundary and entailment) Let O be an EL^{++} ontology, α be an axiom and s be a symbol in α, and then $O \models \alpha$ iff $\forall I \in \text{model}(O)$, $l_s^{\alpha}(I) \subseteq l_s^O(I) \wedge u_s^O(I) \subseteq u_s^{\alpha}(I)$.

For a symbol s in O, the boundary of s is determined by some axioms in O. It can be formalized as follows:

Received 2009-06-30.

Biographies: Fang Jun (1981—), male, doctor, lecturer, junfang@nwpu.edu.cn; Guo Lei (1956—), male, professor, lguo@nwpu.edu.cn.

Foundation item: The Ph. D. Programs Foundation of Ministry of Education of China (No. 20096102120037).

Citation: Fang Jun, Guo Lei, Yang Ning. Optimizing reasoning in EL^{++} ontologies by using boundary-based module [J]. Journal of Southeast University (English Edition), 2009, 25(4): 482 – 485.

Definition 2 (boundary relevance between axiom and symbol) Let O be an EL^{++} ontology, s be a symbol in O, α be an axiom. I_α is a model of α, M^{I_α} is the set of all models of $O \cup \{\alpha\}$ which set interpretation of every symbol t in $\mathrm{Sig}(\alpha)\backslash s$ to t^{I_α}. If $\exists M^{I_\alpha}$, $l(u)_s^O \neq l(u)_s^{O\cup\{\alpha\}}$, $I \in \mathrm{model}(O)$, $I^{I_\alpha} \in M^{I_\alpha}$, then we say α is relevant to $l_s(u_s)$ through O, denoted by $L(U)B\ (s,\ O,\ \alpha) = 1$; else axiom α is irrelevant to $l_s(u_s)$ through O, denoted by $L(U)\ \mathrm{BR}(s,\ O,\ \alpha) = 0$.

Definition 3 (boundary relevant axiom (set)) Let O be an EL^{++} ontology, s be a symbol in O and α be an axiom.

1) If $L(U)\mathrm{BR}(s,\ \varnothing,\ \alpha) = 1$, then we say axiom α is directly relevant to $l_s(u_s)$, and $\mathrm{DL}(U)B(s,\ O)$ is used to denote all these directly relevant axioms in O; if $L(U)\mathrm{BR}(s,\ \varnothing,\ \alpha) = 0$ and $L(U)\mathrm{BR}(s,\ \varnothing,\ \alpha) = 1$, then we say α is indirectly relevant to $l_s(u_s)$ through O. $\mathrm{IL}(U)B(s,\ O)$ is used to denote all the axioms which are indirectly relevant to the boundary of s through any nonempty subset of O.

2) The axiom set relevant to $l_s(u_s)$ in O contains all axioms α which are directly or indirectly relevant to $l_s(u_s)$, denoted by $L(U)R(s,\ O)$.

1.2 Calculation of boundary relevant axiom set

Direct and indirect boundary relevance are used to calculate the boundary relevant axiom set in this subsection. We first introduce the method for testing the direct boundary relevance between axioms and symbols, and then the calculation of the boundary relevant axiom set can be transformed into testing the direct boundary relevance.

Proposition 1 (testing direct boundary relevance) Let α be an axiom and s be a symbol in α. α is changed to α^* by setting the interpretation of s to the empty set (universe set), i. e., $\alpha^* = \alpha_{s^I \to \varnothing(\Delta^I)}$. If $\alpha \models \alpha^*$, then α is not directly relevant to $l_s(u_s)$; else, α is directly relevant to $l_s(u_s)$.

Proof We prove the correctness of testing the direct lower boundary relevance, and then the upper situation can be proved using the similar way. If $\alpha \models \alpha^*$, then $\forall I \in \mathrm{model}(\alpha)$; when all the interpretations of other symbols in $\mathrm{Sig}(\alpha)\backslash s$ remain unchanged, s^I can be the empty set. Thus l_s is not changed. If $\alpha \not\models \alpha^*$, there exists at least one model I^* of $\mathrm{model}(\alpha)$; when the interpretation of other symbols in $\mathrm{Sig}(\alpha)\backslash s$ is given a specific value, s^{I^*} cannot be the empty set. Hence α is directly relevant to l_s.

The calculation of the axiom set relevant to the boundary of s can be divided into computation of the direct and the indirect boundary relevant axiom set. First the direct boundary relevant axiom set is computed using proposition 1; secondly, for every axiom α in the direct boundary relevant axiom set, we calculate axioms which are indirectly relevant to the boundary of s by influencing the interpretations of symbols in α.

Proposition 2 (indirect relevance testing $\Rightarrow$ direct relevance testing) Let α be an axiom, s and t be two symbols in α, and α is directly relevant to the boundary of s. If α is not directly relevant to $l_t(u_t)$, then axioms which are not relevant to $l_t(u_t)$ are not indirectly relevant to the boundary of s by influencing the interpretation of t in α.

Proof We prove the correctness of testing the indirect lower boundary relevance, and the upper situation can be proved similarly. The axiom α is not directly relevant to l_t. So for any model I of α, when t^I is changed to the empty set, the interpretations of other symbols can remain unchanged, and s^I remains the same. If an axiom β is not relevant to l_t, according to definition 2, let O be $\{\alpha\}$, and then $\forall M^{I_\beta}$, t^{I_β} can be an empty set. When $t^{I_\beta} = \varnothing$, s^{I_β} can be s^I. Thus, $\forall M^{I_\beta}, l(u)_s^O(I) = l(u)_s^{O\cup\{\beta\}}(I^{I_\beta})$, $I \in M$, $I^{I_\beta} \in M^{I_\beta}$, and the axiom is not indirectly relevant to the boundary of s by influencing the interpretation of t in α.

1.3 Boundary-based module

In this section, we will show that modules generated by using boundary partition and intersection operation are exact axiom-based modules. It is proved by analyzing the relationship between the boundary relevant axiom set and the axiom-based module.

Proposition 3 (boundary partition: boundary-based module) Let $O \models \alpha$ be a reasoning task, in which O is an ontology and α is an axiom, and s is a symbol in α. 1) If $L(U)$ $\mathrm{BR}(s,\ \varnothing,\ \alpha) = 1$ and $U(L)\,\mathrm{BR}(s,\ \varnothing,\ \alpha) = 0$, then the axiom set relevant to $l_s(u_s)$ in O is the superset of essential the α-module in O, $O_{\mathrm{e}} \subseteq L(U)R(s,\ O)$; 2) If $\mathrm{LBR}(s,\ \varnothing,\ \alpha) = 1$ and $\mathrm{UBR}(s,\ \varnothing,\ \alpha) = 1$, then the axiom set relevant to l_s or u_s in O is the superset of an essential α-module in O, $O_{\mathrm{e}} \subseteq \mathrm{LR}(s,\ O) \cup \mathrm{UR}(s,\ O)$.

Proof $O \models \alpha$ iff $\forall I \in \mathrm{model}(O)$, $l_s^\alpha(I) \subseteq l_s^O(I) \wedge u_s^O(I) \subseteq u_s^\alpha(I)$. When α is only relevant to the lower boundary of s, $u_s^O(I) \subseteq u_s^\alpha(I)$ is always satisfied as $u_s^\alpha(I) = \Delta^I$ or $\Delta^I \times \Delta^I$. So axioms relevant to the lower boundary of s in O provide all the necessary information to test if $l_s^O(I) \subseteq l_s^\alpha(I) \wedge u_s^\alpha(I) \subseteq u_s^O(I)$. Hence, $O_{\mathrm{e}} \subseteq \mathrm{LR}(s,\ O)$. Other situations can be proved analogously.

Based on the above proposition, we compute an exact α-module in O by intersecting all boundary-based modules of symbols in α, which is formalized as follows:

Lemma 2 (intersection operation: intersection boundary-based module) Let $O \models \alpha$ be a reasoning task, in which O is an ontology and α is an axiom, and then the intersection of all the boundary-based modules is an α-module in O. We name it as the intersection boundary-based α-module in O, denoted by $\mathrm{IBMod}(\alpha,\ O)$.

Lemma 2 shows that a reasoning task contains more symbols, and IBMod will not be larger. On the contrary, the volume of the signature-based module becomes bigger when the reasoning task contains more symbols. This means the intersection boundary-based module is suitable for optimizing reasoning, especially for complex tasks which contain many symbols.

1.4 Module-based reasoning

Given a reasoning task $O \models \alpha$, we can first calculate the intersection boundary-based module for $O \models \alpha$, and then execute the reasoning task in $\mathrm{IBMod}(\alpha,\ O)$ instead of the conventional ontology O. Note that since the boundary relevant relationships between symbols and axioms in O is determined, we can compute all boundary relevant axiom sets for symbols offline, and then the intersection boundary-based module task can be easily and quickly obtained.

In Ref. [6], authors show that signature-based module extraction does not help to speed up standard reasoning in

EL^+ since the reasoning algorithm for deciding subsumption in EL^+ is deterministic. The conclusion is also true for EL^{++}. However, we argue that the performance of standard reasoning can be improved by using axiom-based module extraction. For the simplest example, in module-based reasoning, if the generated IBMod(α, O) is empty, then we can directly obtain the result $O \not\models \alpha$ without executing any reasoning process.

2 Implementation and Evaluation

We use Pellet[7] and OWLAPI[8] to implement the boundary-based module extraction method in DL EL^{++}. Experiments are carried out to show that the boundary-based module is suitable for optimizing ontology reasoning in DL EL^{++}. We select several real-world EL^{++} ontologies that have been successfully applied in many applications, and the test suite comprises SUMO, NCI and GO.

The boundary-based method is evaluated on $C \sqsubseteq D$ axioms and $C \sqsubseteq \exists r.D$ axioms. All ontology classification axioms in SUMO are collected for evaluation. As NCI and GO contain too many $C \sqsubseteq D$ and $C \sqsubseteq \exists r.D$ axioms, we randomly generate 5 000 entailed axioms ($O \models \alpha \wedge \alpha \notin O$) and 5×10^5 not entailed axioms($O \not\models \alpha$) for $C \sqsubseteq D$ and $C \sqsubseteq \exists r.D$ separately.

All the following experiments have been carried out on a standard PC: 2.60 GHz Pentium-4 processor and 2 GB of physical memory. The Java memory setting is default when evaluating SUMO; it is set to 1.2 and 1.4 GB when evaluating NCI and GO ontology. We use the function System.nanoTime() to calculate the execution time.

Tabs. 1 and 2 describe the execution time comparison information between original reasoning and module-based reasoning for $C \sqsubseteq D$ and $C \sqsubseteq \exists r.D$ axioms. Original reasoning is the reasoning on the conventional ontology O, while module-based reasoning is the reasoning on the IBMod(α, O). The whole original reasoning execution time is one ontology loading time plus every reasoning time, and the whole module-based reasoning execution time is every module loading time plus every reasoning time. The average execution time for entailed axioms is computed by dividing the whole reasoning execution time for all the generated entailed axioms in the compactness evaluation by the total number of entailed axioms, and the average execution time for not entailed axioms is calculated analogously. They are illustrated in the second and third column.

Tab. 1 Ontology reasoning performance evaluation on $C \sqsubseteq D$

Ontology	Average original (module-based) reasoning time for $O \models \alpha$/ms	Average original (module-based) reasoning time for $O \not\models \alpha$/ms
SUMO	95.24(65.20)	71.45(0)
NCI	1 198.82(106.19)	142.87(0.056)
GO	8 339.53(161.99)	186.21(0.034)

Tab. 2 Ontology reasoning performance evaluation on $C \sqsubseteq \exists r.D$

Ontology	Average original (module-based) reasoning time for $O \models \alpha$/ms	Average original (module-based) reasoning time for $O \not\models \alpha$/ms
NCI	1390.75(138.72)	226.48(0.039)
GO	4615.74(211.64)	266.60(0.037)

Results show that the intersection boundary-based module can speed up ontology reasoning, especially for $O \not\models \alpha$. When $O \models \alpha$, in SUMO, the average original reasoning time and the module-based reasoning time are about the same, and the reason is that the syntax of SUMO is very simple; in NCI and GO, the module-based execution reasoning time can decrease by almost a factor of 10. When $O \not\models \alpha$, the performance of reasoning is significantly improved. In SUMO, the reasoning time is zero as all IBMod for not entailed axioms are empty. In NCI and GO, the module-based reasoning time can decrease by more than a factor of 10^3, and the reason is that IBMod which is empty accounts for a very big proportion.

3 Conclusion and Future Work

In this paper, we present an exact and compact axiom-based module extraction method by analyzing the relationship between axioms and boundaries of interpretations of symbols. We prove that the boundary-based module is exact for reasoning tasks. Experimental results show the performance of reasoning can be significantly improved by using the boundary-based module. In the future, we will develop optimizing techniques to obtain more compact modules. Furthermore, we plan to extend our module to test more complex ontologies such as GALEN and SWEET.

References

[1] Stuckenschmidt H, Klein M. Structure-based partitioning of large concept hierarchies[C]//*The 3rd International Semantic Web Conference*. Hiroshima, Japan, 2004, **3298**: 289 – 303.

[2] Seidenberg J, Rector A. Web ontology segmentation: analysis, classification and use[C]//*The 15th International Conference on World Wide Web*. Edinburgh, Scotland, UK, 2006: 13 – 22.

[3] Noy N F, Musen M A. Specifying ontology views by traversal[C]//*The 3rd International Semantic Web Conference*. Hiroshima, Japan, 2004, **3298**: 713 – 725.

[4] Grau B C, Horrocks I, Kazakov Y, et al. Modular reuse of ontologies: theory and practice[J]. *Journal of Artificial Intelligence Research*, 2008, **31**: 273 – 318.

[5] Grau B C, Horrocks I, Kazakov Y, et al. Just the right amount: extracting modules from ontologies[C]//*The 16th International World Wide Web Conference*. Banff, Alberta, Canada, 2007: 717 – 727.

[6] Suntisrivaraporn B. Module extraction and incremental classification: a pragmatic approach for EL^+ ontologies[C]//*The 5th European Semantic Web Conference*. Tenerife, Spain, 2008, **5021**: 230 – 244.

[7] Sirin E, Parsia B, Grau B C, et al. A practical OWL-DL reasoner [J]. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2007, **5**(2): 51 – 53.

[8] Bechhofer S, Volz R, Lord P. Cooking the semantic web with the OWL API [C]//*The 2nd International Semantic Web Conference*. Sanibel Island, Florida, USA, 2003, **2870**: 659 – 675.

使用边界模块优化 EL^{++} 本体推理

方 俊 郭 雷 杨 宁

（西北工业大学自动化学院，西安 710072）

摘要： 为了优化本体推理，提出了一种在 EL^{++} 本体中的公理模块提取方法. 该方法通过分析符号解释域的边界和公理间的关系计算出给定推理任务的边界模块，对于与推理任务中所有符号相关的公理集合，采用边界分割和交集计算来提取最终的边界模块，并提高模块的紧凑性. 模块的正确性通过计算和推理任务中符号直接和间接相关的公理得到保证. 给出了边界模块的理论基础和求解算法，并通过符号边界和逻辑蕴涵之间的关系证明了模块的正确性. 边界模块提取方法在 EL^{++} 描述逻辑语言上进行了实现,现实的本体实验评估表明，基于边界模块的推理比不采用模块的推理在性能上有很大的提高.

关键词： 边界；模块提取；推理优化；公理模块

中图分类号： TP311