

온톨로지 커널 : 구조 정보를 반영한 온톨로지 정렬*

김성택^{+o} 손정우⁺ 박성배⁺⁺ 박세영⁺⁺

경북대학교 전자전기컴퓨터학부⁺ 경북대학교 컴퓨터공학과⁺⁺

stkim@sejong.knu.ac.kr jwson@sejong.knu.ac.kr sbpark@sejong.knu.ac.kr seyoung@knu.ac.kr

Ontology Kernel : An Ontology Alignment

with Reflecting Structural Information on Ontology

Seongtaek Kim^{+o} Jeong-Woo Son⁺ Seong-Bae Park⁺⁺ Seyoung Park⁺⁺

School of Electrical Engineering and Computer Science, Kyungpook National University⁺

School of Computer Science and Engineering, Kyungpook National University⁺⁺

1. 서론

온톨로지 정렬은 두 온톨로지들 간에 동일한 의미를 가지는 객체 (개념, 속성, 인스턴스) 를 찾는 방법이다. 온톨로지의 구조는 그래프 형태로 표현된다. 그러나 기존 연구에서는 온톨로지의 그래프 구조를 변형하여 이용하기 때문에 온톨로지만의 특징적인 구조를 정확하게 반영하지 못하는 한계가 있다. 본 연구에서는 온톨로지의 구조에 특화된 온톨로지 커널을 활용한 온톨로지 정렬 방법을 제안한다. 온톨로지 커널은 컨볼루션 커널 (convolution kernel) [1] 의 한 종류인 그래프 커널 [2] 을 기반으로 하였으며 비교하고자 하는 두 온톨로지 객체 주변의 온톨로지 그래프 구조를 비교한다.

2. 그래프 커널

컨볼루션 커널의 한 종류인 그래프 커널 [2] 은 그래프의 substructure를 이용하여 그래프 간의 유사도를 구할 수 있도록 고안된 커널이다. 그래프 커널은 substructure를 어떻게 정하는가에 따라 다양하게 나뉠 수 있다. 본 논문에서는 랜덤 워크를 substructure로 사용한 그래프 커널을 사용한다. 이때, 두 그래프의 유사도가 높다는 것은 두 그래프가 공동으로 가지는 랜덤 워크 수가 많다는 것을 말한다. 두 그래프를 G_1, G_2 라고 할 때, 그래프 커널 K_{graph} 는

$$K_{graph}(G_1, G_2) = \sum_{walk_1 \in G_1} \sum_{walk_2 \in G_2} K_{walk}(walk_1, walk_2)$$

으로 나타낸다. 여기서 K_{walk} 는 G_1 의 모든 워크와 G_2 의 모든 워크의 유사도를 구하기 위한 커널로써,

$$K_{walk}(walk_1, walk_2) = \prod_{i=1}^{n-1} K_{step}((v_i, v_{i+1}), (w_i, w_{i+1})) \quad (1)$$

로 나타낸다. 여기서 v 는 G_1 의 노드, w 는 G_2 의 노드를 나타내고, K_{step} 은 워크를 이루는 노드와 엣지의 각각의 유사도를 구하는 커널로써 일반적으로 노드나 엣지에 나타난 레이블을 비교한다. 본 논문에서는 K_{step} 을 이루는 노드와 엣지 커널을 온톨로지 정렬에 맞게 재정의하여 사용한다.

3. 온톨로지 커널

그래프 커널은 두 그래프간의 구조 정보를 반영한 유사도를 계산할 수 있는 방법이지만, 온톨로지 그래프의 특성을 반영하지 못하고 있다. 온톨로지는 개념, 인스턴스, 데이터 값, 데이터 타입을 노드로, 속성을 엣지로 사용하는 그래프 구조이다. 하지만 그래프 커널은 다양한 종류의 노드와 엣지에 대응할 수 없다. 본 논문에서 제안하는 온톨로지 커널은 온톨로지 그래프의 이러한 특성을 반영한다.

온톨로지 커널도 그래프 커널처럼 그래프의 substructure를 이용하여 그래프 간의 유사도를 구하는 것으로, 두 그래프에서 공동으로 가지는 워크의 수를 계산하여 유사도를 구한다. 하지만 온톨로지 커널에

* 본 논문은 지식경제부 산업원천기술개발사업(10035348, 모바일 플랫폼 기반 계획 및 학습 인지 모델 프레임워크 기술 개발)의 지원으로 수행되었습니다.

서는 수식 (1)의 K_{step} 을 재정의함으로써 온톨로지 그래프에 나타나는 다양한 종류의 노드와 엣지를 고려한다. 제안한 온톨로지 커널의 K_{step} 은 아래와 같다.

$$K'_{step}((v_i, v_{i+1}), (w_i, w_{i+1})) = \sum K'_{node}(v_i, w_i) \cdot K'_{node}(v_{i+1}, w_{i+1}) \cdot K'_{edge}((v_i, v_{i+1}), (w_i, w_{i+1}))$$

여기서 $K'_{node}(v_i, w_i)$ 와 $K'_{edge}(v_i, w_i)$ 는 아래와 같이 정의된다.

$$K'_{node}(v_i, w_i) = C(v_i, w_i) \cdot sim(v_i, w_i),$$

$$K'_{edge}((v_i, v_{i+1}), (w_i, w_{i+1})) = C((v_i, v_{i+1}), (w_i, w_{i+1})) \cdot sim((v_i, v_{i+1}), (w_i, w_{i+1})),$$

이때, 함수 $C()$ 는 입력으로 들어온 노드 혹은 엣지가 같은 종류일 경우 1을 다른 종류일 경우 0을 반환하며 함수 $sim()$ 은 주어진 두 노드 혹은 엣지의 레이블 유사도를 계산하는 함수로 edit distance나 사전어를 이용한 다양한 문자열 유사도 함수를 사용할 수 있다.

표 1. 그래프 커널과 온톨로지 커널, 그리고 OAEI 2009에 참가한 시스템의 실험 결과

그래프 커널		온톨로지 커널		aflood		agrMaker		AROMA	
정확률	재현율	정확률	재현율	정확률	재현율	정확률	재현율	정확률	재현율
0.574	0.52	0.942	0.528	0.90	0.81	0.92	0.79	0.85	0.78

4. 실험

그래프 커널과 온톨로지 커널의 성능을 알아보기 위해 OAEI 2009에서 제공하는 benchmarks 데이터를 사용하였다. 실험은 101 온톨로지와 실제 온톨로지 데이터인 301, 302, 303, 304 온톨로지를 이용하여 실험하였다. 표 1은 그래프 커널과 온톨로지 커널, 그리고 OAEI 2009에 참가한 시스템 중 상위 3개 시스템 [3, 4, 5]의 실험 결과를 나타낸다. 온톨로지 커널은 그래프 커널보다 약 64% 나은 정확율을 보였으며 재현율의 경우 2%의 향상을 볼 수 있다. 기존의 온톨로지 정렬방법들과의 비교에서는 제안한 온톨로지 커널이 가장 높은 정확율을 보였다. 제안한 온톨로지 커널은 OAEI에 참가한 시스템과 달리 사전과 같은 외부 정보나 경험적인 규칙들을 사용하지 않는다. 이러한 점을 고려할 때, 온톨로지 커널은 온톨로지 정렬시 새로운 유사도 측정 도구로써 의미가 있다고 볼 수 있다.

5. 결론

온톨로지 정렬은 두 온톨로지들 간에 동일한 의미를 가지는 객체를 찾는 방법이다. 온톨로지 그래프는 개념과 개념간의 관계와 개념과 데이터 간의 관계로 정의한 스키마 그래프와 개념에 따르는 인스턴스 및 이들 간의 관계를 정의한 인스턴스 그래프로 나뉜다. 그래프 커널은 두 그래프간의 구조 정보를 반영한 유사도를 계산할 수 있는 방법이지만, 온톨로지 그래프의 특성을 반영하지 못하고 있다. 본 논문에서는 온톨로지만의 특징을 반영한 온톨로지 커널에 기반한 온톨로지 정렬 방법을 제안한다. 온톨로지 커널은 온톨로지의 그래프 구조를 그대로 활용하여 그래프 간 유사도를 계산하기 때문에 구조정보의 손실을 최소화할 수 있을 뿐만 아니라, 온톨로지 그래프의 특성을 활용하여 성능을 높였다.

참고문헌

[1] M. Collins, and N. Duffy. Convolution Kernels for Natural Language. In *Proc. of NIPS 2001*, pages 625-632, 2001.

[2] T. Gartner, P. Flach, and S. Wrobel. On Graph Kernel: Hardness Result and Efficient Alternatives. In *Proc. of COLT 2003 and 7th Kernel Workshop*, pages 129-143, 2003.

[3] M. H. Seddiqui, and M. Aono. Anchor-Flood: Results for OAEI 2009. In *Proc. of the 8th International Conference of Semantic Web (ISWC'2009), Oct 2009*, pages 127-134, 2009.

[4] I. F. Cruz, F. P. Antonelli, C. Stroe, U. C. Keles, and A. Maduko. Using AgreementMaker to Align Ontologies for OAEI 2009: Overview, Results, and Outlook. In *Proc. of the 8th International Conference of Semantic Web (ISWC'2009), Oct 2009*, pages 135-146, 2009.

[5] J. David. AROMA results for OAEI 2009. In *Proceedings of OM-2009. In Proc. of the 8th International Conference of Semantic Web (ISWC'2009), Oct 2009*, pages 147-151, 2009.