

Factotum SemNet을 활용한 개념간 개연사슬 발견

*양재근^o **강인수 *배재학 **이종혁
*울산대학교 컴퓨터·정보통신공학부
**포항공과대학교 전자컴퓨터공학부 컴퓨터공학과
**첨단정보기술 연구센터

{jgyang^o, jhjbae}@ulsan.ac.kr {dbaisk, jhlee}@postech.ac.kr

Abductive Chain Detection between Concepts with Factotum SemNet

*Jae-Gun Yang^o **In-Su Kang *Jae-Hak J. Bae **Jong-Hyeok Lee

*School of Computer Engineering & Information Technology, University of Ulsan

**Dept. of Computer Science and Engineering, Division of Electrical and Computer Engineering,
POSTECH

and Advanced Information Technology Research Center (AITrc)

요 약

독해 과정에서 문장 의미간 개연성 파악은 저자의 저술의도 파악에 중요한 요소로 작용한다. 개연성 파악을 위해서는 우선 어휘들과 그들의 연관관계로 구성된 개연사슬을 발견할 필요가 있다. 본 논문에서는 Factotum SemNet을 활용하여 개연사슬의 발견을 자동화하였고 이에 대한 도구를 개발하였다. 그 후, 이 도구를 이용한 실험을 통해서 개연사슬 발견의 자동화가 가능함을 알 수 있었다.

1. 서 론

사람은 정보를 전달하기 위하여 글을 쓴다. 글은 한 문장으로 끝나기도 하지만, 보통은 여러 문장으로 이루어진다. 연속된 문장에는 서로 의미적으로 연관된 단어들 이 나타난다. 독자는 이 "연관된 단어들"이 함축하고 있는 의미적 개연성을 파악해서 전체적인 의미를 가능하게 된다. 이와 같은 개연성을 파악하여 글로써 전달하고자 하는 정보를 체계적으로 통합할 수 있다. 글 내용에 대한 이해는 통합된 개연성을 바탕으로 이루어진다.

사람은 광범위한 지식체계를 가지고 있어서 글이나 대화에서 올바른 개연성 파악이 가능하다. 또한, 관심이 없는 요소는 무시하고 서로 관계가 있는 사실들은 걸러낼 수 있는 능력 역시 개연성 파악에 도움이 된다. 이와 같이 문맥이나 독자의 배경, 지적수준 등은 개연성 인지에 영향을 주는 요소들이다[1, 2].

Jim was *hungry*.
He opened the *refrigerator*.

그림 1 개연성 이해를 위한 예문

[그림 1]의 문장에서 hungry와 refrigerator는 직접적인 인과관계에 있어 보이지는 않는다. 하지만 독자는 hungry와 refrigerator 사이의 연관성을 경험적인 상식을

토대로 이 둘 사이의 개연성을 파악한다. 그 후 "Jim이 배가 고파서 음식을 꺼내먹기 위해 냉장고를 열었다"라고 이야기를 이해하게 된다.

개연적으로 관계가 있는 단어들을 서로 연결한 것이 개연사슬이다. 대화나 글의 개연사슬을 올바르게 구성하면 내용의 이해가 보다 쉬워진다. 사람은 상식이 있어서 이를 적당한 수준에서 구성할 수 있다. 개연사슬 발견을 자동화시킬 수 있다면 기계에 의한 글이나 대화의 거시적 이해가 훨씬 용이해질 것이다. 이에 본 논문에서는 Factotum SemNet[3]을 활용해서 개념간 개연사슬의 발견을 자동화하였다.

2. 본 론

2.1 Factotum SemNet

Factotum SemNet은 MICRA사에서 Roget 시소러스 [4]와 Webster 사전을 이용해서 구축한 의미망(Semantic Network)이다. 이 의미망은 [그림 2]처럼 Roget 시소러스의 분류체계를 보완, 재구성하였다. 또한 is-a등의 의미관계표식(Semantic Relational Marker)을 이용하여 어휘의 의미계승을 가능하게 하였다[5]. 기능성 관계(Functional Relation)는 "{" 기호 내부에 표기하고 해당 범주의 동의어 목록은 같은 행에 기술한다.

```

A6.1.4 CONVERSION (R144)
#144. Conversion.
N. ((has_subtype(change, R140))) conversion, transformation.
((has_case: @R7, initial state, final state)).
((has_patient: @R3a, object, entity)).
((result_of) ((has_subtype(process, A7.7))) converting, transforming.
((has_subtype)) processing.
transition.
=>
<change, has_subtype, conversion> <change, has_subtype, transformation>
<conversion, has_case, initial state> <conversion, has_case, final state>
<conversion, has_patient, object> <conversion, has_patient, entity>
<conversion, is-result-of, converting> <conversion, is-result-of, transforming>
<process, has_subtype, converting> <process, has_subtype, transforming>
<conversion, has_subtype, processing>
    
```

그림 2 Factotum SemNet의 구조(일부)

Factotum SemNet에는 400개 이상의 기능성 관계가 있다. 이 관계의 대부분은 비대칭적이어서 한쪽 방향으로만 작용한다. [표 1]에 기능성 관계의 일부를 예시하였다.

표 1 Factotum SemNet의 기능성 관계(일부)

Relation	의미
has-subtype	is-a와 반대의 관계
is-property-of	종속된 특성
is-caused-by	대상의 원인을 지시
has-property	대상의 두드러진 특성
has-part	물리적인 요소
has-high-intensity	특성에 대한 강화자
has-high-level	행위에 대한 함의
is-antonym-of	반대의 의미
is-conceptual-part-of	개념적인 요소
causes _{mental}	동기

2.2 개연사슬

개연사슬(Abductive Chain)이란 개연적인 관계에 있는 단어들의 연결이다. [그림 1]에서 보았던 문장과 설명한 내용은 다음 [그림 3]에서와 같이 세 단계로 나누어 생각 할 수 있다.

- "고프다"를 해결하기 위해서 "먹는다".
- "먹는다"의 대상은 "음식"이 될 수 있다.
- "음식"은 "냉장고"에 "보관"할 수 있다.

그림 3 세분한 예문

[그림 3]의 세 문장에서 주요 어휘를 간추려 펼치면 아래와 같다.

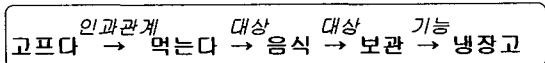


그림 4 기능성 관계를 포함한 개연사슬

이렇게 전개된 어휘사슬이 hungry와 refrigerator 사이의 개연사슬이다. 또한 개연사슬 내의 각 마디를 연결하는 개연관계를 표기하면 [그림 4]와 같다.

즉, 개연사슬은 대상이 되는 두 어휘를 개연적인 어휘

들로 연결시킨 어휘사슬이다. 이 때에 각 마디는 직접적인 인과관계나 개연적인 관계에 있는 단어로 구성된다.

2.3 개연사슬 발견 자동화

본 논문에서는 Factotum SemNet MRD[6]에서 기능성 관계를 탐색하는 방법으로 개연사슬을 발견하였다. 우선, 검색 어휘가 뿌리(Root)에 위치하는 트리구조를 생성한다. 이 트리를 너비우선 탐색하여 나머지 어휘를 찾는다. 개연사슬은 여러 개가 발견될 수 있으므로 트리의 나머지 부분도 탐색한다. 이상의 과정을 [그림 5]에 나타내었다.

```

F_T // Factotum SemNet의 tree구조 표현
node0 // F_T의 root node
node_i // node_0의 depth가 i인 하위 node
word_subject // 편의상, 검색대상인 두 어휘를
word_object // subject, object word로 나눈다.
word_current // 현재 검색된 어휘
depth // F_T에서 해당 node의 깊이
path // 어휘사슬 전체
path_current // 현재 어휘사슬
relation_selection // 탐색에 적용할 기능성 관계 목록
relation_current // 현재 node의 기능성 관계

word_subject로 node_0 생성
repeat (node){
    depth 증가
    CreationSubNode(node_i)
    if word_current = word_object 이면(
        단위결과 출력
    )
}

// 각 node의 Sub node를 생성하는 sub-routine
CreationSubNode(node_i){
    if word_current 검색이 불가능 하다면(
        종료
    )
    if path_current이 path에 없다면(
        if relation_current ∈ relation_selection 이면(
            node_i 생성
            path_current을 path에 추가
        )
    )
}
    
```

그림 5 개연사슬 발견 알고리즘(일부)

본 논문에서는 개연사슬 발견 자동화에 대한 방법을 두 가지 측면에서 살펴보았다: (1) 온톨로지에는 대상 어휘와 함께 인과관계나 개연적인 관계를 포함시켜서 어휘 간 개연성 파악이 쉽도록 한다. 이런 관점에서 Factotum SemNet은 개연사슬 발견에 유용한 자원이다. (2) Factotum SemNet은 약 5천7백여 어휘와 4백여 기능성 관계로 구성된 개념망이다. 이들이 조합된 형태로

서 표현된 개념은 약 16만여 가지이다. 여기에서 탐색된 개연사슬은 사람의 상식에 의거하여 선별해야 할 필요가 있다. 이에 대한 방안으로 본 논문에서는 특정 기능성 관계를 취사선택하는 방법을 이용하였다.

2.4 개연사슬 자동 발견 예

다음은 [그림 1]에 예시한 문장에서 나타나는 hungry와 refrigerator 사이의 개연사슬을 자동 발견한 결과이다.

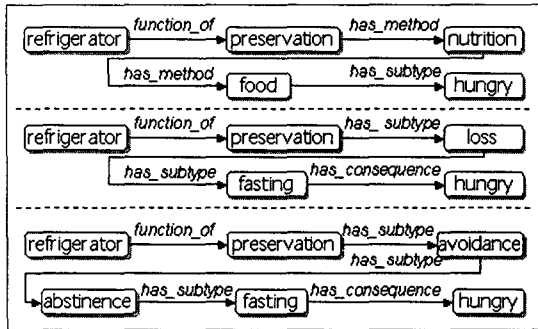


그림 6 hungry와 refrigerator 사이의 개연사슬

[그림 6]의 결과는 인과관계를 나타내는 몇 개의 기능성 관계에 국한시켜 탐색한 결과이다. 그 중 첫 번째 개연사슬인 "refrigerator - preservation - nutrition - food - hungry"는 사람의 상식에 부합하는 경로를 거쳐 대상어휘를 연결시키고 있다. 이처럼 Factotum SemNet을 활용한 개연사슬 발견 자동화가 가능함을 본 실험을 통해서 알 수 있었다.

표 2 자동 발견한 개연사슬의 예

대상어휘	개연사슬
friend offer	friend - friendship - benevolence - gift - offer
puttering settled	puttering - inactivity - indifference - nonobservance - law - established - settled
friend want	friend - friendship - affection - desire - want
gift friend	gift - action - acquaintance - friendship - friend
persuade gift	persuade - demonstration - rally - good-natured - benevolence - gift
ask persuade	ask - tell - communication - persuasion - persuade
kill ruin	kill - killing - destruction - ruin
afraid reluctant	afraid - fear - dissuasion - reluctance - reluctant

앞의 [표 2]는 선행연구[7]를 통해 미리 검증해 보았던 개연사슬을 본 논문에서 구현한 시스템을 통해 재발견한

결과이다.

3. 결 론

독해 과정에서 문장간 개연성 인지는 저자의 저술의도 파악에 중요한 요소이다. 개연성 유추를 위해 본 논문에서는 두 개념을 의미적으로 연결하는 개연사슬 발견을 자동화하였다. 여기에 온톨러지 자원으로 전처리한 Factotum SemNet MRD를 이용하였다. 자동화 과정에서는 이 개념망에서 규정한 기능성 관계(Functional Relation)와 동의어군(Synonym Set)을 탐색하였다. 탐색 경로 구성시 고려 대상 관계들을 취사선택할 수 있게 하여 탐색 결과를 조절 할 수 있게 하였다. 구현한 시스템을 사용하여 사람의 상식에 부합하는 개연사슬들을 얻을 수 있었다. 이를 통해 대화나 글의 의미파악을 위한 개연사슬 발견 자동화가 가능함을 알 수 있었다. 향후, 기능성 관계와 동의어군의 적절한 선택으로 본 논문에서 적용한 방법을 개선하면 현재보다 더 적절한 개연사슬 발견이 가능할 것으로 기대한다.

참고문헌

- [1]Sanda M. Harabagiu. "WordNet-Based Inference of Textual Cohesion and Coherence." FLAIRS-98, May 1998, Sanibel Island, FL.
- [2]Sanda M. Harabagiu. and Dan Moldovan. "A Parallel Algorithm for Text Inference." International Parallel Processing Symposium, IPPS-96, April 1996, Honolulu, Hawaii, pages 441-445.
- [3]MICRA Inc., Factotum SemNet <http://www.micra.com/>.
- [4]Roget's Thesaurus. <http://promo.net/cgi-promo/pg/t9.cgi?entry=22&full=yes&ftpsite=ftp://ibiblio.org/pub/docs/books/gutenberg/>.
- [5]O'Hara, Tom, Janyce Wiebe (forthcoming), "Classifying functional relations in Factotum via WordNet hypernym associations", Proc. Fourth International Conference on Intelligent Text Processing and Computational Linguistics (CiCLing-2003).
- [6]양재균, 강인수, 배재학, 이종혁, "Factotum SemNet의 MRD 변환에 관한 연구", 한국인지과학회 춘계학술대회 논문집, pp. 279-284, 2003.
- [7]양재균, 배재학, 유해영, 이종혁, Factotum SemNet을 활용한 개연규칙 검증. 한국정보처리학회 제19회 춘계학술발표대회 논문집, 제10권, 제1호, pp. 349-352, 2003.