

개연규칙의 활용: Plot Units의 발견

김곤^o 양재곤 배재학
 울산대학교 컴퓨터·정보통신공학부
 {gonkim^o, jgyang, jhjbae}@ulsan.ac.kr

Application of Abductive Rules: Detection of Plot Units

Gon Kim^o Jae-Gun Yang Jae-Hak J. Bae
 School of Computer Engineering and Information Technology, University of Ulsan

요약

개연규칙은 문장 구성성분들의 문장간 개연적 연결상황을 나타내고, 글의 인과성향이나 담화작용을 반영한다. Plot Units는 이야기를 형성하는 줄거리 또는 줄거리에 나오는 여러 사건을 하나로 구성하여 표현한다. 글을 읽고 Plot Units를 파악한다는 것은 그 글의 내용을 이해하고 있다는 것이다. 본 논문에서는 이러한 Plot Units를 찾는 한 방법으로 개연규칙을 생각하였다. 개연규칙들을 Plot Units와 대응시켜 봄으로써, 개연규칙을 활용하여 글의 Plot Units를 발견할 수 있음을 확인하였다.

1. 서론

글을 읽고 이해한다는 것은 단순히 기호를 해석하는 그 자체만으로는 완성되지 않는다. 여기에는 글을 통해 드러내고자 하는 필자의 의도를 파악해야 한다. 또한, 읽는 사람이 자신의 배경지식을 활용하여 그 의미를 추론하고 능동적으로 재구성하기도 한다. 컴퓨터가 글을 읽고 이해하는 경우에는, 기억장치에 이야기의 내용을 표현해야만 한다. 자연언어처리분야에서는 질의응답이나 문서요약과 같은 컴퓨터의 텍스트 이해에 관한 연구가 꾸준히 계속되고 있다.

의미적 자연언어처리에서는 이야기에 등장하는 주인공들간의 상호작용을 표현하는 방법으로 Plot Unit[1]을 활용한다. 이는 등장인물들의 심리적인 상태와 사건을 연결하여 전체적인 줄거리 구성을 이끌어 내는 방법이다. Plot Unit을 파악하는 것은 글의 내용을 이해하는 것이다. 이러한 연구들 중에서, PUGG(the Plot-Unit Graph Generator)[2]는 심상지도(affect state map)로 plot unit graph를 생성한다. 이는 주로 개념적인 부분을 다룬다. BORIS[3]는 이야기 이해 시스템으로, 이야기 표현을 위해 필요한 요소들을 제시한다. MUMBLE[4]은 PUGG에서 만들어진 plot unit graph로 문장을 생성하지만, 문장의 정렬과 중요 개념들의 선택이 불완전하다. 이를 위하여 PRECIS[5]는 PUGG와 MUMBLE 사이에서 문장 및 개념을 선별한다.

본 논문에서는 이야기 표현을 위한 Plot Unit을 찾는 방법으로 개연규칙(Abductive Rules)[6]을 생각하였다. 이를 위하여 구현한 원문해시스템에서 활용한 개연규칙과 Plot Units를 대응해 보았다.

2. Plot Units

Plot Units[1, 2]는 이야기(Narratives)의 표현(Representation)이다. 이는 이야기에 등장하는 주인공들 사이에서 일어나는 상호작용을 표현하고, 원문에서 언급된 사건이나 상황을 묘사한다.

2.1 Plot Unit의 구성

Plot Units는 심상(Affect States)이란 엔터티로 구성된

다. 심상은 등장인물의 관념적인 계획, 목적, 외부 사건에 대한 반응들을 나타낸다. Plot Units는 2개의 심상을 연결한 인과관계 사슬(Causal Links)로 표현된다. 이 사슬은 두 심상의 관계를 motivation(m), actualization(a), termination(t), equivalence(e) 등으로 구분한다.

Plot Units로 표현한 이야기 주인공들의 감정적인 반응들을 일어난 순서대로 나열하여 심상지도를 만든다. 이는 주로 숙련된 encoder들의 수작업으로 이루어 졌다.

2.2 Primitive Plot Units

인과관계 사슬로 연결된 plot units는 Primitive Plot Units와 Complex Plot Units로 나눌 수 있다. Primitive Plot Units는 등장인물 하나 또는 둘 사이의 심상에 관한 모든 가능한 관계들을 표현한다. Complex Plot Units는 Primitive Plot Unit들이 결합된 것이다. 그림 1은 Primitive Plot Units의 예이다.

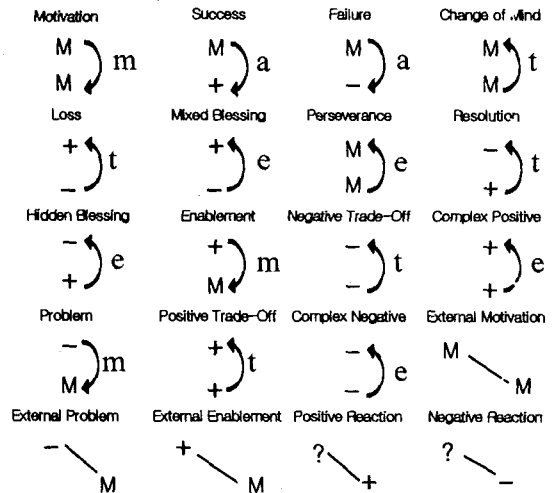


그림 1 Primitive Plot Units

그림 1은 심상 둘을 연결한 15개와 두 등장인물의 쌍으로 표현된 5개의 plot units이다. 이들은 각각 해석하기에 용이하고, 그 의미를 연상할 수 있도록 명명되어 있다.

2.3 Plot Units를 이용한 이야기 요약

심상지도로 만든 후 이야기를 요약하기 위한 단계들은 다음과 같다: (1) 심상지도에 표현된 모든 plot unit들을 인식한다. (2) 최상위의 Plot Units를 가려낸다. (3) 연결을 도식(Plot Unit Graph)한다. (4) 요약할 수 있는 요소들을 식별하기 위해 도식의 구조적 특징을 이용한다.

Plot Unit Graph에서 구조적 특징을 이용하여 연결도가 높은 결정적인 노트들을 식별할 수 있다. 이러한 노트들은 이야기의 중심적인 유닛들을 포함하고 있고, 이들로부터 요약을 생성할 수 있다.

3. 개연규칙: Abductive Rules

개연규칙은 문장간 구성성분들의 개연적인 결속성을 나타내며 문장내 구성성분들이 가지는 OfN(Ontology for Narratives)[7] 정보로서 표현된다.

3.1 OfN(Ontology for Narratives)

OfN은 다음의 7가지 범주로 구성되어 있다: 등장인물(Character), 심상(Affect State), 사건(Event), 상태(State), 시간과 공간의 변화(Delta-{Time, Space}), 담화표지(Discourse Marker). 이렇게 설정한 OfN을 구축하기 위해서 먼저 로젯 시소러스[9]의 범주를 심상, 시간과 공간, 사건, 그리고 상태 등으로 재편성하였다. 등장인물 유형에 속하는 어휘들은 고유명사 자원[10]을 이용하여 선정하였다. 담화표지의 경우는 수사구조의 연구결과[11]를 활용하였다. 이와는 달리 시공의 변화는, 구문분석 후 문장의 구성성분간의 상호작용에 의하여 확인되는 유형인 바, 그 기본유형은 시간과 공간이다.

3.2 개연규칙의 고안

개연규칙은 문단의 화제와 관련이 깊은 문장을 선택하는 과정에서 줄거리 단위, 단서구 용법, 문장 사이의 개연성 등을 감지할 수 있다. 이를 통하여 문단내용에 대한 이해심도를 높인다. 2항 또는 3항인 개연규칙의 일반적인 모습은 다음과 같다.

Ante <= Post [{=>, ==>, ==*}] Cons

여기에서 (1) Ante, Post, Cons는 pred(args)의 형태를 가진다, (2) pred(args)는 OfN에 명시된 개념으로 7가지의 범주로 표현된다, (3) ==> and ==->는 Post와 Ante에 제한사항이 있음을 나타내고, (4) ==*는 담화표지가 있음을 나타낸다. 그림 2는 개연규칙의 예이다.

% (1) 마음이 통하면 주고 싶어진다 affection(sympathetic) <= affection(offer)
% (2) 풀이 죽으면 소극적이 된다 event(inactivity) <= event(descent).
% (3) 장소를 바꾸고 심을 때 여행을 한다 delta(space) <= event(journey) ==-> affection(prospective).
% (4) 싫은 것을 권하면 흥미를 끝지 못한다 affection(advice) <= affection(cause(pleasure)) ==* cue_ phrase(adversative)

그림 2 개연규칙

3.3 개연규칙의 생성

이야기 문장에 일반적으로 적용할 수 있는 개연규칙을 만들기 위해 Dear Abby[8]에서 선택한 23편의 상담문을 분석하였다. 아래의 개연규칙 추출과정을 거쳐, 297개의 개연규칙들을 정리하였다.

- (1) 상담문의 주제 문장을 선택한다. 상담문의 제목과 가장 연관성이 높은 문장을 선택하고, 제목에 내포된 화자의 의도까지를 주제 문장의 범주에 포함시킨다.
- (2) 선택한 주제 문장의 개념과 핵심 단어를 선별한다.
- (3) 주제 문장의 개념, 핵심단어와 연결시킬 다른 문장의 단어를 추적하여 연결관계를 높여나간다.
- (4) 연결시킬 단어는 문장의 주요 구성성분으로 한다.
- (5) 정리된 상담문의 연결관계에 따라 개연규칙의 일반적인 형태로 정리한다.
- (6) 개연규칙에 참여하는 단어가 OfN의 복수 범주에 해당 될 경우에는 상담문의 내용과 가장 연관성이 높은 범주를 선택한다.
- (7) 선정된 OfN 범주에서 대표 개념을 정하고 개연규칙을 단순화한다.

3.4 개연규칙의 활용: 원문이해 시스템

본 논문에서 소개하는 원문이해 시스템은 표층적 원문이해(Shallow Text Understanding)에 바탕을 두고, 현시점에서 가용한 자연언어 처리도구나 자원을 적극적으로 활용하여, 가능한 깊이 원문을 이해할 수 있게 설계된 것이다. 이는 심층적 원문이해를 지향하는 계산가능한 중층적 원문이해(MIDTERM: Mid-Depth Text Understanding)[7]라고 할 수 있다. 원문이해용 계산에 사용하는 자원과 도구는 다음과 같다: 온톨로지 OfN, 견실한 구문분석기(Robust Parser) LGPI+, 문장추상기(Sentence Abstracter) SABOT, 그리고 개연사슬기(Abductive Chainer) SICHA 등.

시스템은 문장추상기와 개연규칙을 적용하여 문단을 구성하는 문장간 개연적 연결상황을 파악하고, 개연사슬기 SICHA로 문장간 연결도가 높은 문단의 화제문들을 선택한다.

4. 개연규칙과 Plot Units의 대응

개연규칙과 Plot Units를 비교하기 위해서 각각의 형식을 이해하고, 그 의미가 서로 연관성을 가지는 쌍을 구해 보았다. 아래 그림 3은 그러한 개연규칙과 Plot Units의 대응 예이다.

Plot Units	개연규칙
Motivation	% 외로울 때는 친구가 생각난다 % boyfriend <= lonely affection(sympathetic) <= event(exclusion).
Success	% 절친해서 일을 도모하다. % close <= make affection(volition(intersocial)) <= event(causation).
Enablement	% 구입할 여력이 있다. % purchased <= able event(relation(possessive)) <= event(connection(cause, effect)).
Problem	% 나빠서 옮기다. % tragedy <= transferred affection(evil) <= event(motion(general)).

그림 3 개연규칙과 Plot Unit의 대응 예시

Ofn은 7가지 유형을 가지고 있으나, Plot Units의 구성은 affect와 event로 이루어져 있다. 따라서, 생성한 개연규칙들 중에서 affect state와 event의 조합으로 만들어진 것만을 대응하였다. 여기에는 267개의 개연규칙들 중에서 169개가 해당된다.

Complex Plot Units의 경우, Primitive Plot Units이 결합된 것이므로 본 논문에서는 개연규칙과 Primitive Plot Units만을 비교하였다. 표 1은 Primitive Plot Units에 개연규칙을 대응시키고, 각 Plot Units에 대응하는 개연규칙의 수를 계산한 것이다.

표 1 개연규칙과 Plot Units의 비교

Primitive Plot Units	Plot Unit Diagram	개연규칙 갯수
1. Motivation	M --(m)--> M	78
2. Success	M --(a)--> +	22
3. Failure	M --(a)--> -	9
4. Change of Mind	M <--(t)-- M	6
5. Loss	+ <--(t)-- -	0
6. Mixed Blessing	+ <--(e)-- -	0
7. Perseverance	M <--(e)-- M	4
8. Resolution	- <--(t)-- +	3
9. Hidden Blessing	- <--(e)-- +	0
10. Enablement	+ --(m)--> M	18
11. Negative Trade-Off	- <--(t)-- -	1
12. Complex Positive	+ <--(e)-- +	3
13. Problem	--(m)--> M	14
14. Positive Trade-Off	+ <--(t)-- +	7
15. Complex Negative	- <--(e)-- -	2
16. External Motivation	M ---- M	1
17. External Problem	- ---- M	0
18. External Enablement	+ ---- M	0
19. Positive Reaction	? ---- +	0
20. Negative Reaction	? ---- -	1
Total of Abductive Rules		169

개연규칙은 Primitive Plot Units의 20가지 분류 중 14가지에 각각 대응하고 있다. 이 중에서 가장 많은 분포를 보인 곳은 Motivation, Success, Enablement 순이다. 대응하는 개연규칙이 없는 Plot Units의 범주도 보인다. 이는 개연규칙의 수와 경우를 보다 더 증가시킨다면 대응률은 향상 될 것으로 기대된다. 표 1은 일반적인 이야기의 줄거리나 사건의 전개를 파악하기 위한 Plot Unit을 찾는 방법으로 개연규칙을 활용할 수 있음을 보인다.

5. 결론

본 논문에서는 주어진 글의 Plot Unit을 찾는 방법으로 개연규칙을 생각하였다. 이를 위하여 원문이해시스템에서 활용한 개연규칙들을 Plot Units와 대응시켜 보았다.

Plot Units는 이야기에 등장하는 주인공들 사이에서 일어나는 상호작용 및 사건 전개를 표현한다. 한편, 개연규칙은 문장간 구성성분들의 개연적인 결속성을 나타내며 그 구성성분들이 가지는 온톨로지 Ofn 정보로서 표현된다. 문단의 화제와 관련이 깊은 문장을 선택하는 과정에서 개연규칙은 줄거리 단위, 단서구 용법, 문장 사이의 개연성 등을 감지할 수 있다. 글을 읽고 Plot Unit을 파악한다는 것은 그 글을 이해하고 있다는 것이다.

개연규칙으로 Plot Unit을 발견할 수 있음을 확인하기 위하여 그 의미가 서로 연관성을 가지는 쌍을 구해 보았

다. 그 결과, 개연규칙은 Primitive Plot Units의 분류 중 70%에 각각 대응되었다. 각 분류마다 대응된 개연규칙의 수는 다르게 나타났다. 특히, Motivation은 가장 높은 대응률과 분포를 보였다. 이는 개연규칙을 이용하여 이야기를 이해하고 표현하기 위한 Plot Unit의 발견이 가능하며, 이야기의 줄거리나 사건을 구성할 수 있음을 보인다.

[참고문헌]

- [1] Lehnert, W.G., "Narrative complexity based on summarization algorithms.", Proc. of IJCAI, pp. 713-716, 1983.
- [2] Lehnert, W.G., Loisel, C.L., "Plot unit recognition for narratives." (Tech. Rep. No. 83-39) Amherst, MA: University of Massachusetts, Department of Computer and Information Science, 1983.
- [3] Lehnert, W.G., Dyer, M. G., Johnson, P. N., Yang, C. J., Harley, S., "BORIS-An experiment in in-depth understanding of narratives." Artificial Intelligence, 20(1), pp. 15-62, 1983.
- [4] McDonald, D. D., "Natural language generation as a computational problem-An introduction." In M. Brady, R. Berwick(Eds.), Computational models of discourse(pp. 209-265). Cambridge, MA: MIT Press, 1983.
- [5] Cook, M., Lehnert, W. G., McDonald, D. D., "Conveying implicit content in narrative summaries." Paper presented at COLING-84, Palo Alto, CA, 1984.
- [6] 김곤, 배재학, "개연성 규칙과 문장추상화를 활용한 문서요약", 한국정보처리학회 춘계학술발표대회 논문집, 제9권 제2호, pp.359-362, 2002.
- [7] Bae, J.-H. J. and Lee, J.-H. "Topic Sentence Selection with Mid-Depth Understanding." Proc. of ICCPOL, pp. 199-204, 2001.
- [8] DearAbby. <http://www.dearabby.com/>.
- [9] Roget's Thesaurus. <http://promo.net/cgi-promo/pg/t9.cgi?entry=22&full=yes&ftpsite=ftp://biblio.org/pub/docs/books/gutenberg/>.
- [10] Proper Names Wordlist. <http://clr.nmsu.edu/cgi-bin/Tools/CLR/clrcat#14>.
- [11] 배재학, "언어학적인 방법론을 취하는 자동 문서요약에 대한 연구." 공학연구논문집, 제 29권 2호, pp.351-363, 울산대학교, 1998.
- [12] R. Barzilay and M. Elhadad, "Using Lexical Chains for Text Summarization", in Proc. ISTS'97 (The Intelligent Scalable Text Summarization Workshop, ACL), Madrid, Spain (Jul. 1997), pp. 10-17.
- [13] D. Marcu, "From Discourse Structures to Text Summaries", in Proc. ACL'97 and EACL'97 Workshop on Intelligent Scalable Text Summarization, Madrid, Spain (Jul. 1997), pp. 82-88, 1997.