

독일어 대화의 자동 분절 방법

- 대화 자동 요약을 위하여

구영은, 홍문표 (성균관대)*

국문요약

대화는 화자 간의 다양한 상호작용들로 구성된다. 대화라는 의사소통을 통해 어떤 상호작용이 발생했는지 파악하는 것은 대화를 이해하는 데에 매우 중요하다. 본 논문은 독일어 대화 요약을 위해 대화 내 발생하는 상호작용 단위로 대화를 분절하고자 하였다. 대화 분절은 대화를 작은 단위로 나누는 과정으로, 관련 있는 문장끼리 묶어서 대화의 구조 및 내용을 파악할 수 있게 한다. 본 논문은 대화 분절에서 발화문의 대화적 특징이 갖는 중요성을 지적하고, 대화 요약에 적합한 단위로 대화를 분절하기 위해 다양한 언어학적 성질을 제안한다. 이를 검증하기 위해 문장 임베딩을 포함한 딥러닝 방법론이 활용된 기계학습 실험을 진행하였다. 그 결과 제안한 언어학적 요소들을 모두 적용하면 자동 분절 정확도가 83.56%로 나타났다. 특히 본 논문에서 강조하는 대화적 특징을 사용하면 기존 연구에서 주장한 음성·어휘·의미적 특징만을 사용할 때에 비해 정확도가 약 5%p 향상하는 것을 알 수 있었다. 이를 통해 인간이 대화를 상호작용 단위로 나누어 이해하는 데에 대화적 자질을 포함한 다양한 언어학적 특징을 활용한다는 점을 발견하였다.

주제어: 대화 분절, 대화 요약, 대화 구조, 기계학습, 임베딩

I. 서론

대화에는 화자 간에 발생하는 다양한 상호작용들이 존재한다. 예를 들어 호텔 예약 대화에는 서로 주고받는 인사, 빈 객실에 대한 정보 제공과 이에 대한 상대방의 반응, 특정 객실을 예약하고자 하는 희망 표출과 이에 대한 상대방의 반응 등과 같은 의사소통을 통한 상호작용들이 발생한다. 대화는 이러한 상호작용

* 제1저자: 구영은(성균관대학교 독어독문학과 박사과정), 교신저자: 홍문표(성균관대학교 독어독문학과 교수)

들로 규정되고 이해될 수 있다. 따라서 화자 간 상호작용 양상을 통해 대화의 주제, 대화 참여자 간의 관계, 대화 주고받음의 결과 등 대화에 관한 다양한 정보를 파악하고 이해할 수 있다.

대화 요약 Dialogzusammenfassung이란 주어진 대화를 핵심 내용이 담긴 짧은 결과물로 변환하는 것을 말한다. 최근 코로나바이러스감염증-19 COVID-19가 장기화되면서 비대면 의사소통의 중요성은 점차 더 부각되고 있다. 이런 상황에서 컴퓨터 기술의 발전으로 데이터 수집이 용이해지고 데이터 보관 플랫폼이 다양하게 개발되면서 대규모의 대화 텍스트가 데이터화되고 있다. 대화 요약은 이러한 방대한 양의 대화 데이터로부터 키워드, 주요 문장 등 중요한 정보를 전달하여 대화 데이터의 가치를 증대시키고 다양한 대화 관련 서비스에 활용될 수 있는 기술이다.

대화를 잘 요약하기 위해서는 대화의 핵심 내용이 무엇인지 잘 파악해야 한다. 이를 위해서 먼저 대화가 어떤 내용들로 구성되어 있는지 파악해야 한다. 특히 여러 개의 발화문으로 구성된 긴 대화가 주어진다면, 관련 있는 발화문끼리 묶어서 대화를 구조화하는 과정이 선행되어야 한다. 대화 분절 Dialogsegmentierung이란 여러 발화문이 연결되어 만들어진 대화를 작은 단위로 쪼개는 과정이며, 대화를 주제 단위로 분절하는 것이 일반적이다. 대화 분절은 긴 문장의 나열을 서로 관련 있는 단위로 나뉘춤으로써 대화에 대한 구체적인 이해를 가능케 하므로, 많은 대화 처리 연구들에 기본이 된다. 대용어 해소 Anaphernresolution(Grosz/Sidner 1986), 정보 검색 Information Retrieval(Oard et al. 2004), 요약(Zechner/Waibel 2000; Murray et al. 2007) 등이 대표적인 활용 분야이다.

대화를 적절한 단위로 분절하기 위해서는 대화에 대한 이해가 선행되어야 한다. 이를 바탕으로 텍스트 분절 Textsegmentierung과는 다른, 대화에 적합한 분절 방법론이 제안되어야 한다. 먼저, 대화는 텍스트와 달리 구어체적 성격이 커서 생략, 발화 실수, 수정, 말더듬 등이 빈번하다. 따라서 대화 분절을 위해서는 발화문의 표층 정보뿐만 아니라, 발화문의 화자 정보나 대화의 진행 과정 등의 대화 상황에 대한 고려도 필요하다. 그리고 텍스트 분절에서 다루는 뉴스 기사, 법률 문서 등의 텍스트와 달리 대화의 경우, 주제만으로는 전체 대화를 이해하고 압축하기 어렵다. 대화 참여자들이 어떤 말을 주고받았는지, 즉 어떤 상호작용이 있었는지에 관한 정보도 필요하다.

따라서 본 논문은 대화 요약에 있어서 화자 간에 발생하는 상호작용의 중요성을 인식하고, 이를 파악할 수 있는 단위로 대화를 분절하는 방법을 제안하고자 한다. 본 논문에서는 대화 분절을 위한 음성·어휘·의미·대화적 측면의 언어학적 요소들을 분석한다. 그리고 제안한 언어학적 요소들을 전산 언어학적 도구를 사용하여 표현하고, 이를 이용한 기계학습 *Machinelles Lernen* 실험을 수행하여 본 논문에서 제안하는 대화 분절 방법론의 타당성을 검증해본다.

본 논문의 구성은 다음과 같다. 우선 2장에서는 대화의 상호작용 단위에 관한 언어학 및 언어 공학적 선행 연구들을 살펴본다. 3장에서는 대화 분절과 관련된 언어학적 자질들을 제안하고, 제안하는 방법론의 자동화를 위해 사용하는 전산 언어학적 방법론과 도구들을 소개한다. 4장에서는 제안한 대화 분절 방법론에 따라 기계학습 실험을 진행하고 그 결과를 살펴본다. 마지막으로 5장에서는 향후 연구 방향에 대하여 논의하며 논문을 마무리한다.

II. 기존 연구

1. 대화의 상호작용 단위에 관한 언어학적 연구

대화분석론에서 대화의 상호작용 단위를 다루는 가장 기본적인 연구로는 인접쌍 *adjacency pair*을 이야기할 수 있을 것이다. 인접쌍은 ‘가장 작은 상호작용적 시퀀스(*minimal interactional sequence*)’이다(Schegloff 2007, 22; Clift 2016, 64). 예를 들어 ‘질문-답변’ 인접쌍에서 두 화자 간에는 정보 교환이라는 상호작용이 발생한다. 이처럼 인접쌍은 두 화자가 어떤 상호작용을 수행했는지 이해할 수 있게 하는 단위이다.

다음으로 대화의 상호작용 단위를 이해할 수 있는 개념으로 대화원형 *Dialogmuster*이 있다. 대화원형은 대화를 구성하는 필수 구성요소로 (Spiegel/Spranz-Fogasy 2001) 대화에서 기대되는 진행 단계들에 해당한다. 이때 대화원형은 누가, 무엇을, 어떻게 행하는지 등을 나타내기 때문에 대화 유형에 따라 다르게 나타난다. 예를 들어 의료 대화는 다음과 같은 대화원형의 흐름을 갖는다(Spranz-Fogasy 2008).

- 첫인사와 대화 개시
- 증상 묘사와 증상 호소
- 진단 내리기
- 치료 계획과 치료 진행
- 대화 종료와 끝인사

수업 대화는 다음과 같은 대화원형의 흐름을 갖는다(박용익 1994a).

- 안내 단계(Einführungsphase)
- 복습 단계(Rekapitulationsphase)
- 주제 전개 단계(Themenbehandlungsphase)
- 예고 단계(Ankündigungsphase)
- 규율 단계(Disziplinierungsphase)

이처럼 Spiegel/Spranz-Fogasy(2001), Spranz-Fogasy(2008)가 제안한 대화원형은 대화의 수행 단계와 관련된 개념으로 대화의 주제 및 목적의 실현 측면에서 대화의 흐름을 표현한다. 반면 Franke(1990)의 최소대화는 화행을 기반으로 대화의 흐름을 표현하려는 시도에서 제안된 개념이다.

최소대화 Minimaldialog란 ‘완결된 하나의 대화 이동 연속체(박용익, 1994a; 1994b)’이다. 즉 대화문법론 Dialoggrammatik에서 최소대화는 의사소통의 최소 단위이다. <표 1>은 Franke(1990)의 최소대화 유형을 한국어로 번역하여 정리한 것이다¹⁾.

Franke(1990)의 최소대화는 발화의 흐름을 발화의 언어행위 측면에서 표현하여 다양한 대화에 적용 가능한 의사소통 단위가 되었다는 의의가 있다. 그러나 시작 화행, 긍정 반응 화행, 반박 화행 등 각각의 화행 유형에 어떤 구체적인 화행이 해당할 수 있는지에 대한 설명을 하고 있지 않아서 실제 대화 처리 기술에 적용하기 위해서는 이에 대한 분석이 추가로 필요하다.

1) 시작 화행(Initialsprechakt), 긍정 반응(Positiver Bescheid), 부정 반응(Negativer Bescheid), 철회 화행(Retraktiver Sprechakt), 수정 화행(Revidierender Sprechakt), 재시작 화행(Re-initiativer Sprechakt), 역시작 화행(Gegen-initiativer Sprechakt), 반박(Zurückweisung).

표 1. Franke(1990)의 최소대화 유형

	화자 1	화자 2	화자 1	화자 2	...
최소대화 1	시작 화행	긍정 반응 화행			
최소대화 2	시작 화행	부정 반응 화행	철회 화행		
최소대화 3	시작 화행	반박 화행	철회 화행		
최소대화 4	시작 화행	역시작 화행	수정 화행	철회 화행	
최소대화 5	시작 화행	부정 반응 화행	수정 화행	철회 화행	
최소대화 6	시작 화행	역시작 화행	재시작 화행	철회 화행	
최소대화 8	시작 화행	부정 반응 화행	재시작 화행	철회 화행	
최소대화 10	시작 화행	반박 화행	재시작 화행		

2. 대화의 상호작용 단위에 관한 언어 공학적 연구

전산 언어학 및 자연어처리 분야에서는 대화 모델 개발을 위한 대화 분절 연구가 활발히 진행되었다. 먼저 대화 분절에 기반이 되는 텍스트 분절에 관한 연구는 크게 두 가지의 방법론을 기반으로 한다. 첫째, 문장에 등장하는 단어의 분포를 바탕으로 분절하는 방식이다. 이 경우 텍스트 내 문장들이 어떤 단어들로 구성되는지, 즉 텍스트의 어휘적 응집성 *lexikalische Kohäsion*을 분석하여 텍스트를 분절한다. 두 문장이 비슷한 어휘를 공유하고 있어 어휘적 응집성이 높은 경우에는 두 문장 간의 관련도가 높다고 볼 수 있다. 따라서 이런 경우에는 두 문장 사이에서 분절하지 않고 연결된 단위로 본다. 어휘적 응집성을 바탕으로 텍스트를 분절하는 대표적인 방법론으로 *TextTiling*(Hearst et al., 1997)과 *LCseg*(Galley et al., 2003)이 있다. *TextTiling*은 분절 경계의 좌우 문맥에 속한 단어들의 단어 빈도를 계산하였고, *LCseg*는 TF-IDF(Term Frequency-Inverse Document Frequency) 가중치를 사용하여 텍스트 분절의 성능 향상을 이끌었다.

둘째, 문장 간의 주제 유사도를 고려하는 분절 방식이 있다. 어휘적 응집성이 있더라도 문장의 주제가 다를 수 있다는 문제와 단어 기반 방법론이 갖는 희소성 문제를 해결하기 위해, 텍스트의 주제를 판단하여 관련 있는 주제의 문장끼리 묶을 수 있다. 주제적 유사도를 바탕으로 하는 텍스트 분절의 대표적인 방법론으로는 *TopicTiling*(Riedl/Biemann 2012)이 있다. *TopicTiling*은 텍스트를 통해 문장들

의 주제 벡터를 추출하는 잠재적 디리클레 할당 Latent Dirichlet Allocation 모델을 이용하여 문장의 주제를 결정하고 이를 바탕으로 텍스트를 분절한다.

대부분의 초창기 대화 분절 연구는 위에서 설명한 텍스트 분절의 방법론을 사용하였다. 즉 발화문에 등장하는 단어의 분포(Galley et al. 2003; Hsueh/Moore 2007; Song et al. 2016)와 발화문 간의 유사도를 활용하여 대화를 분절하였다(Arguello/Rose 2006; Riedl/Biemann 2012). 이후 대화에 초점을 맞춘 분절 모델을 만들기 위해 음성 정보를 활용하였다(Dharanipragada et al. 2000; Galley et al. 2003; Arguello/Rose 2006; Hsueh/Moore 2007). 이외에도 대화는 둘 이상의 화자가 함께 만드는 언어 표현이라는 점을 고려하여 화자 정보를 반영시킨 모델도 존재한다(Galley et al. 2003; Arguello/Rose 2006; Hsueh/Moore 2007).

최근에는 인공 신경망 künstliches neuronales Netz의 발전으로 ELMo(Peters et al. 2018), BERT(Devlin et al. 2018)와 같은 문맥 기반의 임베딩으로 발화문을 풍부하게 표현하고, 이를 바탕으로 대화를 분절하는 연구가 진행되고 있다(Wang et al. 2018; Lukasik et al. 2020; Xing/Carenini 2021). 뿐만 아니라 Bi-LSTM(Hochreiter/Schmidhuber 1997)과 어텐션 기법(Vaswani et al. 2017)을 기반으로 대화 분절이라는 언어 태스크를 잘 이해할 수 있는 구조로 대화 처리 모델을 설계하려는 시도들도 있다(Koshorek et al. 2018; Wang et al. 2018; Lukasik et al. 2020).

III. 방법론

이 장에서는 대화 분절을 위한 다양한 언어학적 특징들을 밝히고, 이를 분류 자질 Merkmal로 활용한 기계학습 기반의 대화 자동 분절 실험을 진행한다. 대화 분절에 영향을 미치는 발화문의 음성, 어휘, 의미, 대화적 특징은 <표 2>와 같이 제한한다. 이어지는 장에서는 제한하는 언어학적 특징들을 유형별로 살펴보고, 이를 활용하여 대화를 자동으로 분절하는 모델을 구축하기 위한 방법론과 전산 언어학적 도구들에 대해 설명한다.

표 2. 대화 분절을 위한 언어학적 자질

자질 번호	자질명		자질값
F1	음성	휴지 길이	대화 코퍼스의 부착 정보
F2		중첩 여부	대화 코퍼스의 부착 정보
F3	어휘	단서 단어	DiMLex(독일어 담화표지 목록) ²⁾
F4		정보량	문장에 등장하는 특정 품사 단어의 개수
F5	의미	문장 유사도	문장 임베딩 벡터의 코사인 유사도(FastText) ³⁾
F6	대화	화행 유형	SWBD-DAMSL Tags ⁴⁾
F7		의사소통 진행 정도	initiating act, responding act, both
F8		인접 화행쌍	구영은(2018)에서 제안한 인접 화행쌍
F9		화자 유형	sp1, sp2, ...
F10		차례 전환	y, n

1. 음성 정보

첫 번째 음성 자질은 발화문과 발화문 사이의 휴지 Pause의 길이이다⁵⁾. 휴지와 분절 간의 가장 직관적인 관계는 휴지의 길이가 긴 위치에서 분절이 일어날 확률이 높다는 것이다. 발화가 바로 이어지지 않고 휴지가 오래 발생한다는 것은 앞뒤 발화 간의 관련성이 상대적으로 떨어지는 것을 의미하기 때문이다.

<표 3>은 본 연구에서 분석한 독일어 대화의 일부이다⁶⁾. 표 가장 우측에 있는 휴지 길이는 발화문과 그다음 발화문 사이의 휴지 길이를 표시한 것이다. <표 3>에 따르면 발화 2와 3 사이에는 1.54초의 휴지가 존재한다. 발화 1~2에서는 두 화자가 구매한 물건을 담은 작은 봉지와 관련있고, 3~5에서는 직원이 손

2) <https://github.com/discourse-lab/dimlex>

3) <https://fasttext.cc/docs/en/crawl-vectors.html>

4) <https://web.stanford.edu/~jurafsky/ws97/manual.august1.html>

5) Sacks et al.(1978)은 화자가 누구인지, 차례 전환이 이루어진 방식이 어떠한지 등에 따라 'lapse', 'gap', 'pause'로 나누어 설명하였다. 이를 스투어트 외(2015)에서는 각각 '경과', '틈', '휴지'로, 서종훈(2013)에서는 각각 '긴 침묵', '묵묵부답', '잘막힌 침'으로 번역하였다.

본 논문은 이를 구별하지 않고 휴지(Pause)라는 용어로 통일하였다.

6) 본 연구에서 분석한 독일어 대화는 독일어 구매 대화 9개(약국 5개, 식물원 4개)이며, 자세한 설명은 IV장에 제시되어 있다.

님에게 제공하는 증정품과 관련된다. 두 발화 시퀀스 사이에 발생한 휴지를 통해 중간에 대화가 분절될 수 있음을 유추할 수 있다.

표 3. 독일어 대화 예시 (1) 휴지 길이와 대화 분절의 관계

발화 번호	화자	발화문	휴지 길이
→ 1	Sp 4	Kleines Täschchen damit Sie nicht?	0
→ 2	Sp 3	Nein, brauche ich nicht.	1.54
→ 3	Sp 4	Paar Tempo aber das kriegen Sie gratis.	0
→ 4	Sp 3	Das freut mich.	0
→ 5	Sp 4	So, bitte schön.	1.14
→ 6	Sp 3	Ich danke Ihnen.	0
→ 7	Sp 4	Ich danke Ihnen.	0
→ 8	Sp 4	Guten Heimweg!	0
→ 9	Sp 4	Tschüss!	0
→ 10	Sp 3	Ciao!	0

두 번째 음성 자질은 발화문이 서로 중첩 overlap 되는지의 여부이다. 발화가 서로 중첩된다는 것은 보통 화자들이 이전 발화에서 얘기하는 내용에 대해 적극적으로 대화를 수행하는 경우에 해당한다.

표 4. 독일어 대화 예시 (2) 중첩 여부와 대화 분절의 관계

발화 번호	화자	발화문	중첩 여부
→ 1	Sp 6	So das macht schön die Nase frei gell man kann wieder atmen.	n
→ 2	Sp 6	Und kann auch dann gut einschlafen nämlich,	y
→ 3	Sp 5	Das will ich hoffen.	n

<표 4>를 보면 발화 2와 발화 3이 중첩되는 것을 알 수 있다. 위의 대화는 비염 스프레이를 구매하러 온 손님과 직원의 대화이며, <표 4>에 제시된 대화 일부는 직원이 비염 스프레이를 사용하고 난 후의 효과에 대해 설명하고 손님이 이에 대한 반응을 보이는 부분이다. 두 화자의 발화가 중첩되는 것을 통해 두

발화가 서로 연관될 가능성이 크고 따라서 두 발화 사이에서 분절하지 않는 것이 적합하다고 유추할 수 있다.

음성 대화를 전사한 대화 코퍼스는 음성인식 결과와 함께 음성이 인식되는 시간이나 복수 개의 음성이 동시에 인식되었음이 주석으로 부착된 경우도 많다. 본 연구에서 분석한 독일어 대화 코퍼스의 경우 휴지 길이와 중첩이 주석되어 있어 이를 활용하여 실험을 진행하였다.

2. 어휘 정보

첫 번째 어휘 정보는 발화문 간의 분절 여부를 암시하는 단서 단어 cue phrases이다. Galley et al.(2003)에 따르면 ‘now’, ‘well’과 같은 표현은 담화의 구조를 암시함으로써 대화를 분절하는 데에 단서가 된다. 본 논문에서는 König(1997)의 독일어 양태첨사 Modalpartikel 17개⁷⁾와 독일어 담화표지 목록인 DiMLex(Stede 2002)의 표현을 합쳐 총 211개의 표현을 단서 단어로 사용하였다. 특히 이러한 표현은 문장 앞 또는 뒷부분에 등장하여 문장 양태를 결정하고 이를 통해 담화를 구조화하기 때문에(구영은 2018), 발화 첫 두 단어와 끝 두 단어가 단서 단어에 해당하는지 분석하였다.

두 번째 어휘 정보는 다음 발화에 등장하는 발화의 정보량이다. 이어지는 발화가 얼마나 많은 정보를 담고 있는지는 발화문 사이에서의 분절 여부를 추측할 수 있는 단서가 된다(Arguello/Rose 2006). 다음에 이어지는 발화가 높은 정보량을 갖는다면 정보량이 낮은 발화문에 비해 새로운 내용을 담고 있을 가능성이 크고, 따라서 이전 발화와 구분될 확률이 높다.

이를 위해 독일 슈투트가르트 대학에서 개발한 TreeTagger⁸⁾를 이용하여 발화문에 품사를 태깅하였다. 이후 전치사, 관사 등의 문법적 유형이 아닌, 내용적 유형에 해당하는 일부 품사⁹⁾의 단어만 추출하여 분포를 계산하였다.

7) 본 논문에서는 aber, auch, bloß, denn, doch, eigentlich, eben, etwa, erst, halt, ja, nun, nur, schon, vielleicht, ruhig, wohl 을 고려하였다.

8) <https://www.cis.lmu.de/~schmid/tools/TreeTagger/>

9) ‘ADJA(attributive adjective)’, ‘ADJD(predicative or adverbial adjective)’, ‘ADV(adverb)’, ‘NN(simple noun)’, ‘NE(proper name)’, ‘VVFIN(finite full verb)’, ‘VVIMP(imperative full verb)’, ‘VVINF(infinitive full verb)’, ‘VVIZU(infinitive+zu)’, ‘VVPP(past participle full verb)’이 해당된다.

3. 의미 정보

대화 분절에 영향을 미치는 의미적 정보는 두 발화문 간의 의미적 유사도이다. 유사도가 높은 경우 두 문장이 서로 깊게 관련될 확률이 높으므로 분절하지 않는 것이 적합하다. 유사도가 낮은 경우 상대적으로 두 문장이 관련도가 떨어지기 때문에 각각 다른 단위로 분절될 수 있다(Song et al 2016).

발화문 간의 유사도는 두 발화문의 문장 임베딩 벡터 사이의 거리를 통해 알 수 있다. 임베딩 *Einbettung*이란 방대한 양의 언어 데이터로부터 자연어의 성질을 학습하여 자연어를 실수 공간에 놓는, 즉 0~1 사이의 실수값으로 변환하는 과정을 말한다. 주어진 데이터에서 어떤 정보를 학습하여 자연어의 성질을 이해하는 지에 따라 다양한 종류의 임베딩 모델이 존재한다. 대표적으로 *Word2Vec*(Mikolov et al. 2013)과 *FastText*(Bojanowski et al. 2017)는 단어와 주변 문맥 간의 관계, *GloVe*(Pennington et al. 2014)는 단어 간의 동시 등장 확률을 통해 자연어의 성질을 학습한다. 이를 바탕으로 임베딩 모델은 각 자연어 표현을 실수값으로 변환하는데, 이 값을 임베딩 벡터라고 부른다. 단어를 실수값으로 변환하는 것을 단어 임베딩 *Worteinbettung*, 문장을 실수값으로 변환하는 것을 문장 임베딩 *Satzeinbettung*이라고 한다.

자연어처리에서는 자연어의 유사도를 그것의 임베딩 벡터 간의 거리를 통해 계산한다. 값이 0에 가까울수록 두 표현은 서로 다른 문맥에서 사용되어, 유사도가 낮음을 의미한다. 반대로 값이 1에 가까울수록 비슷한 문맥에서 사용된 빈도가 높으며, 유사도가 높다고 해석할 수 있다¹⁰⁾.

여러 임베딩 모델 중에서 본 논문은 미등록어 *out-of-vocabulary* 문제에 유연하게 대응 가능하며 쉽게 문장의 임베딩 벡터를 추출할 수 있는 *FastText* 모델을 선택하였고, 페이스북에서 제공하는 사전 학습된 임베딩 모델 *pre-trained embedding model*을 가져와 실험에 활용하였다. 이는 독일어 위키피디아의 본문을 학습시킨 모델이며, 총 1,384,170,636개의 단어를 300차원의 벡터 공간에 매핑하였다¹¹⁾. 발화문 간의 의미적 유사도는 두 발화문의 문장 임베딩 벡터 간 코사인 유사도 *cosine similarity*¹²⁾로 계산하였다.

10) 코사인 유사도를 제외한 대부분의 벡터 유사도 기법은 정규화를 통해 0에서 1 사이의 유사도값으로 표현될 수 있다.

11) <https://fasttext.cc/docs/en/crawl-vectors.html>

4. 대화 정보

첫 번째 대화 정보는 분절 여부를 판단하는 지점 좌우 문맥에 포함된 발화문의 화행이다. 화행이란 의사소통 과정에서 발화자가 갖는 발화 의도를 말한다(강창우 2004; 조국현 2004; 강창우 2007; 구영은 외 2018). 즉 발화문이 어떤 화행 인지는 그 발화를 잘 이해하는 데에 중요할 뿐만 아니라, 화자가 해당 발화를 통해 이어서 어떤 의사소통을 목표하는지 유추할 수 있다. 예를 들어 화자 A가 요청의 화행을 발화했다면 화자 A는 화자 B가 이에 대한 응답을 제공할 것을 기대할 것이다. 만일 화자 B가 이를 수행하지 않으면 화자 A는 자신의 기대를 포기하거나 재도전하게 될 것이다(Franke, 1990).

이처럼 발화문은 대화 내에서 서로 얽혀 존재하는 요소이며, 이에 따라 발화문의 화행은 그것의 앞 혹은 뒤에 어떤 대화의 흐름이 등장하는지 암시하는 대화 정보이다. 따라서 대화의 흐름을 유추할 수 있는 화행 정보는 궁극적으로 각 발화문 간의 분절이 적합한지 판단하는 데에 유용하다. 본 논문에서는 SWBD-DAMSL(Jurafsky et al. 1997)의 화행 태그셋¹³⁾을 바탕으로 발화문의 화행을 태깅하였다.

두 번째 대화 정보로 발화문의 화행을 활용하면 현재 의사소통의 진행 정도를 유추할 수 있다. 인간의 의사소통은 화행의 측면에서 볼 때 크게 2가지로 나눌 수 있다. 상호작용을 시작하는 행위와 시작된 상호작용에 반응하는 행위이다. 더 나아가 해당 상호작용이 의견이나 생각에 관한 것인지 혹은 행위에 관한 것인지로 다시 나뉘볼 수 있을 것이다(구영은/홍문표 2019). 본 논문에서는 SWBD-DAMSL 화행 유형을 각각 시작 행위 initiating act와 반응 행위 responding act로 분류하여, 이를 바탕으로 대화를 분절하는 데에 의사소통의 진행 정도를 활용하였다.

12) 코사인 유사도는 두 벡터 간의 유사도를 구하는 여러 기법 중 가장 널리 사용되는 기법으로 두 벡터 간의 코사인 각도를 이용하여 유사도를 계산한다. 코사인 유사도는 -1에서 1 사이의 값을 가지며, 값이 1에 가까울수록 유사도가 높다고 판단한다. 코사인 유사도 외에도 자카드 유사도 Jaccard similarity, 유클리디언 유사도 Euclidean similarity, 맨하튼 유사도 Manhattan similarity 등의 벡터 유사도 측정 기법들이 있다.

13) SWBD-DAMSL은 영어 전화 대화인 Switchboard corpus에 DAMSL(The Dialogue Act Markup in Several Layers)(Allen/Core 1997) 화행 태그를 부착한 데이터이며, 빈번한 42개의 화행 유형을 밝혔다. SWBD-DAMSL 화행 유형에는 ‘Statement-non-opinion’, ‘Statement-opinion’, ‘Agree’, ‘Reject’, ‘Wh-Question’, ‘Acknowledge’, ‘Uninterpretable’, ‘Non-verbal’, ‘Other’ 등이 있다.

표 5. 독일어 대화 예시 (3) 의사소통 진행 정도와 대화 분절의 관계

발화 번호	화자	발화문	유형
1	Sp 16	Und sprühen das jetzt ab Mai Juni.	both
2	Sp 16	Überwiegend frühmorgens unter die Sträucher.	both
→ 3	Sp 16	Also die auf die Blätter von der Unterseite her.	both
→ 4	Sp 15	Gibt es denn ganz frühmorgens wie sagen Sie?	init
5	Sp 16	Ja frühmorgens so sieben acht Uhr um den Zeitraum.	both
6	Sp 16	Das hat damit zu tun.	both
→ 7	Sp 15	Ja.	resp
→ 8	Sp 15	Soll ich hier jeden Tag draufsprühen?	init
9	Sp 16	Nein nein nein.	resp

<표 5>는 식물원에서 직원과 손님이 잡초, 해충 해결법과 비료에 관해 이야기 하는 대화의 일부이다. 발화 3과 4 사이, 발화 7과 8 사이처럼 발화 다음에 상호 작용이 시작하는 행위가 등장하는 경우에는 새로운 상호작용이 시작될 가능성이 크기 때문에 대화 분절점으로 적절한 경우가 많다.

다음으로 세 번째 대화 정보는 인접 화행쌍 adjacent speech act pair인데, 이는 그 자체로 하나의 상호작용이기 때문에 쌍 중간에서 분절하지 않는 것이 바람직하다. 인접 화행쌍은 의사소통의 최소 단위이다(Schegloff 2007). 따라서 인접 화행쌍에 해당하는 두 발화문은 하나의 단위에 포함되므로 이들 사이에서의 분절은 적절하지 않다.

표 6. 독일어 대화 예시 (4) 인접 화행쌍과 대화 분절의 관계

발화 번호	화자	발화문	인접 화행쌍
1	Sp 16	Und eine von diesen kleinen Portionsfläschchen reicht für fünf Liter Wasser.	-
→ 2	Sp 15	Da brauche ich aber so ein Gerät für so.	sv-aa (주장-동의)
→ 3	Sp 16	Ja am besten schon ja ja klar.	aa-b (동의-맞장구)
⇒ 4	Sp 15	Hmhm.	-

<표 6>¹⁴⁾은 <표 5> 대화의 다른 일부분이다. <표 6>에서 두 화자는 식물에 뿌릴 해충약의 양과 스프레이 기계에 관해 이야기한다. 발화 2와 3, 발화 3과 4는 인접 화행쌍에 속하는 발화쌍으로 조건적 관련성 konditionale Relevanz¹⁵⁾을 갖는 서로 밀접한 발화들이다(Schegloff 1968, 1083). 따라서 해당 발화쌍 사이에서는 분절하지 않는 것이 바람직하다.

마지막으로 네 번째와 다섯 번째 대화 정보는 대화에 참여하는 화자와 관련된 다. 화자 유형은 좌우 문맥의 발화문을 누가 발화했는지에 관한 정보이다. 예를 들어 특정 화자가 대화를 이끌어가는 유형의 대화 도메인에서는 해당 화자가 수행한 두 발화문 사이에서는 분절될 확률이 낮다. 이러한 발화문의 화자 정보는 특히 다자간 대화에서 유용할 것으로 보인다. 화자 차례 전환은 대화 참여자 간의 말차례 구조와 더 나아가 대화 구조를 파악할 수 있게 하므로 대화 분절에 필요한 대화 정보이다.

IV. 실험

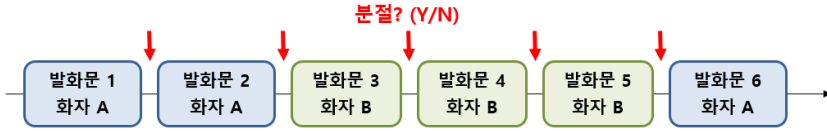
본 연구에서는 대화 분절에 영향을 미치는 언어학적 특징들을 제안하고, 이를 분류 자질로 하는 지도 학습 기반의 기계학습 실험을 수행하였다. 실험은 자바 Java 기반의 기계학습 도구인 웨카 Weka 3.8.4 버전을 활용하였으며, DGD(Datenbank für Gesprochenes Deutsch)¹⁶⁾에서 제공하는 독일 대화 코퍼스 FOLK(Forschungs- und Lehrkorpus gesprochenes Deutsch)의 일부를 학습 데이터로 사용하였다. 학습 데이터에 포함된 발화문은 총 585개이며, 독일어 구매 대화 9개(약국 5개, 식물원 4개)로 구성되어 있다.

14) <표 6>의 인접 화행쌍 예시는 SWBD-DAMSL 화행 유형의 약어로 표현되어 있으며, SWBD-DAMSL 화행 유형의 모든 원어와 약어는 Jurafsky et al.(1997) 또는 <https://web.stanford.edu/~jurafsky/ws97/manual.august1.html>에서 확인할 수 있다. 참고로 ‘sv’는 ‘Statement-opinion’, ‘aa’는 ‘Agree/Accept’, ‘b’는 ‘Acknowledge(Backchannel)’의 약어이다.

15) 조건적 관련성은 둘 이상의 표현 간의 깊은 연결 관계를 나타내는 성질로, Schegloff(1968)은 인접쌍이 갖는 다음의 두 특징을 조건적 관련성으로 설명하였다. 첫째, 인접쌍의 발화들은 단순히 하나가 다른 하나 다음에 등장하는 별개의 두 요소가 아니라, 순서가 있는 쌍에 속한 두 요소라는 점이다. 둘째, 일반적인 발화와는 달리 인접쌍에 속한 발화들은 그중 하나가 대화상에 나타나지 않는다면, 이는 단순히 대화에 없는 정도가 아니라 ‘공식적인 부재(officially absent)’로 여겨진다.

16) https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome

그림 1. 대화 자동 분절 문제 정의



<그림 1>과 같이 본 논문에서 정의한 기계학습 모델은 발화문과 발화문 사이의 모든 지점을 후보 분절점으로 갖고, 각 후보 분절점에 대한 분절 여부를 판단하는 분류 문제를 수행한다.

그림 2. 기계학습 모델의 학습 데이터 예시

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R						
seg	point	context_l	context_r	pause	du	overlap	discourse	discourse	informati	similarity	speech	as	speech	as	conv	pro	conv	pro	adjacent	speaker_l	speaker_r	turn	chn	SEG
1	seg1	Guten	Taj	Hallo!	0	0	0	0	0.05	0.184215	fp	both	both	1	sp1	sp2			1	sp1	sp2	1	0	
2	seg2	Hallo!	Grüße	Sie	0	0	0	1	0.15	0.397494	fp	both	both	1	sp2	sp1			1	sp2	sp1	1	0	
3	seg3	Grüße	Sie	Bitte	schd	0	0	1	0	0.1	0.623222	fp	co	both	both	0	sp1	sp1		0	sp1	sp1	0	1
4	seg4	Bitte	schd	Und	zwar	0	0	0	1	0.15	0.556939	co	ad	both	resp	0	sp1	sp2		0	sp1	sp2	1	1
5	seg5	Und	zwar	Haben	Sie	0	0	1	0	0.15	0.703392	ad	qy	resp	init	0	sp2	sp2		0	sp2	sp2	0	0
6	seg6	Haben	Sie	Ich	war	ai	0	0	0	0.1	0.544094	qy	sd	init	both	0	sp2	sp2		0	sp2	sp2	0	0
7	seg7	Ich	war	ai	ja.	0	0	0	0	0.05	0.212086	sd	b	both	resp	1	sp2	sp1		1	sp2	sp1	1	0
8	seg8	Ja.	Und	habe	0	0	0	0	0.15	0.22479	b	sd	resp	both	0	sp1	sp2		1	sp1	sp2	1	0	
9	seg9	Und	habe	Und	jetzt	0	0	0	0	0.2	0.811786	sd	ba	both	resp	1	sp2	sp1		1	sp2	sp1	1	0
10	seg10	Und	jetzt	Man	hatt	0	0	0	0	0.35	0.876733	ba	sv	resp	both	0	sp1	sp1		0	sp1	sp1	0	0
11	seg11	Man	hatt	ja.	0	0	0	0	0.05	0.252776	sv	b	both	resp	1	sp1	sp2		1	sp1	sp2	1	0	
12	seg12	Ja.	Das	gibt	j	0.22	0	0	0	0.5	0.219324	b	sv	resp	both	0	sp2	sp1		1	sp1	sp1	1	0
13	seg13	Das	gibt	Ach	so.	0	0	0	0	0.05	0.550062	sv	bk	both	resp	1	sp1	sp2		1	sp1	sp2	1	0
14	seg14	Ach	so.	Dann	mu:	0	0	0	0	0.2	0.517379	bk	sv	resp	both	0	sp2	sp1		1	sp1	sp1	1	0

<그림 2>는 기계학습 모델의 학습을 위해 사용하는 학습 데이터의 예시이다. 학습 데이터는 크게 2가지 부분으로 구성되는데, 각 후보 분절점의 유형(레이블/클래스)과 이를 학습하기 위해 사용하는 자질값으로 구성된다. 다시 말해 학습 데이터는 각 후보 분절점의 분절 여부와 분절 여부 분류에 활용하는 언어학적 자질들로 구성된다. 분류 레이블의 경우, 본 논문에서는 지도 학습 supervised learning 기반의 기계학습 실험을 설계하였기 때문에 학습 데이터에 각 후보 분절점의 분절 여부가 레이블로 부착되어 있다. 분류 자질의 경우, 화행 유형은 직접 수작업으로 태깅하였고 그 외의 자질값은 자동으로 부착하였다.

기계학습 알고리즘은 랜덤 포레스트 random forest를 사용하였고, 성능 평가는 10겹 교차 검증 10-fold cross validation 방식을 활용하였다. 10겹 교차 검증은 전체 학습 코퍼스를 10등분 해서 9개를 이용하여 학습하고, 그다음 학습된 기계학습 모델로 나머지 1개의 평가 데이터를 처리하고 그 성능을 평가하는 방식이다. 이 과정을 10개의 세트에 반복적으로 수행하고 그 결과를 모두 합산하

여 산출한 평균값이 최종적인 평가 결과값이다. 평가 지표는 정확도 accuracy를 기반으로 하였다. 전체 후보 분절점 개수 중에 정답과 동일한 레이블로 분류된 후보 분절점의 비율을 계산한다.

표 7. 대화 자동 분절 정확도 (1)

구분	사용한 자질	정확도(%)
실험 1	음성(F1,F2), 어휘(F3,F4), 의미(F5)	78.60
실험 2	음성(F1,F2), 어휘(F3,F4), 의미(F5), 대화(F6~10)	83.56

<표 7>은 제안하는 언어학적 자질들을 사용하여 대화를 자동 분절하는 기계 학습 모델의 성능을 나타낸 것이다. 실험 결과 <표 7>에 따르면 본 논문에서 제안한 대화 분절을 위한 언어학적 요소들을 모두 적용하면 83.56%의 자동 분절 성능을 보였다. 기존 연구들에서 주로 활용한 음성, 어휘, 의미 자질만을 적용했을 때에 비해, 대화 자질을 추가하면 정확도가 약 5%p 향상하는 것을 알 수 있었다¹⁷⁾.

표 8. 대화 자동 분절 정확도 (2)

구분	사용한 자질	정확도(%)
실험 2	음성(F1,F2), 어휘(F3,F4), 의미(F5), 대화(F6~10)	83.56
실험 2-1	음성(F1,F2), 어휘(F3,F4), 의미(F5), 대화(F7,F8,F9,F10) (=실험 2 - F6(화행 유형))	79.05 (-4.51)
실험 2-2	음성(F1,F2), 어휘(F3,F4), 의미(F5), 대화(F6,F8,F9,F10) (=실험 2 - F7(대화 진행 정도))	83.11 (-0.45)
실험 2-3	음성(F1,F2), 어휘(F3,F4), 의미(F5), 대화(F6,F7,F9,F10) (=실험 2 - F8(인접 화행쌍))	83.33 (-0.23)
실험 2-4	음성(F1,F2), 어휘(F3,F4), 의미(F5), 대화(F6,F7,F8,F9) (=실험 2 - F9(화자 유형))	82.21 (-1.35)
실험 2-5	음성(F1,F2), 어휘(F3,F4), 의미(F5), 대화(F6,F7,F8,F10) (=실험 2 - F10(차례 전환))	82.66 (-0.9)

17) 대화 자질 5개에 대한 T-test 분석 결과, 화행 유형(p=0.0001), 인접 화행쌍(p=1.86E-09), 화자 유형(p=0.0028)은 통계적으로 유의미하게 성능을 향상시켰지만, 의사소통 진행 정도(p=0.59), 차례 전환(p=0.63)은 통계적으로 유의미한 성능 향상을 가져오지는 못했다.

<표 8>은 본 논문에서 강조하는 대화적 자질들의 효과를 살펴보기 위해 진행한 실험 결과를 정리한 것이다. 실험 2-1부터 실험 2-5는 앞서 Ⅲ장에서 제안한 대화적 자질 5개를 각각 하나씩 돌아가며 제거한 실험의 결과이다. <표 8>을 보면 제안하는 대화 자질 각각을 하나씩 제거하면 이를 포함하여 대화를 분절할 때에 비해 정확도가 떨어지는 것을 알 수 있다. 따라서 이것들이 대화 자동 분절에 유의미한 자질이라고 볼 수 있다. 특히 대화 자질 중에서도 분절 후보점 좌우 발화의 화행 유형(F6)은 4.51%p로 가장 큰 성능 변화를 보였다.

표 9. 자질별 영향력 순위

순위	자질 유형	자질 번호	자질
1	대화	F6	화행 유형
2	대화	F9	화자 유형
3	대화	F8	인접 화행쌍
4	대화	F7	의사소통 진행 정도
5	음성	F2	중첩 여부
6	어휘	F3	단서 단어
7	대화	F10	차례 전환
8	어휘	F4	정보량
9	의미	F5	문장 유사도
10	음성	F1	휴지 길이

<표 9>는 웨카에서 제공하는 ‘InfoGainAttributeEval’ 기능을 활용하여 분류 자질들의 영향력 순위를 분석한 결과이다¹⁸⁾. 그 결과 대화 자질들이 대부분 높은 순위를 보이는 것을 확인할 수 있었다. 음성 자질 중에서는 중첩(F2)이 높은 영향을 보였다.

18) <표 8>의 세부 실험들은 각 대화 자질을 하나씩 제거하면서 해당 자질이 유의미한 자질인지 확인하기 위해 설계되었다. <표 8>에 제시된 실험 2-1부터 실험 2-5의 각 자동 분절 정확도와 성능 하락의 폭은 제거된 대화 자질 이외의 나머지 9개의 자질들과 그 조합에 의해 도출된다. 반면 <표 9>에 나타난 자질별 영향력은 전체 10개의 자질 조합에서 어떤 자질이 데이터를 잘 구분하고 변별력이 높은 자질인지 계산하는 정보 이득(Information gain) 값에 의해 결정된다. 따라서 <표 8>에서 간접적으로 유추할 수 있는 각 대화 자질의 영향력과 <표 9>의 자질별 영향력 순위는 다소 상이할 수 있다.

이 점을 지적해주신 익명의 심사위원께 감사드립니다.

반면 문장 유사도(F5)는 영향력 순위가 다소 낮았다. 이는 임베딩 모델을 학습할 때 사용한 데이터와 임베딩 모델을 적용하여 임베딩 벡터를 추출한 데이터가 다르기 때문으로 판단된다. 앞서 언급한 바와 같이 본 실험에서 사용한 사전 학습 임베딩 모델은 문어체에 해당하는 위키피디아로 학습되었다. 반면 본 실험의 분석 대상은 구어체적 특징을 띠는 대화이다. 임베딩 모델을 구축하는 데에 사용한 데이터 유형인 텍스트와 구축된 임베딩 모델에 적용한 데이터 유형인 대화가 갖는 어휘적, 통사적 차이로 인해 임베딩 모델의 영향력이 낮았던 것으로 보인다.

V. 결론 및 향후 연구 계획

본 논문은 대화 요약을 위해 대화를 작은 단위로 분절할 때 어떤 언어학적 정보들이 의미 있는지 탐색하였다. 그리고 이를 검증하기 위해 딥러닝 방법론을 활용하여 기계학습 실험을 설계하고 수행하였다. 그 결과 제안한 대화 분절을 위한 언어학적 요소들을 모두 적용하면 83.56%의 자동 분절 성능을 보였다. 특히 본 논문에서 강조한 대화 자질을 추가하면 정확도가 약 5%p 향상하는 것을 알 수 있었다. 이를 바탕으로 기존 연구에서 이루어지던 단어 기반의 대화 분절을 통해 대화를 주제 단위로 분절하는 것보다 대화 자질을 비롯한 다양한 언어학적 자질을 이용하여 대화를 상호작용 단위로 분절하는 것이 효과적임을 확인하였다.

향후에는 본 연구에서 제안한 방법론을 좀 더 다양한 환경에서 평가해보기 위해 현재 수집한 독일어 전화 대화, 상점 대화, 잡담 대화 등을 전처리하여 재실험에 활용할 계획이다. 또한 ELMo와 BERT 같은 문맥 기반의 임베딩 모델이나 Sent2Vec 같은 다른 문장 임베딩 방법론을 활용하여 발화문이 더욱 적절한 임베딩 벡터로 표현될 수 있도록 보완할 예정이다. 뿐만 아니라 각 후보 분절점을 하나씩 분류하는 기존의 실험 방식 대신에 연속되는 후보 분절점을 모델의 입력으로 넣어 각 분절점의 분절 여부를 한꺼번에 판단하는 시퀀스 레이블링 sequence labeling 방식으로 분절 방법론이나 전체 대화를 한꺼번에 입력으로 넣어 분석하는 전역적 global 분류 방식을 검토해볼 것이다. 마지막으로 본 연구의

결과물이 실제 다양한 언어 처리 분야에 적용 가능한지 검증하기 위해, 논문에서 제안한 대화 분절 방법론을 실제 대화 요약이나 정보 추출 태스크에 적용해 볼 계획이다.

향후 연구를 통해 본 연구를 보완 및 발전시킨다면 본 논문에서 제안하는 화자 간 상호작용 단위의 대화 분절은 대화 요약뿐만 아니라 다양한 연구에 활용될 것으로 기대된다. 먼저 N개의 이전 발화로 문맥 Kontext 범위를 고정하는 기존 연구의 방식에서 벗어나, 각 발화에 맞는 문맥 범위를 고려하여 적절한 응답을 생성하는 데에 적용할 수 있다. 뿐만 아니라 본 논문에서 분절하는 단위는 대화의 주제와 더불어 대화의 흐름 측면에서도 관련 깊은 시퀀스이기 때문에 외국어 교육에 활용한다면 외국어 학습자의 자연스러운 대화 구사 능력을 향상하는 데에 효과적일 것으로 기대된다.

참고문헌

- 강창우(2004): 화행 유형의 하위분류 가능성과 그 문제점. 한국독어학회, 독어학, 9, 195-215.
- 강창우(2007): 언어표현과 발화의도의 상관관계. 서울대학교 인문학연구원, 인문논총, 57, 183-209.
- 구영은(2018): 독일어 대화체 문장의 화행 분석과 화행 자동분류에 대한 연구. 석사학위논문, 성균관대학교, 서울.
- 구영은 외(2019): 한국어 수업 대화의 화행 분석과 화행 자동분류를 위한 언어학적 기반 연구. 한국정보과학회, 정보과학회논문지, 45(8), 807-815.
- 구영은/홍문표(2019): 범도메인 대화에서의 화행 자동분류. 한국독어학회, 독어학, 39(1), 25-48.
- 박용익(1994a): 수업대화의 기능단계와 대화이동 연속체원형. 한국텍스트언어학회, 텍스트언어학, 2, 347-374.
- 박용익(1994b): 대화문법이론. 한국독어독문학회, 독일문학, 52, 373-402.
- 서종훈(2013): ‘쉼(pause)’에 대한 국어교육적 고찰. 언어과학회, 언어과학연구, 64, 179-204.
- 스튜어트, 존 외(2015): 소통: 협력적인 의사소통의 방법(서현석 외 역). 서울: 커뮤니케이션북스.
- 조국현(2004): 화행과 청행. 한국독어독문학회, 독일문학, 92, 385-405.

- Arguello, J./Rosé, C.(2006): Topic-segmentation of dialogue. In Proceedings of the Analyzing Conversations in Text and Speech, 42-49.
- Bojanowski, P. et al.(2017): Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5, 135-146.
- Clift, R.(2016): Conversation analysis. Cambridge University Press, United Kingdom.
- Devlin, J. et al.(2019): BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 4171-4186.
- Dharanipragada, S. et al.(2000): Statistical methods for topic segmentation. In Proceedings of the Sixth International Conference on Spoken Language Processing, 516-519.
- Franke, W.(1990): Elementare Dialogstrukturen, De Gruyter, Germany.
- Galley, M. et al.(2003): Discourse segmentation of multi-party conversation. In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, 562-569.
- Grosz, B./Sidner, C. L.(1986): Attention, intentions, and the structure of discourse. Computational linguistics, 12(3), 175-204.
- Hearst, M. A.(1997): Text Tiling: Segmenting text into multi-paragraph subtopic passages. Computational linguistics, 23(1), 33-64.
- Hochreiter, S./Schmidhuber, J.(1997): Long short-term memory. Neural computation, 9(8), 1735-1780.
- Hsueh, P. Y./Moore, J. D.(2007): Combining multiple knowledge sources for dialogue segmentation in multimedia archives. In Proceedings of the 45th annual meeting of the association of computational linguistics, 1016-1023.
- Jurafsky, D. et al.(1997): Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, Institute of Cognitive Science Technical Report.
- König(1997): Zur Bedeutung von Modalpartikeln im Deutschen: Ein Neuanatz im Rahmen der Relevanztheorie. Germanistische Linguistik, 136, 57-75.
- Koshorek, O. et al.(2018): Text Segmentation as a Supervised Learning Task. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), New Orleans, Louisiana, 469-473.
- Lukasik, M. et al.(2020): Text Segmentation by Cross Segment Attention. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing

- (EMNLP), 4707-4716.
- Mikolov, T. et al.(2013): Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 1-12.
- Murray, G. et al.(2007): Automatic segmentation and summarization of meeting speech. In Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics-Human Language Technologies (NAACL-HLT), 9-10.
- Oard, D. W. et al.(2004): Building an information retrieval test collection for spontaneous conversational speech. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, 41-48.
- Pennington, J. et al.(2014): Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 1532-1543.
- Peters, M. et al.(2018): Deep Contextualized Word Representations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2227-2237.
- Riedl, M./Biemann, C.(2012): TopicTiling: a text segmentation algorithm based on lda. In Proceedings of ACL 2012 Student Research Workshop, 37-42.
- Sacks, H. et al.(1978): A simplest systematics for the organization of turn taking for conversation. In J. Schenkein (Ed.), Studies in the organization of conversational interaction. Academic Press, USA, 7-55.
- Schegloff, E. A.(1968): Sequencing in Conversational Openings. American Anthropologist, 70(6), 1075-1095.
- Schegloff, E. A.(2007): Sequence organization in interaction: A primer in conversation analysis, Vol 1, Cambridge university press, United Kingdom.
- Song, Y. et al.(2016): Dialogue Session Segmentation by Embedding-Enhanced TextTiling. Interspeech 2016, 2706-2710.
- Spiegel, C./Spranz-Fogasy, T.(2001): Aufbau und Abfolge von Gesprächsphasen. In K. Brinker, et al. (Eds.), Text- und Gesprächslinguistik. De Gruyter, Berlin/New York, Germany.
- Spranz-Fogasy, T.(2008): The Structure of the Doctor-patient-communication. 의사-환자의 대화행위 구조 분석. 의료커뮤니케이션, 3(1), 39-47.
- Stede, M.(2002): DiMLex: A Lexical Approach to Discourse Markers. In A. Lenci/V. Di Tomaso (Eds.), Exploring the Lexicon - Theory and Computation. Edizioni dell'Orso,

Alessandria, Italy, 1-15.

Vaswani, A. et al.(2017): Attention is all you need. In Proceedings of the Advances in neural information processing systems, 5998-6008.

Wang, Y. et al.(2018): Toward Fast and Accurate Neural Discourse Segmentation. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), Brussels, Belgium, 962-967.

Xing, L./Carenini, G.(2021): Improving Unsupervised Dialogue Topic Segmentation with Utterance-Pair Coherence Scoring. arXiv preprint arXiv:2106.06719, 1-11.

Zechner, K./Waibel, A.(2000): DiaSumm: flexible summarization of spontaneous dialogues in unrestricted domains. In Proceedings of the 18th conference on Computational linguistics, 968-974.

Zusammenfassung

Automatische Segmentierung von deutschem Dialog

– Für die automatische Zusammenfassung eines deutschen Dialogs

KOO Youngeun, HONG Munpyo (Sungkyunkwan Univ.)

Ein Dialog besteht aus verschiedenen Interaktionen zwischen Sprechern. Zu erfassen, welche Interaktionen in einem Dialog stattgefunden haben, ist einer der wichtigsten Aspekte beim Verstehen eines Dialogs. In dieser Arbeit wird versucht, Dialoge in Interaktionseinheiten zu unterteilen, um deutsche Dialoge zusammenzufassen. Eine Dialogsegmentierung ist der Prozess, einen Dialog in kleinere Einheiten aufzuteilen. Die Dialogsegmentierung wird in vielen Studien zur Dialogmodellierung, wie z. B. in der Dialogzusammenfassung und im Information Retrieval, verwendet, weil sie ermöglicht, die Struktur und den Inhalt von Dialogen zu erfassen, indem verwandte Sätze miteinander verbunden werden.

Diese Arbeit betont, wie wichtig Dialoginformationen bei der Dialogsegmentierung für die Dialogzusammenfassung sind, und schlägt verschiedene linguistische Merkmale vor, um den Dialog in für die Dialogzusammenfassung geeigneten Einheiten zu segmentieren.

Darüber hinaus wurde ein Experiment mittels maschinellem Lernen durchgeführt, das Dialoge, mithilfe einer Deep-Learning-Methodik einschließlich Satzeinbettung, automatisch segmentiert. Das Experiment zeigt, dass der vorgestellte Ansatz, der alle vorgeschlagenen linguistischen Merkmale verwendet, für die automatische Dialogsegmentierung eine Korrektheit von 83,56% aufweist. Insbesondere wird die Korrektheit um etwa 5%p verbessert, wenn die Dialoginformationen, die diese Arbeit betont, als Merkmale des maschinellen Lernens hinzugefügt werden. Durch diese Arbeit wurde festgestellt, dass Menschen verschiedene linguistische Merkmale, insbesondere Dialoginformationen, verwenden, um zu verstehen, wie aufeinanderfolgende Äußerungen im Dialog als Interaktionseinheit miteinander verbunden sind.

Schlüsselbegriffe: Dialogsegmentierung, Dialogzusammenfassung, Dialogstruktur, Maschinelles Lernen, Einbettung

필자 이메일 주소 및 직위: sarah8835@skku.edu 학생

skkhmp@skku.edu 교수

논문투고일: 2021.10.29 | 논문심사일: 2021.11.30 | 게재확정일: 2021.12.15