

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명
<div>▶ Intro</div> <div>• 학습열기</div> <div>• 학습목표</div> <div>▶ 학습하기</div> <div>• 워드임베딩 개요</div> <div>• </div> <div>• </div> <div>▶ 적용하기</div> <div>▶ 정리하기</div> <div>• 문제풀기</div> <div>• 학습정리</div>	<div>2) 워드 임베딩의 종류</div> <div>(1) 워드 임베딩의 종류</div> <div>▪ 빈도(Frequency) 기반 워드 임베딩</div> <div>다수 문서에 등장하는 각 단어들의 빈도를 행렬로 표현하거나 가중치를 부여하여 단어의 중요도나 문서간 유사도를 측정하기 위한 임베딩</div> <div>(DTM, TF-IDF)</div> <div>▪ 토픽(Topic) 기반 워드 임베딩</div> <div>주어진 문서에 잠재된 주제(latent topic)를 추론(inference)하기 위한 임베딩</div> <div>(LDA, Latent Dirichlet Allocation)</div> <div>▪ 예측(Prediction) 기반 워드 임베딩</div> <div>주어진 문장이나 단어의 다음 단어 예측, 주변 단어에 대한 예측, Masking 된 단어의 예측등을 위한 임베딩</div> <div>(Word2Vec, FastText, BERT, ELMo, GPT)</div>			<div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.</div> <div>Ex) #번호와 함께 작성</div> <div>#1 버튼을 클릭하면 내용을 제시해 주세요.</div>
				<div>용어사전</div>
				<div>학습 화면에서 새롭게 제시한 용어 중 의미를 표시해야 하는 것이 있다면 작성해 주세요.</div>
내레이션을	<div>학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모션이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.</div>			

1

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명
<div>▶ Intro<ul style="list-style-type: none"><li>• 학습열기</li><li>• 학습목표</li></ul></div> <div>▶ 학습하기<ul style="list-style-type: none"><li>• 워드임베딩 개요</li></ul></div> <div><ul style="list-style-type: none"><li>• </li><li>• </li></ul></div> <div>▶ 적용하기</div> <div>▶ 정리하기<ul style="list-style-type: none"><li>• 문제풀기</li><li>• 학습정리</li></ul></div>		<div>(2) 단어/문장 기반 워드 임베딩</div> <div><ul style="list-style-type: none"><li>▪ 단어 기반 워드 임베딩</li></ul></div> <div>단어(Word) 기반으로 임베딩 수행</div> <div>문맥을 고려하지 않은 상태에서 워드 임베딩을 수행</div> <div>Word2Vec, FastText, Glove 등의 임베딩 방법</div> <div>서로 다른 문맥의 동음이의어가 동일하게 임베딩되는 문제점</div> <div><div><div>Embedding</div><div><div>Static Word Embedding</div><div><div>Word2Vec</div><div>FastText</div><div>Glove</div></div></div><div><div>Contextualized Word Embedding</div><div><div>ELMo</div><div>GPT</div><div>BERT</div><div>GPT2</div></div></div></div><div>단어(Word) 기반</div><div>문장(Sentence) 기반</div></div> <div>출 처 : <a href="https://towardsdatascience.com/from-pre-trained-word-embeddings-to-pre-trained-language-models-focus-on-bert-343815627598">https://towardsdatascience.com/from-pre-trained-word-embeddings-to-pre-trained-language-models-focus-on-bert-343815627598</a></div>		<div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.</div> <div>Ex) #번호와 함께 작성</div> <div>#1 버튼을 클릭하면 내용을 제시해 주세요.</div>
내레이션을		학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모선이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.		<div>용어사전</div> <div>학습 화면에서 새롭게 제시한 용어 중 의미를 표시해야 하는 것이 있다면 작성해 주세요.</div>

2

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명
<div>▶ Intro<ul style="list-style-type: none"><li>• 학습열기</li><li>• 학습목표</li></ul></div> <div>▶ 학습하기<ul style="list-style-type: none"><li>• 워드임베딩 개요</li></ul></div> <div>• </div> <div>• </div> <div>▶ 적용하기</div> <div>▶ 정리하기<ul style="list-style-type: none"><li>• 문제풀기</li><li>• 학습정리</li></ul></div>		<div>▪ 문장 기반 워드 임베딩</div> <div>문맥을 고려하여 문장(Sentence) 기반으로 임베딩을 수행</div> <div>ELMo, GTP, BERT, GTP2 등의 임베딩 방법</div> <div>문장(Sentence) 기반으로 임베딩을 수행하여 언어 모델(Language Model)로도 불리움</div> <div><div><div>Embedding</div><div><div>Static Word Embedding</div><div>Word2VecFastTextGlove</div><div>단어(Word) 기반</div></div><div><div>Language Model</div><div>Contextualized Word Embedding</div><div>ELMoGPTBERTGPT2</div><div>문장(Sentence) 기반</div></div></div></div> <div>출 처 : <a href="https://towardsdatascience.com/from-pre-trained-word-embeddings-to-pre-trained-language-models-focus-on-bert-343815627598">https://towardsdatascience.com/from-pre-trained-word-embeddings-to-pre-trained-language-models-focus-on-bert-343815627598</a></div>		<div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.</div> <div>Ex) #번호와 함께 작성</div> <div>#1 버튼을 클릭하면 내용을 제시해 주세요.</div>
내레이션을		학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모션이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.		<div>용어사전</div> <div>학습 화면에서 새롭게 제시한 용어 중 의미를 표시해야 하는 것이 있다면 작성해 주세요.</div>
3				

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명
<div>▶ Intro</div> <div>• 학습열기</div> <div>• 학습목표</div> <div>▶ 학습하기</div> <div>• 빈도 기반 워드임베딩</div> <div>• </div> <div>• </div> <div>▶ 적용하기</div> <div>▶ 정리하기</div> <div>• 문제풀기</div> <div>• 학습정리</div>	<div>간지</div> <div>빈도 기반 워드임베딩</div>			(기입하지 않습니다.)
				용어사전
				(기입하지 않습니다.)
내레이션	(기입하지 않습니다.)			

4

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명																																																																
<div>▶ Intro<ul style="list-style-type: none"><li>• 학습열기</li><li>• 학습목표</li></ul></div> <div>▶ 학습하기<ul style="list-style-type: none"><li>• 빈도 기반 워드임베딩</li><li>• </li><li>• </li></ul></div> <div>▶ 적용하기</div> <div>▶ 정리하기<ul style="list-style-type: none"><li>• 문제풀기</li><li>• 학습정리</li></ul></div>		<div>2. 빈도(Frequency)기반 워드 임베딩</div> <div>(1) DTM(Document Term Matrix)</div> <div>▪ DTM(Document Term Matrix)이란? 문서는 단어의 집합으로 이루어져 있다는 개념 적용 다수의 문서(Document)에 등장하는 각 단어(Term)들의 빈도를 행렬 구조로 표현 각 문서에 주로 사용되는 단어 파악 가능 문서 간 유사한 단어가 동일하게 사용된 경우 유사한 문서로 판단 가능</div> <div>▪ DTM 예시</div> <table><tr><th>DTM</th><th>단어 #1</th><th>단어 #2</th><th>단어 #3</th><th>단어 #4</th><th>단어 #5</th><th>...</th><th>단어 #N</th></tr><tr><td>문서 #1</td><td>5</td><td>0</td><td>2</td><td>1</td><td>2</td><td>...</td><td>0</td></tr><tr><td>문서 #2</td><td>1</td><td>2</td><td>5</td><td>2</td><td>3</td><td>...</td><td>1</td></tr><tr><td>문서 #3</td><td>3</td><td>5</td><td>0</td><td>2</td><td>4</td><td>...</td><td>2</td></tr><tr><td>문서 #4</td><td>2</td><td>5</td><td>0</td><td>3</td><td>3</td><td>...</td><td>1</td></tr><tr><td>문서 #5</td><td>1</td><td>0</td><td>1</td><td>2</td><td>1</td><td>...</td><td>2</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>문서 #N</td><td>2</td><td>0</td><td>3</td><td>2</td><td>1</td><td>...</td><td>0</td></tr></table>		DTM	단어 #1	단어 #2	단어 #3	단어 #4	단어 #5	...	단어 #N	문서 #1	5	0	2	1	2	...	0	문서 #2	1	2	5	2	3	...	1	문서 #3	3	5	0	2	4	...	2	문서 #4	2	5	0	3	3	...	1	문서 #5	1	0	1	2	1	...	2	...	...	...	...	...	...	...	...	문서 #N	2	0	3	2	1	...	0	<div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.</div> <div>Ex) #번호와 함께 작성</div> <div>#1 버튼을 클릭하면 내용을 제시해 주세요.</div>
DTM	단어 #1	단어 #2	단어 #3	단어 #4	단어 #5	...	단어 #N																																																													
문서 #1	5	0	2	1	2	...	0																																																													
문서 #2	1	2	5	2	3	...	1																																																													
문서 #3	3	5	0	2	4	...	2																																																													
문서 #4	2	5	0	3	3	...	1																																																													
문서 #5	1	0	1	2	1	...	2																																																													
...	...	...	...	...	...	...	...																																																													
문서 #N	2	0	3	2	1	...	0																																																													
				<div>용어사전</div> <div>학습 화면에서 새롭게 제시한 용어 중 의미를 표시해야 하는 것이 있다면 작성해 주세요.</div>																																																																
내레이션을	학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모선이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.																																																																			

5

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명																																																															
<div><div>▶ Intro<ul style="list-style-type: none"><li>• 학습열기</li><li>• 학습목표</li></ul></div><div>▶ 학습하기<ul style="list-style-type: none"><li>• 빈도 기반 워드임베딩</li><li>• </li><li>• </li></ul></div><div>▶ 적용하기</div><div>▶ 정리하기<ul style="list-style-type: none"><li>• 문제풀기</li><li>• 학습정리</li></ul></div></div> <div><div><div>▪ TDM(Term Document Matrix)이란? DTM과 유사하며, DTM의 역행렬(Transpose) 단어를 기준으로 문서에 사용된 빈도를 행렬 구조로 표현</div><div>▪ TDM 예시<table><tr><th>TDM</th><th>문서 #1</th><th>문서 #2</th><th>문서 #3</th><th>문서 #4</th><th>문서 #5</th><th>...</th><th>문서 #N</th></tr><tr><td>단어 #1</td><td>5</td><td>1</td><td>3</td><td>2</td><td>1</td><td>...</td><td>2</td></tr><tr><td>단어 #2</td><td>0</td><td>2</td><td>5</td><td>5</td><td>0</td><td>...</td><td>0</td></tr><tr><td>단어 #3</td><td>2</td><td>5</td><td>0</td><td>0</td><td>1</td><td>...</td><td>3</td></tr><tr><td>단어 #4</td><td>1</td><td>2</td><td>2</td><td>3</td><td>2</td><td>...</td><td>2</td></tr><tr><td>단어 #5</td><td>2</td><td>3</td><td>4</td><td>3</td><td>1</td><td>...</td><td>1</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>단어 #N</td><td>0</td><td>1</td><td>2</td><td>1</td><td>2</td><td>...</td><td>0</td></tr></table></div><div>▪ DTM과 TDM의 한계 대부분의 값이 0이면서 일부만 의미있는 값으로 표현되는 희소 벡터(Sparse Vector) 단어별 가중치를 부여하지 않음으로써 영어에서 'the'와 같이 중요하지 않은 단어가 여러 문서에 공통으로 포함된 경우 유사한 문서로 판단 가능</div></div></div> <div><div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.  Ex) #번호와 함께 작성  #1 버튼을 클릭하면 내용을 제시해 주세요.</div><div>용어사전 학습 화면에서 새롭게 제시한 용어 중 의미를 표시해야 하는 것이 있다면 작성해 주세요.</div></div>				TDM	문서 #1	문서 #2	문서 #3	문서 #4	문서 #5	...	문서 #N	단어 #1	5	1	3	2	1	...	2	단어 #2	0	2	5	5	0	...	0	단어 #3	2	5	0	0	1	...	3	단어 #4	1	2	2	3	2	...	2	단어 #5	2	3	4	3	1	...	1	...	...	...	...	...	...	...	...	단어 #N	0	1	2	1	2	...	0
TDM	문서 #1	문서 #2	문서 #3	문서 #4	문서 #5	...	문서 #N																																																												
단어 #1	5	1	3	2	1	...	2																																																												
단어 #2	0	2	5	5	0	...	0																																																												
단어 #3	2	5	0	0	1	...	3																																																												
단어 #4	1	2	2	3	2	...	2																																																												
단어 #5	2	3	4	3	1	...	1																																																												
...	...	...	...	...	...	...	...																																																												
단어 #N	0	1	2	1	2	...	0																																																												
학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모선이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.																																																																			
내레이션																																																																			

6

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명
<div>▶ Intro</div> <div>• 학습열기</div> <div>• 학습목표</div> <div>▶ 학습하기</div> <div>• 빈도 기반 워드임베딩</div> <div>• </div> <div>• </div> <div>▶ 적용하기</div> <div>▶ 정리하기</div> <div>• 문제풀기</div> <div>• 학습정리</div>	<div>(2) 단어 빈도-역 문서 빈도(TF-IDF, Term Frequency – Inverse Document Frequency)</div> <div>▪ TF(Term Frequency)란? 특정 문서 내에서 단어가 등장한 빈도로 Document Term Matrix 빈도가 높을수록 해당 단어는 특정 문서에서 중요한 역할을 하는 단어일 가능성이 높음</div> <div>▪ DF(문서 빈도, document frequency)란? 특정 단어가 등장한 문서의 수 등장한 문서의 수가 많으면 여러 문서에 두루 등장하는 단어 - 상투어 등장한 문서의 수가 적으면 일부 문서에만 등장하는 희귀 단어 - 핵심어</div> <div>▪ IDF는 코퍼스 성격에 따라 결정 '원자'라는 단어는 일반적인 문서들 사이에서는 자주 등장하지 않기 때문에 IDF 값이 높아지고 문서의 핵심어가 됨 '원자'에 관한 논문을 모아놓은 코퍼스에서 '원자'라는 단어는 상투어가 됨</div>			<div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.</div> <div>Ex) #번호와 함께 작성</div> <div>#1 버튼을 클릭하면 내용을 제시해 주세요.</div>
				<div>용어사전</div> <div>학습 화면에서 새롭게 제시한 용어 중 의미를 표시해야 하는 것이 있다면 작성해 주세요.</div>
	<div>학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모선이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.</div>			
내 레이션				

7

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명
<div>▶ Intro</div> <div>• 학습열기</div> <div>• 학습목표</div> <div>▶ 학습하기</div> <div>• 빈도 기반 워드임베딩</div> <div>• </div> <div>• </div> <div>▶ 적용하기</div> <div>▶ 정리하기</div> <div>• 문제풀기</div> <div>• 학습정리</div>	<div>▪ IDF(역 문서 빈도, document frequency)란?</div> <div>DF에 반비례 하는 값으로 단어의 희귀한 정도에 대한 표현</div> <div>IDF 계산 방법</div> <div>IDF = 문서수 / DF</div> <div>DF가 0인 경우도 있을 수 있으므로</div> <div>IDF = 문서수 / 1 + DF</div> <div>문서수가 많아지면 IDF가 지나치게 커질 수 있으므로</div> <div>IDF = log(문서수 / 1 + DF)</div>			<div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.</div> <div>Ex) #번호와 함께 작성</div> <div>#1 버튼을 클릭하면 내용을 제시해 주세요.</div>
				용어사전
				학습 화면에서 새롭게 제시한 용어 중 의미를 표시해야 하는 것이 있다면 작성해 주세요.
내레이션을	학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모션이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.			

8



과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명
<div>▶ Intro</div> <div>• 학습열기</div> <div>• 학습목표</div> <div>▶ 학습하기</div> <div>• 빈도 기반 워드임베딩</div> <div>• </div> <div>• </div> <div>▶ 적용하기</div> <div>▶ 정리하기</div> <div>• 문제풀기</div> <div>• 학습정리</div>	<div>▪ TF-IDF(단어 빈도-역 문서 빈도, Term Frequency – Inverse Document Frequency)란?</div> <div>특정 문서 내에서 단어가 등장한 단순 빈도와 희귀 단어 등장 빈도를 동시에 고려한 접근 방법</div> <div>특정 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 가중치</div> <div>문서 간 유사도 계산에서도 많이 사용</div> <div>TF-IDF 계산 방법</div> <div>TF-IDF = TF X IDF</div> <div>* IDF = log(문서수 / 1 + DF)</div>			<div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.</div> <div>Ex) #번호와 함께 작성</div> <div>#1 버튼을 클릭하면 내용을 제시해 주세요.</div>
				<div>용어사전</div>
				<div>학습 화면에서 새롭게 제시한 용어 중 의미를 표시해야 하는 것이 있다면 작성해 주세요.</div>
내레이션	<div>학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모선이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.</div>			

9

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명
<div>▶ Intro</div> <div>• 학습열기</div> <div>• 학습목표</div> <div>▶ 학습하기</div> <div>• 빈도 기반 워드임베딩</div> <div>• </div> <div>• </div> <div>▶ 적용하기</div> <div>▶ 정리하기</div> <div>• 문제풀기</div> <div>• 학습정리</div>	<div>▪ Python에서 TF-IDF 적용</div> <div>sklearn패키지의 TfidfVectorizer를 사용</div> <div><pre>from sklearn.feature_extraction.text import TfidfVectorizer corpus = ["나는 학교에 간다",           "나는 친구를 만난다",           "나는 영화보러 간다",           "영화가 재밌다"]  tfidf = TfidfVectorizer().fit(corpus) print(tfidf.transform(corpus).toarray()) print(tfidf.vocabulary_)  [[[0.55349232 0.44809973 0.          0.          0.          0.       0.          0.70203482]   [0.          0.41137791 0.64450299 0.          0.          0.       0.          0.64450299 0.          ]   [0.55349232 0.44809973 0.          0.          0.70203482 0.       0.          0.          ]   [0.          0.          0.          0.70710678 0.          0.70710678       0.          0.          ]]] {'나는': 1, '학교에': 7, '간다': 0, '친구를': 6, '만난다': 2, '영화보러': 4, '영화가': 3, '재밌다': 5}</pre></div>			<div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.</div> <div>Ex) #번호와 함께 작성</div> <div>#1 버튼을 클릭하면 내용을 제시해 주세요.</div>
				용어사전
				학습 화면에서 새롭게 제시한 용어 중 의미를 표시해야 하는 것이 있다면 작성해 주세요.
내레이션을	학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모선이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.			

10

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명
<div>▶ Intro</div> <div>• 학습열기</div> <div>• 학습목표</div> <div>▶ 학습하기</div> <div>• 토픽 기반 워드임베딩</div> <div>• </div> <div>• </div> <div>▶ 적용하기</div> <div>▶ 정리하기</div> <div>• 문제풀기</div> <div>• 학습정리</div>	<div>간지</div> <div>토픽 기반 워드 임베딩</div>			(기입하지 않습니다.)
				용어사전
				(기입하지 않습니다.)
내레이션	(기입하지 않습니다.)			

11

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명
<div><div><div>▶ Intro</div><div><div>• 학습열기</div><div>• 학습목표</div></div><div>▶ 학습하기</div><div><div>• 토픽 기반 워드임베딩</div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div>&lt;</div></div></div>				

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명
<div>▶ Intro<ul style="list-style-type: none"><li>• 학습열기</li><li>• 학습목표</li></ul></div> <div>▶ 학습하기<ul style="list-style-type: none"><li>• 토픽 기반 워드임베딩</li></ul></div> <div>• </div> <div>• </div> <div>▶ 적용하기</div> <div>▶ 정리하기<ul style="list-style-type: none"><li>• 문제풀기</li><li>• 학습정리</li></ul></div>		<div>(1) LDA(Latent Dirichlet Allocation) 수행 과정</div> <div><ul style="list-style-type: none"><li>▪ 코퍼스 내 문서로 부터 단어 추출</li></ul><div>코퍼스 내 문서로 부터 단어를 추출</div><div>빈도 수 상위 단어 기준으로 추출</div></div> <div><div><div>문서 #1</div><div>→</div><div><div>brain</div><div>computer</div><div>data</div><div>dna</div><div>evolve</div><div>gene</div><div>genetic</div><div>life</div><div>nerve</div><div>neuron</div><div>number</div><div>organism</div><div>...</div></div></div><div>코퍼스(Corpus)</div></div> <div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.</div> <div>Ex) #번호와 함께 작성</div> <div>#1 버튼을 클릭하면 내용을 제시해 주세요.</div>		
내레이션을		<div>학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모선이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.</div>		

13

학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모선이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명																																																																											
<div>▶ Intro</div> <div>• 학습열기</div> <div>• 학습목표</div> <div>▶ 학습하기</div> <div>• 토픽 기반 워드임베딩</div> <div>• </div> <div>• </div> <div>▶ 적용하기</div> <div>▶ 정리하기</div> <div>• 문제풀기</div> <div>• 학습정리</div>	<div>▪ 단어의 토픽별 분포 확률</div> <div>각 단어에 대해 토픽별 확률을 계산</div> <div>각 토픽별 계산된 값의 합은 1</div> <table><tr><th>Topic</th><th>Topic #1</th><th>Topic #2</th><th>Topic #3</th><th>Topic #4</th></tr><tr><td>brain</td><td>0.000</td><td>0.000</td><td>0.040</td><td>0.001</td></tr><tr><td>computer</td><td>0.000</td><td>0.000</td><td>0.000</td><td>0.010</td></tr><tr><td>data</td><td>0.000</td><td>0.000</td><td>0.000</td><td>0.020</td></tr><tr><td>dna</td><td>0.020</td><td>0.007</td><td>0.002</td><td>0.000</td></tr><tr><td>evolve</td><td>0.009</td><td>0.010</td><td>0.001</td><td>0.000</td></tr><tr><td>gene</td><td>0.040</td><td>0.007</td><td>0.000</td><td>0.001</td></tr><tr><td>genetic</td><td>0.010</td><td>0.008</td><td>0.000</td><td>0.002</td></tr><tr><td>life</td><td>0.005</td><td>0.020</td><td>0.000</td><td>0.000</td></tr><tr><td>nerve</td><td>0.003</td><td>0.001</td><td>0.010</td><td>0.004</td></tr><tr><td>neuron</td><td>0.004</td><td>0.004</td><td>0.020</td><td>0.001</td></tr><tr><td>number</td><td>0.002</td><td>0.002</td><td>0.002</td><td>0.020</td></tr><tr><td>organism</td><td>0.007</td><td>0.010</td><td>0.001</td><td>0.004</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>합계</td><td>1.000</td><td>1.000</td><td>1.000</td><td>1.000</td></tr></table> <div>Topic #1은 gene이라는 단의 등장 확률이 0.04, dna는 0.02, genetic은 0.01</div> <div>Topic #1은'유전자' 관련 주제로 판단 가능</div>			Topic	Topic #1	Topic #2	Topic #3	Topic #4	brain	0.000	0.000	0.040	0.001	computer	0.000	0.000	0.000	0.010	data	0.000	0.000	0.000	0.020	dna	0.020	0.007	0.002	0.000	evolve	0.009	0.010	0.001	0.000	gene	0.040	0.007	0.000	0.001	genetic	0.010	0.008	0.000	0.002	life	0.005	0.020	0.000	0.000	nerve	0.003	0.001	0.010	0.004	neuron	0.004	0.004	0.020	0.001	number	0.002	0.002	0.002	0.020	organism	0.007	0.010	0.001	0.004	...	...	...	...	...	합계	1.000	1.000	1.000	1.000	<div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.</div> <div>Ex) #번호와 함께 작성</div> <div>#1 버튼을 클릭하면 내용을 제시해 주세요.</div>
	Topic	Topic #1	Topic #2	Topic #3	Topic #4																																																																										
	brain	0.000	0.000	0.040	0.001																																																																										
computer	0.000	0.000	0.000	0.010																																																																											
data	0.000	0.000	0.000	0.020																																																																											
dna	0.020	0.007	0.002	0.000																																																																											
evolve	0.009	0.010	0.001	0.000																																																																											
gene	0.040	0.007	0.000	0.001																																																																											
genetic	0.010	0.008	0.000	0.002																																																																											
life	0.005	0.020	0.000	0.000																																																																											
nerve	0.003	0.001	0.010	0.004																																																																											
neuron	0.004	0.004	0.020	0.001																																																																											
number	0.002	0.002	0.002	0.020																																																																											
organism	0.007	0.010	0.001	0.004																																																																											
...	...	...	...	...																																																																											
합계	1.000	1.000	1.000	1.000																																																																											
내레이션을	<div>학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모선이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.</div>			<div>용어사전</div> <div>학습 화면에서 새롭게 제시한 용어 중 의미를 표시해야 하는 것이 있다면 작성해 주세요.</div>																																																																											

14

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명																								
<div>▶ Intro</div> <div>• 학습열기</div> <div>• 학습목표</div> <div>▶ 학습하기</div> <div>• 토픽 기반 워드임베딩</div> <div>• </div> <div>• </div> <div>▶ 적용하기</div> <div>▶ 정리하기</div> <div>• 문제풀기</div> <div>• 학습정리</div>	<div>▪ 문서별 토픽의 구성 분석</div> <div>문서별 토픽이 차지하는 비중 분석</div> <table><tr><th>Docs</th><th>Topic# 1</th><th>Topic# 2</th><th>Topic# 3</th><th>Topic# 4</th><th>합계</th></tr><tr><td>문서 #1</td><td>0.501</td><td>0.289</td><td>0.001</td><td>0.209</td><td>1.000</td></tr><tr><td>문서 #2</td><td>0.051</td><td>0.399</td><td>0.479</td><td>0.071</td><td>1.000</td></tr><tr><td>문서 #3</td><td>0.374</td><td>0.001</td><td>0.604</td><td>0.021</td><td>1.000</td></tr></table> <div>문서 #1을 보면 Topic#1이 차지하는 비중이 높으므로 문서 #1의 주제는 Topic #1일 가능성이 높음(50.1%)</div> <div>Topic #1은 '유전자' 관련 주제이므로 문서#1의 주제는 '유전자'로 판단 가능</div>			Docs	Topic# 1	Topic# 2	Topic# 3	Topic# 4	합계	문서 #1	0.501	0.289	0.001	0.209	1.000	문서 #2	0.051	0.399	0.479	0.071	1.000	문서 #3	0.374	0.001	0.604	0.021	1.000	<div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.</div> <div>Ex) #번호와 함께 작성</div> <div>#1 버튼을 클릭하면 내용을 제시해 주세요.</div>
	Docs	Topic# 1	Topic# 2	Topic# 3	Topic# 4	합계																						
	문서 #1	0.501	0.289	0.001	0.209	1.000																						
문서 #2	0.051	0.399	0.479	0.071	1.000																							
문서 #3	0.374	0.001	0.604	0.021	1.000																							
<div>학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모선이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.</div>			<div>용어사전</div> <div>학습 화면에서 새롭게 제시한 용어 중 의미를 표시해야 하는 것이 있다면 작성해 주세요.</div>																									
내레이션																												

15

과정명	인공지능 자연어 처리	회차명	인공지능 자연어 처리를 위한 워드임베딩 알아보기	화면설명
<div>▶ Intro</div> <div>• 학습열기</div> <div>• 학습목표</div> <div>▶ 학습하기</div> <div>• 토픽 기반 워드임베딩</div> <div>• </div> <div>• </div> <div>▶ 적용하기</div> <div>▶ 정리하기</div> <div>• 문제풀기</div> <div>• 학습정리</div>	<div>▪ Python에서 LDA 적용</div> <div>lda 패키지의 LDA 클래스를 사용</div> <div>lda 패키지 설치</div> <div>pip install lda</div> <div>lda 패키지를 이용한 LDA 적용</div> <div><pre>import lda import lda.datasets X = lda.datasets.load_reuters() vocab = lda.datasets.load_reuters_vocab() titles = lda.datasets.load_reuters_titles() model = lda.LDA(n_topics=20, n_iter=1500, random_state=1) model.fit(X) topic_word = model.topic_word_</pre></div>			<div>학습 화면에 구성된 학습 내용 중 움직임이 있어야 하거나 색깔을 순차적으로 표시하는 등 시간 차를 두어 제시해야 하는 것에 대해 설명을 기입해 주세요.</div> <div>Ex) #번호와 함께 작성</div> <div>#1 버튼을 클릭하면 내용을 제시해 주세요.</div>
				<div>용어사전</div> <div>학습 화면에서 새롭게 제시한 용어 중 의미를 표시해야 하는 것이 있다면 작성해 주세요.</div>
	<div>학습자가 학습 화면에 구성된 내용을 보고 있다고 가정하고, 학습 설명으로 함께 들어야 할 음성 내용을 자세하게 기입해 주세요. (모선이 있을 경우 #번호를 내레이션 앞에 붙임) #1 제목 위계의 순서를 살펴보세요.</div>			
내레이션을				

16