|  |  |
| --- | --- |
| 교육 제목 | 자연어 처리 Embedding |
| 교육 일시 | 11월 30일(화) |
| 교육 장소 | 재택 |
| **교육 내용** | |
| 전체 내용 | * 단어(Words) * 단어는 discrete symbol & categorical value 형태이지만,   우리의 머리속에서는 다르게 동작   * 어휘는 계층적 의미 구조를 지니고 있으며, 이에 따라 단어 사이의 유사성을 지님 * One-hot 인코딩으로 표현된 값은 유사도나 모호성을 표현할 수 었다.   Dense vector로 표현하는 것이 유리   * Word Embedding * 딥러닝의 시대에 들어와 신경망의 이러한 특성을 활용하여 단어를 연속적인 값으로 표현하고자 하는 시도가 이어짐 * 이전에 비해 훌륭한 Dense vector를 얻을수 있게 되어, 단어의 필요한 특징을 잘 표현할 수 있게 되었음 * 유사도 등의 연산에 유리함 * WordNet * 단어의 계층적 구조를 파악할 수 있음 * 동의어 집합(synset)을 구할 수 있음 * 단어 사이의 유사도를 계산할 수 있음   단점   * 특정 도메인(또는 수집데이터)에 특화된 수치를 계산하고 싶을 때 * 신조어나 사전에 등록되지 않은 단어에는 약하기 때문에 Data-driven Methods가 필요 * Data-driven Methods * Local Representation   해당 단어 그 자체만 보고, 특정값을 맵핑하여 단어를 표현하는 방법 = descrete representation   * Distrubuted Representation   그 단어를 표현하고자 주변을 참고하여 단어를 표현하는 방법  = continuous representation |