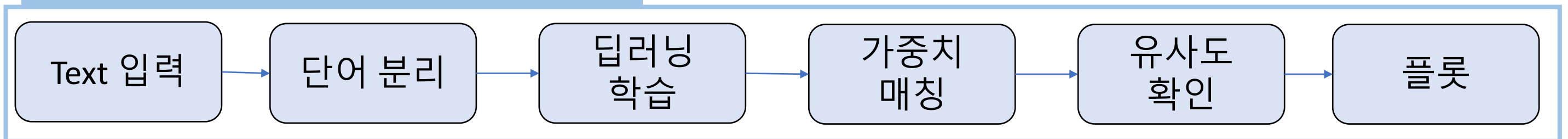


딥러닝 기반 Word embedding

지도교수 : 이인성 교수님

참여학생: 2015039028 이주호
2014039058 이용욱

1. 연구내용



2. 연구원리 및 구조

< 함수1. WordInput >

- Text file에서 단어 및 문장 별로 분리

< 함수2. t_data_from_x_data >

- Skip-gram방식에 맞게 입,출력 값을 불러옴

< 함수3. twolayernet >

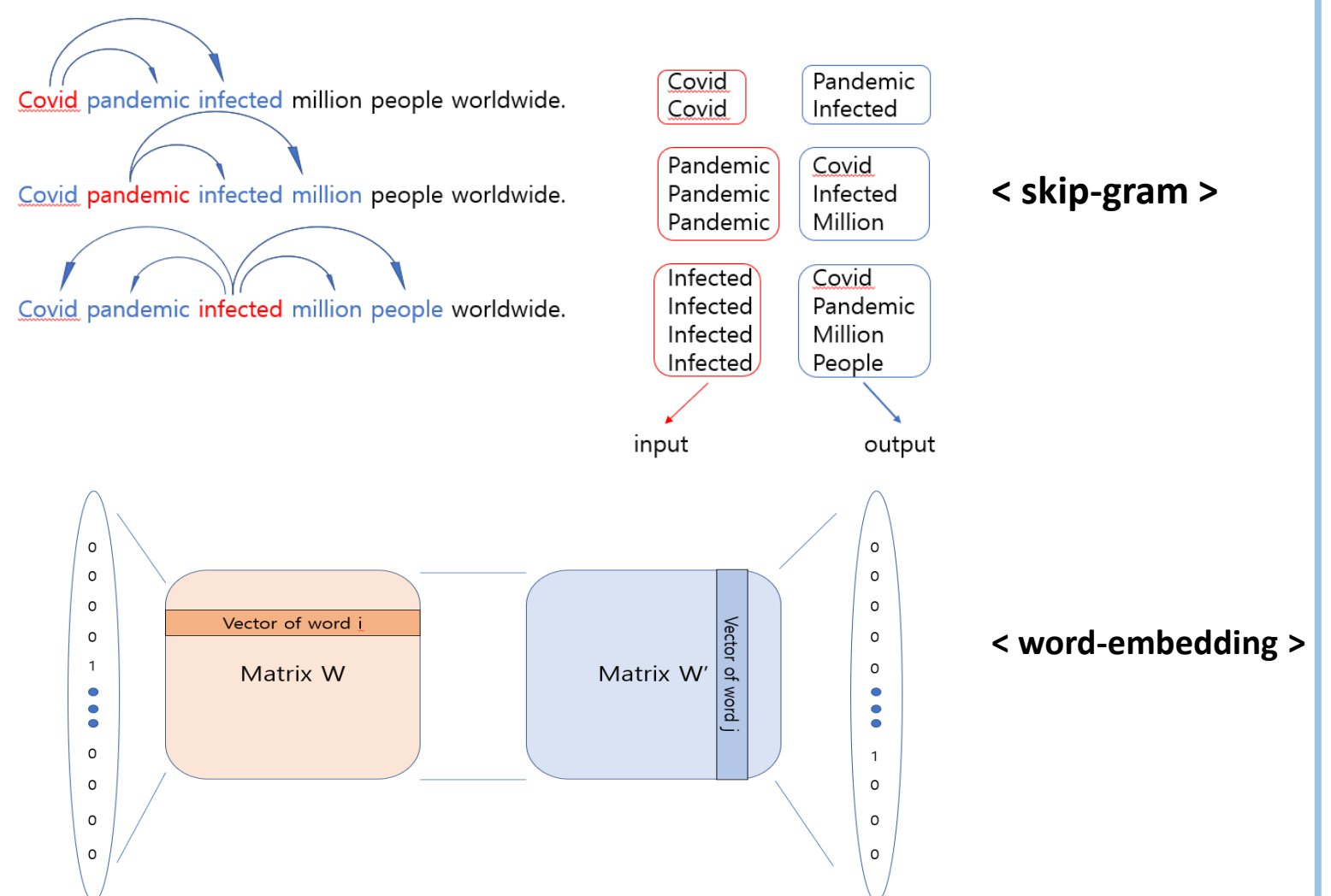
- 단어장에 맞게 network를 생성

< 함수4. cos_sim >

- 코사인유사도를 통해 단어 유사도 계산

< 함수5. training_param_add >

- 추가로 학습할 시 늘어난 데이터에 맞게 가중치 추가



3. 결과

```
In [12]: print(voc[4])
people

In [13]: get_recommendations_predict(net,wi,idx = 4, num = 10)

covid 0.14 %

twice 0.14 %

following 0.14 %

p 0.14 %

colder 0.14 %

symptoms 0.13 %

abnormalities 0.13 %
```

```
In [14]: print(get_recommendations_cos(word_vec_total,wi,4))
('people', ['try', 'arrest', 'colder', 'done', 'possible', 'tissues', 'later', 'keep', 'actual', 'analysis'])
```

```
In [15]: print(get_recommendations_dis(word_vec_total,wi,4))
('people', ['try', 'done', 'colder', 'arrest', 'later', 'keep', 'host', 'tissues', 'analysis', 'reviewed'])
```

```
In [16]: word, word_cos = get_recommendations_cos(word_vec_total,wi,4)
word2, word_dis = get_recommendations_dis(word_vec_total,wi,4)
```

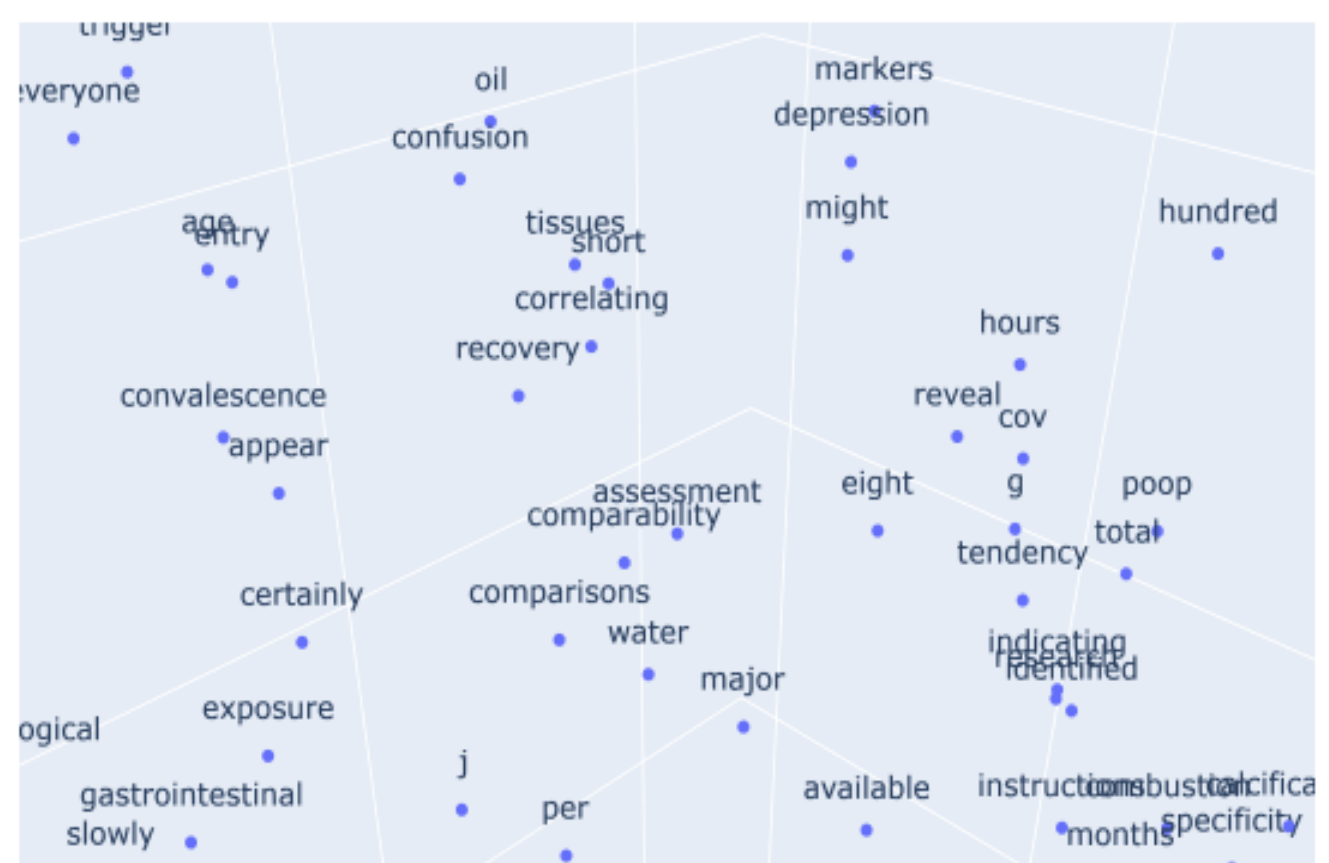
```
for word in word_cos:
    if word in word_dis:
        print(word)
```

```
try
arrest
colder
done
tissues
later
keep
analysis
```

- One-hot encoding방식으로 단어를 입력
- 특정단어(Covid19) 관련 기사들을 모아 학습
- 학습 후 얻은 가중치를 통해 특정단어 주변에 올 단어 예측 (Window size =2 , 결과값 %로 출력)

- Cos유사도, 유클리드 거리 방식을 통해 특정단어와 유사도가 높은 단어 추출
- 추출된 단어 간 겹치는 단어 출력

- PCA, Plot를 통해 고차원 데이터를 3차원 데이터로 시각화



< 시각화 >