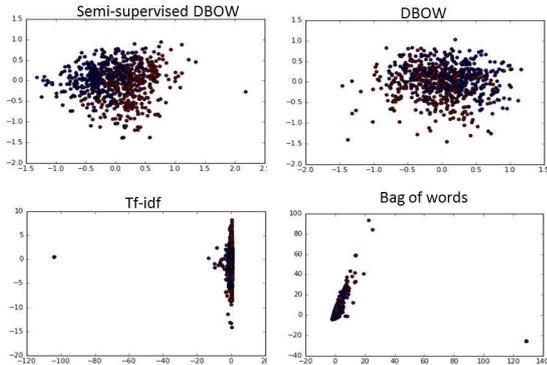


■ 연구논문 요약문1

<p>논문제목</p>	<p>Semi-supervised distributed representations of documents for sentiment analysis</p>
<p>게재정보</p>	<p>Neural Networks, Vol 119, 2019</p>
<p>개요</p>	<p>- 본 연구에서는 문서를 적합한 수치 데이터로 변환 할수 있는 문서 표현 방법론을 개발함. 실제 데이터에서는 모든 텍스트가 감성 정보를 갖고 있는 것은 아니기 때문에 일부 데이터의 감성 정보를 반영 할 수 있는 반교사 문서 표현 방법론을 개발함.</p> <p>- 개선된 방법론의 목적 함수는 다음과 같음. 첫 번째 수식에서 3가지의 목적 함수를 함께 최적화 하는데 각각 단어 벡터의 학습, 문서 벡터의 학습, 문서의 감성 정보 반영과 관련된 부분임. 문서 표현을 학습 시키면서 단어 표현 벡터와 함께 학습 시킴.</p> $\frac{1}{T} \sum_{t=1}^T \sum_{1-l \leq j \leq l} \log (P(w_{t+j} w_t))$ $\frac{1}{T} \sum_{t=1}^T \log (P(w_t d_k))$ $\frac{1}{N_{nb}} \sum_{i=1}^{N_{nb}} \sim (d_{k_i}, d_k) \log (P(d_{k_i} d_k))$ <p>[수식 1] 단어 벡터, 문서 벡터, 감성 정보 반영 목적함수</p>
<p>연구결과</p>	<p>- 구해진 문서 표현의 성능을 보기 위해서 PCA를 통하여 2차원으로 낮추어 시각화를 수행함. 기존 문서 표현 방법론인 Distributed bag of words (DBOW), Bag of words (BoW), Term frequency - inverse document frequency (TF-idf)와 비교를 진행함. 그 결과는 다음 그림과 같음.</p>  <p>[그림 2] 문서 표현 방법론들의 시각화 결과</p> <p>- 본 연구진이 개발한 임베딩 방법론이 더 우수한 것을 확인 할 수 있음.</p>
<p>활용분야 및 기대효과</p>	<p>- 본 연구를 통해 반교사 임베딩 방법론을 개발하면, 기존에 라벨이 있는 데이터가 적어서 기계학습 모델을 구축 하지 못했던 많은 데이터들을 활용 할 수 있음.</p>