

단어 임베딩을 이용한 사진 유사도 평가 및 가상현실기반 인지재활 시스템

최 권 택

강남대학교 소프트웨어응용학부

Image similarity evaluation using word embedding and virtual reality based cognitive rehabilitation system

KwonTaeg Choi

Division of Software Application, Kangnam University, YongIn-si 16979, Korea

[요 약]

고령화에 따른 치매 노인의 증가로 사회 및 경제적 비용이 크게 증가해 이를 해결하기 위한 연구가 이루어지고 있다. 본 논문은 치매 대상자의 기억과 추억이 반영되어 있는 사진들을 수집해 영상 인식 기술과 군집화 알고리즘을 사용한 가상현실 기반의 인지 재활 시스템을 제안한다. 기존의 객관적인 사진이나 그림을 사용하는 방법을 탈피해 개인의 경험과 추억을 가상현실 공간에 재현하는 융합 콘텐츠라는 점에서 기존 방법과 구별된다. 사진 2022장을 실험 데이터 셋으로 사용해 제안하는 군집화 알고리즘의 효율성을 평가하였다. 사진 영상을 의미 기반의 단어벡터로 표현해 군집화 함으로써 문맥적으로 유사한 사진들이 군집화되는 것을 확인했다. 또한 가우시안 혼합모델로부터 최소 거리가 유지되도록 사진을 선별하는 방법이 문맥적으로 떨어진 사진들을 선택하는데 매우 효율적임도 확인했다. 또한 제안한 알고리즘을 사용한 가상현실 기반의 인지재활 시스템을 제시한다.

[Abstract]

As the dementia elderly increase due to aging, socio-economic costs have increased greatly and research is being done to solve this problem. This paper proposes a virtual reality-based cognitive rehabilitation system using image recognition technology and clustering algorithm by collecting pictures reflecting memories and memories of subjects. It is distinguished from existing methods in that it is a converged content that reproduces individual experiences and memories in the virtual reality space by using existing objective photos or pictures. We evaluated the effectiveness of the proposed clustering algorithm using 2022 photographs as an experimental data set. By expressing images with semantic-based word vectors, we confirmed that contextually similar photographs were clustered. We also found that the method of selecting photos to maintain the minimum distance from the Gaussian mixture model is very efficient for selecting contextually separated photos. In addition, a virtual reality-based cognitive rehabilitation system using the proposed algorithm is presented.

색인어 : 영상 인식, 컴퓨터 비전, 단어군집화, 가상현실, 인지 재활

Key word : Image Recognition, Computer Vision, Word Embedding, Virtual Reality, Cognitive Rehabilitation

<http://dx.doi.org/10.9728/dcs.2019.20.11.2121>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 01 October 2019; **Revised** 04 November 2019

Accepted 25 November 2019

***Corresponding Author; KwonTaeg Choi**

Tel: +82-031-280-3660

E-mail: kwontaeg.choi@kangnam.ac.kr

1. 서론

한국을 포함해 전 세계적으로 고령화에 따른 노인 인구의 증가로 사회적 측면에서 그리고 경제적 측면에서 다양한 문제가 대두되고 있다. 특히 치매로 인한 사회적 경제적 비용이 크게 증가해 국가에서 책임져야 하는 문제로까지 여겨지고 있다.

치매를 치료할 수 있는 다양한 방법이 있지만 약물적 중재로 인한 부작용 때문에 경증 치매 환자를 대상으로 비약물적 중재에 관련된 상업화된 서비스가 이미 제공되고 있고, 다양한 연구 또한 꾸준히 진행되고 있다. 특히 컴퓨터 기반의 재활 시스템 [1]~[2]이 개발되어 병원 및 재활 센터를 중심으로 일상 생활에 필요한 인지 훈련을 수행하고 있다.

컴퓨터 기반의 치매 예방 및 인지 재활 치료는 환자의 특성과 정도에 따라 난이도, 시간, 속도 등이 조절 가능하고, 훈련 과정에서 발생한 치료 시간, 대처 상황에 대한 데이터 기록이 가능해 향후 분석에 활용할 수 있는 장점이 있다. 또한 즉각적인 피드백 시스템이 가능하기 때문에 전문가 개입 없이도 사용자 스스로 난이도를 선택해 반복적으로 훈련이 가능하다는 장점이 있다.

여기에 정보기술이 발전하면서 그리고 효과적인 재활 치료를 위해 증강현실[3], 가상현실[4]~[8], 혼합현실[9] 같은 실감 기술이 재활 치료에 적용되고 있다. 특히 키넥트 센서, Wii 모션 센서, 증강현실 기술을 사용해 지루한 재활 치료과정에 스스로 몰입할 수 있도록 하고 있다. 여기에 가상현실 기술이 빠르게 대중화되면서 가상현실을 이용한 인지 재활 치료에 관한 연구가 최근 관심을 받고 있다. 일부 가상현실 기반의 인지 재활 시스템은 이미 상업화되었고, 연구소와 병원을 중심으로 임상적으로 사용되면서 그 효과가 입증되고 있다.

가상현실을 이용한 재활 치료는 최근에 이루어진 것은 아니고 이미 1990년 후반에 재활치료에 가상현실을 적용하는 시도가 있었다. 그러나 당시에는 가상현실 기술이 발전하지 않았기 때문에 하드웨어 제작과 콘텐츠 개발에 한계가 있었다. 그러나 최근 4년 동안 구글, 페이스북, SONY, HTC 등 정보기술 분야를 선도하는 기업들이 가상현실 기술을 대중화 시키면서 많은 발전이 이루어졌다.

가상현실 관련 하드웨어와 소프트웨어, 그리고 서비스 인프라가 정비되어 대중화되기 시작한 2016년 초기에는 노인들이 가기 어렵거나 힘든 장소 또는 과거의 장소를 가상현실로 재현해 생생하게 보여줌으로써 스트레스 해소 및 우울감을 개선하는 분야에 사용되었다. 즉 직접적으로 인지 재활 치료에 사용되지는 않았다. 그러나 관련 기술의 발전으로 최근에는 치매나 인지 재활 치료가 가능한 가상현실 콘텐츠를 개발하기 위해 기업과 병원 및 재활 센터들이 협력해서 연구하고 있고, 의미 있는 결과도 지속적으로 늘어나고 있다[3]~[6].

지금까지 살펴보았듯이 인지 재활 치료는 컴퓨터, 모션 센서, 증강현실, 가상현실, 혼합현실 기술을 적극적으로 도입하고 있다. 이를 통해 치료 대상자의 수준과 반응 정도를 고려해



그림 1. 한국형 전산화 인지재활 프로그램 CoTras
Fig. 1. Korean Computer-Based Cognitive Rehabilitation Program(CoTras)

콘텐츠가 조정되어 맞춤형 재활 치료가 가능하다고 하지만 재활 훈련에 사용되는 콘텐츠는 그림 1처럼 대부분 자신에 관한 내용이 아닌 치료 대상자들이 쉽게 이해할 수 있는 보편적이고, 객관적인 내용들로 구성되어 있어 몰입감을 느끼며 치료에 집중하는데 한계가 있다.

인지 재활 치료 방법 중 회상 요법은 개인의 기억과 추억을 인제 재활 과정에 적극적으로 도입해 과거 경험을 기반으로 치료의 효과를 높이는 방법이다. 이러한 회상 요법은 치료 과정에 대부분 필수적으로 들어가는 항목으로 국내의 다수의 연구논문들을 통해 그 효과성이 입증되고 있다[10]. 일반적인 사실을 담고 있는 사진, 예를 들면 과거에 존재 했던 물건, 인물, 풍경 등을 사용해 치료 대상자 과거의 추억과 경험을 이끌어 낸다. 이는 치료 대상자의 흥미를 유발하기 쉽고, 기억능력을 활성화시켜 치료에 집중함으로써 훈련의 효과를 높게 된다. 그러나 회상 요법은 전문가의 도움이 반드시 필요하고 치료 대상자와 많은 대화가 필요로 하기 때문에 대상자 혼자서 재활 치료를 수행할 수 없는 한계가 있다.

본 논문에서는 과거의 기억과 추억을 회상하는데 도움을 주기 위한 콘텐츠로 개인이 그동안 촬영한 사진을 이용한 가상현실 기반 인지 재활 훈련 시스템을 제안한다. 일반적인 인물, 물건을 기반으로 콘텐츠를 구현한 기존의 인지 재활치료 시스템과는 다르게 대상자의 경험과 추억이 반영된 개인 사진이 몰입감이 뛰어난 가상현실 공간에 제시되기 때문에 치료에 더욱 집중할 수 있게 된다.

개인의 경험과 추억을 기록할 수 있는 미디어는 글, 소리, 사진, 동영상 등 다양한 매체가 있지만 사진은 생산하기 쉽고, 다양한 소셜 네트워크 서비스들 통해 다른 사람들과 손쉽게 소통할 수 있는 수단으로 사용되어 그 어느 시대보다 개인의 과거 기억이 누적되어 있다. 무엇보다도 영상 처리, 영상 인식, 컴퓨터 비전 분야에서 매우 오랫동안 영상 이해에 관한 기술이 축적되었고 최근 딥러닝 기술의 발달로 그 어느 때보다 사진에 대한 문맥적 이해가 가능하게 되었다.

최근 개인의 사진을 자동으로 분석해서 가상현실 공간에서 인지 재활 콘텐츠 개발에 관한 연구가 있었다[4]. 이 논문에서는 영상 인식 기술을 사용해 인물 정보, 음식사진여부를 예측하

고 사진 촬영 시 GPS(Global Positioning System) 정보, 시간 정보를 추출해 이들을 군집화해서 다양한 시나리오를 기반으로 인지 재활 콘텐츠를 구성했다. 그러나 영상에서 인식 가능한 정보가 인물, 나이, 성별, 음식 사진 여부이기 때문에 이러한 분류에 해당하지 않는 개인의 경험과 추억을 기록한 사진은 사용할 수 없는 한계가 있다.

본 논문에서는 영상 인식 기술을 사용해 인물사진, 음식사진 등으로 명확하게 구분하는 방법이 아닌 사진의 문맥적 의미를 인식하고, 사진간 유사도를 비교해 군집화 할 수 있는 새로운 방법을 제시하고 이를 활용해 인지 재활 치료에 사용할 수 있는 가상현실 콘텐츠 시스템을 제안하고자 한다.

II. 제안 방법

2-1 제안하는 시스템 개요

제안하는 시스템의 전체 개요와 주요 알고리즘은 그림2와 같다. 크게 세 부분으로 구성되어 있는데, 우선 개인이 가지고 있는 사진을 구글 vision API(Application Programming Interface))를 사용해 문맥적 의미 단어를 추출한다. 다음으로 이렇게 추출된 인식 단어 집합을 가지고 대표적인 워드임베딩 방법인 word2vec[11]~[13] 알고리즘을 사용해 추출 단어를 특정 차원의 벡터로 학습시킨다. 보통 하나의 사진은 다수의 단어 벡터로 이루어져 있기 때문에 이를 평균화해서 사진 한 장을 벡터로 표현한 뒤, 이를 가우시안 혼합 모델(GMM, Gaussian Mixture Model)을 사용해 비지도 학습 방식으로 사진들을 군집화 한다. 그리고 적용 시나리오를 고려해 서로 다른 군집에 해당하는 사진과 단어가 선택되어 가상현실 공간에 제시된다. 치료 대상자는 제시된 문제를 해결하면서 인지 강화훈련을 하게 된다.

그림 2에서는 두 가지 구체적인 사례를 제시하고 있다. 첫 번째 사례 a)에서는 한 장의 사진과 단어가 주어지는데 대상자는

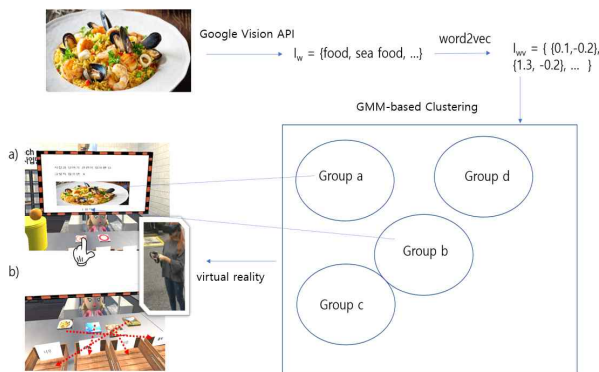


그림 2. 제안하는 인지 재활 시스템 개요

Fig. 2. Outline of the proposed cognitive rehabilitation system

사진과 단어가 관련되어 있는지 아닌지를 판단해야 한다. 사용자는 HMD(Head-mounted Display)를 착용하고 가상현실 공간에서 손을 사용해 Yes 혹은 No버튼을 선택해야 한다. 그림의 경우 사진이 속한 그룹이 아닌 다른 그룹에서 단어가 선택되었기 때문에 사용자는 No 버튼을 선택해야 한다.

두 번째 사례 b)에서는 4개의 사진이 주어져 있고, 이를 단어가 표시된 사진 케이스에 정확히 배치해야 한다. 다수의 그룹에서 발생하는 관계를 파악해야 하기 때문에 첫 번째 사례보다 종합적인 인지 과정이 필요하게 된다. 그림의 경우 4개의 사진을 각각 4,2,3,1 번째 사진 케이스에 배치해야 한다. 이와 같이 치료 대상자는 자신의 기억과 추억을 회상해가면서 인지 재활치료에 참여하게 된다.

이러한 시스템을 개발하는데 있어서 해결해야 할 기술적 문제는 유사한 사진들이 한 번에 사용자에게 제시되면 안된다는 점이다. 예를 들어 첫 번째 사례에서 사진이 c 그룹에서 제시되고, 단어는 b 그룹에서 제시 될 때 b와 c 그룹이 문맥적으로 유사한 영상으로 군집화 되어 있다면 사진과 단어가 관련되어있다고 인식하고 사용자는 No가 아닌 Yes 버튼을 선택할 수 있다. 이러한 오류가 발생하지 않도록 하기 위해서는 동시에 제시되는 사진들이 구별 되도록 군집화가 되어야 한다. 다음절에서 사진을 비지도 학습 방법을 사용해 효과적으로 군집화 할 수 있는 방법에 대해서 설명한다.

2-2 문맥 정보를 이용한 이미지 임베딩

1) 영상 단어 임베딩

사진을 군집화 하는 일반적인 방법은 시간, 위치 정보 등 메타 정보와 에지 및 색상 정보를 특징으로 추출해 이를 군집화 한다[14]. 그러나 이러한 방법은 메타 정보를 사용할 수 없거나 문맥적으로 같은 내용인데 촬영 환경에 따라 에지 정보와 색상 정보가 다를 수 있기 때문에 다른 그룹으로 분류될 수 있다. 따라서 본 논문에서는 사진의 에지나 색상 같은 특징을 사용하지 않고 문맥을 직접 인식하는 방법을 사용한다.

사진 영상으로부터 문맥정보를 추출하기 위해서 우선 구글 vision API를 사용해 하나의 영상을 다수의 단어 집합으로 표현한다. 영상 I로 부터 n개의 단어(w1~wn)가 추출되었다면 이를 I_w로 표기하고 식 (1) 처럼 정의할 수 있다.

$$I_w = \{ w_1, w_2, w_3, \dots, w_n \} \quad (1)$$

구글 vision API는 영상 하나에 대해서 다수의 단어를 추천하는데 이때 고정된 크기의 단어를 추천하지 않는다. 확실하게 해당 단어의 의미를 포함하고 있을 경우만 단어를 추천하기 때문에 영상마다 식(1)의 원소 개수는 다르다.

다음으로 사진에서 인식된 전체 단어를 모아서 워드임베딩 방법인 word2vec 알고리즘을 사용해 학습시킨다. 이 알고리즘은 함께 출연하는 단어의 빈도수가 커질수록 그 단어들 간의 거

리가 작아지도록 학습시킨다. 단어의 빈도수가 낮은 경우 해당 단어는 다른 영상과는 관련이 적다는 것을 의미하기 때문에 여기서는 빈도수 1인 단어는 학습에서 제외하였다. 이런 단어를 제외하는 것이 어떤 영향을 끼치는지에 대해서는 실험부분에서 설명한다. 이렇게 영상 I를 word2vec 알고리즘으로 변환한 것을 아래 식 (2)처럼 I_{wv} 로 정의한다.

$$I_{wv} = \{wv_1, wv_2, wv_3, \dots, wv_n\} \tag{2}$$

여기서 wv_i 은 단어 w_i 에 대한 워드임베딩 공간에서의 특징 벡터를 의미한다. 식 (1)의 I_w 가 n개의 단어로 구성되어 있다면 I_{wv} 도 n개의 특징 벡터로 구성된다. 실험부분에서 식 (2)에 의해서 wv 들이 어떠한 분포를 가지고 있는지 시각화해서 문맥적으로 관련된 단어들끼리 서로 군집화되는지 살펴본다.

2) 영단어 임베딩에 대한 평균 벡터

식 (2)에 의해 영상 I가 다수의 단어 벡터의 집합으로 정의되는데 유사도를 평가해 영상을 군집화 하기 위해서는 I_{wv} 를 하나의 벡터로 표현해야 한다. 이는 단어의 집합으로 이루어진 문서를 하나의 특징 벡터로 표현하는 문제와 동일하다.

그러나 문서의 경우 문서를 구성하고 있는 단어가 매우 많기 때문에 가우시안 혼합 모델을 사용해 하나의 특징 벡터로 표현할 수 있지만, 제안하는 시스템에서 사용하는 영상의 경우 n이 매우 작기 때문에 이러한 방법을 사용할 수 없다. 더구나 가우시안 혼합 모델을 사용할 경우 적합한 k값을 결정하는 것도 어려운 문제다.

이러한 두 가지 이유로 본 논문에서는 단어 벡터 wv 에 대해서 평균벡터를 계산하고 이를 영상 I에 대한 특징 벡터 I' 로 정의 하였다. 따라서 문맥정보를 포함한 영상에 I에 대한 특징 벡터 I' 는 식 (3) 처럼 정의 된다.

$$I' = \frac{\sum_{i=1}^n wv_i}{n} \tag{3}$$

그리고 i번째 영상 I_i 와 j 번째 영상 I_j 간의 문맥정보를 고려한 공간에서의 거리는 식 (4)처럼 정의 할 수 있다.

$$distance(I_i, I_j) = \| I'_i - I'_j \| \tag{4}$$

따라서 두 영상이 같은 그룹에 속하면 식(4)의 값이 작아지고, 서로 다른 그룹이면 식(4)의 값이 커진다.

2-3 가우시안 혼합 모델을 이용한 영상군집화

제안하는 시스템의 목적은 문맥적으로 서로 다른 영상들이 분리될 수 있도록 군집화는 것이다. 즉 I_i 와 I_j 가 유사한 문맥을 가진 영상이라면 식 (4)의 값은 작아야 하고, 다른 문맥을 가진

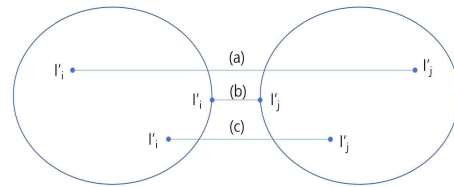


그림 3. GMM을 이용한 영상 군집화 문제
Fig. 3. Image Clustering Problem Using GMM

영상이면 값이 커야한다. 식 (3)으로 표현된 I' 는 식 (5)처럼 k 개의 그룹으로 분할된 가우시안 혼합 모델로 표현될 수 있다.

$$p(I) = \sum_{j=1}^k \phi_j N(\mu_j, \Sigma_j) \tag{5}$$

여기서 ϕ 는 가중치, μ 는 평균, Σ 는 공분산을 의미한다.

그러나 가우시안 혼합 모델을 사용해 제안하는 시스템을 구현할 때 두 가지 요소를 고려해야 한다. 제안하는 인지 재활 시스템에서는 서로 다른 문맥을 가진 사진들이 문제로 제시되어야 한다. 이를 위해 각 그룹에서 각각 임의로 사진 I_i 와 I_j 를 선택해야 하는데 이때 최악의 경우 서로 다른 그룹에 속하지만 식 (4)의 값은 매우 작을 수 있고, 이는 사진들이 문맥적으로 유사할 수 있음을 의미한다.

이 문제는 그림 3을 통해 확인 할 수 있다. 예를 들어 그림처럼 I' 가 2개의 그룹으로 군집화되어 있다고 가정하자. 두 개의 그룹에서 각각 I'_i 와 I'_j 를 선택할 때 다행히 (a)처럼 다 차원 공간의 양쪽 끝에 있는 I'_i 와 I'_j 가 선택된다면 이들의 거리가 크기 때문에 I_i 와 I_j 는 문맥적으로 서로 다른 영상일 수 있다. 그러나 최악의 경우 (b) 처럼 I'_i 와 I'_j 두 개가 선택된다면 이들의 거리가 매우 작다. 이럴 경우 I_i 와 I_j 가 문맥적으로 유사할 확률이 높다.

본 논문에서는 이러한 문제를 해결하기 위해 그림 3 (c)처럼 두 개의 이미지를 선택할 때 I'_i 와 I'_j 의 거리가 어느정도 유지될 수 있는 제약 조건을 사용하였다. 따라서 서로 다른 그룹에서 이미지 I_i, I_j 를 선택하는 것은 식 (6)에 의해 결정된다.

$$select(I_i, I_j), \text{ where } \| I'_i - I'_j \| \geq \theta \tag{6}$$

식 (6)에서 θ 는 그룹간 유지되어야 하는 최소 거리이다. θ 가 0이면 랜덤하게 I_i, I_j 는 선택하는 것을 의미한다. 반면 θ 가 너무 크면 조건을 만족하는 I_i 와 I_j 를 찾지 못 할 수도 있다. 실험부분에서 각각의 그룹에서 랜덤하게 이미지를 선택하는 방법과 식 (6)을 사용해 선택한 방법의 차이를 분석한다.

가우시안 혼합 모델을 적용할 때 고려해야 하는 두 번째 문제는 몇 개의 그룹으로 분할해야 하는지를 나타내는 k 값이 사전에 주어져야 한다는 점이다.

여기서는 치료 대상자에게 인지 훈련을 위해 제시되는 사진과 단어 유형에 따라 3가지로 분류하고 각 유형에 따라 k 값을 어떻게 결정할 수 있는지 설명한다.

- 1:1 유형

이 유형은 그림 2에서 a)처럼 한 장의 사진과 단어가 제시된다. 주로 사진의 의미가 맞는지 틀린지 판단하는 문제 유형이다. 이를 위해서는 I'를 2개의 그룹으로 분할해서 첫 번째 그룹과 두 번째 그룹에서 식 (6)을 만족하는 두 개의 사진 I_1, I_2 를 선택한다. 그리고 사진영상은 I_1 를 사용하고, 단어는 I_2 의 단어중 하나를 제시하면 사진과 단어가 서로 다른 그룹이기 때문에 사용자는 no를 선택해야 한다. 반면 사진 I_1 를 제시하고 단어도 I_1 의 단어중 하나를 제시하면 치료 대상자는 yes를 선택해야 한다. 따라서 1:1 유형에서는 k를 2로, 가우시안 혼합 모델을 계산하면 된다.

- 1:N 유형

이 유형에서는 1개의 단어와 N개의 사진이 제시되어 단어에 해당하는 사진을 선택하는 경우이다. N이 클 경우 많은 사진들을 비교 분석해야 하고 문제의 수준이 어려워 질 수 있다. 일반적인 인지 재활 시스템에서 이런 유형은 주로 4개의 이미지가 주어진다. 특별한 이유가 있기 보다는 보통 사람들은 4지 선다에 익숙하기 때문이다. 이를 위해 4개의 I_1, I_2, I_3, I_4 를 선택해야 하는데 이때 4개 영상에서 모두 식 (6)이 만족되어야 한다. 단어는 4개 영상중 하나에서 추출하고, 4개의 영상은 모두 제시하면 된다. 따라서 1:N 유형의 문제가 제시되는 경우 k를 4로 해서 가우시안 혼합 모델을 계산할 수 있다. 사진을 제시할 때 난이도를 낮추기 위해 2개 혹은 3개의 사진만을 제시할 수도 있다. 이 때 k 값은 각각 2와 3을 사용하면 된다.

- N:N 유형

마지막 세 번째 유형은 그림 2에서 b)처럼 다수의 단어와 다수의 사진을 비교하는 방식이다. N개의 단어와 N개의 이미지가 동시에 주어지고 사용자는 각 단어와 관련 있는 사진을 적절히 배치해야 한다. 1:N 유형과 동일한 방식으로 N개의 이미지와 이에 해당하는 단어를 제시하면 된다. 역시 N이 클 경우 문제의 난이도가 높아지고 문맥적으로 유사한 영상인데 서로 다른 그룹으로 분할될 확률이 높아지기 때문에 N은 최대 4로 지정했다. 이는 k값을 4로 해서 가우시안 혼합 모델을 계산함을 의미한다. 난이도에 따라 3개의 사진과 단어를 제시해 이들 간 관계를 인지하도록 유도할 수 있다.

다음 절에서는 1:N 유형과 N:N 유형이 가상현실 공간에서 인지 재활 치료에 어떻게 적용될 수 있는지 설명한다.

2-4 가상현실 응용

컴퓨터 기반 인지 재활 치료 시스템은 모니터와 키보드, 혹은 전용 컨트롤러를 사용하기 때문에 제한된 공간에서 제한된 손동작을 반복적으로 사용해 인지 훈련을 수행한다. 따라서 물입감을 가지고 오랫동안 치료에 집중하기 어렵다.

그러나 가상현실은 360도 공간을 모두 활용할 수 있고 화면



그림 4. 1:N 유형에 대한 가상현실 적용 사례
Fig. 4. Application of virtual reality for 1: N type



그림 5. N:N 유형에 대한 가상현실 적용 사례
Fig. 5. Application of virtual reality for N : N type

크기에 대한 제약도 없다. 공간 연출을 포함한 다양한 시나리오도 가능하다. 더구나 사용자의 몸 전체를 움직이며 진행할 수 있어 사용자의 흥미를 더욱 이끌어낼 수 있다.

여기서는 2가지 구현된 사례와 응용 가능한 사례를 살펴본다. 1:1 유형의 경우는 사진과 단어가 하나씩 주어지는 경우로 주로 yes 혹은 no를 선택하는 간단한 경우라 여기서는 가상현실에 더 유용한 1:N 유형과 N:N 유형에 대해서 살펴본다. 사진에서 단어를 추출할 때 구글 vision API는 영어단어를 추천하기 때문에 가상현실 공간에서 출력할 때는 영어를 한글로 변경해 출력하였다.

그림 4는 1:N 유형에 대한 구현 사례이다. 주어진 단어에 해당하는 사진을 골라 박스에 넣어야 한다. 기존 인지 재활 시스템에서는 손가락 클릭으로 훈련 과정이 진행된다면 가상현실에서는 컨트롤러를 사용해 여러 사진을 잡고 놓는 동작도 필요하고, 사진을 돌려보는 동작도 하게 된다. 이러한 과정에서 과거의 추억을 회상할 수 있다. 답안을 제출할 때도 사진을 선택하는게 아니라 옆에 있는 박스에 담은 행동을 해야 하기 때문에 인지 훈련뿐만 아니라 신체 활동 강화에도 도움이 된다.

그림 5는 N:N 유형의 문제이다. 주어진 4장의 사진을 적절한 사진 박스에 분류해서 답아야 한다. 이러한 유형은 기존 인

지 재활 시스템에서는 마우스를 이용해 관련된 사진과 단어를 연결한다. 1:N 유형처럼 컨트롤러를 사용해 4개의 사진을 집고 놓는 동작도 필요하고, 사진을 비교하는 작업도 필요하다. 사진들을 분류할 때 사진을 넣은 박스의 위치를 어떻게 설정하느냐에 따라 다양한 신체 재활이 가능하다. 간단히 4개의 박스를 좌우로 배치했지만 치료 목적에 따라 무릎을 일부 사용하도록 사진 박스를 상하로 배치할 수도 있다. 이러한 배치에 따라서 몸의 특정 부위를 강화해야 하는 경우 효과적으로 응용될 수 있다.

가상현실 기반으로 인지 재활 치료 시스템을 개발할 때 고려해야 할 사항은 안전성이다. 가상현실 기술은 사용자의 이동도 파악할 수 있지만 인지재활 치료과정에서 안전상의 문제가 발생할 수 있기 때문에 제자리에서 이동만 가능한 상황으로 한정해야 한다. 또한 몸을 많이 움직여야 할 경우 치료 대상자의 상태에 따라서 무리한 요구 일수도 있기 때문에 치료 대상 상태와 인지 및 신체 수준에 따라 적절한 범위에서 움직일 수 있도록 사진이나 사진 박스등이 배치되도록 해야 한다.

III. 실험

제안한 방법론에 대한 분석과 성능 검증을 위해 4년간 스마트폰으로 촬영한 사진 2022장을 사용해 실험을 진행하였다. 사진은 최소 700x800에서 최대 3000x4000까지 다양한 해상도를 가지며 별도의 선별과정이나 영상처리를 수행하지 않았다. 따라서 2022장의 사진에는 일상 생활 사진, 여행사진, 인물 사진, 캡처한 사진등 다양한 영상으로 이루어져 있다. 사진 분포에 따라 제안하는 알고리즘의 특성이 어떻게 변하는지 분석하기 위해 전체 사진은 촬영 시기와 촬영 장소를 기준으로 Set#1(1000장)과 Set#2(1002장)로 나누어 실험을 진행하고 모든 알고리즘은 python으로 작성되었다.

실험은 크게 세 부분으로 이루어져 있다. 첫 번째는 2022장의 사진을 구글 Vision API를 사용해 사진 영상을 단어 집합으로 표현했을 때 통계적 의미를 찾을 수 있는지에 대한 실험이다. 두 번째 실험은 word2vec 알고리즘을 사용해 단어들이 워드임베딩 공간에서 어떠한 분포를 갖는지 확인하는 실험이다. 마지막 실험은 다수의 워드벡터들로 이루어진 영상을 하나의 단일 벡터로 바꾸어 문맥적으로 유사한 영상들끼리 군집화가 가능한지 그 분포를 확인하는 실험이다.

1) 영상인식을 이용한 사진의 문맥 단어 추출

2022장의 사진을 구글 vision API로 분석할 때 사진에서 최대 10개의 추천 단어가 제시되도록 했다. 따라서 한 장의 사진은 최대 10개의 단어로 표현되는데 일부 영상의 경우 1개의 단어만 제시되기도 했다. 2022장의 사진 중 47장의 영상은 영상 문제로 분석되지 않아 영상 인식에 성공한 사진은 Set#1과 Set#2에서 각각 972장과 1002장이다. 표 1에 Set#1, Set#2에 대

표 1. Set#1과 Set#2 사진에 대한 인식 단어 수

Table 1. Recognition word count for Set # 1 and Set # 2 photos

# of recognized words		10	9	8	7	6
ratio	Set#1	57.9%	7.6%	8.0%	7.2%	5.8%
	Set#2	59.1%	7.4%	6.9%	5.2%	6.4%
인식 count		5	4	3	2	1
ratio	Set#1	4.1%	3.9%	3.2%	1.3%	0.7%
	Set#2	4.0%	3.9%	3.0%	2.1%	1.2%

해서 인식된 단어 수에 대한 분포를 정리하였다.

표 1에서 알 수 있듯 실험 데이터셋에 상관없이 비슷한 분포를 보인다. 평균적으로 58.5%의 사진은 10개 단어로 인식되었고, 89.8%는 5개 이상의 단어가 인식되었다.

인식된 단어수가 적은 영상을 word2vec 알고리즘을 사용해 학습해야 하는지 결정하기 위해 우선 인식 단어 수가 1개인 영상을 분석하였다. 해당 영상을 살펴보면 대부분 구체적인 의미를 파악할 수 없어 추상적인 한 개의 단어로 인식한 경우가 많았다. 그러나 이러한 단어의 발생 빈도는 매우 높았다.

예를 들어 Set#1에서 하나의 영상이 한 개의 단어로 추천된 영상에서 인식된 단어들은 child(85), font(73), girl(458), brand(9), technology(26)이고, Set#2에서는 product(219), yellow(47), outdoor structure(15), interior design(29), space(17), product(219), institution(21), electronic device(37), event(15), window(44), recreation(120), room(19)이다. 여기서 괄호 안의 숫자는 Set#1과 Set#2에서 발생한 단어 빈도수이다. Set#1에서 girl이라는 한 개의 단어로 인식된 사진의 경우 girl 단어는 전체 972장의 사진에서 458장의 사진에서 인식되었다. 발생 빈도가 매우 높기 때문에 word2vec 알고리즘으로 단어를 학습할 때 인식된 단어 수에 상관없이 모든 영상을 학습에 사용하였다. 따라서 어떤 영상은 10개의 단어 집합으로 표현되고, 어떤 영상은 1개의 단어로 표현 된다.

다음 실험으로 Set#1, Set#2의 단어 분포가 얼마나 다른지 살펴보기 위해 단어 분포를 분석하였다. Set#1에서 총 인식된 단어 수는 8234개이고, 중복 단어를 제외하면 593개의 단어가 인식되었다. 인식단어가 많아 보이는 이유는 구글 vision API의 경우 음식을 음식으로 인식하기도 하지만 정확히 어떤 종류의 음식인지 그 종류까지 인식하는 경우가 많아 구별되는 단어가 많은 것이다. 결과적으로 Set#1 972장의 사진에서 문맥 정보 593개가가 인식되었다. Set#2에서는 인식 단어 수 8441개, 중복 단어를 제외하면 677개의 단어가 인식되었다.

Set#1과 Set#2에서 중복 단어를 제외하면 인식 단어 수는 883개 인데 공통으로 나오는 단어는 387개로 전체 단어의 43.8%에 해당한다. 나머지 56.2%는 서로 다른 단어들이다. 구체적으로 Set#1에만 존재하는 단어는 206개로 전체 단어의 23.3%, Set#2에만 존재하는 단어는 290개로 전체 단어의 32.8%이다.

이러한 수치들의 의미를 살펴보기 위해 우선 공통 단어가 아

년 56.2%의 단어들의 발생 빈도수를 살펴보면 대부분 빈도수가 낮다. 구체적으로 Set#1에서는 56.7%가 빈도수가 1이고, Set#2에서는 46.5%가 빈도수가 1이다. 이렇게 발생 빈도수가 낮은 경우 word2vec 알고리즘을 사용해 학습하는 과정에서 제외될 확률이 높아진다. 반면 43.8%의 공통된 단어들은 주로 인물, 음식, 촬영 장소와 관련되어 있다. 어떤 장소에서 촬영해도 빈번하게 발생하는 단어들이다.

결과를 종합해 보면 Set#1, Set#2 사진에서 56.2%는 서로 다른 영역에서 촬영된 사진들이다. 실제로 인물 사진은 Set#1, Set#2 모두에 많이 존재하고, 풍경사진은 Set#2에 많이 존재하고, Set#1에는 단체 활동 관련된 사진이 많이 존재한다. 음식 사진의 경우 Set#1에 많이 존재하지만 Set#2에는 많지 않다.

2) 워드임베딩 공간 분석

word2vec 같은 워드임베딩 알고리즘에서 단어와 단어의 관계는 단어의 빈도수를 기반으로 결정된다. 따라서 워드임베딩 알고리즘을 수행하기 전에 빈도수에 대한 특성을 파악해 학습 파라미터를 결정할 필요가 있다.

그림 6은 Set#1에 있는 593개 단어에 대한 빈도수 히스토그램이다. 가장 빈도수가 높은 단어는 458개의 영상에서 발생했다. 이렇게 빈도수가 높은 상위 20개의 단어는 girl, fun, product, recreation, vacation, sky, water, leg, beauty, smile, tree, long hair, human hair color, light, child, leisure, lighting, brown hair, sea, food이다. 많은 사진들이 인물, 음식, 풍경을 포함하기 때문이다.

반면 405번째 단어부터는 빈도수가 모두 1로 188개의 단어가 이에 해당한다. 이는 전체 단어 593개 기준으로 31.7% 해당한다. 빈도수가 1인 경우 다른 단어의 발생 관계를 파악할 수 없기 때문에 실제 분석은 빈도수가 2이상인 404개 단어에 대해서 워드임베딩 알고리즘을 적용했다.

이렇게 31.7%의 단어가 분석에서 제외된다면 많은 단어가 학습 시 사용되지 않아 보이지만 Set#1의 경우 전체 단어 수인 8234개에서 생각해 보면 빈도수 1인 단어 188개를 제외할 경우 분석에 사용되는 단어 빈도수에 대한 정보는 전체 데이터의

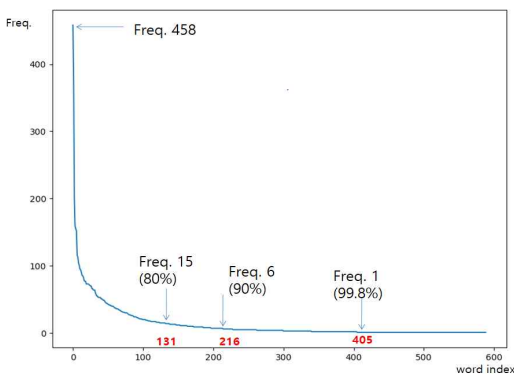
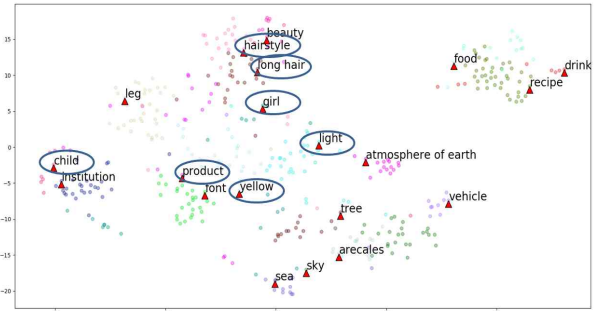
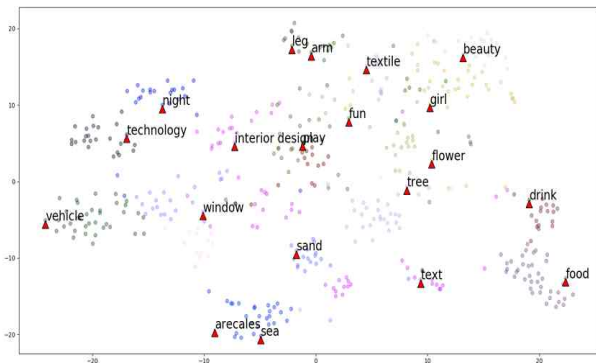


그림 6. Set#1(589개 단어)에 대한 빈도수 히스토그램
Fig. 6. Frequency histogram for Set # 1 (589 words)



(a) 20개로 군집화된 워드 벡터 공간(Set#1, 405단어)



(b) 20개로 군집화된 워드 벡터 공간(Set#2, 479단어)

그림 7. 워드 임베딩 시각화

Fig. 7. Word embedding visualization

99.8%가 된다. 그림에서 단어 발생 빈도수가 6보다 큰 216개의 단어를 사용해 학습이 이루어지는 경우 전체 데이터의 90%를 워드임베딩 알고리즘으로 분석하게 된다. 이러한 실험 결과로부터 논문에서는 빈도수 1인 단어만 제외하였다.

Set#2의 경우 가장 빈도수가 높은 20개의 단어는 girl, fun, tree, plant, product, sky, vacation, water, leisure, leaf, blue, recreation, woody plant, beauty, child, skin, cheek, long hair, smile, branch이다. Set#2는 야외 관광지에서 촬영한 사진이 많아 관련 단어들이 증가한 것으로 보인다. 빈도수 1인 단어의 수는 198개로 전체 단어 수 677개 기준으로 29.2% 해당한다. Set#1에서 같이 매우 높은 비율이다. Set#1에서와 마찬가지로 빈도수 1인 단어를 제외하고 479개 단어만 사용해 학습하였다. 이는 전체 단어 8441개에서 빈도수 1인 단어 198개를 제외하면 전체 데이터의 97.6%에 해당한다.

종합해 보면 어떤 데이터셋을 사용하느냐에 상관없이 빈도수 1인 단어를 제외하고 학습시켜도 대략 98.7% 이상의 데이터를 포함하기 때문에 데이터셋에 따라 파라미터 값을 조절할 필요는 없다.

다음 실험으로 빈도수 1인 단어를 제외하고 Set#1과 Set#2에 대해 단어임베딩 알고리즘을 적용하고 단어들이 어떻게 분포되어 있는지 검증하였다. 이를 위해 고차원 특징 시각화 방법을 사용해 사진이 문맥적으로 유사한 단어들로 군집화 되는지 살

펴보았다. 이를 위해 word2vec 알고리즘을 사용할 때 단어들은 100차원 특징 벡터로 표현하고 100차원의 단어 벡터들을 가우시안 혼합 모델 알고리즘을 사용해 다수의 그룹으로 군집화했다. 그리고 100차원 공간을 2D 그래프로 확인할 수 없기 때문에 PCA(Principal Component Analysis) 알고리즘을 사용해 2차원으로 축소하여 그래프로 출력했다. 가우시안 혼합 모델 알고리즘 수행 시 k 값으로 20을 사용하였다. k 값은 군집화 정도를 시각적으로 확인하기 위한 값으로 값 자체에 큰 의미는 없다.

그림 7은 단어벡터에 대한 시각화 결과이다. 그림에서 20개 그룹에 표시된 빨간 세로 마크와 그 위에 표기된 텍스트는 그룹의 중심점과 중심점에 가장 가까운 단어가 아닌 해당 그룹에서 가장 빈도수가 높은 단어와 그 위치를 의미한다. 가장 빈도수가 높은 단어가 해당 그룹을 대표하는 단어로 표시하여 명확하게 어떤 그룹인지 파악할 수 있도록 시각화했다.

그림 7에서 군집화 되어 있는 단어를 보면 Set#1과 Set#2가 어떤 사진으로 구성되어 있는지 확인 할 수 있다. Set#1, Set#2에서 공통적으로 군집화된 단어 영역은 girl, beauty, leg, food, drink, arecales, sea, tree, vehicle이다. 촬영된 사진이 주로 인물, 음식, 장소사진이고, 배경으로 나무와 자동차가 많이 촬영되어 vehicle, tree 또한 양쪽에서 모두 각각의 그룹으로 분할되었다. Set #1과 Set#2 모두에는 스마트폰 화면을 캡처한 영상이 상당수 포함되어 있는데 이는 Set#1에서 font, Set#2에서 text로 인식되었다. 공통적으로 나타나지 않는 그룹들은 Set#1과 Set#2가 다른 장소에서 촬영되었기 때문인데 이 역시 시각화를 통해서 확인 할 수 있다.

어떤 단어들이 서로 유사한 단어로 군집화 되었는지 파악하기 위해 많은 단어로 군집화된 그룹부터 적은 수의 단어로 군집화된 그룹을 선택해 해당하는 그룹의 단어를 전부 나타내면 표 2와 같다. food 그룹의 경우 61개의 단어가 유사한 단어로 군집화 되었고, 가장 작은 child 그룹은 8개의 단어가 군집화 되었다. 표 2에 나타나듯이 food 그룹에는 주로 음식관련 단어들이, arecales 그룹에는 야외 촬영 단어들이, child 그룹에는 주로 단체 활동사진에서 나타내는 단어들이 포함되었다. 표 2를 보면 유사한 문맥을 갖는 단어끼리 군집화 된 것을 어느 정도 확인할 수 있다.

이렇게 문맥적으로 관련된 단어들이 함께 묶이는 듯이 보이지만 이것이 문맥적으로 유사한 사진들이 함께 묶인다는 것을 의미하지는 않는다. 이유는 하나의 사진이 하나의 그룹에 속하지 않고 보통 다수의 그룹에 속한다. 예를 들어 270번 사진은 yellow, light, eye, hand, child, ear, sunlight, girl, organ, finger 단어로 이루어져 있는데, 이들 10개의 단어는 특정 1개 그룹에 해당하지 않고 child, yellow, hairstyle, product, girl, long hair, light 등 7개의 그룹에 속한다. 그림 7 a)에 타원으로 표기된 부분이 이에 해당한다. 하나의 사진이 하나의 그룹에만 속할 경우 사진들을 워드임베딩된 특징만 가지고도 문맥적으로 유사한 사진끼리 군집화 할 수 있는데 시각화 결과에서 볼 수 있듯 하나의 단어 그룹으로 군집화 되지 않는다.

표 2. Set#1에서 단어 군집화 결과

Table 2. Word Clustering Results in Set # 1

representative word	count	word
food	61	meal, dish, cuisine, lunch, eating, asian food, breakfast, comfort food, recipe, appetizer
arecales	32	sky, house, city, property, palm tree, town, residential area, neighborhood, transport, real estate, condominium mixed use, architecture
font	27	text, line, screenshot, web page, document, paper, number, media, software, website
child	8	people, child, day, social group, youth, community, crowd, public space

표 3. 워드 사진에서 인식된 단어 수

Table 3. Number of words recognized in the picture

# of group	1	2	3	4	5	6	7	8
Set#1	78	170	278	220	140	69	16	0
Set#2	104	247	274	229	113	32	3	1

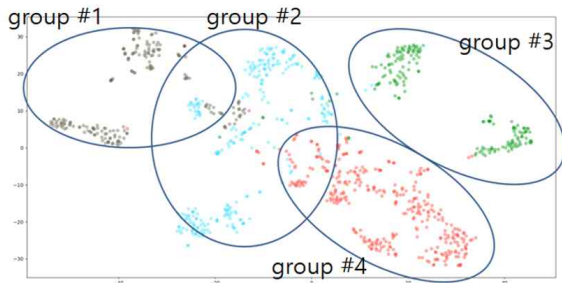
명확한 분석을 위해 사진 한 장 한 장이 10개의 단어 그룹 중 몇 개의 그룹에 속해있는지 계산해 표 3에 나타내었다. Set#1에서 한 장의 사진이 한 개의 단어 그룹에 속하는 경우는 78장이다. 그러나 최대 7개의 단어 그룹으로 분할되는 사진도 16장이나 된다. Set#2에서는 한 장의 사진이 8개의 단어 그룹에 걸쳐 나타나기도 한다. 표 3을 보면 대다수의 사진은 2~5개 정도의 그룹에 동시에 속한다. 따라서 워드임베딩 공간 분포만을 가지고 영상을 군집화 할 수 없음을 확인할 수 있다.

3) 이미지 벡터 스페이스 분석

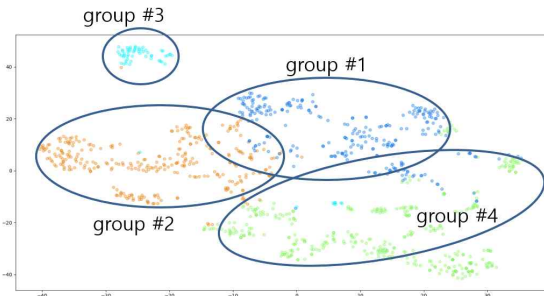
하나의 사진은 최대 10개의 단어벡터로 이루어져 있기 때문에 본 논문에서는 영상 하나를 하나의 특징 벡터로 표현하기 위해 단어벡터에 대한 평균 벡터를 사용했다. 이렇게 평균 벡터를 사용해 사진들을 군집화해서 유사한 사진끼리 군집화 되는지 분석하기 위해 식 (3)으로 계산한 I'에 대해서 가우시안 혼합 모델을 사용해 k개로 군집화하고 시각화하기 위해 PCA를 사용해 2차원으로 축소하였다. Set#1과 Set#2 각각에 대한 결과는 그림 8과 같다.

실험에서 가우시안 혼합 모델을 계산 할 때 k=4를 사용했다. 본 논문에서 다루고 있는 1:N이나 N:N 유형에서 선택해야 하는 사진이 많아지면 문제의 난이도가 어려워지고 일반적으로 4지선다에 익숙하기 때문이라고 설명했다.

그림 8 (a)를 보면 group#2 와 group#4는 사진들이 대체로 하나의 그룹으로 분포되어 있으나 group#1과 group#3은 두 개의



(a) 4개로 군집화 하기(#Set1,972개 평균벡터 분포)



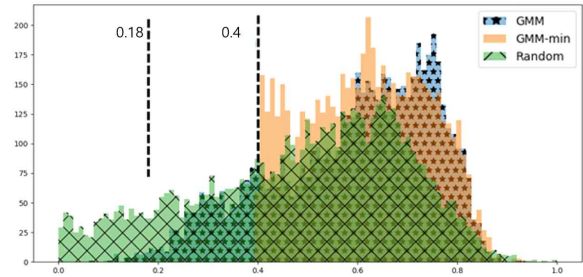
(b) 4개로 군집화 하기(#Set2,1003개 평균벡터 분포)

그림 8. 이미지 군집화 결과
Fig. 8. Image Clustering Results

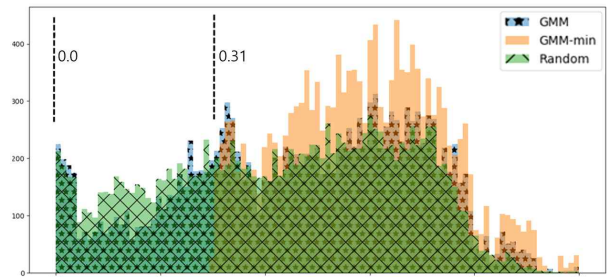
분할된 영역으로 나누어져 있다. 이는 group#1과 group#2의 경우 단어임베딩 공간에서는 서로 다른 영역이고, 문맥 공간에서는 같은 영역임을 의미한다. 즉 단어만 보면 다른 그룹이지만 전체 사진에서 발생하는 문맥을 보면 같은 문맥을 같이 사진이라는 것을 의미한다. 그림 8 (b)에서는 4개의 그룹이 각각 하나의 그룹으로 분포되어 있는데 group#3의 경우 분포 영역이 넓지 않다. 이는 특정 단어들로 사진들이 구성되어 있을 의미한다.

마지막 실험은 두 개의 그룹에서 두 개의 이미지를 선택할 때 최소 거리가 유지될 수 있도록 한 식 (6)의 제약 조건 성능을 검증한다. 그림 8 (b)에서 볼 수 있듯 group#2와 group#3은 양 끝 쪽에 있는 영상의 경우 문맥적으로 매우 가깝기 때문에 이러한 영상들이 선택되면 안된다. 제안하는 방법으로 사진을 군집화 할 때의 성능을 검증하기 위해 세가지 방법을 비교하였다. 랜덤방식은 전체 사진에서 무작위로 두 개의 사진을 선택하는 방법이다. GMM 방법은 Γ 를 가우시안 혼합 모델로 군집화 한 후 각 그룹에서 랜덤하게 사진을 선택하는 방법이다. 세 번째 방법은 두 번째 방법에서 식 (6)의 제약 조건을 사용한 경우이다. k 값에 따라 어떠한 특성을 보이는지 분석하기 위해 $k=4$, $k=6$ 일 때 1000번의 반복실험을 하고 선택된 사진간 거리를 측정하였다. 그림 9는 실험결과에 대한 히스토그램 분포이다.

그림 9 (a)를 보면 Random 방법은 두 이미지 거리가 0에 근접하는 경우가 매우 빈번하게 발생한다. 반면 GMM 방법은 거리 최소값의 분포가 0.18 부근으로 이동했기 때문에 Random 방법보다는 문맥적으로 다른 사진이 선택될 확률이 높아진다. 그러나 본문에서 설명했듯이 서로 근접한 가장자리에 있는 사진



(a) $k = 4$



(b) $k = 6$

그림 9. Random, GMM, GMM-min 방법 비교
Fig. 9. Comparison of Random, GMM, GMM-min Methods

들이 선택될 수 있다. 반면 GMM-min 방법은 \min 값을 0.4로 하면 최소 거리가 0.4 이상이 되기 때문에 선택된 두 이미지의 문맥간 거리가 커지게 된다. 따라서 제안하는 방법은 분포가 오른쪽으로 치우쳐져 있다.

그림 9 (b)는 k 가 (a) 경우 보다 조금 큰 경우 인데, 랜덤 방법이나 GMM 방법이나 큰 차이가 없다. GMM도 최소 거리가 0 가까이 몰려 있다. 그 이유는 6개의 그룹으로 나누어져 있을 때 그룹간 거리가 너무 가까워져서 유사한 두 개의 영상들이 선택될 확률이 높기 때문이다. 반면 제안하는 알고리즘인 GMM-min 방법에서는 최소 거리가 0.31이다. 분할되는 그룹이 많으면 그룹간 거리가 가까워지기 때문에 이에 따라 θ 값은 작아져야 한다. 제안하는 시스템에서는 사진이 최대 네장까지 제시되기 때문에 θ 값으로 0.4를 사용했다. 그림 9로부터 제안하는 알고리즘의 성능이 뛰어난 것을 확인할 수 있다.

IV. 결론

본 논문에서는 word2vec 알고리즘을 이용한 유사도 기반 군집화 알고리즘을 제안하고 이를 가상현실기반 인지재활 시스템에 응용하였다. 이를 위해 사진 영상에서 단어 목록을 추출한 다음 word2vec 알고리즘을 사용해 다차원 단어 벡터로 변환하였다. 하나의 사진이 다수의 단어 벡터로 구성되어 있기 때문에 이를 평균화하고 가우시안 혼합 모델링을 사용해 전체 사진 영상을 군집화 하였다. 그리고 서로 다른 두 개의 그룹에서 효과적인 사진 영상을 선택하기 위해 최소 거리가 유지되도록 하는 GMM-min 방법을 제안하였다. 2022장의 사진 영상을 사용해 이미지를 문맥적으로 군집하는 것이 효과적임을 분석

하였다. 향후 인지재활 치료를 위해 가상현실 기술을 융합한 연구는 치료의 효과성을 높일 수 있기 때문에 임상 실험과 더불어 다양한 연구가 필요해 보인다.

감사의 글

본 연구는 2019년도 강남대학교 교내 연구비 지원 사업에 의하여 이루어진 연구로서, 관계부처에 감사드립니다.

참고문헌

[1] Y. G. Kim, "Development of a Korean Computer-based Cognitive Rehabilitation Program (CoTras) for Patients with Cognitive Disabilities and the Validation of it's Effects", Pusan National University in South Korea, 34-9,2011.

[2] W. H. Kim and J. H. Chang, "A system for efficient checkup and management of dementia using a smart-phone", *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, 7(1): 577-585,2017.

[3] E. Richard, V. Billaudeau, P. Richard and G. Gaudin, "Augmented Reality for Rehabilitation of Cognitive Disabled Children: A Preliminary Study", *Virtual Rehabilitation*, 2017.

[4] KwonTaeg Choi, "Image Recognition and Clustering for Virtual Reality based on Cognitive Rehabilitation Contents," *Journal of Digital Contents Society*, Vol. 18, No. 7, pp. 1249~1257, 2017.

[5] C. S. Kim, and Y. H. Kwon, "Therapeutic Virtual Reality Program in Chronic Stroke Patients Recovery of Upper Extremity and Neuronal Reorganization", *Journal of Special Education & Rehabilitation Science*, 44(1), 87-106, 2005.

[6] Flynn, D., Van Schaik P., Blackman, T., and Femcott, C., "Developing a Virtual Reality-Based Methodology for People with Dementia: A Feasibility Study", *CyberPsychology*, 6(6), 591-611,2003.

[7] Larson EB, Feigon M, Gagliardo P3 and Dvorkin AY, "Virtual reality and cognitive rehabilitation: a review of current outcome research", *NeuroRehabilitation*, 34(4):759-72, 2014

[8] M. I. Khan, A. Prado, and S. K. Agrawal, "Effects of Virtual Reality Training With Trunk Support Trainer (TruST) on Postural Kinematics Sign In or Purchase", *IEEE Robotics and Automation Letters*, 2(4), Oct. 2017.

[9] Park E, Yun BJ, Min YS, Lee YS, Moon SJ, Huh JW, Cha H, Chang Y, and Jung TD, "Effects of a Mixed Reality-based Cognitive Training System Compared to a Conventional Computer-assisted Cognitive Training System on Mild

Cognitive Impairment: A Pilot Study.", *Cogn Behav Neurol*. 32(3),172-178,Sep,2019

[10] Young Ran Kim, "The Effects of the Reminiscence Therapy Program on the Perception of the Elderly Suffering from Dementia", *Journal of the Korean Academy of Health and Welfare for Elderly*, Vol. 7, No. 1, 19-36,2015.

[11] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient Estimation of word Representations in Vector Space", arXiv:1301.3781v3, Sep. 2013.

[12] Jey Han Lau, Timothy Baldwin, "An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation", *Proceedings of the 1st Workshop on Representation Learning for NLP*, Berlin, Germany, pp. 78-86.

[13] Kim, H. K.; Kim, H.; Cho, S., "Bag-of-concepts: Comprehending document representation through clustering words in distributed representation", *Neurocomputing*, vol. 266, pp. 336-352, 2017.

[14] S. Reddy et al., "Image Browsing, Processing, and Clustering for Participatory Sensing: Lessons from a Dietsense Prototype", *Proc. EmNets*, pp. 13 - 17,2007.



최권택(KwonTaeg Choi)

2001년 : 한림대학교 컴퓨터공학과(이학 학사)

2006년 : 연세대학교 컴퓨터과학과(이학 석사)

2011년 : 연세대학교 컴퓨터과학과 (공학박사-컴퓨터비전, 패턴인식)

2011년~2015년: LG전자

2016년~현 재: 강남대학교 컴퓨터미디어정보공학부 교수

2017년~현 재: 강남대학교 소프트웨어응용학부 교수

※ 관심분야 : 가상현실, 모바일컴퓨팅, 기계학습, HCI