

시맨틱 출력 코드를 사용한 제로샷 학습

Mark Palatucci
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
mpalatuc@cs.cmu.edu

학장 Pomerleau
Intel Labs
Pittsburgh, PA 15213
dean.a.pomerleau@intel.com

Geoffrey Hinton 토
론토 대학교 컴퓨터 과학과 토론토, 온
타리오 M5S 3G4, 캐나다
hinton@cs.toronto.edu

Tom M. Mitchell 기
계 학습 부서 Carnegie Mellon
University Pittsburgh, PA
15213
tom.mitchell@cs.cmu.edu

추상적인

훈련 세트에서 생략된 Y 의 새로운 값을 예측해야 하는 분류자 $f: X \rightarrow Y$ 를 학습하는 것이 목표인 제로샷 학습의 문제를 고려합니다. 이를 달성하기 위해 새로운 클래스로 추정하기 위해 Y 의 의미론적 속성에 대한 지식 기반을 활용하는 의미론적 출력 코드 분류기(SOC)의 개념을 정의합니다. 우리는 이러한 유형의 분류기에 대한 형식주의를 제공하고 PAC 프레임워크에서 이론적 속성을 연구하여 분류기가 새로운 클래스를 정확하게 예측할 수 있는 조건을 보여줍니다. 사례 연구로서 우리는 신경 디코딩 작업을 위한 SOC 분류기를 구축하고 해당 단어에 대한 훈련 예제 없이도 신경 활동의 기능적 자기 공명 이미지(fMRI)에서 사람들이 생각하는 단어를 종종 예측할 수 있음을 보여줍니다.

1. 소개

기계 학습 알고리즘은 컴퓨터 비전, 사기 탐지 및 뇌 이미지 분석과 같은 많은 영역에서 학습 분류기에 성공적으로 적용되었습니다. 일반적으로 분류기는 Y 에 대한 모든 가능한 값을 포함하는 레이블이 지정된 훈련 데이터 세트와 때때로 레이블이 지정되지 않은 추가 훈련 데이터 세트가 주어지면 목표 함수 $f: X \rightarrow Y$ 를 근사화하도록 훈련됩니다.

제로샷 학습에 대한 연구는 거의 수행되지 않았으며 클래스 변수 Y 에 가능한 값에는 훈련 예제에서 생략된 값이 포함됩니다. 이것은 특히 Y 가 많은 값을 가질 수 있고 모든 값에 대해 레이블이 지정된 예제를 얻는 비용이 높은 도메인에서 중요한 문제 설정입니다. 한 가지 명백한 예는 컴퓨터가 인식하기를 원하는 수만 개의 객체가 있는 컴퓨터 비전입니다.

또 다른 예는 신경 활동 디코딩에서 목표는 사람의 신경 활동 이미지를 관찰하여 사람이 생각하고 있는 단어나 물체를 결정하는 것입니다. 영어로 가능한 모든 단어에 대한 신경 훈련 이미지를 수집하는 것은 다루기 어렵기 때문에 실용적인 신경 디코더를 구축하려면 훈련 세트에 있는 단어를 넘어서는 단어를 인식하도록 외삽하는 방법이 있어야 합니다.

이 문제는 분류기 훈련 중에 단어를 명시적으로 포함하지 않고 단어를 인식하는 것이 바람직한 자동 음성 인식의 문제와 유사합니다. 어휘 독립성을 달성하기 위해 음성 인식 시스템은 일반적으로 음소 기반 인식 전략을 사용합니다(Waibel, 1989). 음소는 언어의 단어를 구성하기 위해 결합될 수 있는 구성 요소입니다. 음성 인식 시스템은 상대적으로 작은 음소 집합을 활용하여 성공합니다.

음소의 조합으로 단어를 나타내는 대규모 지식 기반과 함께 인식기.

신경 활동 디코딩에 유사한 접근 방식을 적용하려면 신경 활동에서 단어 의미의 구성 요소 부분을 추론하는 방법을 발견해야 합니다. 뇌가 의미론적 정보를 인코딩하는 방법에 대한 명확한 합의는 없지만(Plaut, 2002), 신경 활동의 지식 기반 역할을 할 수 있는 몇 가지 제안된 표현이 있으므로 신경 디코더가 가능한 많은 단어 세트를 인식할 수 있습니다. 해당 단어가 훈련 세트에서 생략된 경우에도 마찬가지로 작업합니다.

이 문서에서 묻는 일반적인 질문은 다음과 같습니다.

대규모 개념 클래스 집합의 의미론적 인코딩이 주어지면 훈련 집합에서 생략된 클래스를 인식하는 분류기를 구축할 수 있습니까?

우리는 이 질문을 해결하기 위한 공식적인 프레임워크와 신경 활동 디코딩 작업에 대한 구체적인 예를 제공합니다. 특정 단어에 대한 훈련 예제 없이도 사람이 생각하는 단어를 인식할 수 있는 분류기를 구축하는 것이 가능함을 보여줍니다.

1.1 관련 작업

제로 샷 학습 문제는 기계 학습 커뮤니티에서 거의 주목을 받지 못했습니다. 일부 작업은 Larochelle et al. (2008)은 제로 데이터 학습에 대해 훈련 세트에서 생략된 숫자의 새로운 클래스를 예측하는 능력을 보여주었습니다. 컴퓨터 비전에서 객체 클래스 간에 기능을 공유하는 기술이 조사되었지만(Torralba & Murphy, 2007; Bart & Ullman, 2005) Lampert et al. (2009) 새로운 객체의 시각적 속성 예측 및 Farhadi et al. (2009) 객체 인식을 위한 시각적 속성 예측 사용.

신경 영상 커뮤니티에서 Kay et al. (2008)은 (시각 피질 활동에서) 사람이 가능한 많은 이미지 세트에서 보고 있는 새로운 시각적 장면을 해독하는 능력을 보여 주었지만 이미지 내용 자체를 인식하지 못했습니다.

우리 작업과 가장 유사한 작업은 Mitchell(2008)입니다. 말뭉치 통계에서 파생된 의미론적 기능을 사용하여 영어 명사에 대한 신경 활동 패턴을 생성합니다. 대조적으로 우리 작업에서는 새로운 신경 이미지가 주어지면 단어 해독에 초점을 맞추고 가능한 많은 단어 집합에서 단어를 예측하려고 합니다. 또한 말뭉치 통계에 더해 사람의 라벨링에서 파생된 의미론적 특징도 고려합니다. 또한, 우리는 제로 샷 학습자를 위한 형식주의를 도입하고 훈련 세트에서 누락된 새로운 클래스를 인식하는 능력에 대한 이론적 보증을 제공합니다.

2 의미론적 지식을 이용한 분류

이 섹션에서는 의미론적 지식을 사용하여 새로운 클래스로 추정하는 제로 샷 학습자의 개념을 공식화합니다. 제로샷 학습자는 다양한 형태를 취할 수 있지만, 우리는 의미론적 지식 기반에서 파생된 중간 기능 세트를 활용하는 그러한 모델을 제시합니다. 직관적으로 우리의 목표는 각 클래스를 단순히 원자 레이블로 취급하는 것이 아니라 가능한 많은 클래스를 특징짓는 의미론적 특징의 벡터를 사용하여 나타내는 것입니다. 우리의 모델은 입력 데이터와 시맨틱 기능 간의 관계를 학습합니다. 새로운 입력 데이터에 대한 클래스 레이블을 복구하기 위해 2단계 예측 절차에서 이 학습된 관계를 사용할 것입니다. 새로운 입력 데이터가 주어지면 모델은 해당 입력에 해당하는 일련의 의미론적 특징을 예측한 다음 지식 기반에서 예측된 특징 집합과 가장 일치하는 클래스를 찾습니다. 중요하게도, 이 절차는 해당 클래스가 의미론적 지식 기반에 포함된 경우(즉, 클래스에 사용할 수 있는 입력 공간 표현이 없지만 해당 클래스의 기능 인코딩이 의미론적 지식 기반에 존재하는 경우) 새로운 클래스의 입력 데이터에 대해서도 작동합니다.).

2.1 형식주의

정의 1. 시맨틱 특징 공간 p 차원의 시맨틱 특징 공간은 각 p 차원이 의미 속성 값을 인코딩하는 메트릭 공간입니다. 이러한 속성은 본질적으로 범주형이거나 실제 값 데이터를 포함할 수 있습니다.

정의 2. 의미론적 지식 베이스 M개의 예의 의미론적

동물의 지식 기반은 많은 동물에 대한 시맨틱 인코딩 및 레이블을 포함합니다.

$$\mathbb{P} : Fp \rightarrow Y$$

이 질문에 대답할 때 우리의 목표는 PAC 스타일 경계를 얻는 것입니다. 우리는 높은 확률로 새로운 클래스를 복구하면서 의미론적 속성의 예측에서 얼마나 많은 오류를 허용할 수 있는지 알고 싶습니다. 그런 다음 이 오류 범위를 사용하여 분류기의 첫 번째 단계에서 해당 수준의 오류를 달성하는 데 필요한 예제 수에 대한 범위를 얻습니다. 아이디어는 분류기의 첫 번째 단계 $S(\cdot)$ 가 의미론적 속성을 잘 예측할 수 있다면 두 번째 단계 $L(\cdot)$ 이 새로운 클래스에서 인스턴스에 대한 올바른 레이블을 복구할 수 있는 좋은 기회를 갖게 된다는 것입니다.

제로 샷 학습의 일반 이론을 향한 첫 번째 단계로 의미론적 출력 코드 분류기의 인스턴스화를 고려할 것입니다. 의미론적 특징이 이진 레이블이고 첫 번째 단계 $S(\cdot)$ 는 PAC 학습 가능한 선형 분류기 모음(특징당 하나의 분류기)이고 두 번째 단계 $L(\cdot)$ 은 해밍 거리 메트릭. 이러한 가정을 통해 선형 분류기에 대한 기존 PAC 이론과 근사 최근접 이웃 검색에 대한 이론을 활용할 수 있습니다. 최근접 이웃 분석의 대부분은 Ciaccia와 Patella(2000)의 작업과 유사합니다.

우리는 먼저 의미론적 특징에 대한 예측이 주어지면 허용할 수 있는 오류의 양을 제한하려고 합니다. 이 범위를 찾기 위해 우리는 F 를 지식 기반 K 에서 점의 의미론적 특징 공간의 분포로 정의합니다. 분명히 의미론적 공간의 점(클래스)은 서로 등거리에 있지 않을 수 있습니다. 한 포인트가 다른 포인트와 멀리 떨어져 있을 수 있으므로 고유한 ID(레이블)을 복구할 수 있는 기능을 유지하면서 이 포인트에 대한 의미론적 특징을 예측할 때 더 많은 오류를 허용할 수 있습니다. 반대로 의미론적 공간에서 다른 지점에 가까운 지점은 오류 허용 오차가 낮습니다. 요컨대, 오류에 대한 허용 오차는 의미 공간의 다른 지점과 관련하여 특정 지점에 상대적입니다.

다음으로 예측 q 를 원시 입력 예제 $x \in X_d$ 에 적용된 $S(\cdot)$ 맵의 출력으로 정의합니다. 의미 공간. 점 q 에 대한 상대 분포 R_q 를 q 에서 q 까지의 거리가 어떤 거리 z 보다 작을 확률로 정의합니다.

$$R_q(z) = P(d(q, q) \leq z)$$

이 경험적 분포는 F 에 따라 다르며 q 에서 z 만큼 떨어진 F 에서 샘플링된 점의 비율입니다. 이 분포를 사용하여 η_q 로 정의되는 q 의 가장 가까운 이웃까지의 거리에 대한 분포를 정의할 수도 있습니다. $G_q(z) = P(\eta_q \leq z)$

이는 Ciaccia(2000)에서 다음과 같이 제시됩니다.

$$G_q(z) = 1 - (1 - R_q(z))^n$$

여기서 n 은 분포 F 에서 가져온 실제 포인트의 수입니다. 이제 τ_q 를 원시 입력 예 x 에 대한 예측 q 가 속한 클래스의 실제 의미론적 인코딩으로부터의 거리로 정의한다고 가정합니다. 직관적으로 입력 x 에 대해 추론하는 클래스는 예측 q 에 가장 가까운 지점이 될 것이므로 실제 클래스까지의 거리 τ_q 가 q 와 가장 가까운 이웃 간의 거리보다 클 확률 γ 를 원합니다. 의미 공간에서 q 의 실제 클래스를 나타내는 점보다 q 에 더 가까운 스퍼리어스 이웃이 있음을 의미합니다.

$$P(\tau_q \geq \eta_q) \leq \gamma$$

재정렬하면 이것을 분포 G_q 로 표현한 다음 τ_q 에 대해 풀 수 있습니다.

$$\begin{aligned} P(\eta_q \leq \tau_q) &\leq \gamma \\ G_q(\tau_q) &\leq \gamma \end{aligned}$$

$G_q(\cdot)$ 가 가역적이라면 원하는 γ 에 대한 값 τ_q 를 즉시 복구할 수 있습니다. 일부 분포의 경우 $G_q(\cdot)$ 는 1:1 함수가 아니므로 반전이 없을 수 있습니다. 그러나 $G_q(\cdot)$ 는 누적 분포 함수이기 때문에 절대 감소하지 않습니다. 따라서 함수 G^{-1} 을 정의합니다.

다음과 같이: $G^{-1}(\gamma) = \arg\max_{\tau_q} G_q(\tau_q) \leq \gamma$.

따라서 $L(\cdot)$ 에 대해 가장 가까운 이웃을 사용하면 $\tau_q \leq G^{-1}(\gamma)$ 인 경우 최소 $1 - \gamma$ 확률로 올바른 클래스를 복구합니다. 이 오류 범위를 달성하려면 $S(\cdot)$ 의 총 오류가 τ 로 정의하는 $G^{-1}(\gamma)$ 보다 작아야 합니다. 이 분석에서 우리는 p 개의 이진 의미 기능과 해밍 거리 메트릭이 있다고 가정하므로 τ 는 총계를 정합니다.

이진 기능을 예측할 수 있는 실수의 수입니다. 각 사맨틱 기능은 광고 차원 원시 입력 공간에서 선형 분류기를 사용하여 PAC 학습 가능하다는 가정에 유의하십시오. 분석을 단순화하기 위해 각 p 의미 기능을 독립적으로 학습된 것으로 취급합니다. PAC 가정에 의해, 각 p 학습된 가설의 실제 오류(즉, 분류기가 실수할 확률)는 p 의미론적 특징에 대한 예상 오류 수는 다음과 같습니다.

최대 τ q / 피. 또한 최대 τ 를 만들 확률

이항 분포: $\text{BinoCDF}(\tau_{\max}/p); p, \tau_{\max}$

VC-dimension1 (Mitchell, 1997)에 대한 표준 PAC 경계를 활용하여 각 가설에 대해 원하는 오류율을 얻을 수 있습니다. 실제 오류가 있는 $(1 - \delta)$ 확률로 가설을 얻으려면

가장해야 τ q / p = G $1(\gamma)/p$ 이면 분류기에는 Mq, δ 의 여러 예가 필요합니다.

$$Mq, \delta \geq \frac{\pi}{\tau_{\max} q} 4 \log(2/\delta) + 8(d + 1) \log(13p/\tau_{\max}) \tag{1}$$

p개의 분류자(특징 예측자) 각각이 이렇게 많은 예를 통해 학습되면 확률 $(1 - \delta)$ p로 모든 특징 예측자가 원하는 오류율을 달성할 것입니다. 그러나 이것은 원시 입력에 오류율 p 에 대해 학습을 달성할 확률일 뿐입니다. 실제 오류율을 학습자 p 에 대한 확률로 설정합니다. 따라서 의미론적 출력 코드 분류기가 새로운 클래스를 예측하는 데 필요한 세 가지 확률적 이벤트가 있으며 이러한 이벤트의 총(공동) 확률은 다음과 같습니다.

$P(\text{실제 오류가 } \leq \tau \text{인 } p \text{개의 특징 예측자가 있습니다.}) = \prod_{i=1}^p P(\tau_i \leq \tau)$

$P(\text{최대 } \tau \text{ 실수} \mid \text{실제 오류 } \leq \tau \text{인 } p \text{개의 특징 예측자가 있습니다.}) = \prod_{i=1}^p P(\tau_i \leq \tau \mid \text{실수})$

$P(\text{진정한 클래스 복구} \mid \text{최대 } \tau \text{ 실수})$

그리고 τ 이후 $\tau = G \frac{1}{p}(\gamma)$, 총 확률은 다음과 같이 지정됩니다.

$$(1 - \delta) \pi \cdot \text{비노CDF}(G \frac{1}{p}(\gamma); p, G \cdot 1(\gamma)/p) \cdot (1 - \gamma) \tag{2}$$

요약하면 두 분류기 단계에 대해 원하는 오류 매개변수 $(1 - \gamma)$ 및 $(1 - \delta)$ 가 주어지면 방정식 2는 새로운 클래스를 정확하게 예측할 총 확률을 제공합니다. γ 에 대한 값이 주어지면 각 기능 예측기에 필요한 것을 계산할 수 있습니다. 수식 1에 지정된 대로 기능 예측자가 Mq, δ 원시 입력 예제로 훈련된 경우 총 확률을 얻을 수 있습니다.

우리가 아는 한, 방정식 1과 2는 분류자가 새로운 클래스를 예측할 수 있는 조건을 제공하는 첫 번째 공식 보증을 지정합니다.

4 사례 연구: 참신한 생각의 신경 해독

이 섹션에서는 신경 디코딩 작업에서 의미론적 출력 코드 분류기를 경험적으로 평가합니다. 목적은 훈련 중에 해당 단어의 fMRI 이미지 예를 포함하지 않고 사람의 신경 활동에 대한 fMRI 이미지에서 사람이 생각하고 있는 새로운 단어를 해독하는 것입니다.

4.1 데이터세트

우리는 Mitchell(2008)의 동일한 fMRI 데이터 세트를 활용했습니다. 이 데이터 세트에는 60개의 서로 다른 구체적인 단어(12개의 서로 다른 범주에서 5개의 예)를 보는 동안 9명의 인간 참가자로부터 관찰된 신경 활동이 포함되어 있습니다. 몇 가지 예로 곰, 개, 고양이, 소, 말과 같은 동물과 트럭, 자동차, 기차, 비행기, 자전거와 같은 차량이 있습니다. 각 참가자에게는 단어와 단어가 나타내는 구체적인 대상의 작은 그림이 표시되었습니다. 참가자들은 뇌 활동의 이미지가 기록되는 동안 몇 초 동안 이러한 물체의 속성에 대해 생각하도록 요청 받았습니다.

각 이미지는 뇌의 약 20,000개 위치(즉, вок셀)에서 신경 활동을 측정합니다. 각 단어에 대해 6개의 fMRI 스캔을 수행했습니다. 우리는 Mitchell(2008)에서 설명한 동일한 시간 평균을 사용하여 각 참가자에 대해 60개 단어 각각에 대한 단일 평균 뇌 활동 패턴을 생성했습니다.

1d 차원에서 선형 분류기의 VC 차원은 $d + 1$ 입니다.

시맨틱 출력 코드 분류기의 언어에서 이 데이터 세트는 원시 입력 공간 예제의 컬렉션 D를 나타냅니다.

또한 이 60개 단어에 대한 두 개의 의미론적 지식 기반을 수집했습니다. 첫 번째 시맨틱 지식 예지 기반인 corpus5000에서 각 단어는 Google Trillion-Word-Corpus2에서 가장 빈도가 높은 5000개의 단어와 동시 발생 벡터로 표현됩니다 .

두 번째 시맨틱 지식 기반인 human218은 Amazon.com의 Mechanical Turk 인간 계산 서비스를 사용하여 생성되었습니다. 60개의 단어에 대해 218개의 의미적 특징이 수집되었으며, 신경 활동 부호화에 대한 심리적 추측을 반영하기 위해 질문이 선택되었습니다. 예를 들어, 크기, 모양, 표면 특성 및 일반적인 용도와 관련된 질문입니다.

예시 질문에는 인공적인가?가 포함됩니다. 그리고 당신은 그것을 잡을 수 있습니까?. Mechanical Turk 서비스 사용자는 각 단어에 대해 1에서 5까지의 척도로 이러한 질문에 답했습니다(확실히 그렇지 않지만 확실히 예임).

4.2 모델

우리의 실험에서는 다중 출력 선형 회귀를 사용하여 의미적 출력 코드 분류기의 $S(\cdot)$ 맵을 학습합니다. $X \in \mathbb{R}^{d \times p}$ 를 각 행이 특정 단어의 이미지이고 d 가 fMRI 이미지의 차원 수인 fMRI 예제의 출력 벡터라고 가정합니다. 여기서 p 는 해당 단어에 대한 의미론적 특징의 수입니다(예: human218 지식 기반의 경우 218). 우리는 가중치 $W \in \mathbb{R}^{d \times p}$ 의 행렬을 학습합니다. 이 모델에서 각 출력은 독립적으로 처리되므로 하나의 행렬 작업으로 모든 출력을 신속하게 해결할 수 있습니다(수천 개의 시맨틱 기능 포함).

$$y = (X^T X + \lambda I)^{-1} X^T y \tag{3}$$

여기서 I 는 항등 행렬이고 λ 는 교차 검증 점수 함수(Hastie et al., 2001)3를 사용하여 자동으로 선택된 정규화 매개변수입니다. 새로운 fMRI 이미지 x 가 주어지면 이미지에 가중치를 곱하여 이 이미지에 대한 의미론적 특징의 예측 f 를 얻을 수 있습니다. $f = x \cdot W$

의미론적 출력 코드 분류기의 두 번째 단계인 $L(\cdot)$ 에서는 단순히 1-최근접 이웃 분류기를 사용합니다. 즉, $L(f)$ 는 피쳐 예측을 수행하고 유클리드 거리 (L_2) 메트릭에 따라 주어진 지식 기반에서 가장 가까운 지점을 반환합니다 .

4.3 실험

위에서 설명한 모델과 데이터 세트를 사용하여 이제 세 가지 중요한 질문을 제기하고 답변합니다.

- 1. 훈련 세트에 어떤 클래스도 나타나지 않는 두 클래스를 구별하기 위해 분류기를 만들 수 있습니까?

이 질문에 답하기 위해 우리는 leave-two-out-cross-validation을 수행했습니다. 구체적으로, 우리는 58개의 fMRI 이미지와 각 단어에 대한 의미론적 특징 사이의 매핑을 학습하기 위해 방정식 3의 모델을 훈련했습니다. 첫 번째 보류 이미지에 대해 학습된 가중치 행렬을 적용하여 의미 특징의 예측을 얻은 다음 1-최근접 이웃 분류기를 사용하여 예측 벡터를 두 보류 단어의 실제 의미 인코딩과 비교했습니다. . 레이블은 fMRI 이미지에 대한 예측에 가장 가까운 인코딩으로 단어를 선택하여 선택되었습니다. 그런 다음 두 번째 보류 이미지를 사용하여 동일한 테스트를 수행했습니다. 따라서 교차 유효성 검사의 각 반복에 대해 두 가지 개별 비교가 이루어졌습니다. 이 프로세스는 모든 = 1,770개의 가능한 leave-two-out 조합에 대해 반복되어 총 3,540개의 비교가 이루어졌습니다.

60
2

표 1은 두 가지 다른 시맨틱 기능 인코딩에 대한 결과를 보여줍니다. 우리는 human218 의미론적 기능이 corpus5000 기능보다 훨씬 뛰어난 성능을 보였고, 9명의 참가자에 대한 평균 정확도는 각각 80.9%와 69.7%였습니다. 그러나 두 기능 세트에 대해 9명의 참가자 각각에 대해 두 개의 새로운 클래스를 구별할 수 있음을 알 수 있습니다.

2 벡터는 단위 길이로 정규화되며 a, the, is와 같은 100개의 불용어를 포함하지 않습니다.
3 각 작업에 대한 교차 검증 점수를 계산하고 평균을 최소화하는 매개변수를 선택합니다.
모든 출력 작업에서 손실.

표 1: 9명의 fMRI 참가자에 대한 leave-two-out-cross-validation의 백분율 정확도(P1-P9로 표시됨). 값은 훈련 세트에서 생략된 두 개의 fMRI 이미지를 구별할 때 3,540회 시도에 대한 분류기 백분율 정확도를 나타냅니다.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	평균	코퍼스5000	79.6	67.0	69.5	56.2
77.7	65.5	71.2	72.9	67.9	69.7	human218	90.3	82.9	86.6	71.9	89.5	75.3	78.0	77.7	76.2
80.2															

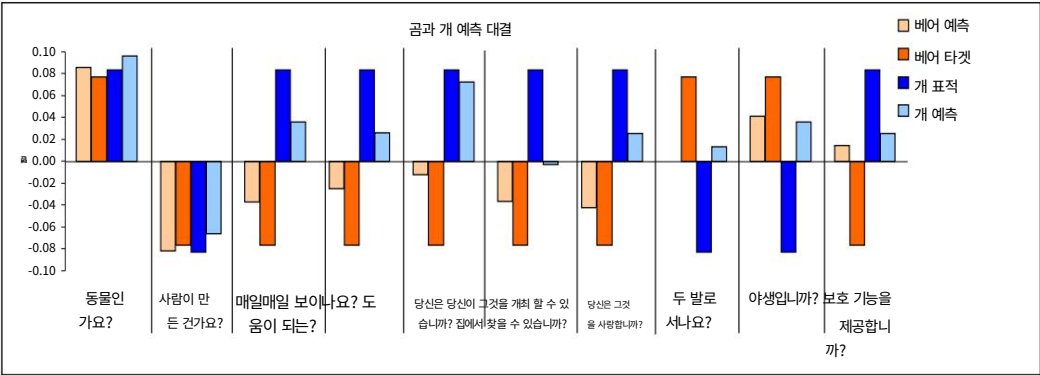


그림 1: 공과 개라는 단어에 대한 human218 지식 기반의 10가지 의미론적 특징. 실제 인코딩은 공과 개에 대한 fMRI 이미지가 훈련 세트에서 제외되었을 때 예측된 인코딩과 함께 표시됩니다.

2. 분류기는 밀접하게 관련된 소설 클래스를 어떻게 구별할 수 있습니까?

그림 1은 human218 데이터 세트의 10가지 의미론적 질문(특징)을 보여줍니다. 그래프는 다른 58개 단어에 대해 훈련되었을 때 공과 개 모두에 대해 예측된 특성 값과 함께 참 값을 보여줍니다. 우리는 모델이 공과 개가 공통적으로 가지고 있는 많은 주 요 기능(예: 동물입니까?)을 예측하는 방법을 학습할 수 있음을 확인했습니다. 뿐만 아니라 매일 보십니까?와 같이 두 가지를 구 별하는 것? 그리고 당신은 그것을 잡을 수 있습니까? 이 두 가지 새로운 단어에 대해 신경 데이터에서 예측된 특징이 실제 단어 에 가장 가깝습니다.

3. 가능한 많은 단어 세트에서 단어를 해독할 수 있습니까?

의미론적 출력 코드 분류기가 두 개의 새로운 단어에 대한 뇌 이미지를 구별하는 데 성공했기 때문에 이제 많은 후보 단어 세트에 서 새로운 단어를 구별하는 훨씬 더 어려운 문제를 고려합니다. 이 능력을 테스트하기 위해 우리는 59개 단어에 대한 이미지 및 의미론적 특징에 대해 방정식 3을 사용하여 훈련하는 leave-one-out-cross-validation을 수행했습니다. 그런 다음 60번째 단어의 홀드아웃 이미지에 대한 특징을 예측한 다음 많은 후보 단어 집합에서 1-최근접 이웃 분류를 수행했습니다.

우리는 두 가지 다른 단어 집합을 테스트했습니다. 첫 번째는 fMRI 데이터를 수집한 60개의 구체적인 명사 모음인 mri60이었습 니다. 여기에는 59개의 훈련 단어와 단일 단어가 포함되었습니다. 두 번째 집합은 noun940으로, Wilson(1988)과 Snodgrass(1980)에서 수집한 친근감, 구체성, 상상 가능성이 높은 940개의 영어 명사 모음입니다. 이 단어 집합의 경우 각 교 차 유효성 검사 반복에서 940개의 집합에 진정한 홀드아웃 단어를 추가했습니다. 우리는 corpus5000 및 human218 기능 세트를 모두 사용하여 이 실험을 수행했습니다. 네 가지 실험의 순위 정확도 결과(60회 이상의 교차 검증 반복)가 그림 2에 나 와 있습니다.

human218 기능은 평균 및 중간 순위 정확도 측정 모두에서 다시 말문치5000을 훨씬 능가하며 두 기능 세트 모두 기회를 훨 씬 증가합니다. mri60 단어의 총 540개 프레젠테이션 중 12개(참가자 9명당 60개 프레젠테이션)에서 human218 기능은 훈 련 세트의 다른 59개 단어보다 하나의 홀드아웃 단어를 예측했습니다. 확률 수준(9/540)보다 약간 높지만, 모델이 의심할 여지 없이 해당 단어와 유사한 특성 값을 예측하는 쪽으로 편향되어 있기 때문에 모델이 모든 훈련 단어보다 홀드아웃 단어를 선택한 다는 사실은 주목할 만합니다. 훈련. 명사 940단어에서 모델은 941개의 대안 세트에서 human218 기능에 대해 총 26번, corpus5000 기능에 대해 22번 올바른 단어를 예측했습니다. 일부 주제의 경우 모델이 올바른 항목을 올바르게 선택했습니다.

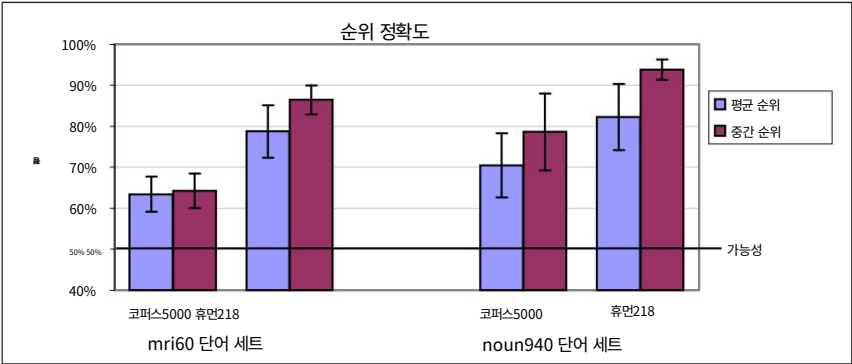


그림 2: 2개의 서로 다른 시맨틱 기능 세트에 대한 9명의 참가자에 대한 평균 및 중간 순위 정확도. 원래 60개의 fMRI 단어와 940개의 명사 세트가 모두 고려되었습니다.

표 2: 굵게 표시된 단어에 대해 촬영된 새로운 fMRI 이미지의 상위 5개 예측 단어(참가자 P1에서 가져온 모든 fMRI 이미지). 괄호 안의 숫자는 영어로 된 941개의 구체적인 명사에서 선택한 올바른 단어의 순위를 포함합니다.

곰발 스크루드라이버	가차	트럭	셀러리	(5)	(6)	(1)	(1)	(1)	(1)	(2)	(21)	곰발 스크루드	집	바지
트 슈퍼마켓	외부	화와	발	공	구멍	구멍	구멍	구멍	구멍	구멍	구멍	구멍	구멍	구멍

시간의 10% 이상 941 세트에서 단어. 단어를 정확하게 예측할 확률 정확도는 0.1%에 불과합니다. 즉, 540개의 모든 프레젠테이션에서 올바른 예측이 하나 미만일 것으로 예상됩니다.

그림 2에서 볼 수 있듯이 중간 순위 정확도는 종종 평균 순위 정확도보다 상당히 높습니다. noun940 단어에 human218 기능을 사용하면 평균 순위 정확도는 각 참가자에 대해 90% 이상인 반면 평균은 일반적으로 약 10% 낮습니다. 이는 여러 단어가 지속적으로 잘못 예측된다는 사실 때문입니다. 동물, 신체 부위, 음식, 도구 및 차량 범주의 단어 예측은 일반적으로 잘 수행되는 반면 가구, 인공 품목 및 곤충 범주의 단어는 종종 잘 수행되지 않습니다.

올바른 단어가 가장 근접한 일치나 아닌 경우에도 예측된 특징과 가장 일치하는 단어는 보류된 단어와 매우 유사한 경우가 많습니다. 표 2는 참가자 P1에 대한 8개의 서로 다른 fMRI 이미지에 대한 상위 5개의 예측 단어를 보여줍니다(즉, 941 세트에서 의미론적 특징의 예측 벡터에 가장 가까운 5개 단어).

5. 결론

시맨틱 출력 코드 분류기로 알려진 제로 샷 학습 알고리즘에 대한 형식을 제시했습니다. 이 분류기는 소설 클래스와 훈련 세트 모두에 공통적인 기능을 인코딩하는 의미론적 지식 기반을 활용하여 훈련 세트에서 생략된 새로운 클래스를 예측할 수 있습니다. 우리는 또한 이 분류자가 새로운 클래스를 예측하는 조건을 보여주는 최초의 공식적인 보증을 증명했습니다.

우리는 인간 라벨링 및 말뭉치 통계에서 파생된 의미론적 지식 기반을 사용하여 신경 디코딩 작업에 대해 이 의미론적 출력 코드 분류기를 시연했습니다. 우리는 이 분류기가 그 사람의 신경 활동에 대한 기록된 fMRI 이미지에서 우연보다 훨씬 높은 정확도로 사람이 생각하고 있는 단어를 예측할 수 있음을 보여주었습니다. 거의 1,000개의 대안 중에서 단어를 선택하십시오.

우리는 분류자가 인식하기를 원하는 모든 단어에 대해 뇌 활동의 훈련 이미지가 필요하지 않다는 것을 보여주었습니다. 이러한 결과는 신경 디코딩의 최신 기술을 크게 발전시키고 대규모 어휘 뇌-컴퓨터 인터페이스를 향한 유망한 단계입니다.

참조

- Bart, E., & Ullman, S. (2005). 교차 일반화: 기능 교체를 통해 단일 예제에서 새로운 클래스를 학습합니다. 컴퓨터 비전 및 패턴 인식, 2005. CVPR 2005. IEEE Computer Society Conference on, 1, 672–679 vol. 1.
- Ciaccia, P., & 슬개콜, M. (2000). PAC 최근접 이웃 쿼리: 고차원 및 미터법 공간에서 대략적이고 제어된 검색입니다. 데이터 공학, 국제 회의, 244.
- Dietterich, TG, & Bakiri, G. (1995). 오류 수정을 통한 다중 클래스 학습 문제 해결 출력 코드. 인공 지능 연구 저널.
- Farhadi, A., Endres, I., Hoiem, D., & Forsyth, D. (2009). 속성으로 개체를 설명합니다. 컴퓨터 비전 및 패턴 인식(CVPR)에 관한 IEEE Computer Society 회의 간행물.
- Hastie, T., Tibshirani, R., & Friedman, JH (2001). 통계 학습의 요소. 뛰는 것.
- Kay, KN, Naselaris, T., Prenger, RJ, & Gallant, JL (2008). 인간의 뇌 활동에서 자연 이미지를 식별합니다. 자연, 452, 352–355.
- Lampert, CH, Nickisch, H., & Harmeling, S. (2009). 클래스 간 속성 전송을 통해 보이지 않는 객체 클래스를 감지하는 방법을 학습합니다. 컴퓨터 비전 및 패턴 인식(CVPR)에 관한 IEEE Computer Society 회의 간행물.
- Larochelle, H., Erhan, D., & Bengio, Y. (2008). 새로운 작업에 대한 제로 데이터 학습. AAAI 컨퍼런스 인공 지능에.
- Mitchell, T., 외. (2008). 명사의 의미와 관련된 인간의 뇌 활동을 예측합니다. 과학, 320, 1191–1195.
- 미첼, TM (1997). 기계 학습. 뉴욕: McGraw-Hill.
- Mitchell, TM, Hutchinson, R., Niculescu, RS, Pereira, F., Wang, X., Just, M., & Newman, S. (2004). 뇌 이미지에서 인지 상태를 해독하는 방법을 학습합니다. 기계 학습, 57, 145–175.
- Plaut, DC (2002). 의미론의 차등 양식별 전문화: 전산 계정 사신경 실어증. 인지 신경 심리학, 19, 603–639.
- Snodgrass, J., & Vanderwart, M. (1980). 표준화된 260개의 그림 세트: 이름 일치, 이미지 일치, 친숙함 및 시각적 복잡성에 대한 규범. 실험 심리학 저널: 인간 학습 및 기억, 174–215.
- Torralba, A., & Murphy, KP(2007). 멀티클래스 및 멀티뷰 개체 감지를 위한 시각적 기능 공유. IEEE 트랜스. 패턴 항문. 마하. Intell., 29, 854–869.
- van der Maaten, L., & Hinton, G. (2008). t-SNE를 사용하여 데이터 시각화. 저널 오브 머신 학습 연구, 9(11월), 2579–2605.
- 와이벨, A. (1989). 음성 인식을 위한 시간 지연 신경망의 모듈식 구성. 신경 계산, 1, 39–46.
- 윌슨, M. (1988). MRC 심리 언어 데이터베이스: 기계 판독 가능 사전, 버전 2. 행동 연구 방법, 6–11.