**과제 2: 추론을 통한 문제해결 [5팀]**

로고이(가) 표시된 사진

자동 생성된 설명

과목명: 인공지능(목 7-9)

교수명: 한경수 교수님

팀원: 20190274 정택원

20201092 정유진

20210833 남민지

목차

1. 문제 도메인 선정3

**2. 문제 선택 배경3**

**3. 지식베이스(KB) 구축3**

4. 질문에 대한 답변 추출5

**5. 코드 설명**

**6. 향후 발전 가능성**

7. 참고문헌

8. 개인 보고서

**1. 문제 도메인 선정**

본 팀은 사용자의 상황을 입력 받아, A+의 성적을 받을 확률을 추론한다. 추론 알고리즘으로는 Gibbs 샘플링 알고리즘을 사용한다.

**2. 문제 선택 배경**

수강하는 과목에서 받을 수 있는 최고 학점은 A+이다. 학기가 지남에 따라 A+성적을 부여할 수 있는 기준 또한 달라진다. 무엇보다 중요한 것은 학생으로서 얼마나 성실히 수업에 참여했는가 이다. 본 팀은 성실함을 포함한 여러가지 변수들을 정의하고 사용자로부터 변수(이하 노드)의 참 거짓 값을 입력 받아, 학생들이 최고 학점을 받을 확률을 추론해 주는 것이 흥미로울 것 같아 위와 같은 주제를 선정하게 되었다.

**3. 지식베이스(KB) 구축**

본 팀의 추론 방식으로 Gibbs샘플링을 이용하기 위해 베이즈 네트워크를 직접 제작하였다.

도표, 텍스트, 평면도, 기술 도면이(가) 표시된 사진

자동 생성된 설명

🡪 ‘교수님의 수제자’: 평소 행실이 우수하여 교수님의 마음 한 켠에 자리잡은 학생이다.

🡪 ‘높은 아이큐’: 학습을 하는 데에 있어 높은 아이큐는 큰 영향을 미친다.

🡪 ‘체력 좋음: 건강한 육체는 건강한 마음을 갖게 한다.

🡪 ‘선배와 친분’: 교수님의 마음 한 켠을 이미 자리잡은 선배들을 소개받을 수 있다.

🡪 ‘외부 대외활동’: 진로 탐색 학점제, 다양한 학술제 등을 교수님과 함께 진행할 수 있다.

🡪 ‘족보 보유’: 선배와 친분이 있다면 현재 수강하고 있는 교수님의 문제 제출 경향을 선배들을 통해 미리 파악할 수 있다.

🡪 ‘사전 지식 보유’: 높은 아이큐로 청소년기에 많은 독서와 경험을 할 수 있고, 외부 대외활동을 통해 수업의 내용을 뛰어넘는 고급 과정을 알 수 있다.

🡪 ‘매주 예복습’: 체력이 좋지 않으면 3시간의 수업을 수강하고 20시간을 취침해야 피로가 풀리는 학생이 존재할 수 있다. 체력이 좋으면 수업 직후 예복습이 가능하다.

🡪 ‘자료 3회독’: 매주 예복습을 통해 수업 내용에 대해 탐구욕이 솟구쳐 시험을 치루기 전까지 제공된 자료를 3회독 이상 할 수 있다.

🡪 ‘에이플’: 가장 많은 조상을 갖는 노드로, 본 팀이 추론하고자 하는 노드이다.

**4. 질문에 대한 답변 추출**

사용자로부터 받을 수 있는 입력은 2가지가 존재한다.

🡪 첫번째는 아무런 입력을 받지 않는다. 아무런 입력을 받지 않으면, A학점의 실제 비율인 35%와 근접하게 추론을 할 수 있도록 사전/사후 확률을 조정하였다. 결과적으로 약 36%의 확률을 추론한다.

🡪 두번째는 베이즈 네트워크에 존재하는 노드들의 값을 사용자로부터 입력 받는다. 모든 노드에 대해 True, False값 중 하나를 입력 받고 해당 상황에 대해 A학점을 받을 확률을 추론한다.

‘에이플’의 확률이 아닌 다른 변수의 확률 추론 또한 가능하다. 추론을 진행할 때에는 추론하고자 하는 변수와 증거변수는 달라야 한다.

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명[그림 n] 증거변수가 주어지지 않은 상황에서의 샘플링

🡪 위 셀에서 주어진 베이즈 네트워크를 정의하였다.

🡪 아래 셀에서 기각 샘플링을 통해 추론한 진 확률과 10,000개의 샘플을 이용한 Gibbs샘플링

결과를 출력한다.

텍스트, 폰트, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명

[그림 n] 증거변수가 주어진 상황에서의 샘플링

🡪 교수님의 수제자, 높은 아이큐, 체력 좋음 변수의 값을 입력 받아 10,000개의 샘플을 이용한 Gibbs샘플링 결과를 출력한다.

**n. 현실과의 괴리**

Gibbs샘플링의 수렴은 에르고드성을 충족해야 하며, 베이즈 네트워크에 결정론적 조건부 분포가 없어야 한다. 하지만, 수업에 일정 횟수 이상 결석을 하게 된다면 성적은 결정론적으로 F를 맞게 된다. 본 팀이 추론하고자 하는 A+성적은 당연히 받을 수 없다.

결정론적인 조건부 분포를 갖는 노드를 정의하고, 실제 추론과 다른 점을 찾아 수정하는 과정을 소개한다.

스케치, 도표, 그림, 기술 도면이(가) 표시된 사진

자동 생성된 설명

[그림 n] 결정론적 조건부 분포를 포함한 베이즈 네트워크

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

[그림 n] 결정론적 조건부 분포를 포함한 추론 결과

🡪 결정론적 조건부 분포를 포함한 결과는 정확하지 않음을 알 수 있다. ‘정시기상’ 변수가 False 값을 가질 경우 Gibbs샘플링 도중 다른 변수로의 이동(에르고드성 충족x)이 불가능하다.

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

[그림 n] probability.py 코드수정

🡪 또한 정규화 과정에서 ZeroDivisionError가 발생하여 코드를 수정해야한다.

샘플링 중 다양한 샘플을 생성하지 못하는 문제를 발견하고 절대적인 결정론적 조건부 분포를 제거하고 변수를 수정하여 새로운 베이즈 네트워크를 정의한다.

도표, 스케치, 그림, 기술 도면이(가) 표시된 사진

자동 생성된 설명

[그림 n] 결정론적 조건부 분포를 제거한 베이즈 네트워크

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

[그림 n] 결정론적 조건부 분포를 제거한 추론 결과

🡪 결정론적 조건부 분포를 포함한 추론 결과보다 더 나은 확률을 갖는다. ‘정시출석’ 변수 값이 True, False여도 결정론적 조건부 분포를 제거하였기 때문에, 본 팀의 의도대로 에르고드성을 띄어 어느 곳에 치우치지 않는 샘플들이 생성될 수 있다.