

수학 독서캠프 활동1.

<인공지능과 스팸메일 걸러내기>

1.인공지능이 조건부 확률로 스팸메일을 걸러내는 알고리즘을 만들고 이를 이용하여 스팸메일을 분류하는 과정을 이해하고 조별로 활동함.

2.인공지능이 스팸메일에 많이 들어 있는 단어(키워드)를 이용하여 스팸메일일 확률을 계산하는 과정을 이해함.

(1)스팸메일 분류 키워드, 스팸메일 분류 알고리즘

② 스팸메일 분류와 인공지능 알고리즘

❖ 스팸메일

- 1970년대 한 영국 코미디 프로그램에서 ‘스팸’이라는 단어를 우스꽝스럽게 사용
- 지나친 광고에 대한 표현에서 비롯된 단어
- 수신자의 눈길을 끌기 위해 스팸메일 제목에 많은 노력을 기울임
→ 독창적이고 확인하고 싶도록 제목을 만들어야 수신자들이 스팸메일을 열어 보기 때문
- 스팸메일함에 자동 분류되는 스팸메일 → 확률 활용
- 스팸메일과 제목의 특징 분석 활동 가능

❖ 스팸메일 제목의 특징

- 특수 기호 사용
- 특정 문구 사용
- 스팸메일에 많이 사용되는 기호 및 문구 분석과 스팸 분류 필터 필요

❖ 스팸메일 분류 알고리즘

- 메일에 사용된 단어를 조사하여 스팸메일에서 자주 사용되는 단어인지 아닌지를 분석
→ 인공지능은 기계 학습 기법을 이용하여 사전에 스팸메일에 사용된 단어를 학습
→ 스팸메일의 특징을 파악하는 규칙을 만들어 냄
→ 규칙을 바탕으로 이메일을 조사하여 스팸메일의 특성을 보이면 스팸메일함으로 이동
- 기계 학습 알고리즘의 정교화를 통해, 스팸 메일 판별 능력을 키우는 것이 중요

(2) 스팸메일 분류 알고리즘 예시

② 스팸메일 분류와 인공지능 알고리즘

❖ 스팸메일 분류 알고리즘 예시

- ‘할인’ 단어가 포함된 새로운 메일을 받을 때 인공지능은 이 메일을 광고 메일로 판단할 수도 있고 일반 메일로 판단할 수도 있음
- 각각의 가능성에 대한 확률을 구할 수 있다면 인공지능은 확률이 큰 쪽을 결과로 판단
- ‘할인’ 단어가 포함될 때 광고 메일일 확률 = 3/4
- ‘할인’ 단어가 포함될 때 일반 메일일 확률 = 1/4
- 1/4은 3/4보다 작기 때문에 인공지능은 ‘할인’ 단어를 포함한 새 메일을 받으면 광고 메일이라고 예측하고 분류

(3) 주어진 데이터에서 조건부 확률을 상대도수로 이용하여 스팸 메일일 확률을 계산함

1) 데이터 제시

- 키워드 분류 후 표 생성 시 상대도수 계산에 유용함

구분	주요 단어	구분	구분	주요 단어	구분
1	할인, 가격	일반	6	구매, 연체	일반
2	당첨, 기회	광고	7	가격, 기회	일반
3	가격, 구매	일반	8	할인, 구매	광고
4	구매, 기회	일반	9	할인, 당첨	광고
5	구매, 할인	광고	10	연체, 구매	일반

2) 인공지능이 키워드(‘할인’과, ‘기회’)와 조건부 확률을 이용하여 두 키워드를 모두 포함한 메일을 받았을 때, 그 메일이 광고 메일(스팸)일 확률을 계산하는 과정.

(a) 인공지능이 계산하는 조건부 확률 제시

❖ 확률에 따른 광고 메일 분류

- ‘할인’과 ‘기회’의 두 단어를 모두 포함한 메일을 받았을 때, 그 메일이 광고 메일일 확률

구분	주요 단어	구분	구분	주요 단어	구분
1	할인, 가격	일반	6	구매, 연체	일반
2	당첨, 기회	광고	7	가격, 기회	일반
3	가격, 구매	일반	8	할인, 구매	광고
4	구매, 기회	일반	9	할인, 당첨	광고
5	구매, 할인	광고	10	연체, 구매	일반

- 가정1: 자료의 크기가 크지 않으면 특정 조건 여러 개를 동시에 만족하는 확률은 각 조건의 확률을 곱하여 계산
- 가정2: (받은 메일이 광고 메일일 때, 그 메일에 ‘할인’과 ‘기회’의 두 단어를 모두 포함할 확률)
= (받은 메일이 광고 메일일 때, ‘할인’의 단어를 포함할 확률) × (받은 메일이 광고 메일일 때, ‘기회’의 단어를 포함할 확률)
- 가정3: (받은 메일이 ‘할인’과 ‘기회’의 두 단어를 모두 포함할 때, 그 메일이 광고 메일일 확률)
= (받은 메일이 광고 메일일 확률) × (광고 메일일 때, ‘할인’과 ‘기회’를 모두 포함할 확률
P1: 그 메일이 광고 메일일 확률

(b) 조건부 확률 계산 과정

❖ 확률에 따른 광고 메일 분류

- ‘할인’과 ‘기회’의 두 단어를 모두 포함한 메일을 받았을 때, 그 메일이 일반 메일일 확률
 - 왼쪽 표에서 받은 메일이 일반 메일일 때, ‘할인’의 단어를 포함하는 메일이 1개 → 1/6
 - 오른쪽 표에서 받은 메일이 일반 메일일 때 #6 ‘기회’의 단어를 포함하는 메일이 2개 → 2/6
 - 받은 메일이 일반 메일일 때, ‘할인’과 ‘기회’의 두 단어를 모두 포함할 확률
 $1/6 \times 2/6 = 1/18$

구분	일반 메일	광고 메일	계	구분	일반 메일	광고 메일	계
‘할인’ 포함	1	3	4	‘기회’ 포함	2	1	3
‘할인’ 미포함	5	1	6	‘기회’ 미포함	4	3	7
계	6	4	10	계	6	4	10

- 받은 메일이 광고 메일일 확률 = 6/10
- 가정 6에 의한 받은 메일이 두 단어를 포함할 때 광고 메일일 확률
 $P2 = 6/10 \times 1/18 = 1/30$
- 확률 비교 결과
 - 광고 메일일 확률
 $P1 = 4/10 \times 3/16 = 3/40$
 - 일반 메일일 확률
 $P2 = 6/10 \times 1/18 = 1/30$
 - 계산 결과에 따라 예측 및 분류 실행
 - 복잡한 확률 원리를 방법대로 계산할 수 있도록 지도 및 확률 학습의 유용성 전달
- 지도 시 유의점
 - 조건부 확률의 용어 및 기호는 사용 지양
 - 활용되는 수학적 원리를 이해하도록 지도
 - 확률의 활용 사례를 찾아보는 활동 구성

수학독서캠프 활동 1 : 인공지능의 스팸메일 걸러내기 이해

인공지능 수학 탐구

인공지능과 문제해결
- 자기평가 -
()학년()반이름:()

※ 인공지능 수학 탐구에서 나는 어떻게 공부하였나요?

1. 주제는?

2. 나의 역할은?

3. 재미있었던 점은?

4. 어려웠던 점과 어려운 것을 극복한 점은?

5. 알게 된 것은?

6. 더 알고 싶은 것은?

7. 프로젝트에 대한 나의 참여 점수를 매기면?

상

중

하

 • 이유는?

인공지능 수학 탐구

인공지능과 문제해결
- 교사평가 -
()학년()반이름:()

주제 정하기	가치가 있는 주제를 선정하였는가?
인공지능 조사하기	조사한 인공지능이 적절하였는가?
	조사한 인공지능에 대한 설명과 수학적 원리가 잘 정리되었는가?
문제해결하기	조사한 인공지능으로 문제를 해결하는 방법이 적절하였는가?
발표하기	발표 자료를 설득력 있게 발표하였는가?

모둠 이름	평가			
	주제 정하기	인공지능 조사하기	문제해결하기	발표하기

인공지능 수학 탐구

인공지능과 문제해결
- 모둠 구성원 동료평가 -
()학년()반이름:()

※ 평가대상자1:()반 이름:() 역할:

평가 영역	평가 기준	평가		
		상	중	하
참여도	모둠별 탐구 활동에 적극적으로 참여하였는가?			
기여도	모둠별 탐구 활동에 참신한 아이디어를 제시하였는가?			
기타 의견	평가대상자의 활동에 대하여 자유롭게 적으시오.			

인공지능 수학 탐구

인공지능과 문제해결
- 모둠별 동료평가 -
()학년()반이름:()

모둠 이름	창의성			발표력		
	상	중	하	상	중	하

각 모듬의 발표를 학생들도 평가할 수 있도록 실시