

Max Marginal Relevance (MMR) 알고리즘



Max Marginal Relevance (MMR) 알고리즘은 정보 검색과 문서 요약 과정에서 다양성과 관련성을 모두 고려하여 결과를 선택하는 데 사용됩니다. 이 알고리즘은 주로 **결과의 중복성을 줄이면서도 가장 관련성 높은 정보를 제공하는 것을 목표로 합니다.**

MMR 알고리즘을 이해하는 데 도움이 되는 비유 중 하나는 "레스토랑 메뉴 선택"입니다.

당신이 레스토랑에 가서 메뉴를 고르고 있는데, 여러분의 **취향**과 **메뉴의 다양성**을 모두 고려해야 합니다. 여러분은 피자를 좋아하지만, 피자만 먹기보다는 다른 종류의 요리도 시도해 보고 싶습니다. MMR 알고리즘은 여러분이 선택할 수 있는 메뉴 항목(문서)들 중에서 여러분의 취향(쿼리의 관련성)과 다른 이미 선택한 메뉴 항목들과의 차별성(문서 간의 다양성)을 고려하여 **최적의 선택**을 돕습니다.

주요 개념

1. **관련성(Relavance)**: 검색 쿼리나 주제와 문서의 관련성을 평가합니다. 이는 보통 문서가 주어진 쿼리와 얼마나 잘 일치하는지를 나타내는 점수로 표현됩니다.
2. **다양성(Diversity)**: 이미 선택된 문서와의 유사성을 평가하여, 선택 과정에서 문서 간의 다양성을 보장합니다. 이는 문서 간의 유사성을 계산하여, 이미 선택된 문서와 비슷한 새 문서의 선택 가능성을 낮춥니다.

Max Marginal Relevance (MMR) 알고리즘을 이해하기 위한 쉬운 비유와 예시를 제공해 보겠습니다.

과정

MMR 알고리즘은 이런 식으로 작동합니다.

- 첫 번째 단계에서는 **가장 관련성 높은 항목을 선택**합니다.
- 이후의 각 단계에서는 현재 선택된 항목들과 **관련성은 높으면서도 가장 차별화된 항목을 찾아 선택**합니다. 이는 λ 값에 의해 조절되며, 이 값이 클수록 관련성을, 작을수록 다양성을 더 중시합니다.

이렇게 MMR 알고리즘은 단순히 관련성만을 고려하는 대신 다양성도 함께 고려하여 더 폭넓고 흥미로운 선택을 제공합니다.

예시

예를 들어, 영화 추천 시스템에서 MMR 알고리즘을 적용한다고 가정해 봅시다. 사용자가 "액션 영화"에 대해 관심을 표시했다고 합시다.

1. **관련성 평가:** 시스템은 모든 영화 중에서 액션 장르에 속하는 영화들을 높은 관련성 점수를 부여합니다.
2. **다양성 평가:** 이미 추천된 액션 영화가 있을 경우, 시스템은 비슷한 액션 영화를 다시 추천하는 대신 다른 유형의 액션 영화(예: 스파이 액션, 고대 전쟁 액션 등)를 추천하여 다양성을 보장합니다.
3. **최적의 선택:** 시스템은 관련성이 높으면서도 기존에 추천된 영화들과는 다른 새로운 유형의 액션 영화를 사용자에게 추천합니다.

알고리즘의 수식

MMR 알고리즘은 다음과 같은 수식으로 정의됩니다.

$$\text{MMR} = \arg \max_{D_i \in R \setminus S} \left[\lambda \cdot \text{Sim1}(D_i, Q) - (1 - \lambda) \cdot \max_{D_j \in S} \text{Sim2}(D_i, D_j) \right]$$

- D_i 는 선택 가능한 문서,
- R 는 검색 결과 집합,
- S 는 이미 선택된 문서 집합,

- Q 는 사용자 쿼리,
- $Sim1$ 은 문서와 쿼리 사이의 유사성 함수,
- $Sim2$ 은 문서 간의 유사성 함수,
- λ 는 관련성과 다양성 사이의 균형을 조정하는 파라미터입니다.

알고리즘의 응용

MMR 알고리즘은 다음과 같은 경우에 유용합니다.

- **검색 엔진 최적화:** 사용자의 쿼리에 대해 관련성 높고 다양한 결과를 제공합니다.
- **문서 요약:** 대량의 문서 집합에서 중복을 최소화하면서 주요 정보를 추출하여 요약합니다.
- **추천 시스템:** 사용자에게 다양하면서도 관련성 높은 아이템을 추천합니다.