

NER 기술을 활용한 Gale-Shapely 매칭 알고리즘

윤성준¹, 조승호², 양조은³, 김영종⁴, SSUlay
 송실대학교 소프트웨어학부¹, 송실대학교 소프트웨어학부²
 송실대 소프트웨어학부³, 송실대 소프트웨어학부⁴
 sj990212@gmail.com¹, seunghyo7742@naver.com², jxxxxharu@gmail.com³,
 youngjong@ssu.ac.kr⁴

Utilizing NER for Gale-Shapley Matching Algorithm

Sungjun Yun¹, Seunghyo Jo², JoEun Yang³, Youngjong Kim⁴
 School of Software, Soongsil University¹
 School of Software, Soongsil University²
 School of Software, Soongsil University³
 School of Software, Soongsil University⁴

요 약

Gale-Shapley 알고리즘의 한계를 보완하기 위해 개체명 인식(Named Entity Recognition, 이하 NER) 기술을 활용한 보완된 알고리즘 방법을 제안한다. 이를 통해 개인의 성향과 관심사를 고려하여 매칭 결과를 더욱 정확하고 효율적으로 도출할 수 있다. 이 방법은 다수의 그룹원을 가진 두 그룹 간의 매칭, 특히 1:N 매칭의 문제를 해결하는 방식을 제안한다.

1. 서론

Gale-Shapley 알고리즘은 [1]의 논문에서 소개되었으며, 안정적인 매칭 알고리즘으로 인정받고 있다. 이 알고리즘은 수학적 안정성과 최적의 결과를 보장하며, 두 그룹 간의 이중 매칭 문제에서 개체들이 상호 선호하는 그룹원끼리 짝을 지어 공정하고 효율적인 매칭을 이룰 수 있다.

그러나, 그룹원들이 많이 있는 경우 각 그룹원이 모든 다른 그룹원에 대한 선호도 순위를 정해야 하므로 실용성이 떨어진다. 이러한 문제는 내국인 학생과 외국인 교환학생 간의 매칭 과정에서도 나타난다. 예를 들어, 140명의 교환학생이 있는 경우 내국인 학생이 이들에 대한 선호도 순위를 모두 부여해야 하는데 이는 상당히 번거로운 과정이다.

이를 개선하기 위해 그룹원들에게 일부 선호 순위만 받아 나머지 선호도 순위는 개체명 인식 (Named Entity Recognition, 이하 NER) 기술을 활용하여 키워드를 추출하고 자동으로 채워주는 방식을 제안한다. 이를 통해 다양한 상황에 유연하게 대처할 수 있으며, 다수의 그룹원을 가진 두 그룹 간의 매칭 문제에 대한 새로운 접근법과 해결책을 제시할 수 있다.

2. 요구사항

- 각 그룹원들이 선택한 k명의 선호도 순위 데이터가 필요하다.
- NER 기술을 적용하여 취향이나 관심사와 같은 '키워드'를 추출하고, 이를 기반으로 그룹원이 선택한 k명과 다른 그룹원 간의 유사도를 분석하여 나머지 그룹원의 선호 순위를 자동으로 채워야 하기 때문에 충분한 양의 그룹원들이 작성한 자기소개글 데이터의 수집과 전처리 과정이 필요하다.

3. 제안방식

이 논문에서는 새로운 매칭 알고리즘을 제안하며, <그림 1>과 같은 순서도를 따라 설명한다. 순서도에 사용된 용어들은 옆에 표기되어 있다. 본 알고리즘은 그룹 X와 Y를 매칭하기 위해 다음과 같은 과정을 거친다.

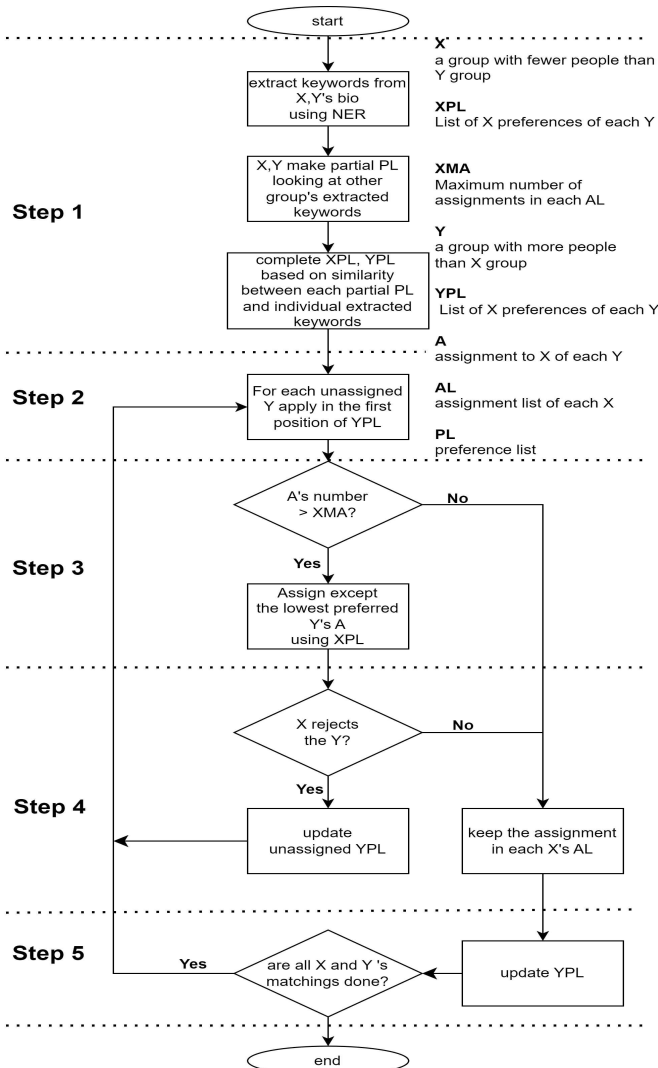
먼저, X와 Y의 사전 제출된 자기소개서(bio)를 이용하여 BERT 모델을 개체명 인식(NER)을 위해 fine-tuning한다. 이를 통해 학생들의 자기소개글(bio)에서 개인적인 취향, 관심사, 취미 등을 나타내는 개체명을 추출한다.

추출된 키워드를 기반으로 원하는 상대 그룹원을 선택하고, 선택된 그룹원들을 부분 선호도 순위 리스트

트에(partial preference list) 추가한다.

Gale-Shapley 알고리즘에 필요한 모든 사용자의 선호도 순위를 완성하기 위해, 유사도 분석을 통해 그룹원이 선택한 선호 그룹원의 키워드와 유사한 키워드를 가진 그룹원을 우선으로 하여 나머지 선호도 순위를 정한다. 이후, 기존 Many to one 매칭 문제에 사용되는 변형된 Gale-Shapley 알고리즘을 적용한다. 각각의 Y 그룹원들의 우선순위 1순위에 해당하는 X 그룹원에 할당한다. 만약 X 그룹원의 최대 할당수(XMA)를 초과하게 되면, XPL를 사용하여 가장 우선순위가 낮은 Y 그룹원을 제외하고 모든 Y 그룹원을 X 그룹원의 할당 리스트(AL)에서 할당한다.

할당되지 않은 Y 그룹원은 YPL의 다음 우선순위에 해당하는 X 그룹원에 다시 할당되도록 우선순위가 갱신한다. 할당된 Y 그룹원의 경우 X 그룹원의 할당 테이블에 저장된다. 이 과정을 Y 그룹원들이 모두 할당될 때까지 즉, 모든 매칭이 완료될 때까지 반복한다.



<그림 1. 제안된 알고리즘 flow diagram>

4. 결론

본 연구에서는 기존의 Gale-Shapley 알고리즘이 다수의 그룹원의 데이터에 적용 됐을 때 가지는 한계점을 보완하고, 그룹원들의 취향이나 관심사를 고려한 1:N 매칭 방식을 제안하였다. 기존 알고리즘은 모든 개체의 선호 선호도 순위를 요구하는 한계를 가진 반면, 제안 방식에서는 각 그룹원에게 일부 상위 K명에 대한 선호 순위만 입력받고, 나머지 선호도 순위는 각 그룹원의 취향이나 관심사를 담은 자기소개글에서 NER기술을 이용하여 키워드를 추출하고, 이를 기반으로 유사도를 분석하여 선호도 순위를 자동으로 채운다.

제안된 방식은 그룹원들이 작성한 자기소개글에서 NER 기술을 통해 그룹원의 취향이나 관심사를 자동으로 분석하며, 추출한 키워드를 모든 그룹원들에게 공개하여 상위 K명에 대한 선호 순위를 매기는 것을 가능하게 하였다.

따라서 본 연구의 결과는 그룹 간 선호 입력 비용을 줄이며, 내재적인 취향과 관심사를 더욱 잘 반영하고 상호 소개 기능을 가능케 하는 효율적인 매칭 방법으로, 향후 다양한 분야에서의 응용이 기대된다. 이러한 새로운 접근 방식은 대규모 그룹 매칭 문제를 해결하는 데 기여할 것으로 예상된다. 이를 통해 기존 Gale-Shapely 알고리즘의 한계를 극복하고, 더욱 공정하고 효율적인 매칭 시스템을 구축할 수 있을 것으로 기대된다.

참고문헌

[1] Gale, D., and L.S. Shapley. "College Admissions and the Stability of Marriage." The American Mathematical Monthly, vol.69, no.1, 1962, pp.9 - 15.
 [2] Mikheev, Andrei, Marc Moens, and Claire Grover. "Named entity recognition without gazetteers." Ninth Conference of the European Chapter of the Association for Computational Linguistics. pp.1-8 1999.