

고객의 암묵적 이상형을 반영하여 배우자 선택기준을 동적으로 조정하는 온라인 매칭 시스템: 의사결정나무의 활용을 중심으로

A Matchmaking System Adjusting the Mate-Selection Criteria based on a User's Behaviors using the Decision Tree

박 윤 주 (Yoon-Joo Park) 서울과학기술대학교 글로벌경영학과

요 약

본 연구는 고객이 명시적으로 기재한 배우자 선택기준을 고객의 행동에 기반한 암묵적 선호도를 활용하여 동적으로 조정하는 온라인 데이트 시스템을 제안한다. 최근 미혼남녀를 소개시키는 온라인 데이트시스템에 대한 사용이 증대되면서, 많은 온라인 데이트 서비스 업체들이 더욱 정확한 매칭 서비스를 제공하기 위하여 고객에게 직접 배우자 선정기준을 기재하도록 요구하고 있다. 그러나, 많은 고객들이 자신의 이상형을 명확히 밝히지 않거나, 상황에 따라서 이상형을 조정하기 때문에 고객의 배우자 선택기준을 정확히 파악하는데 어려움이 있다. 이에 본 연구는 고객이 명시적으로 기재한 이상형뿐만 아니라, 고객이 과거 상대 이성에게 보낸 메시지 기록을 분석하여 파악된 암묵적 이상형을 반영하여 매칭을 수행하는 온라인 데이트 시스템을 제안한다. 제안된 시스템은 고객의 행동 데이터가 일정 수 이상으로 축적되면, 이를 데이터마이닝의 의사결정나무 기법으로 분석하여 고객의 암묵적 이상형을 파악하였으며, 이를 동적으로 반영하여 배우자 선택기준을 조정한다. 제안된 매칭 시스템은 Decision Tree based Matchmaking System(DTMS)이라고 명명하였으며, 본 연구에서는 DTMS 시스템을 설계하고 사용자 인터페이스에 대한 프로토타입을 구현하였다.

키워드 : 배우자 선정기준, 명시적 선호도, 암묵적 선호도, 의사결정나무, 온라인데이트

I. 서 론

매칭 시스템은 추천시스템의 하나로, 고객

† 이 연구는 서울과학기술대학교 교내 학술연구비 지원으로 수행되었습니다.

본 논문은 지난 2012년 한국경영정보학회 추계학술대회에서 우수논문상을 수상했으며, Information Systems Review 편집위원회에 의해 12월 23일 게재확정된 논문임을 알려드립니다.

의 개인정보, 이상형 정보, 행동패턴 등을 분석하여, 해당 고객에게 적합한 데이트 상대를 추천하는 시스템이다. 다수의 온라인 데이트 업체들이 매칭 시스템에 기반하여 추천서비스를 제공하고 있으며(Deepens, 2010), 미국시장 점유율 1위 업체인 match.com의 경우에는 주선자(match maker)에 전혀 의존하지 않고 순수하게 매칭 시스템만으로 추천(matchmaking)을 수행하고 있다(Carrico *et al.*, 2009). 국내

에서도 정오의 데이트, 이음 등의 업체가 순수 온라인 기반 매칭 서비스를 제공하고 있으며, 듀오와 같은 온오프라인 결합형태의 결혼정보업체들도 배우자 추천에 매칭 시스템을 부분적으로 활용하고 있다(신은경, 2002).

많은 온라인 데이트 서비스 업체들은 고객에게 더욱 정확한 매칭 서비스를 제공하기 위하여 직접 배우자 선정기준을 기재하도록 요구하고 있다. 즉, 고객 스스로가 자신이 선호하는 상대 이성의 연령, 직업, 학력, 신장 등에 대한 배우자 선택기준을 기재하는 것이다. 그러나 이러한 배우자 선택 기준이 명확하지 않은 고객의 경우, 이를 명시적으로 기재하는데 어려움이 있을 수 있다. 또한, 일부 고객들은 배우자 선택기준과 같은 사적인 정보를 기재하는 것을 부담스러워하거나, 기피하는 경향이 있으며(Nichols, 1997; Oard and Kim, 1998; Kelly and Teevan, 2003), 고객이 명시적으로 기재한 이상형과 고객의 행동이 실제로 일치하지 않는 경우도 빈번하게 발생한다(Pizzato, Chung, Rej, Koprinska, Yacef and Kay, 2010). 뿐만 아니라, 사람의 배우자 선택기준은 고정된 것이 아니라, 상황이나 가치관의 변화에 따라서 조정된다. Regan(1998)은 사람들이 제한된 시간 안에 매칭이 되어야 한다는 압박을 느낄 때, 자신의 선택기준을 조정하여 주어진 환경 하에서 매칭 성사율을 높이도록 타협한다는 연구 결과를 발표한 바 있다.

본 연구는 고객이 명시적으로 기재한 배우자 선택기준을 고객의 과거 행동에 기반한 암묵적 이상형 정보를 반영하여 동적으로 조정하는 온라인 데이트시스템을 제안한다. 제안된 시스템은 고객의 행동 데이터가 분석 가능한 수준으로 축적되면, 데이터마이닝의 의사결정나무 기법으로 데이터를 분석하여 고객의 암묵적 이상형을 파악하고, 이를 시스템에 피드백(feedback)하여 고객의 다음 번 데이

트 상대 추천에 반영한다. 본 연구에서 제안한 매칭시스템은 Decision Tree based Matchmaking System(DTMS)이라고 명명하였다. DTMS는, 과거 행동 데이터가 부족한 신규 고객이나 소극적으로 활동하는 고객(inactive users)에 대해서는 고객과 상대방의 명시적인 매칭 기준에 기반하여 매칭을 수행한다. 이때, 서로의 매칭 기준을 상호 충족하는 배우자가 없거나 부족하다면, 고객의 이상형과 가장 유사한 차선의 상대를 추천한다. 그러나 고객의 과거 행동 데이터가 분석 가능한 수준으로 축적되면, DTMS는 의사결정나무 C4.5 기법(Quinlan, 1993)으로 데이터를 분석하여 고객의 암묵적 매칭 기준을 파악한다. 즉, 각 고객별이 암묵적으로 중요하게 고려하는 기준과 상대적으로 타협이 용이한 기준을 파악하여, 일대일 방식으로 매칭 기준을 조정하는 것이다. 이러한 암묵적 매칭 기준은 명시적 매칭 기준과 더불어 시스템에 반영되어, 다음 번 추천 시에 적용된다.

본 연구에서는 DTMS 시스템을 설계하고, 사용자 인터페이스에 대한 프로토타입을 구현하였다. 제안된 DTMS 시스템은 각 고객별로 매칭 기준을 동적으로 조정하여, 고객의 이상형 변화 및 암묵적 선호도가 추천에 반영될 수 있도록 하였다. 또한, 고객의 명시적 이상형과 암묵적 이상형을 복합적으로 사용하여, 과거 데이터가 부족한 신규고객에 대한 cold start 문제를 해소하였다. 뿐만 아니라, 온라인 데이트시스템의 중요 요소인 고객 및 추천 대상자의 상호 만족도를 함께 고려하여 호혜적인 추천이 이뤄지도록 하였다.

본 논문의 나머지 부분은 다음과 같이 구성된다. 우선, 제 II장에서는 온라인 데이트 시스템에 대한 선행연구를 조사하였다. 제 III장에서는 제안된 DTMS 시스템의 주요개념 및 알고리즘을 소개하였으며, 제 IV장에서는 DTMS의 사용자 인터페이스에 대한 프로토

타입을 제시하였다. 마지막으로 제 V장에서는 결론 및 향후 연구방안을 기술하였다.

II. 관련 연구

상품추천에 대한 많은 기존 연구가 존재하는 것에 반하여(박윤주, 2011; 김재경 등, 2009; 김경재 등, 2005), 사람을 추천해주는 매칭 시스템에 특화된 기존 연구는 매우 제한적으로 수행되고 있다. 국외의 경우, 매칭 시스템에 대한 초기 연구로 Brozovsk and Petricek(2007)이 기존의 협업필터링 기법을 온라인 데이트 분야에 적용한 연구가 있다. 이 연구는 고객의 프로필이나 이상형 정보를 전혀 사용하지 않고, 해당 고객이 과거에 상대방들에게 부여한 선호도 점수만을 활용하여 추천을 수행한다. 이 연구는 추천시스템의 주요한 분야로 온라인 매칭을 인식하였다는 점에서 의의가 있으나, 매칭 시스템에 요구되는 상호 만족성을 고려하지 못하는 등 온라인 데이트 시스템의 특징을 시스템에 충분히 반영하지 못하는 한계가 있다. Krzywicki et al.(2010)도 협업필터링에 기반한 온라인 데이트 시스템을 제안하였으며, 이들은 사용자들 간의 메시지가 긍정적 이었는지, 부정적 이었는지를 고려하여 매칭을 수행하였다. 즉, 사용자가 상대방에게 긍정적인 회신을 받은 경우에만 양자간의 교류를 인정하는 방식으로 행동 데이터를 수집하였으며, 상호 호혜성이 시스템에 반영되도록 하였다. 그러나 이러한 협업 필터링 기법은 매칭 시스템에서 중요하게 인식되는 고객의 프로필 및 이상형에 대한 정보를 활용하지 못하는 한계가 있다.

반면, 고객의 인물정보에 기반하여 추천을 수행하는 내용기반(content-based) 추천기법은 고객의 나이, 직업, 연봉 등 다양한 인물정보 및 해당 요소들에 대한 배우자 선택 기준을 활용하여 매칭을 수행할 수 있다. 사용자의

인물정보를 매칭에 반영한 연구로는 Hitsch et al.(2010)이 특정 그룹의 고객군과 이들이 선호하는 상대 이성의 인물 특성 사이의 연관성을 연구하였으며, 그 외에도 Fiore and Donath(2005)은 매칭 가능성이 높은 인물 특성을 분석하였다. 그러나 이러한 기존 연구들은 온라인 매칭 시스템의 주요 요소인 양자간의 상호 만족도를 고려하고 있지는 못하다. 국내 연구로는 박윤주(2012)가 고객의 인물정보 및 배우자 기준정보를 활용하여, 고객의 이상형과 확률적으로 가장 유사한 상대방을 선별하는 매칭 시스템을 제안하였다. 이 시스템은 추천 받는 고객과 추천되는 상대방의 만족도를 함께 고려하여 상호 호혜적인 추천이 이뤄지도록 하였다. 그러나 위의 연구들은 고객의 명시적인 선호도를 기준으로 추천을 수행하였으며, 고객의 행동에 기반한 암묵적인 선호도를 고려하고 있지는 못하다.

이러한 한계를 극복하기 위하여, Pizzato et al.(2010)은 고객의 명시적 프로필과 과거 행동데이터를 동시에 활용하여 매칭을 수행하는 방법을 제안하였다. 즉, 이 연구는 고객의 프로필은 명시적으로 기재하도록 요구하지만, 고객의 배우자 선택기준은 기재를 요구하지 않으며, 대신하여 과거 상대방에게 보낸 메시지 정보를 분석하여 고객의 암묵적 선호도를 파악하는 것이다. 그러나 이 연구는 과거 메시지 전송기록이 충분하지 않은 신규고객이나 소극적인 활동고객(inactive users)의 선호도를 파악하는 데에는 제약이 있어, cold-start 문제가 발생할 수 있다(Schein et al., 2002).

III. Decision Tree based Matching System(DTMS)

본 장에서는 고객의 명시적인 매칭 기준과 더불어, 고객의 행동에 기반한 암묵적인 매칭 기준을 반영하여 데이트상태를 추천해주

는 Decision Tree based Matchmaking System (DTMS)을 제안한다. 제 3.1절에서는 고객의 명시적인 매칭 기준에 기반하여 매칭을 수행하는 방법을 기술한다. 다음으로, 제 3.2절에서는 고객이 과거 상대방들에게 보낸 메시지 기록을 분석하여 고객의 암묵적 배우자 선택기준을 파악하는 방법을 설명한다. 마지막으로 제 3.3절에서는 명시적 이상형에 암묵적 이상형 정보를 반영하여 배우자 기준을 동적으로 조정하는 방법 및 DTMS의 전체적인 구조를 제안한다.

3.1 명시적 매칭 기준을 활용한 데이트 상대 추천

본 연구는 고객 C로부터 자신의 나이, 직업, 학력, 연봉, 신장에 대한 프로필 속성 p와, 동일항목들에 대한 선호 이상형의 속성 e를 직접 입력 받는다고 가정한다. 이때, 나이, 연봉, 신장, 학력에 대한 고객 프로필과 이상형 프로필은 간격변수(interval variable)를 사용하였고, 직업속성은 명목변수로 표현하였다. 따라서 남성 참가자 m_i 와 여성 f_j 는 각각 다음과 같은 데이터 셋을 갖는다. 만약, 고객이 특정 항목에 대해서 선호하는 이상형이 없을 경우에는 e값을 null로 설정하였다.

$$C(m_i) = \{p_{age}(m_i), p_{job}(m_i), p_{education}(m_i), p_{salary}(m_i), p_{height}(m_i), e_{age}(m_i), e_{job}(m_i), e_{education}(m_i), e_{salary}(m_i), e_{height}(m_i)\}$$

$$C(f_j) = \{p_{age}(f_j), p_{job}(f_j), p_{education}(f_j), p_{salary}(f_j), p_{height}(f_j), e_{age}(f_j), e_{job}(f_j), e_{education}(f_j), e_{salary}(f_j), e_{height}(f_j)\}$$

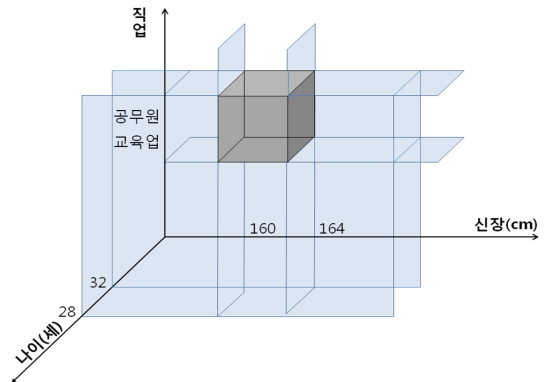
(단, $e(f_j) = [e_{min}(f_j), e_{max}(f_j)]$, $e(f_j) = [e_{min}(f_j), e_{max}(f_j)]$)

이러한 고객의 이상적 배우자 기준은 기

본적으로 5차원으로 표현될 수 있다. 이를 시각적으로 나타내기 위하여, 임의의 남성 참가자 m_i 의 매칭 기준이 28~32세의 공무원 또는 교육업 종사자인 160~164cm의 여성으로, 상대방의 연봉과 학력은 고려하지 않는다고 가정하자. 그러면, 매칭 기준 중 2개의 요소가 null 값을 가지므로, m_i 의 매칭 기준은 <그림 1>과 같은 3차원의 입방체로 표현될 수 있다. 이때, 입방체 내부에 포함되는 프로필을 갖는 상대 여성들은 모두 m_i 의 명시적인 이상형에 부합되는 고객들이 된다. 즉, 식 (1)과 같이 상대방 f_j 의 프로필 $p_{attribute}(f_j)$ 가 고객의 이상형 $e_{attribute}(m_i)$ 의 폐구간 안에 포함될 경우, f_j 는 m_i 의 매칭 기준을 충족한다고 판단한다. 단, 직업은 명목변수이므로 $p_{job}(f_j)$ 이 $e_{job}(m_i)$ 과 일치하는지를 확인하였다.

$$\text{Min}[e_{attribute}(m_i)] < p_{attribute}(f_j) < \text{Max}[e_{attribute}(m_i)] \quad (1)$$

(attribute = age, education, salary, height)



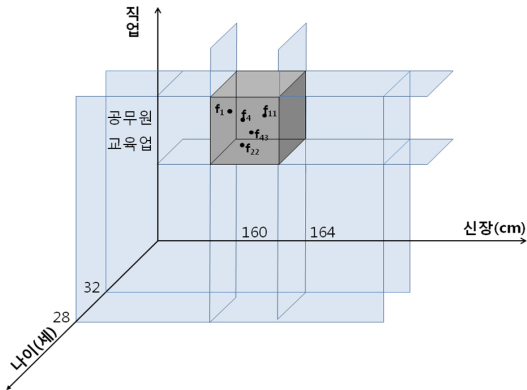
<그림 1> 남성 참가자 m_i 의 배우자 선택기준 사례

다음으로는, 고객의 이상형에 부합되는 상대 참가자들 중에서, 해당 고객의 프로필을 선호하는 매칭 상대를 선별한다. 즉, <그림 1>의 입방체 내부에 포함되는 모든 상대 참가자들은 남성 참가자 m_i 의 이상형에 부합되지만, 이들 중에서 일부의 여성들만이 m_i 의 프

로필을 선호할 것이다. 따라서 남성 참가자의 프로필도 상대 여성들의 매칭 기준을 충족시켜야만 상호 호혜적인 매칭이 수행될 수 있다. 이를 위하여, 본 연구는 고객 f_j 의 프로필 $p_{\text{attribute}}(m_i)$ 가 상대방의 이상형 $e_{\text{attribute}}(f_j)$ 의 폐구간 안에 식 (2)과 같이 포함되었을 때, m_i 도 f_j 의 매칭 기준을 충족시킨다고 판단하였다. 위에서 언급한 바와 같이, 직업은 명목 변수이므로 $p_{\text{job}}(m_i)$ 이 $e_{\text{job}}(f_j)$ 과 일치하는지를 확인하였다. <그림 2>는 남성 참가자 m_i 와 상호 매칭 기준을 충족하는 여성 참가자 5명 ($f_1, f_4, f_{11}, f_{22}, f_{43}$)의 프로필 값을 입방체 내부에 점으로 표시하였다.

$$\text{Min}[e_{\text{attribute}}(f_j)] < p_{\text{attribute}}(m_i) < \text{Max}[e_{\text{attribute}}(f_j)] \quad (2)$$

(attribute = age, education, salary, height)

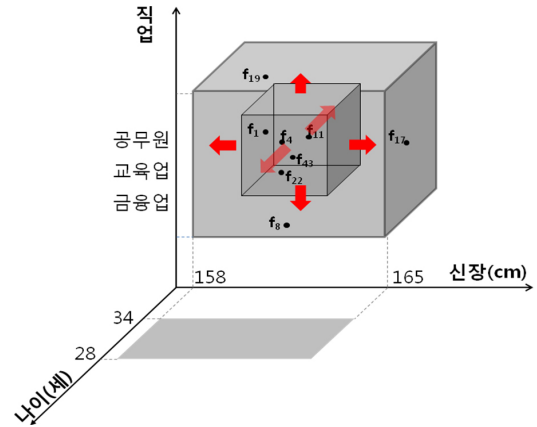


<그림 2> 고객 m_i 와 상호 호혜적인 매칭 상대 선별 사례

만약, 이러한 상호 호혜적인 매칭 기준을 충족하는 상대의 수 n 이 고객이 희망하는 추천 수 a 명 이상이라면, 제안된 DTMS는 이들 중에서 임의로 a 명을 선별하여, m_i 의 데이트 상대로 추천하게 된다. 그러나 매칭 상대의 수 n 이 고객이 희망하는 추천 수 a 명 보다 작을 경우에는 고객의 배우자 선택기준을 완화해야만, 희망하는 a 명의 데이트 상대를 추

천 받을 수 있다.

제안된 DTMS는 매칭 상대의 수 n 이 고객이 희망하는 추천 수 a 미만일 경우, 자동으로 고객의 매칭 기준을 확장하여, 차선의 매칭 상대를 a 명 만큼 추천하도록 한다. 이때, 고객의 행동 데이터가 충분히 수집되지 못했을 경우, DTMS는 모든 속성의 중요도를 동일하게 고려하여 매칭기준을 확장한다. 즉, 나이, 직업, 학력, 연봉, 신장이 각 고객에게 갖는 중요도가 모두 동일하다는 가정하에 매칭 기준을 확장하는 것이다. <그림 3>은 남성참가자 m_i 가 희망하는 추천 수는 8명이지만, 매칭 기준을 상호 충족하는 상대방의 수가 5명뿐일 때, m_i 의 매칭 기준이 확장되어 f_{19}, f_8, f_{17} 세 명을 추가적으로 추천하는 모습을 나타내고 있다.



<그림 3> 행동 데이터 수가 부족할 경우, 배우자 선택기준 완화 사례

위와 같이 매칭 기준을 완화하기 위하여, 본 연구는 고객의 이상형과 상대방의 실제 프로필 사이의 거리를 산정하였다. 우선, 각 변수 별 단위효과를 제거하기 위하여 식 (3)과 같이 데이터의 표준화 작업을 수행한다. 단, 직업은 명목변수이므로 표준화 작업에서 제외한다.

$$p'_{attribute}(m_i) = \frac{p_{attribute}(m_i) - p_{min}}{p_{max} - p_{min}} \quad (3)$$

$$(0 < p'_{attribute}(m_i) < 1)$$

$$e'_{attribute}(m_i) = \frac{e_{attribute}(m_i) - e_{min}}{e_{max} - e_{min}}$$

$$(0 < e'_{attribute}(m_i) < 1)$$

다음으로, 고객 m_i 의 명시적 이상형 $e_{attribute}(m_i)$ 과 상대 참가자 f_j 의 실제 프로필 $p_{attribute}(f_j)$ 과의 유클리디안 거리 $D_{ER_{m_i \rightarrow f_j}}$ 를 식 (4)와 같이 산출한다. 이때, 직업 속성에 대한 거리 $e_{job}(m_i) - p_{job}(f_j)$ 는 동일 직업이면 0, 타 직업이면 1로 계산하였다.

$$D_{ER_{m_i \rightarrow f_j}} = \sqrt{\sum_{attribute=1}^{Noofattributes} (e'_{attribute}(m_i) - p'_{attribute}(f_j))^2} \quad (4)$$

이렇게 산정된 거리 $D_{ER_{m_i \rightarrow f_j}}$ 가 0이면 m_i 의 명시적 이상형과 f_j 의 프로필이 완전히 일치하는 것이며, 마찬가지로 $D_{ER_{f_j \rightarrow m_i}}$ 가 0이면 상대방 f_j 의 명시적 이상형에 고객 m_i 가 부합됨을 의미한다. 본 연구는 $D_{ER_{m_i \rightarrow f_j}}$ 와 $D_{ER_{f_j \rightarrow m_i}}$ 가 모두 0인 참가자들을 우선적으로 매칭한다. 그러나 이러한 매칭 상대의 수 n 이 희망 추천수 α 미만이면, $D_{ER_{f_j \rightarrow m_i}}$ 와 $D_{ER_{m_i \rightarrow f_j}}$ 의 조화평균 값이 적은 순으로 매칭 상대방을 추천하여, 주어진 환경에서 고객과 상대방의 이상형을 호혜적으로 만족시킬 수 있는 차선의 매칭을 수행한다.

그러나 이러한 획일적인 매칭 기준 확장방식은 고객별로 조정 가능한 매칭 기준이 다르다는 점을 반영하지 못하는 한계가 있다. 즉, 어떤 고객은 배우자의 연봉에 대한 조건을 완화할 수는 있지만, 나이에 대한 조건은 고수하려고 한다. 반면, 다른 고객은 연봉 조건은 고수하려고 하지만, 신장에 대한 기준은 조정 가능하다. 이러한 고객별 차이를 고

려하지 않고 획일적으로 매칭 기준을 조정할 경우, 고객은 조정된 매칭 기준에 불만족할 가능성이 높다. 뿐만 아니라, 고정된 명시적 매칭 기준을 적용하면, 동일한 매칭 pool 내에서 매칭되는 상대방이 동일하게 되므로, 점차 고객의 명시적 이상형과 다른 상대방들이 추천되는 문제가 있다.

다음으로, 제 3.2절에서는 고객의 과거 행동 데이터를 분석하여 암묵적 배우자 선택 기준을 파악하고, 이에 기반하여 각 고객별로 매칭 기준을 조정하는 방법을 제안한다.

3.2 암묵적 배우자 선택기준 분석

본 절에서는 고객이 상대방을 추천 받은 후, 실제로 추천 받은 상대방에게 관심이 있을 경우에는 메시지는 전송한다고 가정한다. 이때, 메시지를 보냈는지에 대한 결과 데이터를 r 로 표현하였다. 즉, 남성 참가자 m_i 가 여성 참가자 f_j 에게 메시지를 보냈는지에 대한 결과 데이터 $r_{m_i \rightarrow f_j}$ 는 m_i 가 f_j 를 추천 받은 후 메시지를 보냈으면 “Y”, 추천을 받았으나 메시지를 보내지 않았으면 “N”, 추천을 받은 적이 없으면 null로 설정된다.

다음으로, 고객의 행동 데이터 셋(behavioral dataset) R 을 구성한다. 임의의 남성 참가자 m_i 의 행동 데이터 셋 $R(m_i)$ 의 독립변수로는 m_i 가 과거에 추천 받았던 상대 참가자 f_j 의 프로필 속성인 $p_{age}(f_j)$, $p_{job}(f_j)$, $p_{education}(f_j)$, $p_{salary}(f_j)$, $p_{height}(f_j)$ 를 사용하였고, 종속변수는 메시지 전송결과 $r_{m_i \rightarrow f_j}$ 를 사용하였다.

$$R(m_i) = \{p_{age}(f_j), p_{job}(f_j), p_{education}(f_j), p_{salary}(f_j), p_{height}(f_j), r_{m_i \rightarrow f_j}\}$$

예를 들어, 임의의 남성참가자 m_i 가 전체 여성참가자들 중에서, 10명의 여성참가자 f_i

부터 f_{10} 까지를 추천 받았고, 이들 중 f_1, f_3, f_5 에게만 메시지를 전송하였다고 하자. 그러면, m_t 의 행동 데이터셋 $R(m_t)$ 는 <표 1>과 같이 구성된다.

<표 1> 고객 m_t 의 메시지 전송 데이터

Data	IV_{age}	IV_{job}	$IV_{education}$	IV_{salary}	IV_{height}	DV
1	$P_{age}(f_1)$	$P_{job}(f_1)$	$P_{education}(f_1)$	$P_{salary}(f_1)$	$P_{height}(f_1)$	Y
2	$P_{age}(f_2)$	$P_{job}(f_2)$	$P_{education}(f_2)$	$P_{salary}(f_2)$	$P_{height}(f_2)$	N
3	$P_{age}(f_3)$	$P_{job}(f_3)$	$P_{education}(f_3)$	$P_{salary}(f_3)$	$P_{height}(f_3)$	Y
4	$P_{age}(f_4)$	$P_{job}(f_4)$	$P_{education}(f_4)$	$P_{salary}(f_4)$	$P_{height}(f_4)$	N
5	$P_{age}(f_5)$	$P_{job}(f_5)$	$P_{education}(f_5)$	$P_{salary}(f_5)$	$P_{height}(f_5)$	Y
6	$P_{age}(f_6)$	$P_{job}(f_6)$	$P_{education}(f_6)$	$P_{salary}(f_6)$	$P_{height}(f_6)$	N
7	$P_{age}(f_7)$	$P_{job}(f_7)$	$P_{education}(f_7)$	$P_{salary}(f_7)$	$P_{height}(f_7)$	N
8	$P_{age}(f_8)$	$P_{job}(f_8)$	$P_{education}(f_8)$	$P_{salary}(f_8)$	$P_{height}(f_8)$	N
9	$P_{age}(f_9)$	$P_{job}(f_9)$	$P_{education}(f_9)$	$P_{salary}(f_9)$	$P_{height}(f_9)$	N
10	$P_{age}(f_{10})$	$P_{job}(f_{10})$	$P_{education}(f_{10})$	$P_{salary}(f_{10})$	$P_{height}(f_{10})$	N

위와 같은 방식으로 수집된 고객의 레코드 수가 기준치인 β 이상으로 축적되면, 제안된 DTMS는 이를 의사결정나무 기법으로 분석한다. 본 연구는 β 값을 10-fold cross validation을 수행할 수 있는 최소값인 10으로 설정하였고, 의사결정나무 C4.5 알고리즘(Quinlan, 1993)을 사용하여 분석을 수행하였다.

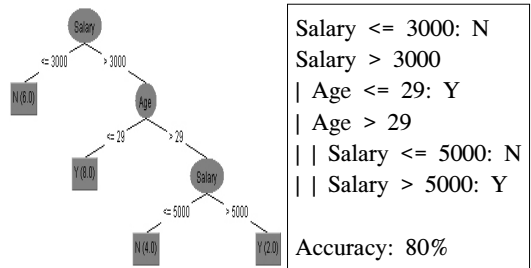
이러한 분석 결과로, <그림 4>(a)와 같은 의사결정나무가 수립되며, 이 나무와 매칭되는 규칙 셋(rule set)이 <그림 4>(b)와 같이 도출된다. <그림 5>는 본 연구에서 생성한 가상의 남성참가자 m_t 의 행동 데이터를 분석하여 수립한 의사결정나무이다. 남성참가자 m_t 는 명시적으로 25세~29세의, 교육업/일반사무직의 직업을 가진, 전문대나 4년제 대학을 졸업한 여성으로 연봉과 신장은 무관하다고 이상형을 기재하였다. 그는 과거에 20명의 상대 여성 참가자를 추천 받았고, 이들 중 10명에게 메시지를 보냈으므로, 기준치인 10개 이

상의 행동 데이터 레코드가 수집되어, 암묵적 선호도를 분석하였다. 이렇게 m_t 의 암묵적 매칭 기준을 분석한 결과, 그는 연봉 3000만 원 이상의 29세 미만의 여성, 또는 29세 이상인 여성이라도 연봉이 5000만 원 이상인 여성에게 메시지를 전송한다는 것을 알 수 있다. 이는 연봉을 고려하지 않는다는 그의 명시적 기준과는 차이가 있으며, 직업이나, 학력 요소보다 연봉과 나이를 매칭에 중요하게 고려하고 있음을 나타낸다.

본 연구는, 고객의 암묵적 배우자 선택기준을 위의 의사결정나무의 종단마디(leaf node)가 Y인 규칙들로부터 추출한다. <그림 5>(b)의 사례에서는 종단마디(leaf node)가 Y인 규칙이 다음과 같이 두 개 추출되었다.

암묵적 규칙 1: Salary(연봉) > 3000 and Age(나이) <= 29

암묵적 규칙 2: Salary(연봉) > 5000 and Age(나이) > 29



(a) 의사결정나무

(b) 규칙 셋(set)

<그림 4> 가상의 남성 참가자 m_t 의 행동 데이터를 학습하여 생성된 의사결정나무 사례

이러한 암묵적 규칙은 데이터 수가 일정 수준 증가할 때마다, 오프라인으로 정기적으로 수행되도록 하여, 시스템의 실시간 반응 속도(response time)을 줄이면서도 고객의 변화된 최신의 이상형이 매칭에 반영될 수 있도록 한다.

다음 제 3.3절에서는, 이렇게 파악된 고객의 암묵적 이상형을 반영하여, 배우자 선택 기준을 조정하는 방법을 설명하겠다.

3.3 배우자 선택기준의 조정

고객의 명시적 배우자 선택 규칙을 Explicit Rule(ER)이라고 하고, 암묵적 선택 규칙들을 Implicit Rule(IR)이라고 하자. 제안된 DTMS는 제 3.1절에서 설명한 유클리디안 거리 계산법으로, 상대방의 실제 프로필과 ER 및 IR의 거리를 각각 산출한다. 고객 m_i 의 ER로부터 상대방 f_j 의 프로필간의 거리인 $D_{ER_{m_i \rightarrow f_j}}$ 은 제 3.1절의 식 (4)에서 기술한 바와 동일하다. 고객 m_i 의 IR로부터 상대방 f_j 의 프로필간의 거리인 $D_{IR_{m_i \rightarrow f_j}}$ 는 암묵적 규칙들 간 거리를 각각 계산한 후, 이들 중 최소값을 사용한다. 이를 위하여, 우선 고객 m_i 의 암묵적 이상형 속성 $h_{attribute}(m_i)$ 과 상대 참가자 f_j 의 실제 프로필 $p_{attribute}(f_j)$ 의 단위효과를 제거하기 위하여, 식 (5)와 같이 표준화 작업을 수행한다.

$$p'_{attribute}(m_i) = \frac{p_{attribute}(m_i) - p_{min}}{p_{max} - p_{min}} \quad (5)$$

$$(0 < p'_{attribute}(m_i) < 1)$$

$$h'_{attribute}(m_i) = \frac{h_{attribute}(m_i) - h_{min}}{h_{max} - h_{min}}$$

$$(0 < h'_{attribute}(m_i) < 1)$$

이 표준화된 값들간 유클리디안 거리 $D_{IR}(m_i, f_j)$ 를 식 (6)을 통하여 산출한다.

$$D_{IR}(m_i, f_j) = \sqrt{\sum_{attribute=1}^{Noofattributes} (h'_{attribute}(m_i) - p'_{attribute}(f_j))^2} \quad (6)$$

이때, 고객별 암묵적 배우자 선택규칙이

제 3.2절에서 기술한 바와 같이 다수 개가 도출될 경우, 각 암묵적 규칙의 거리를 계산한 후, 이들 중 최소값을 $D_{IR_{m_i \rightarrow f_j}}$ 의 값으로 사용한다.

$$D_{IR_{m_i \rightarrow f_j}} = \text{Min} [D_{IR_1}(m_i, f_j), D_{IR_2}(m_i, f_j), \dots, D_{IR_K}(m_i, f_j)]$$

(k = 암묵적 규칙 수)

위의 과정을 통하여 산출된 $D_{ER_{m_i \rightarrow f_j}}$ 과 $D_{IR_{m_i \rightarrow f_j}}$ 를 반영하여, 고객의 이상형과 상대방의 프로필간의 거리를 종합적으로 산출한다. 고객이 직접 기재한 명시적 이상형과의 상대방간 거리인 $D_{ER_{m_i \rightarrow f_j}}$ 은 전체 거리에 1의 가중치로 반영하였다. 그러나 의사결정나무를 통하여 파악된 암묵적 이상형과의 거리인 $D_{IR_{m_i \rightarrow f_j}}$ 는 의사결정나무의 정확도(accuracy) 성능인 ρ_{m_i} 만큼만 가중치를 반영하였다. 이는, 고객 m_i 의 의사결정나무의 예측 성능이 상대적으로 부정확한 경우, 전체 시스템이 미치는 영향을 줄이기 위함이다. 따라서 암묵적 이상형과의 거리 $D_{IR_{m_i \rightarrow f_j}}$ 은 전체 거리에 최소 0에서부터 최대 1까지의 가중치로 반영될 수 있으며, 의사결정나무의 정확도가 1(= 100%)일 경우에는 가중치도 1로 반영된다. 식 (7)은 고객 m_i 의 이상형과 상대방 f_j 의 프로필간의 종합 거리 $D_{m_i \rightarrow f_j}$ 를 산출하는 공식을 나타낸다.

$$D_{m_i \rightarrow f_j} = D_{ER_{m_i \rightarrow f_j}} + (\rho \times D_{IR_{m_i \rightarrow f_j}}) \quad (7)$$

(ρ : 의사결정나무의 정확도, $0 < \rho < 1$)

마찬가지 방식으로, 상대 참가자 f_j 의 이상형과 고객 m_i 의 프로필간의 거리 $D_{f_j \rightarrow m_i}$ 를 산출한다.

다음으로는, 고객과 상대방의 상호 거리의

평균값을 산출한다. 본 연구는 고객뿐만 아니라, 매칭되는 상대방도 추천에 대해서 대칭적으로 만족하는지를 중요한 요소로 고려하였다. 따라서 $D_{m_i \rightarrow f_j}$, $D_{f_j \rightarrow m_i}$ 가 대칭적일수록 높은 평균값이 산출되고, 비대칭일 때에는 낮은 평균값을 도출하는 조화평균을 사용하여 양자간 거리의 평균인 $D_{m_i \leftrightarrow f_j}$ 를 식 (8)과 같이 산출하였다.

$$D_{m_i \leftrightarrow f_j} = \frac{2}{\frac{1}{D_{m_i \rightarrow f_j}} + \frac{1}{D_{f_j \rightarrow m_i}}} \quad (8)$$

마지막으로, 양자간 거리의 평균인 $D_{m_i \leftrightarrow f_j}$ 가 적은 순으로 a명의 상대방을 고객에게 추천하였다.

<그림 5>는 제안된 DTMS 시스템의 전체적인 알고리즘을 기술하고 있다.

Step 1: 데이터 표준화
 Step 2: 각 고객별 과거 메시지 전송기록 수(x)가 기준치 β 이상인 경우, 암묵적 배우자 선택기준 도출 (단, β 가 일정수준 증가할 때마다, 암묵적 기준을 오프라인에서 재 도출)
 Step 3: 고객 m_i 의 이상형과 상대방 f_j 의 거리 계산
 1) $x_{m_i} < \beta$ 일 때,
 : 고객의 명시적 이상형과 상대 프로필간 거리 산출

$$D_{m_i \rightarrow f_j} = D_{ER_{m_i \rightarrow f_j}}$$

 2) $x_{m_i} \geq \beta$ 일 때,
 : 고객의 명시적/암묵적 이상형과 상대 프로필간 거리 산출

$$D_{m_i \rightarrow f_j} = D_{ER_{m_i \rightarrow f_j}} + (\rho \times D_{IR_{m_i \rightarrow f_j}})$$

 Step 4: 상대방 f_j 의 이상형과 고객 m_i 의 거리 계산
 1) $x_{f_j} < \beta$ 일 때,
 : 상대방의 명시적 이상형과 고객 프로필간 거리 산출

$$D_{f_j \rightarrow m_i} = D_{ER_{f_j \rightarrow m_i}}$$

2) $x_{f_j} \geq \beta$ 일 때,
 : 상대방의 명시적/암묵적 이상형과 고객 프로필간 거리 산출

$$D_{f_j \rightarrow m_i} = D_{ER_{f_j \rightarrow m_i}} + (\rho \times D_{IR_{f_j \rightarrow m_i}})$$

 Step 5: 고객과 상대방의 상호 거리 평균값 $D_{m_i \leftrightarrow f_j}$ 계산

$$D_{m_i \leftrightarrow f_j} = \frac{2}{\frac{1}{D_{m_i \rightarrow f_j}} + \frac{1}{D_{f_j \rightarrow m_i}}}$$

 Step 6: 고객 m_i 와의 상호 거리 $D_{m_i \leftrightarrow f_j}$ 가 적은 순으로, 상대방을 지정된 수 a만큼 추천

<그림 5> DTMS의 전체 알고리즘

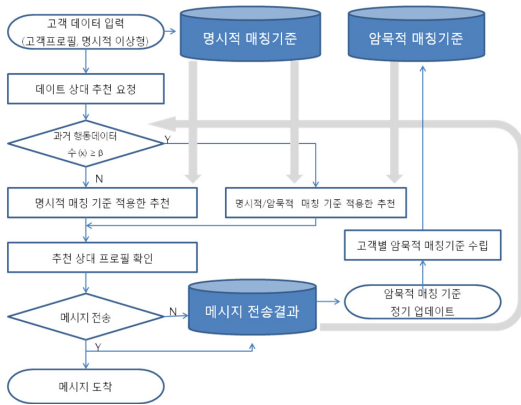
IV. 프로토타입(Prototype) 설계

본 장에서는 Decision Tree based Matchmaking System(DTMS)의 운영방식에 대한 흐름을 설명하고, 사용자 인터페이스에 대한 프로토타입을 제안한다.

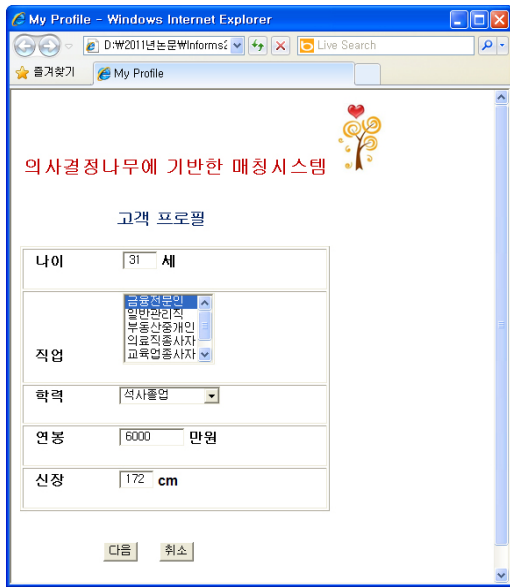
DTMS의 전체적인 순서도는 <그림 6>에 제시한 바와 같다. 우선, DTMS는 고객으로부터 프로필 및 이상형 정보를 입력 받는다. 고객은 자신의 프로필을 <그림 7>의 사용자 인터페이스를 통하여 입력하고, 자신의 명시적인 이상형 정보를 <그림 8>과 같은 인터페이스를 통하여 입력한다. 이렇게 입력된 사용자 정보는 <그림 6>에서와 같이, 명시적 매칭 기준을 관리하는 데이터베이스에 저장된다.

다음으로, 고객은 추천 받고자 하는 데이터상대의 수를 지정한 후, 추천을 요청한다. <그림 9>는 고객이 자신의 희망하는 추천 수를 입력하는 인터페이스를 나타내고 있다. 이때, DTMS는 고객과 상대방의 상호 이상형을 충족하는 매칭 상대의 수가 고객의 희망 데이터상대 수보다 적을 경우, 매칭 기준

이 확장된다는 점을 고객에게 공지한다.



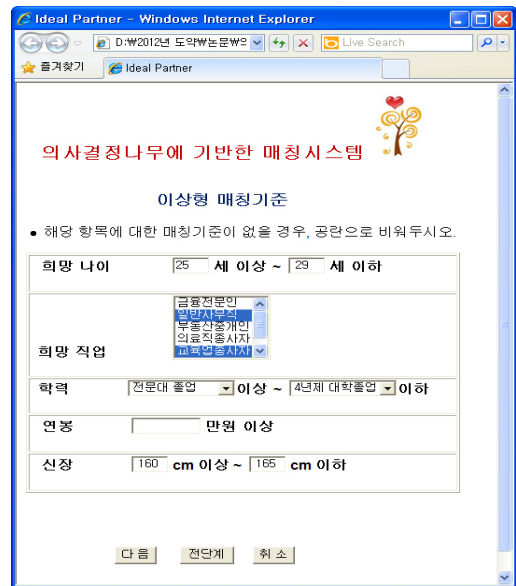
<그림 6> DTMS 시스템의 순서도



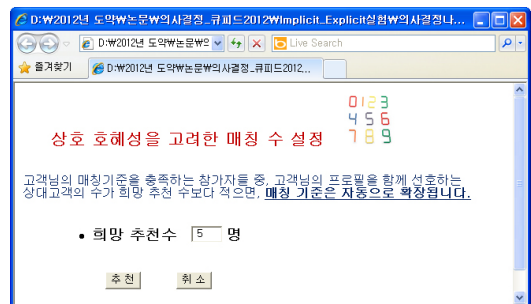
<그림 7> 고객프로필 입력

고객이 희망 추천 수를 지정한 후, <그림 9>의 ‘추천’ 버튼을 클릭하면, DTMS는 해당 고객의 과거 행동데이터 수가 얼마나 축적되었는지를 확인한다. 만약 행동데이터 수가 기준치인 β미만일 경우, 명시적 매칭 기준만을 적용하여 매칭을 수행한다. <그림 10>은

고객의 명시적 매칭 기준에 기반하여, 고객이 요청한 5명의 데이트 상대가 추천된 결과를 나타내고 있다. 추천 결과에는 상대방의 프로필 및 추가적인 사진정보가 포함되며, 추천된 상대방에게 관심이 있을 경우 ‘메시지 보내기’ 버튼을 눌러서 메시지를 전송할 수 있다. 이렇게 고객이 메시지를 전송한 결과는 <그림 6>과 같이 메시지 전송결과를 관리하는 데이터베이스에 저장되어, 추후 암목적 선호도를 분석하는데 활용된다. 또한, 고객은 자신의 명시적인 매칭 기준을 수정



<그림 8> 고객의 명시적 이상형 기준 입력



<그림 9> 고객의 희망 추천수 입력

할 수 있다. <그림 10>의 ‘변경’ 버튼을 누르면, <그림 8>의 고객의 이상형을 입력하는 인터페이스가 재 활성화되어 명시적 이상형을 수정할 수 있으며, 이를 반영하여 다시 데이트상대를 추천 받는다.



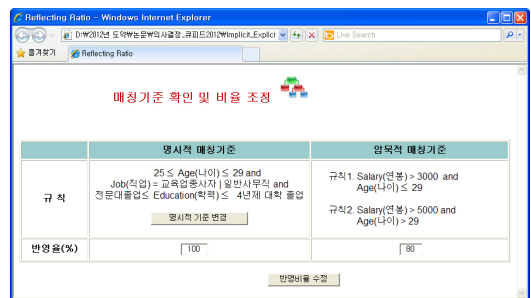
<그림 10> 고객의 명시적 매칭 기준에 의한 추천결과

반면, 행동데이터 수가 기준치인 β 이상일 경우, DTMS는 고객의 명시적 기준과 행동에 기반한 암묵적 기준을 모두 적용하여 매칭을 수행한다. <그림 11>은 고객의 과거 행동 데이터가 기준치 이상으로 축적되었을 때, 명시적인 매칭 기준과 더불어 암묵적인 매칭 기준이 추천에 반영된 결과를 나타내고 있다. 재 3.3절에서 소개한 바와 같이, 명시적인 매칭 기준은 100% 반영되지만, 암묵적인 기준은 의사결정나무의 정확도(accuracy)만큼만 전체 모형에 반영된다는 점을 고객에게 공지한다. 만약, 고객이 자신의 명시적/암묵적 매칭 기준을 확인하고자 하면, <그림

11>의 ‘매칭기준 확인’ 버튼을 누른다. 그러면, <그림 12>와 같은 팝업창이 활성화되며, 명시적 기준과 더불어, 의사결정나무에서 추출한 암묵적 매칭 기준을 확인할 수 있다. 또한, 고객은 ‘반영비율 수정’ 버튼을 눌러서, 명시적/암묵적 매칭기준이 전체 모형에 반영되는 비율을 직접 수정할 수 있다.



<그림 11> 고객의 명시적/암묵적 매칭 기준에 의한 추천결과



<그림 12> 매칭 기준 확인 및 반영비율 수정

V. 결 론

본 연구는 고객이 명시적으로 기재한 배우자 선택기준을 고객의 과거 행동에 기반한 암묵적 이상형 정보를 반영하여 동적으로 조정하는 온라인 데이트 시스템을 제안하였다. 제안된 시스템은 과거에 상대방에게 메시지를 전송한 기록이 적은 신규고객이나 소극적으로 활동하는 고객(inactive users)에게는 명시적인 이상형 기준을 적용하지만, 메시지 전송기록이 축적되면 이를 분석하여 고객의 암묵적 이상형 기준을 매칭에 함께 반영한다.

본 연구에서 제안한 Decision Tree based Matchmaking System(DTMS)는 다음의 세 가지 특징을 갖는다. 첫째, 각 고객별로 매칭 기준을 동적으로 조정하여, 고객의 변화하는 이상형 기준이 매칭에 반영되도록 하였다. 둘째, 고객의 명시적 이상형과 암묵적 이상형을 복합적으로 사용하여, 과거 데이터가 부족한 신규고객에 대한 cold start 문제를 해소하였다. 마지막으로, 온라인 데이트 시스템의 중요 요소인 고객 및 추천 대상자의 상호 만족도를 함께 고려하여 호혜적인 추천이 이뤄지도록 하였다. 이러한 연구를 통하여, 향후 온라인 데이트 서비스가 고객 스스로도 인지하지 못하고 있는 내재된 이상형을 더욱 정확하고, 체계적으로 반영할 수 있을 것으로 기대된다. 또한, 매칭기준 조정이 필요한 고객에 대하여, 고객별로 중요하게 고려하는 요소 및 상대적으로 조정이 용이한 요소를 파악하여, 획일적인 매칭 기준 확장방식에 따른 고객 불만족을 줄일 수 있을 것으로 생각된다.

본 연구의 한계점은, 가상의 데이터를 활용하여 실험하였기 때문에, 시스템의 예측 성능을 평가하지 않았다는 것이다. 고객이 실제로 선호할만한 상대방을 얼마나 정확히 추천하는지에 대한 성능평가는 매칭 시스템

연구의 중요한 요소이다. 그러나 가상의 데이터를 활용한 예측성능의 평가는 무의미하기 때문에, 본 논문에서는 이에 대한 연구를 수행하지 않았다. 향후, 실제 데이터를 활용한 예측성능 평가 및 이에 기반한 모형의 보완 작업을 수행할 필요가 있겠다.

참 고 문 헌

- 박윤주, “상호 대칭적 만족성을 고려한 온라인 데이트시스템”, 지능정보연구, 제18권, 제2호, 2012, pp. 179-194.
- Brožovsk, L. and V. Petříček, “Recommender system for online dating service”, *Proceedings of Znalosti*, 2007, pp. 1-12.
- Carrico, T. M., K. B. Hoskins, and J. C. Stone, *Matching Process System And Method*, United States Patent Application 20090164464.
- Deepens, Z., *Global Footprint as World's First Platform-Agnostic Online Dating Service*, Marketwire, 2010.
- Fiore, A. T. and J. S. Donath, *Homophily in online dating: when do you like someone like yourself?*, In CHI, ACM, New York, 2005, pp. 1371-1374.
- Hitsch, G., A. Horta_csu, D. Ariely, “What makes you click?: mate preferences in online dating”, *Quantitative Marketing and Economics*, Vol.8, No.4, 2010, pp. 393-427.
- Kelly, D. and J. Teevan, “Implicit feedback for inferring user preference: a bibliography”, *SIGIR Forum*, Vol.37, No.2, 2003, pp. 18-28.
- Krzywicki, A., W. Wobcke, and X. Cai, “Interaction-Based Collaborative Filtering Methods for Recommendation in Online Da-

- ting”, *Lecture notes in computer science*, Vol.6488, 2010, pp. 342-356.
- Nichols, D. M., *Implicit rating and filtering*, In *Proceedings of the Fifth DELOS Workshop on Filtering and Collaborative Filtering*, 1997, pp. 31-36.
- Oard, D. W. and J. Kim, “Implicit feedback for recommender systems”, In *AAAI Workshop on Recommender Systems*, 1998, pp. 81-83.
- Pizzato, L., T. Chung, T. Rej, I. Koprinska, K. Yacef, and J. Kay, *Learning user preferences in online dating*, Technical Report 656, University of Sydney, 2010.
- Pizzato, L., T. Rej, T. Chung, I. Koprinska, and J. Kay, *RECON: a reciprocal recommender for online dating*, Proceedings of the fourth ACM conference on Recommender systems, Barcelona, Spain, 2010, pp. 207-214.
- Quinlan, J. R., *C4.5: Program for Machine Learning*, San Mateo, CA: Morgan Kaufmann, 1993.
- Regan, P. C., “What if You Can’t Get What You Want? Willingness to Compromise Ideal Mate Selection Standards as a function of Sex”, *Mate Value, and Relationship Context, Personality and Social Psychology Bulletin*, Vol.24, No.12, 1998, pp. 1294-1303.
- Schein A., A. Popescul, L. Ungar, and D. Pennock, *Methods and Metrics for Cold-Start Recommendations*, Proceedings of the 25th ACM SIGIR Conference, 2002, pp. 253-260.
- (Patent-듀오정보 주식회사) 신은경, 배우자 주선 시스템(System of spouse mediation), 10-2002-0087616(2002-11-23).

A Matchmaking System Adjusting the Mate-Selection Criteria based on a User's Behaviors using the Decision Tree

Yoon-Joo Park*

Abstract

A matchmaking system is a type of recommender systems that provides a set of dating partners suitable for the user by online. Many matchmaking systems, which are widely used these days, require users to specify their preferences with regards to ideal dating partners based on criteria such as age, job and salary. However, some users are not aware of their exact preferences, or are reluctant to reveal this information even if they do know. Also, users' selection standards are not fixed and can change according to circumstances. This paper suggests a new matchmaking system called Decision Tree based Matchmaking System (DTMS) that automatically adjusts the stated standards of a user by analyzing the characteristics of the people the user chose to contact. AMMS provides recommendations for new users on the basis of their explicit preferences. However, as a user's behavioral records are accumulated, it begins to analyze their hidden implicit preferences using a decision tree technique. Subsequently, DTMS reflects these implicit preferences in proportion to their predictive accuracy. The DTMS is regularly updated when a user's data size increases by a set amount. This paper suggests an architecture for the DTMS and presents the results of the implementation of a prototype.

Keywords: Matchmaking Criterion, Explicit Preference, Implicit Preference, Decision Tree, Online Dating

* Business Administration Department, Seoul National University of Science and Technology

◎ 저 자 소 개 ◎



박 윤 주 (yjpark@seoultech.ac.kr)

고려대학교 컴퓨터학과에서 학부 및 석사학위를 취득하였으며, 한국과학기술원에서 경영공학 박사학위를 취득하였다. 이 후, New York University의 Stern Business School에서 초빙연구원으로 근무하였으며, 삼성생명 정보기획부에서 과장으로 근무한 바 있다. 현재는 서울과학기술대학교 기술경영융합대학 글로벌경영학과에서 조교수로 재직 중이다. 기존 연구는 Artificial Intelligence in Medicine, Expert Systems with Applications, Expert Systems 등의 논문지에 게재되었으며, IEEE TKDE에 게재 승인되었다. 주요 연구분야는 데이터마이닝을 이용한 질병 예측, 개인화 시스템, 그리고 온라인 매칭시스템 등이다.

논문접수일 : 2012년 11월 23일

게재확정일 : 2012년 12월 23일