ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2015 September: 21(3): 37~52 http://dx.doi.org/10.13088/jiis.2015.21.3.37

산학연 협업 활성화를 위한 R&D 네트워크 연결 예측 연구

박미연
연세대학교 기술정책협동과
(na800717@naver.com)

이상헌
연세대학교 정보산업공학과
(sangheon@yonsei.ac.kr)

김국성
연세대학교 정보산업공학과
(gooksong89@gmail.com)

심홍매
연세대학교 정보산업공학과
(hongmae1102@gmail.com)

김우주
연세대학교 정보산업공학과 교수
(wkim@yonsei.ac.kr)

최근 전세계적으로 R&D 네트워크 및 산학연 협력 등을 강화하고 있는 추세이다. 네트워크 및 협업연구 부문에 대한 지원이 증가하면 학제간 융합 연구를 통한 새로운 이론의 창출과 새로운 학문・사업 분야로의 확장 가능성을 높일 수 있다.

우리나라도 정부의 R&D 과제 수행을 통해 형성된 R&D 네트워크를 효율적으로 지원할 수 있는 전략의 필요성이 증대되고 있다. 그럼에도 불구하고 우리나라는 국가 R&D 사업 참여자에 대한 개별인력정보와 일반화된 통계 자료에만 의존하여 네트워크 관점에서의 정책은 미흡한 실정이다. 이에 따라 R&D 사업에 참여하는 각 주체들 간의 관계를 분석하고 산학연 R&D 네트워크를 기반으로 향후 발생할 수 있는 네트워크의 변화를 예측하고자 한다. R&D 네트워크 변화 예측을 위해 Common Neighbor 모형과 Jaccard's Coefficient 모형을 기반 모델로서 채택하고자 하며, 이들의 한계점을 보완하고 Link Prediction 정확도를 향상시킨 새로운 예측 모형을 제안하고 이들간의 비교분석 결과를 도출하고자 한다. 이와 같은 연구 결과는 향후 R&D 네트워크의 변화에 대한 효과적인 예측을 통해 선제적인 산학연 사업 지원 전략을 수립하고, 융합 R&D사업 등을 효과적으로 지원할 수 있는 국가 정책을 도모하기 위한 방안을 제시한다는 점에서 의의가 있다. 본 연구 결과 가중치의 적용은 Common Neighbor 모형과 Jaccard's coefficient 모형 모두에서 긍정적인 성과를 나타냈는데 상대적으로는 가중치가 적용된 Common Neighbor 모형에서의 정확도가 더 개선된 것으로 도출되었다. 즉, Common Neighbor 모형에서는 4,136개 중 650개를 예측한 반면, 가중치를 적용한 Common Neighbor 모형에서는50개의 정답이 증가한 700개를 예측하는 효과를 보였다. 한편, 상대적으로 Jaccard 계수의 경우는 약간의 성능 개선은 있으나 그 차이가 미미한 것으로 나타났다.

주제어 : 산학연, 네트워크 분석, 활성화, 국가 R&D, 관계 예측

논문접수일 : 2015년 7월 20일 논문수정일 : 2015년 9월 17일 게재확정일 : 2015년 9월 17일
투고유형 : 국문일반 교신저자 : 김우주

1. 서론

1.1 연구의 배경 및 목적

최근 한국경제의 새로운 패러다임으로 '창조경제'가 제시되고 있으며, 창조경제의 패러다임에서 산학연 협력은 더욱 강조될 것이다. 특히, 신기술 및 신상품 창조의 새로운 유형의 도구로 대두되고 있는 산학연 네트워크는 각기 특성이 다른 주체들의 연계와 협력을 통해 시너지 효과

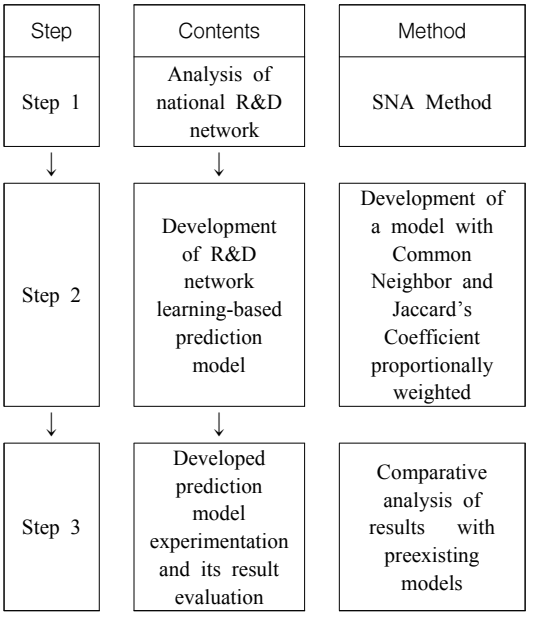
를 도출할 수 있는 강력한 방법이다. 산학연 협력사업에 있어서도 그 본질이 참여주체간의 역할분담과 협력이기 때문에 네트워크가 얼마나 잘 구성되어 있는지에 따라 성과가 상당히 달라질 수 있다(Yang, 2006). 이에 따라 산학연 협업 활성화를 위한 네트워크 예측 연구의 필요성은 날로 증가하고 있다.

특히, 정부 3.0 추진에 따라 국가 R&D 과제의 추진 절차 투명성에 대한 관심이 증가하였고, 과제 기획과 평가 및 과제 참여에 대한 현황 모니터링의 필요성이 증대되고 있다. 과제 연구 참여자는 과제 기획 및 평가 위원에 의해 선정되므로 과제 기획, 평가, 연구 참여 관계에 대한 개인 단위의 상호 연관 관계를 분석하여 선제적으로 관리하는 정책이 필요하다. 본 논문은 국가 R&D 과제를 통해 형성된 네트워크를 분석하고 향후 추가로 발생될 수 있는 네트워크를 예측하고자 한다. 네트워크 예측을 하게 되면 미래에 어느 분야의 네트워크 활성화가 취약한지 등을 용이하게 파악할 수 있게 된다. 이를 통해 국가는 효율적으로 정부 지원금을 투입하고 맞춤형 정책 지원 등을 용이하게 할 수 있게 된다. 즉, 융복합 기술 개발을 위해서는 균등한 지원보다 산학연 네트워크 파악과 예측을 통해 효율적인 정책 마련이 필요하다는 것이다. 따라서 본 연구는 R&D사업을 분석대상으로 산학연 네트워크를 분석하고 R&D네트워크 변화 예측 방법을 연구하고자 한다.

분석 대상 사업은 대표적인 중장기 연구개발 사업으로 산학연 컨소시엄으로 과제가 수행되며 신산업, 정보통신산업, 주력산업 분야에 대해 과제별 특성에 따라 3년부터 5년까지 지원되는 사업이다.

1.2. 연구분석 틀

본 연구에서는 산학연 협업 활성화를 위한 R&D네트워크 변화 예측을 위하여 사회망 분석(SNA)과 가중학습모형 기반의 관계 예측 방법론을 이용하고자 하며 전체적인 연구 수행 방안은 아래 <Figure 1>에서 설명하고 있다.



〈Figure 1〉 Step-by-Step research framework

<Figure 1>에서와 같이 먼저 1단계로는 국가 R&D 네트워크를 개괄적으로 분석하여 현재 산학연 네트워크 현황을 살펴보고, 2단계에서는 네트워크 변화 예측에 많이 활용되고 있는 기존 예측모형인 Common Neighbor 모형과 Jaccard's Coefficient 모형(Coulon, 2005)을 기반으로 네트워크의 관계 유형의 변화 기여도를 반영할 수 있는 가중 예측 모형과 이러한 가중치를 학습할 수 있는 학습 방법을 제시하고자 한다. 한편 3단계

에서는 제안하는 가중 모형과 기존 예측 모형의 성과 평가를 수행하여 새로운 예측 모형의 타당성을 실험을 통해 검증하고자 한다.

본 연구의 구체적 수행을 위해 네트워크 데이터 처리를 위해 MySQL을, SNA를 위해서는 Gephi를 사용하였다.

1.3 R&D 네트워크 데이터 기초 통계 분석

본 연구의 분석 대상은 국가 R&D사업 중 2009년부터 2011년까지의 기간 동안 지원된 신규과제 총 845개 과제이며 관련 기관은 총 2,974개에 달한다. 과제 관련자는 기술로드맵위원회, 기술위원회, 기획위원회, 기획실무위원회, 과제선정평가위원회, PD, 참여연구원 등으로 총 29,053명에 달한다. 먼저 관계자 분석 대상의 기초 통계량은 다음의 <Table 1>과 같다.

위 <Table 1>에 따르면 2009년에는 431개 과제에 대해 12,116명이 기술로드맵위원회, 기술위원회, 기획위원회, 기획실무위원회, 과제선정평가위원회, PD, 참여연구원 등으로 참여하였으며, 2010년에는 271개 과제에 대해 12,504명, 2011년에는 143개 과제에 8,252명의 관계자가 있는 것으로 나타났다. 이에 따라 전체 845개 과제에서의 29,053명의 관계자들의 관계 유형별로 보면 기획 관계자가 4,320명, 평가관계자는 1,860명, 수행 관계자는 24,699명인 것으로 나타났다.

2. 이론적 배경 및 선행연구 분석

2.1 관계 예측(Link prediction)

미래에 발생하는 사건을 알고자 하는 인간의 욕망은 항상 존재하여 왔고, 이러한 욕망은 미래를 좀 더 합리적이고, 과학적으로 예측하고자 하는 노력으로 나타났다. 본 연구에서는 R&D 네트워크의 관계 변화를 예측하고자 하며 이를 토대로 이상적인 R&D네트워크 구축을 위해 정책적 의사결정에 도움을 주고자 한다.

임의의 네트워크의 변화는 두 가지 관점에서 발생할 수 있는데 하나는 참여하는 노드(node)의 변화이고 다른 하나는 참여중인 노드 간의 관계 변화일 것이다. 이 중 본 연구는 관계 변화에 초점을 맞추고자 하며 이는 곧 노드 간의 연결(link)의 생성 및 소멸로서 인지할 수 있다. 이러한 연결의 생성 여부를 예측하는 분야를 연결(link) 예측이라 하며 특정 시점에 연결되지 않은 노드 간 유사도를 측정하고 이 유사도를 미래 시점의 노드 생성 지표를 사용하여 예측하는 방법론들이 일반적으로 활용되어 왔다 (Coulon, 2005). 노드 간 유사도 측정 방법은 크게 노드 이

〈Table 1〉 Statistical analysis based on subjects

year	number of projects	ratio of each year's project	total number of involved officials	number of planning officials	number of assessment officials	number of research officials
2009	431	51.01%	12,116	1,992	760	9,993
2010	271	32.07%	12,504	2,075	786	10,311
2011	143	16.92%	8,252	1,366	484	6,636
Total	845	100.00%	29,053	4,320	1,860	24,699

웃 기반 유사도 측정 방법과 노드 간 경로기반 유사도 측정 방법으로 분류할 수 있으며 일반적으로 사용되는 노드 이웃 기반 유사도 측정 방법론으로는 Common Neighbor, Jaccard's Coefficient, Adamaic/Adar, Preferential Attachment 등이 있고, 경로 기반 유사도 측정 방법론으로는 Katz β, Hitting time, rooted PageRankα, SimRankγ 등이 있다.

연결 예측(link prediction)에 대해서는 다양한 논의가 있어왔는데, 사회연결망에 대한 연결 예측 문제에 대한 연구로는 David Liben Nowell and Jon Kleinberg의 연구가 가장 대표적이다(Liben-Nowell and Kleinberg, 2007). 이 연구는 사회 연결망은 매우 동적인 개체이며, 기존 Common Neighbor, Jaccard's Coefficient, Adamaic/Adar, Preferential Attachment 등의 모형은 노드의 유형 및 시계열 분석 등을 하지 않았다는 한계점이 있음을 밝히고 있다.

한편, Mena Chalco 외 3인은 브라질의 계량서지학 공저 네트워크 분석을 통해 연구자들간의 연구 활동에 대한 구조와 동적 변화를 분석하였다(Mena Chalco et al., 2014). 이 연구는 농업, 생물, 인문학, 예술학 등 8개 분야에 대한 공저 네트워크를 분석한 것이며 네트워크 협력성이 높은 분야는 생물과학 분야인 것 등을 시각화하여 보여주었다. 또한, 이 연구는 시간 변화에 따라 공저 네트워크가 부단히 발전하고 있다고 분석하며, 학제간의 종속성과 의존성 등에 대해 계량화하였다. Lu and Zhou는 link prediction 알고리즘이 발전하면서 다양한 모형과 유형이 개발되고 있다고 기술하고 있다(Lu and Zhou, 2011). 이와 관련하여 알고리즘이 크게 3가지로 적용되고 있는데, 네트워크 구현, 진화하는 매커니즘에 대한 평가, 네트워크 분류하는데 적용되고 있다고 분석하였다.

Tylenda 외 2인은 기존 link prediction 연구가 시간적 변화를 고려하지 않고 한 시점의 네트워크만을 분석함에 따라 끊임없이 변화하는 네트워크의 특성을 고려하지 않았다고 분석하였다(Tylenda et al., 2009). 이에 따라 시계열 정보를 link prediction 예측에 반영하여 보다 더 정확한 예측을 하였다. 그러나, 국내 산학연 R&D 네트워크에 대해 노드 유형 및 시계열 등을 반영하여 연결 예측 분석을 한 연구는 미비한 실정이며, 특히 연구분야별 네트워크 동적 변화에 대한 파악을 통해 네트워크 활성화 방안을 제시한 연구는 전무한 상황이다.

2.2 네트워크 분석(network analysis)

본 연구의 이론적 배경은 '사회 네트워크 분석'에 기초한다. '사회 네트워크'란 사람 또는 특정 노드들이 연결되어 있는 관계망을 지칭한다. 네트워크는 각 개체(노드)에 의해 형성되기 때문에, 개별 개체의 선택에 의해 연결관계 혹은 전체 윤곽이 변화한다.

사회 네트워크 분석의 연결망의 형태에 대해서는 다양한 논의가 있어왔는데, 그 중 Granovetter의 연구가 가장 대표적이다(Granovetter, 1983). Granovetter는 연결망을 강한 연결(Strong ties)과 약한 연결(Weak ties)로 구분한다. 강한 연결과 약한 연결은 노드 관계의 강도에 따라 구별되는데, 관계의 강도는 시간의 양, 감정의 강도, 친밀도, 그리고 상호 도움의 정도로 정의된다고 분석했다. 가장 잘 알려진 그의 연구는 취업자가 취업에 관한 정보를 얻는 과정에서 정보를 알려 준 사람을 접촉수로 연결망의 강도를 정의하였다. 다음에서는 특히 본 연구에서 분석하고자 하는

R&D 분야와 관련된 선행연구를 살펴보고자 한다.

Ahuja 외 3인은 조직 내 직접 대면하여 연구개발을 수행하는 R&D 그룹이 아닌, 상이한 조직 사이 공통된 목표를 가진 사람들로 자발적으로 구성된 네트워크 형 R&D 그룹을 Virtual R&D 그룹이라 정의하고 이들의 네트워크 특징을 분석하였다. 이러한 Virtual R&D 그룹에 있어 성과를 창출하는 주요 요인은 개인적 역할 특징(individual role characteristics)과 구조적 위치(structural position)로 구분할 수 있는데, 개인적 역할 특징은 기능적 역할(functional role), 직위, 커뮤니케이션 역할로 구분하고, 구조적 위치는 개인적 집중도(individual centrality)로 세분화하였다. 이를 통해 위 요인들은 모두 네트워크 활동에 있어 매우 중요한 영향을 끼치고, 개인의 R&D 성과에 긍정적인 영향을 주고 있음을 보였다(Ahuja et al., 2003).

Coulo는 기술혁신 연구 활동에 대해 SNS 분석이 어떻게 수행되고 있는가를 기존 연구를 통해 분석하였다. 이를 통해 대부분의 연구는 기관 수준에서의 실증 분석에 치중되어 있고, 네트워크의 성과보다 조직 내 네트워크를 결정하는 주요 요인을 분석하는데 집중되어 있음을 보였다. 구체적으로 분석 대상 기존 연구 중 46%가 기관, 기업의 네트워크를30%가 성과(논문, 특허)를 활용하여 네트워크를 분석하였다(Coulon, 2005).

기존 네트워크 분석은 주로 사회적 근접성(social proximity)이 지식확산에 영향을 미치는 프로세스에 대한 연구 또는 네트워크 구조가 기술혁신 성과에 영향을 미치는 프로세스에 대한 연구에 국한되어 있었다. 구체적으로 특허, 개인, 기관 사이 상호작용 및 관계 프로세스가 새로운 지식 등의 창출과 같은 성과에 어떠한 영향을 미치는가에 집중하고 있다.

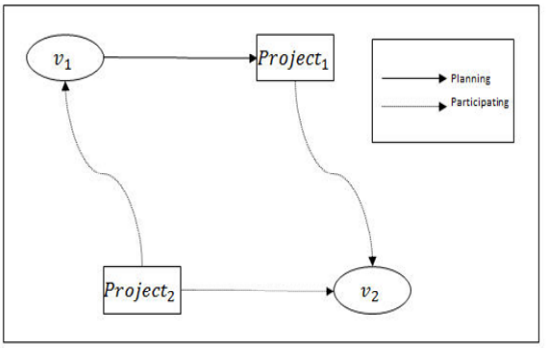
이와 같은 기존 연구들은 다양한 분야에 대해 현상을 분석하고, 특징을 도출하기 위한 방안으로 사회 네트워크 분석 이론 및 방법론을 활용하였으나, 특정한 연구 구성원간의 경향성 분석에 그치고 있다.

3. R&D 네트워크 연결 예측 모형의 개발

본 장에서는 R&D 네트워크의 특성을 고려한 네트워크 연결 예측 모형을 제안한다. 3.1절에서 R&D 네트워크의 정의와 그 표현방법을 정의하고, 3.2절에서 R&D 과제 관계자들 간의 관계유형 가중치 측정 방법론을 제안하며, 3.3절에서 측정한 관계유형 가중치 기반 가중 유사도 측정 방법을 제안한다.

3.1 R&D 네트워크의 정의와 표현

본 절에서는 R&D 네트워크 연결 예측에 앞서

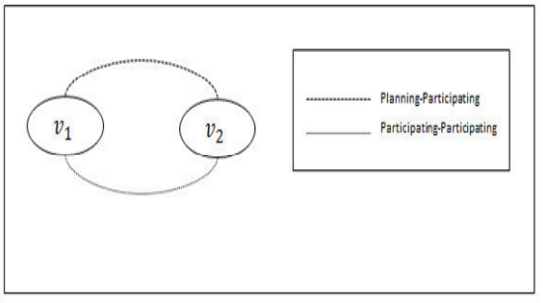


〈Figure 2〉 R&D Network example

R&D 네트워크 표현 방법을 정의하고자 한다. R&D 네트워크는 매년 신규 선정되는 과제들에 참여하는 과제 관계자들과 관계자들 간의 관계를 표현하기 때문에 시간에 따라 변화하며, 여러 종류의 관계자 간 관계 유형을 가지는 특성을 가진다.

위의 <Figure 2>는 가상의 R&D 네트워크를 도식화 한 것이다. <Figure 2>는 v_1이 기획한 *Project*$_1$에 v_2가 참여연구원으로 참여하고, v_1와 v_2가 *Project*$_2$에 참여연구원으로 참여하는 것을 나타낸다.

본 연구는 R&D 네트워크 관계자 간 관계 생성 여부 예측을 목적으로 하기 때문에, 관계자 간 R&D 네트워크(과제들이 제거된)를 분석 대상으로 한다. 아래의 <Figure 3>은 과제 노드들이 제거된 관계자 간 R&D 네트워크를 도식화 한 것이다.



〈Figure 3〉 Modified R&D Network example (projects removed)

<Figure 3>은 v_1와 v_2가 기획자-참여자 관계와 참여자-참여자 관계로 과제를 수행했음을 나타낸다. 관계자 간 R&D 네트워크는 관계자들이 가지는 여러 유형의 관계들을 표현한다.

따라서 본 연구에서 다루는 R&D 네트워크는 관계자들 간의 관계 유형을 구분할 수 있어야 하며 임의의 두 관계자들 사이의 여러 관계들을 표현할 수 있어야 하고, 관계의 발생 시점도 구분하여야 하기 때문에 다음과 같은 labeled multigraph로 표현하고자 한다.

〈Table 2〉 Definition of terms related to R&D Network

Notation	Definition
G	R&D network G $G = (V, E)$
V	Set of vertices(nodes) $V = \{v_1, v_2, \cdots, v_n\}$
E	Set of edges $E \subseteq \{((v_s, v_t); t, l) \mid v_s, v_t \in V, v_s \neq v_t, t \in T, l \in L\}$
T	Set of time stamps $T = \{t-1, t, t+1\}$
L	Set of edge labels $T = \{t-1, t, t+1\}$
G_t	R&D network in time t $G_t = (V, E_t)$
E_t	Set of edges in time t $E_t = E \wedge hasYear.(= t)$
P_t	Set of node pairs in time t $P_t = \bigcup_{\forall e \in E_t} getPair(e)$
$\Gamma_t(v)$	Set of neighbor nodes of node in time t
$L_t^{v_s,v_t}$	Set of labels within pair in time t $L_t^{v_s,v_t} = \bigcup_{\forall e \in E_t^{v_s,v_t}} getLabel(e)$
$E_t^{v_s,v_t}$	Set of edges which have source node as and target node as in time t $E_t^{v_s,v_t} = E_t \wedge hasSource.(= v_s) \wedge hasTarget.(= v_t)$

이상에서의 R&D 네트워크 정의를 활용하여 R&D 네트워크 연결 예측 문제를 정의하면 다음과 같다. 연결 예측에서는 임의의 두 관련자 v_s, v_t가 t+1 시점에 관계를 형성할 지

$(E_{t+1}^{v_s,v_t} \neq \emptyset)$를 과거의 일정 기간 [*t-k*, *t*] 동안의 R&D 네트워크의 부분 그래프 집합인 $G_{[t-k,t]}=\{G_t, G_{t-1}, \dots, G_{t-k}\}$를 기반으로 예측하고자 한다.

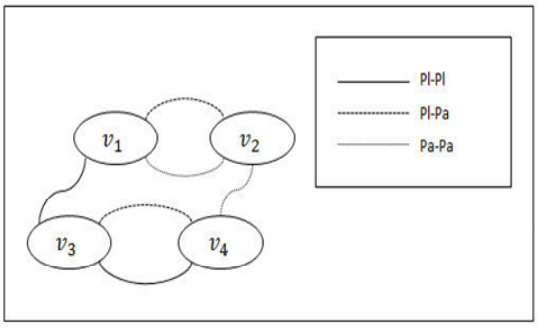
3.2 연결 생성 기여 가중치 측정

임의의 쌍 (v_s, v_t)이 특정 시점 *t*+1에 나타날지의 여부는 과거 연구들에서 이들 두 노드 v_s, v_t가 $G_{[t-k,t]}$에서 가지는 네트워크 기반의 유사도 측정 방법들을 활용하였으며 common neighbor 수를 기반으로 유사도를 측정하는 것이 보편적이었다. 그러나 이러한 방식은 시간적 흐름에 따른 관계 유형들의 새로운 연결 생성에의 기여도를 고려하지 못하는 단점을 가지고 있었다. 따라서 본 연구는 관계 유형들이 새로운 연결 생성에 기여하는 연결 생성 기여 가중치 측정 방법론을 제안하고자 한다.

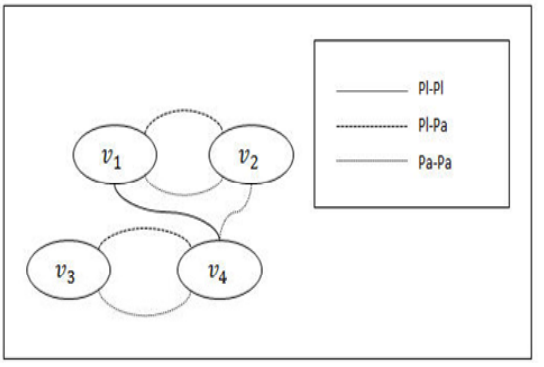
특정 시점$(T = t)$에 노드 v_s와 v_t 사이에 새로운 관계가 형성될 가능성은 이전 시점$(T = t-1)$의 노드 , v_s, v_t와 노드 v_s, v_t의 공통 이웃 노드 v의 관계 유형들에 영향을 받는다고 가정한다.

만약 *k*=1이라 하면, 임의의 관계 유형 *l*의 연결 생성 기여 가중치를 아래의 수식과 같이 잠재 관계 형성 노드 쌍과 *t* - 1 시점의 공통이웃 사이의 엣지(edge) 집합 내 관계 유형 *l*출현 횟수 대비 *t* 시점에 새로 형성된 관계(노드 쌍)과 *t* - 1 시점의 공통이웃 사이의 엣지(edge) 집합 내 관계 유형 *l* 출현 횟수의 비율로 측정할 수 있다.

$$wel_{t-1,t}(l) = \frac{\sum_{\forall(v_s,v_t)\in P_t - P_{t-1}} \sum_{\forall v\in \Gamma_{t-1}(v_s)\cap\Gamma_{t-1}(v_t)} contains\left(L_{t-1}^{v_s,v}, l\right) + contains\left(L_{t-1}^{v,v_t}, l\right)}{\sum_{\forall(v_s,v_t)\in V\times V - P_{t-1}} \sum_{\forall v\in \Gamma_{t-1}(v_s)\cap\Gamma_{t-1}(v_t)} contains\left(L_{t-1}^{v_s,v}, l\right) + contains\left(L_{t-1}^{v,v_t}, l\right)} \quad (1)$$



〈Figure 4〉 R&D Network in time $T = t-1$



〈Figure 5〉 R&D Network in time $T = t$

<Figure 4>는 $t-1$ 시점의 가상의 R&D 네트워크를 도식화 한 것이며, <Figure 5>는 t 시점의 가상의 R&D 네트워크를 도식화 한 것이다.

예를 들어, '기획-참여' 관계유형의 연결 생성 기여 가중치는 잠재 관계 형성 노드 쌍과 $t-1$ 시점의 공통이웃 사이의 엣지(edge) 집합 내 관계 유형 *l* 출현 횟수 대비 t 시점에 새로 형성된 관계(노드 쌍)과 $t-1$ 시점의 공통이웃 사이의 엣지(edge) 집합 내 관계 유형 *l* 출현 횟수의 비율이다.

'기획-참여' 관계유형의 잠재 관계 형성 노드

쌍과 $t-1$ 시점의 공통이웃 사이의 엣지(edge) 집합 내 관계 유형 l 출현 횟수는 $t-1$ 시점의 R&D 네트워크에 존재하지 않는(t시점에 관계가 새로 형성될 잠재성이 있는) 모든 노드 쌍들 $(\forall(v_s, v_t) \in V \times V - P_{t-1})$과 $t-1$ 시점의 공통이웃$(\forall v \epsilon \Gamma_{t-1}(v_s) \cap \Gamma_{t-1}(v_t))$ 사이의 엣지 집합 내 '기획-참여' 관계유형 출현 횟수이다.

$t-1$ 시점의 R&D 네트워크에 존재하지 않는 노드 쌍 집합은 $\{(v_1, v_4), (v_2, v_3)\}$이다. 노드 쌍 (v_1, v_4)에 대한 $t-1$ 시점의 공통이웃은 $\{v_2, v_3\}$ 이고, 노드 쌍 (v_2, v_3)에 대한 $t-1$ 시점의 공통이웃은 $\{v_1, v_4\}$이다. 노드 쌍들과 공통이웃 사이의 엣지 집합 내에 '기획-참여' 관계유형은 4번 출현하므로, '기획-기획' 관계유형의 잠재 관계 형성 노드 쌍과 $t-1$ 시점의 공통이웃 사이의 엣지(edge) 집합 내 관계 유형 l 출현 횟수는 4이다.

t 시점에 새로 형성된 관계(노드 쌍)과 $t-1$ 시점의 공통이웃 사이의 엣지(edge) 집합 내 관계 유형 l 출현 횟수는 t 시점에 새로 형성된 ($t-1$ 시점에 존재하지 않았던) 관계들($\forall(v_s, v_t) \in P_t - P_{t-1}$)과 $t-1$ 시점의 공통이웃 사이의 엣지 집합 내 '기획-참여' 관계유형 출현 횟수이다.

t 시점에 새로 형성된 관계는 $P_t - P_{t-1} = \{(v_1, v_2), (v_1, v_4), (v_2, v_4), (v_3, v_4)\} - \{(v_1, v_2), (v_1, v_3), (v_2, v_4), (v_3, v_4)\} = \{(v_1, v_4)\}$이며, 노드 쌍 (v_1, v_4)에 대한 잠재 공통이웃은 $\{v_2, v_3\}$ 이다. 노드 쌍과 공통이웃 사이의 엣지 집합 내에 '기획-참여' 관계유형은 2번 출현하므로, '기획-기획' 관계유형의 잠재 관계 형성 노드 쌍과 $t-1$ 시점의 공통이웃 사이의 엣지(edge) 집합 내 관계 유형 l 출현 횟수는 2이다.

즉, '기획-참여' 관계 유형의 연결 생성 기여 가중치는 잠재 관계 형성 노드 쌍과 $t-1$ 시점의 공통이웃 사이의 엣지(edge) 집합 내 관계 유형 l 출현 횟수 대비 t 시점에 새로 형성된 관계(노드 쌍)과 $t-1$ 시점의 공통이웃 사이의 엣지(edge) 집합 내 관계 유형 l 출현 횟수의 비율이므로 0.5가 된다.

3.3 가중 유사도 측정 방법

David Liben Nowell and Jon Kleinberg 은 연결되지 않은 노드 쌍들에 대한 공통 이웃(Common Neighbor) 개수 기반 유사도 측정 모형들을 통해 미래 시점$(T = t + 1)$에 연결될 가능성이 높은 노드 쌍들을 예측하였으며, 그 중 Common Neighbor와 Jaccard's Coefficient 모형이 정밀도(precision)가 가장 높았다.

Nowell의 연구를 바탕으로 본 연구는 Common Neighbor, Jaccard's Coefficient 모형을 기반 모형으로 하여 측정한 관계유형 별 연결 생성 가중치를 반영할 수 있는 유사도 측정 방법을 제안한다.

3.3.1. Common Neighbor 모형 기반 가중 유사도 측정 방법

t+1시점의 노드 쌍 (v_s, v_t)에 대한 Common Neighbor 모형 기반 유사도 측정 방법은 두 노드 v_s, v_t의 t시점 공통 이웃 노드의 개수로 유사도를 측정한다. 이 방법은 두 노드 v_s, v_t와 t시점 공통 이웃 사이의 경로 개수로 해석 가능하다.

$$CN_{t+1}(v_s, v_t) = |\Gamma_t(v_s) \cap \Gamma_t(v_t)| \qquad (2)$$

t+1시점의 노드 쌍 (v_s, v_t)에 대한 Common

Neighbor 모형 기반 가중 유사도 측정 방법은 두 노드 v_s, v_t와 t시점 공통 이웃 사이의 경로에 있는 관계 유형들의 과거 연결 생성 기여 가중치를 반영한다.

$$WCN_{t+1}(v_s, v_t) = \sum_{\forall v_{cn} \in \Gamma_t(v_s) \cap \Gamma_t(v_t)} \left(\frac{\sum_{l_1 \in L_t^{v_s, v_{cn}}} wel_{t-1,t}(l_1)}{\left| L_t^{v_s, v_{cn}} \right|} + \frac{\sum_{l_2 \in L_t^{v_{cn}, v_t}} wel_{t-1,t}(l_2)}{\left| L_t^{v_{cn}, v_t} \right|} \right) \quad (3)$$

즉, t+1시점의 노드 쌍 (v_s, v_t)에 대한 Common Neighbor 모형 기반 가중 유사도는 t 시점의 모든 공통 이웃들에 대한 노드 v_s와 공통 이웃 v_{cn} 사이의 연결 생성 기여 가중치 평균과 노드 v_t와 공통 이웃 v_{cn} 사이의 연결 생성 기여 가중치 평균의 합산으로 측정한다.

3.3.2. Jaccard's Coefficient 모형 기반 가중 유사도 측정 방법

t+1시점의 노드 쌍 (v_s, v_t)에 대한 Jaccard's Coefficient 모형 기반 유사도 측정 방법은 두 노드 v_s, v_t의 t시점 이웃 노드 개수 대비 공통 이웃 노드의 개수의 비율로 유사도를 측정한다. 이 방법은 두 노드 v_s, v_t와 이웃 노드 사이 경로 개수 대비 두 노드 v_s, v_t와 t시점 공통 이웃 사이의 경로 개수 비율로 해석 가능하다.

$$JA_{t+1}(v_s, v_t) = \frac{|\Gamma_t(v_s) \cap \Gamma_t(v_t)|}{|\Gamma_t(v_s) \cup \Gamma_t(v_t)|} \quad (4)$$

t+1시점의 노드 쌍 (v_s, v_t)에 대한 Jaccard's Coefficient 모형 기반 가중 유사도 측정 방법은 두 노드 v_s, v_t와 t시점 이웃 노드 사이의 경로에 있는 관계 유형들의 연결 생성 기여 가중치를 반영한다.

$$WJA_{t+1}(v_s, v_t) = \frac{\sum_{\forall v_{cn} \in \Gamma_t(v_s) \cap \Gamma_t(v_t)} \left(\frac{\sum_{l_1 \in L_t^{v_s, v_{cn}}} wel_{t-1,t}(l_1)}{\left| L_t^{v_s, v_{cn}} \right|} + \frac{\sum_{l_2 \in L_t^{v_{cn}, v_t}} wel_{t-1,t}(l_2)}{\left| L_t^{v_{cn}, v_t} \right|} \right)}{\sum_{\forall v_1 \in \Gamma_t(v_s)} \left(\frac{\sum_{l_1 \in L_t^{v_s, v_1}} wel_{t-1,t}(l_1)}{\left| L_t^{v_s, v_1} \right|} \right) + \sum_{\forall v_2 \in \Gamma_t(v_t)} \left(\frac{\sum_{l_2 \in L_t^{v_t, v_2}} wel_{t-1,t}(l_2)}{\left| L_t^{v_t, v_2} \right|} \right)} \quad (5)$$

즉, t+1시점의 노드 쌍 (v_s, v_t)에 대한 Jaccard's Coefficient 모형 기반 가중 유사도는 t 시점의 모든 이웃들에 대한 노드 v_s, v_t와 이웃 노드 사이의 연결 생성 기여 가중치 평균 합산 대비 t 시점의 모든 공통 이웃들에 대한 노드 v_s와 공통 이웃 v_{cn} 사이의 연결 생성 기여 가중치 평균과 노드 v_t와 공통 이웃 v_{cn} 사이의 연결 생성 기여 가중치 평균의 합산 비율로 측정한다.

4. 가중치 기반 연결 예측 모형의 성능 평가

4.1 분석 데이터의 설정

R&D 네트워크 연결 예측 모형의 검증을 위해 2009년부터 2011년까지 매년 신규 선정된 과제들에 다양한 형태로 참여하는 과제 관계자들 간의 관계를 수집하였는데, 최소 1번의 과제 기획 경험을 가지며, 전체 기간에 어떤 역할로라도 모두 등장하는 관계자들만을 대상으로 한다. 인적 네트워크 관점에서 보면 어느 시점에서든 관계

형성이 없는 관계자는 제외하고 수집하였다.

이렇게 수집된 관계자들에 대한 기술영역별 통계를 살펴보면 다음과 같다. 총 472명의 과제 관계자가 추출되었고, 다음의 표는 각 분야 별로 분야 내부 및 외부로의 연계 정도를 보여주고 있다.

분석대상으로 결정된 472명의 연구자들에 대한 R&D 네트워크 기초 통계는 다음의 <Table

〈Table 3〉 Statistical analysis of researcher by technological domains

Label	Main field of research	Number of researchers	Number of researchers associated with			
			Only one field of research		More than one field of research	
A1	BcN	18	9	(50.00%)	9	(50.00%)
A2	IT Convergence	133	2	(1.50%)	131	(98.50%)
A3	LED/Light	35	6	(17.14%)	29	(82.86%)
A4	RFID/USN	30	1	(3.33%)	29	(96.67%)
A5	S/W	31	9	(29.03%)	22	(70.97%)
A6	Metallic Material	23	8	(34.78%)	15	(65.22%)
A7	Nano-base	31	6	(19.35%)	25	(80.65%)
A8	Display	34	13	(38.24%)	21	(61.76%)
A9	Digital TV/Broadcast	45	8	(17.78%)	37	(82.22%)
A10	Robot	37	16	(43.24%)	21	(56.76%)
A11	Bio-	15	1	(6.67%)	14	(93.33%)
A12	Semiconductor	58	13	(22.41%)	45	(77.59%)
A13	Production-base	38	8	(21.05%)	30	(78.95%)
A14	Production system	49	9	(18.37%)	40	(81.63%)
A15	Textile and Apparel	25	15	(60.00%)	10	(40.00%)
A16	Medical Devices and Appliances	24	1	(4.17%)	23	(95.83%)
A17	Mobile communication	30	8	(26.67%)	22	(73.33%)
A18	Automobile	48	12	(25.00%)	36	(75.00%)
A19	Shipbuilding and Marine Engineering	24	9	(37.50%)	15	(62.50%)
A20	Knowledge/Information Intelligence Service	30	8	(26.67%)	22	(73.33%)
A21	Knowledge/Information Intelligence Security	13	6	(46.15%)	7	(53.85%)
A22	Next Generation Computing	20	7	(35.00%)	13	(65.00%)
A23	Clean-base	14		-	14	(100.00%)
A24	Plant Engineering	6		-	6	(100.00%)
A25	Home Network/Information Electronic Appliances	50	5	(10.00%)	45	(90.00%)
A26	Chemical Process Material	37	18	(48.65%)	19	(51.35%)
Aggregate	Total fields of research	472	198	(41.95%)	274	(58.05%)

4 >와 같다.

〈Table 4〉 Statistics of links and node pairs in R&D network

T	$\|E_t\|$	$\|P_t\|$
$t-1$	9,814	5,549
t	12,320	6,254
$t+1$	7,263	4,687

4.2 가중치 기반 모형 실험 및 성과 비교

R&D 네트워크 연결 예측 실험은 다음과 같은 과정으로 진행한다.

1) $G_{[t-1,t]}$를 기반으로 모든 관계유형 $\forall l \in L$에 대한 연결 생성 기여 가중치를 학습한다.

2) t+1 시점에 대해 잠재적으로 발생할 수 있는 모든 신규 노드 쌍인 $\forall(v_s, v_t) \in V \times V - (P_{t-1} \cup P_t)$에 대해 유사도를 측정한다.

3) t+1시점에 실제로 신규 발생한 노드 쌍 수인 K를 이용하여 유사도 상위 K개(K개$(K =, (P_{t+1} - (P_{t-1} \cup P_t))$, t+1 시점에 새로 형성된 관계 수)의 노드 쌍들을 신규 발생될 것이라 예측한다.

4) 각각 K개의 원소를 갖는 추정 노드 쌍 집합과 실제 발생 노드 쌍 집합을 비교하여 연결 예측 기법의 성능을 분석 평가한다.

이상의 연결 예측 실험 과정을 통해 기존 연구에서 사용한 Common Neighbor 및 Jaccard계수의 성능과 이들 각각의 방법에 연결 유형 가중치를 반영한 가중 Common Neighbor 및 가중 Jaccard 계수 방법의 성과는 아래의 <Table 5>와 같이 나타났다.

Weighted Common Neighbor 모형은 2011년에 새로 형성되는 4,136개의 관계들 중 700개의 관계들을 정확히 예측하였으며 16.92% 정밀도를 보였다. 한편 Weighted Jaccard's Coefficient 모형은 2011년에 새로 형성되는 4,136개의 관계들 중 822개의 관계들을 정확히 예측하였으며 19.87%의 정밀도를 보여주었다.

〈Table 5〉 Relationship prediction experiment results

Predictor	K-value	#(true positive)	Precision(%)
Common Neighbor	4,136	650	15.72
Weighted Common Neighbor	4,136	700	16.92
Jaccard's Coefficient	4,136	810	19.58
Weighted Jaccard's Coefficient	4,136	822	19.87

가중치의 적용은 Common Neighbor 모형과 Jaccard's coefficient 모형 모두에서 긍정적인 성과를 나타냈는데 상대적으로는 Common Neighbor 모형에서의 성과 개선 효과가 두드러지게 나타나고 있는데 50개의 정답 증가와 정밀도 측면에서 1퍼센트 포인트 이상 개선 효과를 보이고 있다. 반면 상대적으로 Jaccard 계수의 경우는 약간의 성능 개선은 있으나 그 차이가 미미하게 나타났다.

그 이유를 잠재적으로 추정하면 가중치의 학습이 Common Neighbor 방법을 기준으로 이루어졌으며, 오히려 Common Neighbor 방법에 비해 Jaccard 계수가 가지는 장점을 Common Neighbor 기반의 가중치를 결합함으로써 그 특징을 둔감하게 만든 측면이 있다.

5. 결론

5.1 연구의 정책적 함의

본 연구는 국가 R&D 사업에 참여하는 각 주체들 간의 관계를 분석하여 현존하는 국가 R&D 지식 네트워크 현황을 살펴보고, 네트워크 예측 정확도를 높이기 위한 연구를 하였다. 이를 위해 관계 유형들에 대한 연결 생성 기여 가중치를 개발하였고 연결 생성 기여 가중치 기반 유사도 측정 방법을 개발하였다. 가중치의 적용은 Common Neighbor 모형과 Jaccard's coefficient 모형 기반으로 하였으며, 결과는 기존 모형보다 예측 정확도를 일정부분 높이는 효과를 도출하였다. 특히, 가중치를 적용한 Common Neighbor 모형에서의 성과 개선 효과가 두드러지게 나타남을 확인하였다.

본 연구는 효율적으로 산학연 사업을 지원하고, 융합 R&D 사업 등을 효과적으로 지원할 수 있는 국가 정책 마련의 기초 연구이다. 또한, 산학연 협업 네트워크의 과제 수행자간 관계를 사전에 예측하여 선제적으로 관리 및 정책적 대응을 가능하게 한다는 점에서 의의가 있다.

5.2 연구의 한계점 및 향후 연구 방향

본 연구의 연결 예측 대상은 산업통상자원부의 산업융합원천기술개발사업 연구 기획자들의 행태를 예측하는데 초점을 맞춤으로써 기획 경험이 있는 연구자들에 그 대상을 한정하고 있다. 산업융합원천기술개발사업이 신기술을 개발하는데 있어서 산학연 협력관계 전체를 대변한다고 할 수는 없지만 3년간의 기획 및 평가 대상자를 대상으로 하였기 때문에 타 사업에서도 유사성을 가질 것으로 사료된다. 그러나, 분석과정에서 최첨단 기술분야인 특정 산업분야는 인적자원의 협소성을 고려한 예측이 필요하였으나 이에 대한 분석이 정교하지 않았기 때문에 향후 이에 대한 대응 방안 연구가 필요할 것이다.

향후 모든 과제 관계자들에 대한 관계 예측으로 그 범위를 넓히고자 하며, 또한 Jaccard 계수 기반의 가중치 획득 방법을 연구하여 그 효과를 더욱 제고 시킬 수 있도록 연구를 진행하고자 한다.

참고문헌(References)

Ahuja, M. K., D. F. Galletta, and K. M. Carley, "Individual Centrality and Performance in Virtual R&D Groups: An Empirical Study," *Management Science*, Vol.49, No.1(2003), 21~38.

Coulon, F., "The use of Social Network Analysis in Innovation Research: A Literature Review," Lund University, 2005.

Granovetter, M., "The Strength of Weak Ties: A Network Theory Revisited," *Sociological Theory*, Vol.1, No.1(1983), 201~233.

Liben-Nowell, D. and J. Kleinberg, "The Link Prediction Problem for Social Networks," *Journal of The American Society for Information science and technology*, Vol 58, No.7(2007), 1019~1031.

Lu, L. and T. Zhou, "Link prediction in complex networks: a survey," *Physica A: Statistical Mechanics and its Applications*, Vol.390, No.6(2011), 1150~1170.

Mena Chalco, J. P., L. A. Digiampietri, F. M. Lopes, and R. M. Cesar, "Brazilian bibliometric coauthorship networks," *Journal of the*

Association for Information Science and Technology, Vol.65, No.7(2014), 1424~1445.

Tylenda, T., R. Angelova, and S. Bedathur, “Towards time-aware link prediction in evolving social networks,” *Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, (2009).

Yang, S. - c., "Study on the policy of industry-education-research collaboration: A case of Daejun District,” *Master's Dissertation*, Choongnam University Graduate School of Public Administration, 2006.

Abstract

Predicting link of R&D network to stimulate collaboration among education, industry, and research

Mi-yeon Park* · Sangheon Lee** · Guocheng Jin** · Hongme Shen** · Wooju Kim***

The recent global trends display expansion and growing solidity in both cooperative collaboration between industry, education, and research and R&D network systems. A greater support for the network and cooperative research sector would open greater possibilities for the evolution of new scholar and industrial fields and the development of new theories evoked from synergized educational research.

Similarly, the national need for a strategy that can most efficiently and effectively support R&D network that are established through the government's R&D project research is on the rise.

Despite the growing urgency, due to the habitual dependency on simple individual personal information data regarding R&D industry participants and generalized statistical data references, the policies concerning network system are disappointing and inadequate.

Accordingly, analyses of the relationships involved for each subject who is participating in the R&D industry was conducted and on the foundation of an educational-industrial-research network system, possible changes within and of the network that may arise were predicted.

To predict the R&D network transitions, Common Neighbor and Jaccard's Coefficient models were designated as the basic foundational models, upon which a new prediction model was proposed to address the limitations of the two aforementioned former models and to increase the accuracy of Link Prediction, with which a comparative analysis was made between the two models. Through the effective predictions regarding R&D network changes and transitions, such study result serves as a stepping-stone for an establishment of a prospective strategy that supports a desirable educational-industrial-research network and proposes a measure to promote the national policy to one that can effectively and efficiently sponsor

* Process of technology policy cooperation, Yonsei University
** Department of Information and Industrial Engineering, Yonsei University
*** Corresponding author : Wooju Kim
Department of Information and Industrial Engineering, Yonsei University
50 Yonsei-ro, Seodaemun-Gu, Seoul, 120-749, Korea
Tel :+82-2-2123-5716, Fax: +82-2-3647807 E-mail: wkim@yonsei.ac.kr

integrated R&D industries.

Though both weighted applications of Common Neighbor and Jaccard's Coefficient models provided positive outcomes, improved accuracy was comparatively more prevalent in the weighted Common Neighbor. An un-weighted Common Neighbor model predicted 650 out of 4,136 whereas a weighted Common Neighbor model predicted 50 more results at a total of 700 predictions. While the Jaccard's model demonstrated slight performance improvements in numeric terms, the differences were found to be insignificant.

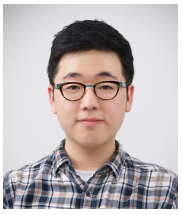
Key Words : Industry-university, Network analysis, activation, National R&D, link prediction

Received : July 20, 2015 Revised : September 17, 2015 Accepted : September 17, 2015
Type of Submission : Normal Track Corresponding Author : Wooju Kim

저 자 소 개

박 미 연

연세대학교 기술정책협동과정에서 박사과정을 수료하였으며, 2006년에 동국대학교에서 정치학으로 석사학위를 취득한 바 있다. 현재는 국회 법제사법위원회 소속 정책비서관으로 근무 중이다. 주요 연구 관심 분야는 빅데이터, 국가 R&D 법규 및 정책 등이며, 기존 연구는 The Journal of Society for e-Business Studies에 기재된 “A study of PD system effectiveness based on R&D network analysis” 등이 있다.



이 상 헌

현재 연세대학교 정보산업공학과 석박사 통합과정에 재학 중이며, 2013년에 연세대학교 정보산업공학 학사학위를 취득한 바 있다. 주요 연구 관심분야는 데이터 마이닝, 시맨틱 웹 마이닝 등이다.

김 국 성

2012년 6월에 중국 연변과학기술대학교에서 경영정보학과에서(복수 컴퓨터과학과) 학부를 마치고, 2015년 연세대학교 정보산업 공학과 석사 학위를 취득하였다. 주요 연구 관심 분야는 시맨팁 웹 마이닝이다.

심 홍 매

현재 연세대학교 정보산업공학과 석사과정에 재학 중이며, 2011년에 중국 해양대학교 컴퓨터과학기술학 학사과정을 취득한 바 있다. 주요 연구관심분야는 데이터 마이닝, 시맨틱 웹 마이닝 등이다.

김 우 주

현재 연세대학교 정보산업공학과 교수로 재직 중이며, 1994년에 KAIST에서 경영과학과 박사 학위를 취득한 바 있다. 주요 연구 관심 분야는 시맨틱 웹, 지식 관리 및 인공지능 웹 서비스 등이며, 기존 연구는 Decision Support Systems, Expert Systems with Applications, An International Interdisciplinary Journal, International Journal of Distributed Sensor Network 등의 논문지에 게재되었다.