

비즈니스 프로세스 수행자들의 Social Network Mining 에 대한 연구 Mining Social Networks from business process log

송민석¹, W.M.P. van der Aalst², 최인준¹

포항공과대학교 산업공학과¹, 아인트호벤공대 정보시스템학과²

Abstract

Current increasingly information systems log historic information in a systematic way. Not only workflow management systems, but also ERP, CRM, SCM, and B2B systems often provide a so-called "event log". Unfortunately, the information in these event logs is rarely used to analyze the underlying processes. Process mining aims at improving this problem by providing techniques and tools for discovering process, control, data, organizational, and social structures from event logs. This paper focuses on the mining social networks. This is possible because event logs typically record information about the users executing the activities recorded in the log. To do this we combine concepts from workflow management and social network analysis. This paper introduces the approach and presents a tool to mine social networks from event logs.

1. 서론

소시오메트리(sociometry)는 그래프나 매트릭스폼으로 인간 관계 및 집단 구조를 표현하는 방법이다 [6,7,8]. 이는 Jacob Levy Moreno가 처음 사용한 용어로, 그는 이를 거주자들의 거주 지역 할당에 활용하였다. 그 결과 거주자들 사이의 이동량을 획기적으로 줄일 수 있었다. 소시오메트리와 관련된 연구는 현재 사회 과학 분야에서 활발하게 진행되고 있고, 여러 분야에 다양하게 활용되고 있다.

기업의 조직 구조 관점에서도 소시오메트리의 응용이 가능하다. 업무의 수행에 참여하는 업무 수행자들의 연관 관계를 분석함으로써 이들의 지위나 역할, 집단의 응집성 등에 대한 정보를 얻을 수 있다. 본 연구에서는 이를 위해서, 프로세스 수행 정보로부터 사회연결망(social network)을 도출하는 방법에 대한 연구를 수행한다.

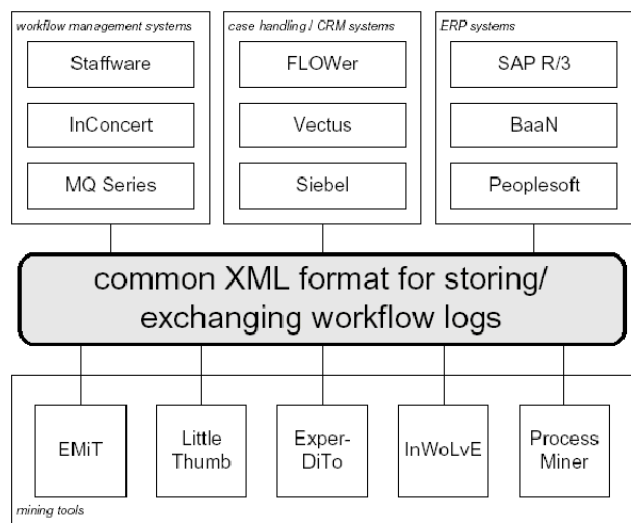
기본적으로 회사에는 업무 수행을 위한 많은 시스템과 이러한 시스템으로부터 생성된 많은 정보가 존재한다. 이러한 정보를 통합하여 프로세스 로그의 형태로 저장하고, 저장된 로그를 분석하여 사회연결망을 도출한다. 이렇게 도출된 사회연결망은 분석을 통해 기업의 조직 구조를 이해하거나, 프로세스를 개선하는데 활

용될 수 있다.

본 논문의 구성은 다음과 같다. 2 장에서는 본 연구의 기반 연구 분야인 프로세스 마이닝에 대해서 간단히 소개한다. 3 장에서는 사회연결망에 대한 설명과 도출 방법에 대해서 설명한다. 4 장에서는 3 장에서 제안한 방법을 지원하는 시스템의 구현에 대해서 설명한다. 5 장에서는 본 연구와 관련된 연구를 살펴 보고, 6 장에서 본 논문의 결론을 맺는다.

2. 프로세스 마이닝 (Process Mining)

프로세스 마이닝은 회사의 업무 처리 기록에서 유용한 정보를 추출해 내는 것을 목적으로 한다[2]. 업무 처리 기록에는 각각의 이벤트가 기록이 되는데, 각 이벤트에는 관련 작업(activity), 관련 인스턴스(instance), 수행한 작업자(performer)가 기록 되고, 각 이벤트는 발생한 순서대로 기록 된다고 가정한다. 실제로 기업의 정보 시스템인 ERP, CRM, Workflow 에서 형태는 서로 다르지만, 앞에서 언급한 내용을 포함한 정보를 제공한다 [1].



[그림 1] 프로세스마이닝 연구 프레임워크 [2]

[그림 1]은 프로세스 마이닝 연구의 기본 프레임워크를 나타낸다. 여러 정보시스템에서 기록된 업무 수행 로그를 XML 기반의 표준 데이터 포맷으로 바꾸어 저장을 한다. 저장된 프로세스 로그는 프로세스 마이닝 툴인 EMiT,

Little Thumb, ExperDiTo, InWoLvE, Process Miner 등을 통해서 분석할 수 있다 [2]. 현재 프로세스 마이닝에 대한 연구는 주로 구조화된 프로세스 모델을 찾는 연구에 초점을 맞추고 있다 [2,3]. 그 밖에도 기본적인 수행 결과 분석이나 조직 관점에서의 연구가 수행되고 있다. 본 논문에서 수행하는 연구는 조직 관점의 분석, 특히 사회연결망의 응용을 위한 기반 연구에 해당한다.

3. 조직 관계 마이닝

이 장에서는 조직 관계 마이닝을 위해 사용할 사회연결망 분석(social network analysis)에 대해서 알아보고, 프로세스 로그에서 어떻게 사회연결망을 도출하는가에 대해서 설명한다.

3.1. 사회연결망 분석

사회연결망 분석은 매우 광범위하게 응용되고 있다. 하나의 예로, 미국 911 테러 사건의 테러리스트들 사이의 관계 분석을 위해 사회연결망 툴인 InFlow(<http://www.orgnet.com/>)를 사용하였다. 사회연결망 연구의 기본 아이디어는 사람이나 조직을 노드로 표현하고 이들 사이의 관계가 있는 경우 해당하는 노드를 연결해서 네트워크를 구성하고 분석하는 것이다. 이때 노드 사이의 관계 정도에 따라서 연결선의 가중치 값을 설정할 수도 있다. 이러한 사회연결망은 기업의 조직과 직원 사이의 관계 분석에 응용이 될 수 있다.

네트워크를 만드는 방법으로, 예전에는 설문 조사를 많이 이용하였다. 설문 조사를 통해 노드 사이에 관계를 조사하고, 네트워크를 구성하고 분석을 하였다. 하지만 요즘에는 인터넷의 발달로 야기된 정보를 바탕으로 한 네트워크 구성이 많아지고 있고 있다. 예를 들어 BuddyGraph (<http://buddygraph.com>), MetaSight (<http://www.metasight.co.uk>) 등의 소프트웨어는 이메일 교환 기록을 바탕으로 사회연결망을 만들고 분석을 시작한다. 또한 웹로그를 이용한 네트워크의 작성도 시도되고 있다. 본 논문에서는 프로세스 로그를 사용한다.

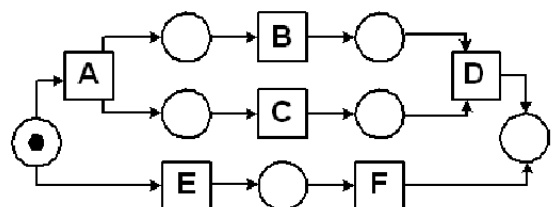
한편 사회연결망의 분석에 있어서는 집중도(centrality)와 노드의 위치가 중요하다. 집중도에는 크게 로컬 집중도와 글로벌 집중도가 있다. 로컬 집중도는 하나의 노드와 직접 연결되어 있는 노드 사이에서의 집중도를 나타내고, 글로벌 집중도는 전체 네트워크를 바탕으로 하나의 노드의 집중도를 나타낸다 [4]. 보통 네트워크에서 집중도가 높게 나타나는 노드가 중요한 역할을 하게 된다. 또한 노드의 위치가 중요하다. 서로 비슷한 위치에 속한 노드는 비슷한

역할을 하는 경우가 많다. 두개의 클러스터를 연결하고 있는 노드도 중요하다. 이밖에도 여러가지 분석 방법이 있는데, 이를 이용하여 네트워크를 구성하는 노드들의 관계를 분석하게 된다.

이러한 사회연결망은 기업의 조직 사이의 관계나 직원들 사이의 관계 분석에 활용될 수 있다. 이를 위해 기업의 정보 시스템에 저장되어 있는 작업 기록을 활용한다. 예를 들어 워크플로우 시스템인 Staffware에서는 작업의 시작과 종료에 대한 기록을 하고 있고, SAP 과 같은 ERP 시스템은 모든 트랜잭션을 기록하며, B2B 시스템과 CRM 시스템도 작업자의 작업에 대한 기록을 남긴다 [1,2]. 이러한 기록으로부터 사회연결망을 생성하여, 분석하는 것은 큰 의미를 지닐 수 있다.

3.2. 사회연결망 생성방법

이 장에서는 프로세스 로그로부터 사회연결망을 구성하는 방법에 대해서 설명한다. 이해를 돕기 위해 [그림 2]의 프로세스 모델에서 생성된 [표 1]의 프로세스 로그를 예제로 사용한다. [표 1]을 인스턴스에 따라서 정리해 보면 [표 2]와 같다.



[그림 2] 예제 프로세스 모델

[표 1] 프로세스 로그의 예

Process instance	Activity identifier	Performer
1	A	John
2	A	John
3	A	Lucia
3	B	Mary
1	B	Mary
1	C	John
2	C	John
4	A	Lucia
2	B	Mary
2	D	Mary
5	E	Alex
4	C	Mary
1	D	Mary
3	C	Peter
3	D	Lucia
4	B	Peter
5	F	Alex
4	D	Lucia

[표 2] 인스턴스별 프로세스 수행 기록

Instance	Logs
1	(A,John),(B,Mary),(C,John),(D,Mary)
2	(A,John),(C,John),(B,Mary),(D,Mary)
3	(A,Lucia),(B,Mary),(C,Peter),(D,Lucia)
4	(A,Lucia),(C,Mary),(B,Peter) (D,Lucia)
5	(E,Alex),(F,Alex)

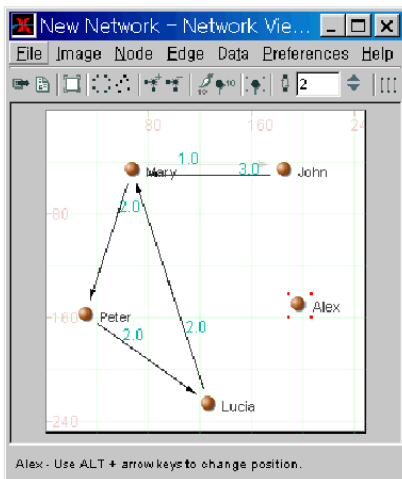
(1) 작업의 전달 (Handover of work)

선행 작업자가 선행 작업이 종료되어 후행 작업자에게 전달을 할 경우, 선행 작업자와 후행 작업자 사이의 관계로 네트워크를 구성한다. 예를 들어 [표 2]를 살펴 보면, John 과 Mary 사이에 세번의 작업의 전달이 있다. 반면에 John 과 Lucia 는 한번의 작업의 전달도 없다. 이런 작업의 전달의 빈도에 따라서 [표 3]과 같은 매트릭스를 구성할 수 있다.

[표 3] 작업의 전달로 구한 매트릭스

	John	Alex	Lucia	Peter	Mary
John	1	0	0	0	3
Alex	0	1	0	0	0
Lucia	0	0	0	0	2
Peter	0	0	2	0	0
Mary	1	0	0	2	1

[표 3]을 바탕으로 사회연결망을 구하면, [그림 3]과 같은 결과가 나온다.



[그림 3] 작업의 전달로 생성된 사회연결망

(2) 계약 관계 (Subcontracting)

세개의 작업이 있고, 첫번째 작업을 작업자 a 가 수행하고, 두번째 작업을 작업자 b 가 수행하고, 세번째 작업을 다시 작업자 a 가 수행한 경우에 a 와 b 사이에는 계약 관계가 있다고 할 수 있다. 예를 들어, 인스턴스 1 의 경우에서처럼 작업 A 와 작업 C 를 John 이 수행하고 그 사이에 있는 작업 B 를 Mary 가 수행한 경우에 John 과 Mary 사이에는 계약 관계가 있다고 할 수 있다. 계약 관계의 경우도 작업의 전달 경우처럼, 빈도 매트릭스를 구하고, 사회연결망

을 도출할 수 있다.

(3) 공동 작업 (Working together)

이 경우는 작업의 선후행 관계는 무시하고, 작업자들이 같은 인스턴스에 참가하는 빈도를 계산한다. 예를 들어 [표 2]를 보면 Jonh 은 두개의 인스턴스에 참가를 하는데, 항상 Mary 와 함께 일을 한다. 이 경우에 John 과 Mary 에는 2 의 상관 관계가 있다. 앞의 경우와 마찬가지로 빈도 매트릭스와 사회연결망을 구한다.

(4) 비슷한 업무 참여 (Doing similar task)

비슷한 작업에 참가한 사람들은 비슷한 역할을 가지고 있을 수 있다. 예를 들어 앞의 예에서 작업 A 는 John 과 Lucia 만 수행을 한다. 따라서 이 둘 사이에는 관계가 있다고 할 수 있다. 이 기준에 따라서 네트워크를 구성하기 위해서는 우선 [표 4]와 같이 작업자가 어떤 작업을 수행했는지에 대한 표를 구한다. 그리고, 이 표에서 사람들 사이의 거리를 구한다. 거리를 구하는 방법으로 유클리디안 디스턴스, 해밍 디스턴스, 상관계수 등을 사용할 수 있다. 거리가 구해지면 이를 바탕으로 사회연결망을 도출한다.

[표 4] 수행자와 작업자 관계 매트릭스

	A	B	C	D	E	F
John	2	0	2	0	0	0
Alex	0	0	0	0	1	1
Lucia	2	0	0	2	0	0
Peter	0	1	1	0	0	0
Mary	0	3	1	2	0	0

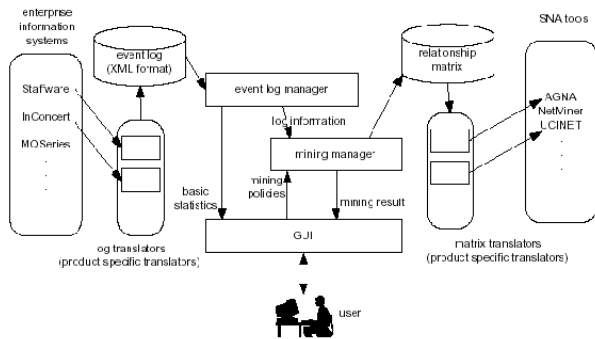
(5) 작업의 위탁 (Reassignment)

업무 수행 기록에는 이벤트 타입에 대한 기록이 존재한다. 예를 들어 작업자 a 가 작업을 완료 했다는 기록에는 '완료'라는 이벤트가 기록 될 수 있고, 작업자 a 가 작업자 b 에게 작업을 위탁하게 되면, '위탁'(reassignment)이라는 기록이 남을 수 있다. 이러한 여러 이벤트 타입 중에 '위탁' 이벤트를 사용해 네트워크를 구성한다. 작업자 a 와 작업자 b 가 위탁 관계에 있을 경우에 이들 사이에 연관 관계가 있다고 할 수 있다. 이 경우도 앞의 경우처럼 빈도 매트릭스를 구하고, 사회연결망을 도출한다.

4. 프로토타입 시스템

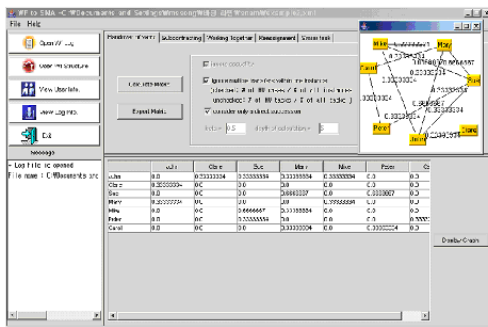
본 장에서는 앞에서 제시한 방법론에 따라서 구현된 프로토타입 시스템인 MiSoN(Mining Social Network)에 대해서 설명한다. [그림 4]는 시스템의 아키텍처를 나타낸다. 시스템은 트랜잭션을 기록하는 Staffware, InConcert, MQSerie

등의 데이터를 읽어서 표준 프로세스 로그로 변환하는 기능, 변환된 데이터를 읽어 3장에서 제시한 방법에 따라 사회연결망을 생성하는 기능, 생성된 사회연결망을 디스플레이 하는 기능, 결과를 사회연결망 분석툴인 AGNA, NetMiner, UCINET 의 파일 형태로 변환하는 기능을 가지고 있다.



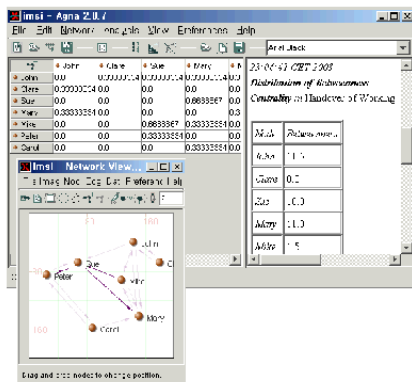
[그림 4] MiSoN 시스템 아키텍처

[그림 5]는 MiSoN의 주요 화면이다. 사용자는 앞에서 정의한 다양한 분석 방법을 정의할 수 있고, 이에 따라서 다양한 형태의 사회연결망을 만들 수 있다.



[그림 5] MiSoN의 화면

[그림 6]은 이 결과를 분석툴인 AGNA를 사용하여 분석하는 화면이다. 이런 분석툴의 사용을 통해 다양한 종류의 분석이 가능하다.



[그림 6] AGNA를 이용한 사회연결망 분석

5. 관련 연구

요즘 본 논문의 기본적인 연구 배경이 된 프로세스 마이닝에 대한 연구가 활발히 진행되고 있다. 이에 관련된 전반적인 소개는 [2,3]을 참조할 수 있다. 본 논문에서 제시한 방법에 대한 보다 자세한 정의는 [4]를 참조할 수 있다. 사회연결망에 대한 연구도 활발하게 진행 중이다 [5,6,7,8]. 워크플로우 프로세스를 사회연결망을 통해 분석하고자 하는 노력이 있었는데, 이들은 사회연결망을 구성하기 위해서 주로 설문 조사를 이용하고 있다.

6. 결론

본 논문에서는 프로세스 로그에서 사회연결망을 도출해 내는 방법을 제시하고, 그 방법을 지원하는 시스템의 구현 결과를 기술 하였다. 본 논문에서 제안한 방법을 이용해서 도출한 사회연결망은 비즈니스 프로세스 수행에 참가하는 작업자 사이의 관계 분석에 활용될 수 있다. 향후 연구로, 본 논문의 연구 결과를 실제 기업 사례에 적용해 보고 있으며, 또한 본 논문에서 제안한 시스템을 포함한 통합 프로세스 마이닝 툴의 구현을 진행 중이다.

참고문헌

1. W.M.P. van der Aalst and K.M. van Hee. "Workflow Management: Models, Methods, and Systems." MIT press, Cambridge, MA, 2002.
2. W.M.P. van der Aalst, B.F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and A.J.M.M. Weijters. "Workflow Mining: A Survey of Issues and Approaches." Data and Knowledge Engineering, 47(2):237-267, 2003.
3. W.M.P. van der Aalst and A.J.M.M. Weijters, editors. "Process Mining", Special Issue of Computers in Industry, Volume 53, Number 3. Elsevier Science Publishers, Amsterdam, 2004.
4. W.M.P. van der Aalst and M. Song "Mining Social Networks: Uncovering interaction patterns in business processes", International Conference on Business Process Management (BPM), 2004 (to appear)
5. L.C. Freeman. "Centrality in Social Networks: Conceptual Clarification." Social Networks, 1:215-239, 1979.
6. J.L. Moreno. "Who Shall Survive?" Nervous and Mental Disease Publishing Company, Washington, DC, 1934.
7. J. Scott. Social Network Analysis. Sage, Newbury Park CA, 1992.
8. S. Wasserman and K. Faust. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge, 1994.