IBM SPSS Modeler Social Network Analysis 17.1
사용자 안내서

IBM

참고

이 정보와 이 정보가 지원하는 제품을 사용하기 전에, 반드시 23 페이지의 『주의사항』의 정보를 읽으십시오.

제품 정보

이 개정판은 새 개정판에 별도로 명시하지 않는 한, IBM(r) SPSS(r) Modeler의 버전 17, 릴리스 1, 수정 0 및 모든 후속 릴리스와 수정에 적용됩니다.

목차

서론 v
IBM Business Analytics 소개 v
기술 지원 v

제 1 장 IBM SPSS Modeler Social Network Analysis 1
IBM SPSS Modeler Social Network Analysis 정보 1
소셜 네트워크 분석 정보 2
네트워크 표시 2
네트워크 설명 3
IBM SPSS Modeler Social Network Analysis 노드 5
노드 탭 5
출력 미리보기 6
데이터 분석. 6
응용 프로그램 7
데이터 구조. 7
스크립팅 속성 8

제 2 장 그룹 분석 11
그룹 분석 개요 11
사회 유사성 결정 11
그룹으로 파티션 12
그룹 및 그룹 멤버 설명 13
그룹 분석 노드 13
그룹 분석을 위한 데이터 지정 14
그룹 분석을 위한 작성 옵션 설정 14
그룹 분석 통계 보기 15
그룹 분석의 출력 15

제 3 장 확산 분석 17
확산 분석 개요 17
확산 분석 예제 17
확산 분석 노드 18
확산 분석을 위한 데이터 지정 19
확산 분석을 위한 작성 옵션 설정 19
확산 분석 통계 보기 20
확산 분석의 출력 20

주의사항 23
상표. 25

색인. 29

서론

IBM® SPSS® Modeler Social Network Analysis는 사람 간의 관계에 대한 정보를 소셜 네트워크에서 개인의 역할을 설명하는 필드로 처리하여 예측 모형에 사회 정보를 포함하도록 합니다. 이 매뉴얼은 IBM SPSS Modeler 환경에서 IBM SPSS Modeler Social Network Analysis 노드 사용에 대해 설명하여 사용자의 스트림에 노드를 포함할 수 있게 해줍니다. 노드 출력이 개인의 측정을 나타내는 필드와 결합될 경우 개인 결과의 프로파일이 더 완전해집니다.

IBM Business Analytics 소개

IBM Business Analytics 소프트웨어는 의사 결정자가 비즈니스 성능을 개선하기 위해 신뢰하는 완벽하고 일관되며 정확한 정보를 제공합니다. 비즈니스 지능, 예측 분석, 금융 성과와 전략 관리 및 분석 응용 프로그램의 종합 포트폴리오는 현재 성과와 앞으로의 결과를 예측하는 능력에 분명하고 즉각적이면서 실행 가능한 통찰력을 제공합니다. 풍부한 업계 솔루션, 입증된 사례 및 전문 서비스가 결합되어 어떠한 크기의 조직이라도 생산성을 극대화하고 자신있는 자동 결정을 내릴 수 있으며 더 나은 결과를 가져올 수 있습니다.

이 포트폴리오의 일부인 IBM SPSS Predictive Analytics 소프트웨어를 통해 조직은 미래의 사건을 예측하고 더 나은 비즈니스 결과를 얻기 위한 통찰력에 대해 적극적인 조치를 할 수 있습니다. 전 세계의 기업, 정부 및 학계 고객들은 고객을 매료시키고 유지하며 성장하게 만드는 동시에 불공정 행위를 줄이고 위험을 낮추는 IBM SPSS 기술의 경쟁 이점을 활용합니다. 일상 업무에서 IBM SPSS 소프트웨어를 활용한다면 예측형 기업으로 거듭날 수 있습니다. 즉 비즈니스 목표 달성을 위해 의사 결정의 방향을 정하고 이를 자동화하며 측정 가능한 경쟁 우위를 달성할 수 있습니다. 자세한 내용을 보거나 담당자에게 문의하려면 http://www.ibm.com/spss 사이트를 방문하십시오.

기술 지원

기술 지원은 유지 관리 고객에게 제공됩니다. IBM Corp. 제품 사용 및 지원된 하드웨어 환경 중 하나에 대해 설치하는 데 도움이 필요한 경우 기술 지원부로 문의하십시오. 기술 지원에 문의하려면 http://www.ibm.com/support의 IBM Corp. 웹 사이트를 참고하십시오. 지원을 요청하려면 본인의 신상과 소속 조직(회사) 및 지원 동의서를 제시해야 합니다.

제 1 장 IBM SPSS Modeler Social Network Analysis

IBM SPSS Modeler Social Network Analysis 정보

모델화 행동의 많은 접근법은 개인에 초점을 맞춥니다. 이러한 접근법은 개인에 대한 다양한 정보를 사용하여 행동을 예측하기 위한 행동의 핵심 표시기를 사용하는 모델을 생성합니다. 개인이 행동의 발생과 연관된 핵심 표시기에 대한 값을 가지고 있을 경우, 그 개인은 행동을 예방하기 위해 설계된 특별한 주의의 대상이 될 수 있습니다.

고객이 한 회사와의 관계를 종료하는 이탈 모델에 대한 접근법을 생각해 봅시다. 고객을 유지하는 비용은 고객을 교체하는 비용보다 훨씬 낮으며, 이탈 징후의 위험이 있는 고객을 식별할 수 있는 능력이 생깁니다. 분석가는 인구 통계학적 정보 및 각 개인 고객의 최근 호출 패턴을 비롯하여 고객을 설명하기 위해 여러 가지 핵심 성과 표시기를 자주 사용합니다. 이러한 필드에 근거한 예측 모형은 이탈 위험이 높은 사람을 식별하기 위해 과거에 이탈한 고객의 호출 패턴과 일치하는 고객 호출 패턴의 변화를 사용합니다. 위험하다고 식별된 고객은 고객을 유지하려는 노력의 일환으로 추가적인 고객 서비스 또는 서비스 옵션을 받습니다.

이런 방법은 고객의 행동에 큰 영향을 줄 수 있는 사회 정보를 간과합니다. 회사에 대한 정보와 다른 사람이 하는 행동에 대한 정보는 사람에게 영향을 주는 관계 전반에 흐릅니다. 그 결과 다른 사람과의 관계는 그러한 사람에게 한 사람의 결정과 행동에 영향을 줄 수 있게 합니다. 개인 측정만 포함하는 분석은 예측력을 가진 중요한 요인이 생략되어 있습니다.

IBM SPSS Modeler Social Network Analysis는 관계 정보를 모델에 포함될 수 있는 추가 필드로 처리하여 이 문제를 다룹니다. 이렇게 파생된 핵심 성과 표시기는 개인의 사회 특성을 측정합니다. 사회적 속성과 개인 기반 측정을 결합하여 개인에 대한 더 나은 개요를 제공하여 결과적으로 모델의 예측 정확성이 향상될 수 있습니다.

IBM SPSS Modeler Social Network Analysis는 두 가지 기본 구성 요소로 구성됩니다.

- IBM SPSS Modeler Social Network Analysis 노드는 스트림에 사회 분석적 기법을 포함시키는 IBM SPSS Modeler 환경에 추가됩니다.
- IBM SPSS Modeler Server Social Network Analysis는 IBM SPSS Modeler Server에 노드 지정 처리를 추가합니다. IBM SPSS Modeler Server Social Network Analysis는 수많은 개인 및 관계가 들어 있을 수 있는 대량의 네트워크 데이터를 심층 분석을 위해 상대적으로 작은 필드 수로 효과적으로 처리합니다.

예를 들어 IBM SPSS Modeler Social Network Analysis는 네트워크에서 특정인의 이탈로 가장 영향을 받는 개인을 식별합니다. 뿐만 아니라 네트워크에서 이탈 위험성이 커지는 개인의 그룹을 발견할 수 있습니다. 이러한 효과를 위해 모델에서 핵심 성과 표시기를 통합함으로써 전체 성능을 향상시킬 수 있습니다.

© Copyright IBM Corp. 2012, 2015

소셜 네트워크 분석 정보

소셜 네트워크는 개인의 집합과 이들 간의 관계로 구성되어 있습니다. 소셜 네트워크 분석은 이러한 관계를 조사하여 사회 구조의 일부로서 개인과 그룹을 설명합니다. 개인은 서로 상호작용하며 이러한 상호작용 패턴은 관련된 개인에 대한 통찰력을 제공합니다. 관계는 네트워크에 걸쳐 정보를 흐르게 하며 한 개인이 다른 사람에게 영향을 미칠 수 있게 합니다. 관계 정보 세트의 중요성은 소셜 네트워크 분석을 다른 접근법과 다른 것으로 만듭니다. 각 개인을 따로 초점을 맞추는 대신 연구 단위는 두 개인과 그들의 관계로 구성된 한 쌍입니다.

네트워크 내의 관계는 방향 또는 비방향으로 분류될 수 있습니다. *방향 관계*의 경우, 한 개인은 관계의 시발자 또는 소스로 식별되며 다른 개인은 수신자 또는 목적지로 식별됩니다. 예를 들어 전화 호출은 한 사람이 다른 사람에게 호출하는 방향 관계입니다. 반대로 *비방향 관계*에 대한 소스와 목적지의 역할은 정의할 수 없습니다. 이 사례에서 양측은 관계에 동등하게 참여합니다. 서로 말하는 것은 비방향 관계의 한 예입니다.

관계를 구별하는 또다른 특성은 관계가 이분적인지 계량적인지에 대한 것입니다. *이분적 관계*에서 사용할 수 있는 유일한 정보는 두 개인 사이에 관계가 존재하는지에 대한 여부입니다. 네트워크 내의 모든 쌍에서 관계는 존재하거나 존재하지 않습니다. 반대로 *계량 관계*는 관계의 강도를 나타내는 가중치가 들어 있습니다. 가중치를 사용하여 관계를 서로 비교할 수 있습니다.

"관계 유형" 테이블은 방향 및 척도를 기준으로 관계를 상호 분류한 예를 보여줍니다. 방향 관계에서 조는 관계의 소스이며 *매리*는 목적지입니다. 비방향 관계에서는 누가 관계를 시작했는지에 대한 표시가 없습니다. 계량 관계는 대화의 길이를 관계 가중치로 사용하는 반면, 이분적 관계는 발생했거나 발생하지 않았습니다.

표 1. 관계 유형.

방향	척도	예제
비방향	이분적	조와 매리가 서로 대화했습니다.
비방향	계량	조와 매리가 20분 동안 서로 대화했습니다.
방향	이분적	조가 매리에게 전화했습니다.
방향	계량	조가 매리에게 전화해 20분 동안 대화했습니다.

소셜 네트워크 분석 필드에 대한 자세한 내용은 이 분야를 포괄적으로 다룬 도서를 참조하십시오[1].

네트워크 표시

소셜 네트워크는 일반적으로 *소시오그램* [2]을 사용하여 나타냅니다. 이 유형의 시각 표시에서 개인은 공간 내의 점 또는 노드에 해당합니다. 점을 연결하는 선(가장자리)은 개인 간의 관계를 나타냅니다. 관계가 방향적일 경우 가장자리는 화살표를 포함하여 방향을 나타냅니다. 관계가 가중치를 가질 경우 가장자리의 레이블이 값을 나타냅니다. 다음 그래프는 7명의 개인에 대한 네트워크를 표시합니다.

1. Wasserman, S., and K. Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

2. Moreno, J. L. 1934. *Who Shall Survive?: Foundations of Sociometry, Group Psychotherapy, and Sociodrama*. Washington, D.C.: Nervous and Mental Disease Publishing Co..

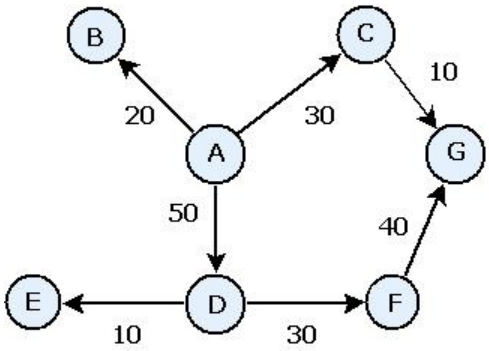


그림 1. 예제 소셜 네트워크

네트워크가 호출 길이를 나타내는 관계 가중치를 사용하여 개인이 건 전화 호출을 나타낸다고 가정해 봅시다. 이 사례에서 A는 세 사람에게 전화하여 대부분의 시간을 D와 이야기하는 데 썼습니다.

이 네트워크는 실제 보게 되는 것보다 훨씬 작습니다. 그러나 간단한 소시오그램에 나타난 개념은 모든 *크기*와 복잡성을 가진 네트워크를 일반화합니다.

네트워크 설명

네트워크, 그룹 및 개인에 대한 정보는 상호 비교가 가능하고 예측 모형에 포함될 수 있는 설명적 특징에서 추출해야 합니다. 네트워크는 분석 가능한 핵심 성과 표시기의 유한 세트로 추출되어야 합니다. 예를 들어, 네트워크 또는 네트워크 내의 노드 그룹을 서로 비교하려고 할 수 있습니다. 또는 네트워크 내의 개인을 서로 비교하거나 가장 중요한 개인을 식별하고 싶을 수 있습니다.

소셜 네트워크를 설명하기 위해 흔히 사용되는 두 측도는 **밀도**와 **차수**입니다. 두 통계는 연결성을 반영하지만 전자는 전체 네트워크 또는 네트워크 하위 그룹에 초점을 맞추는 반면 후자는 네트워크 내의 개인을 특징화합니다.

네트워크 밀도

네트워크에 있는 모든 노드 세트에는 유한한 수의 가능한 관계가 있습니다. 각 노드는 모든 다른 노드와 관계의 소스 또는 대상이 될 수 있습니다. 세 노드인 A, B, C로 이루어진 네트워크를 생각해 봅시다. 다음 테이블은 노드 사이의 모든 가능한 방향 관계를 나열합니다.

표 2. 세 노드에 대해 가능한 방향 관계.

소스	대상
A	B
A	C
B	A
B	C
C	A
C	B

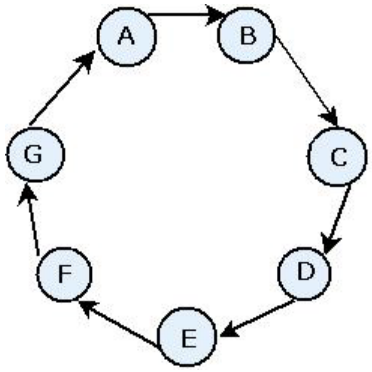
각 노드는 다른 두 노드와 관계의 소스입니다. 그러나 실제에서는 모든 가능한 관계가 사실 존재하지 않을 수 있습니다. 일부 노드는 다른 노드와의 방향 관계가 전혀 없을 수 있습니다. 또한 일부 방향 관계는 상호적이지 않을 수 있습니다.

밀도 통계는 실제 존재하는 네트워크 내의 가능한 관계의 비율을 나타냅니다. 값의 범위는 0부터 1까지이며, 하한은 관계가 없는 네트워크에 해당하고 상한은 모든 가능한 관계가 있는 네트워크를 나타냅니다. 값이 1에 가까울수록 네트워크가 더 조밀하며 네트워크 내의 노드 응집력이 더 높습니다.

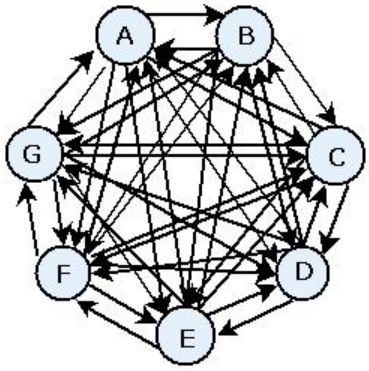
조밀한 네트워크 내의 정보는 희박한 네트워크 내의 정보보다 쉽게 흐를 수 있습니다. "희박하거나 조밀한 네트워크" 테이블은 7명의 노드로 구성된 두 가지 네트워크를 보여줍니다. 희박한 네트워크는 노드 사이에서 가능한 42개 관계 중 7개만 포함하며, 밀도는 0.17입니다. 반대로 조밀한 네트워크는 모든 가능한 관계가 포함되어 1.0의 밀도를 가집니다.

표 3. 희박하거나 조밀한 네트워크

희박한 네트워크



조밀한 네트워크



희박한 네트워크에서 노드 A에서 노드 G로 흐르는 정보는 반드시 5명의 다른 노드를 지나야 합니다. 반대로 조밀한 네트워크에서 정보는 노드 A에서 노드 G로 바로 갈 수 있습니다.

노드의 차수

네트워크에서 중요한 개인은 가장 중요한 관계에 관여된 사람일 경우가 많습니다. 이러한 개인은 다양한 소스로부터 정보를 얻으며 그 정보를 여러 명의 다른 개인에게 퍼뜨립니다. 반대로 소수의 관계에 참여하는 개인은 네트워크 내의 여러 사람에게 직접적인 영향을 줄 수 없습니다.

노드의 *정도*는 해당 노드와 연관된 관계의 전체 수로 정의되며 네트워크 참가자 사이의 비교가 가능합니다. 높은 차수값을 가진 개인은 낮은 값을 가진 사람보다 더 활동적입니다. 차수는 관계의 방향을 무시하여 노드 활동성의 전체 측도를 제공합니다.

방향 관계의 경우 관계의 수를 기록할 때 노드가 소스인지 대상인지에 초점을 맞출 수 있습니다. 노드의 *내도*는 특정 노드가 대상인 관계의 수입니다. 반대로 *외도*는 노드가 소스인 관계의 수입니다. 다음 테이블은 "소셜 네트워크의 예" 그림에서 각 노드의 차수, 내도 및 외도 값을 보여줍니다.

표 4. 차수, 내도 및 외도 값의 예.

노드	차수	내도	외도
A	3	0	3
B	1	1	0
C	2	1	1
D	3	1	2
E	1	1	0
F	2	1	1
G	2	2	0

내도는 명성의 측도로 취급되는 경우가 많습니다. 값이 높은 내도는 해당 노드에서 끝나는 관계가 더 많은 경우에 해당합니다. 즉, 이러한 개인은 아주 많은 수의 다른 개인의 연락을 받습니다. 다른 많은 노드는 노드와 관계를 시작합니다. 반대로 외도는 중심성의 측도로 취급됩니다. 값이 높을수록 해당 노드에서 시작하는 관계가 더 많은 경우에 해당합니다. 이러한 개인은 아주 많은 다른 개인에게 연락합니다.

예제 네트워크의 노드의 경우, 차수값은 노드 A와 D가 가장 활발한 반면 노드 B와 E는 활동성이 가장 떨어짐을 나타냅니다. 외도값은 노드 G가 가장 큰 명성을 가졌음을 보여줍니다. 외도값을 봤을 때 노드 A가 가장 중심적입니다.

IBM SPSS Modeler Social Network Analysis 노드

IBM SPSS Modeler와 제공되는 많은 표준 노드와 함께 IBM SPSS Modeler Social Network Analysis 노드로도 작업하여 스트림에 소셜 네트워크 분석의 결과를 포함할 수 있습니다. "IBM SPSS Modeler Social Network Analysis 노드" 테이블은 소스 팔레트에 저장되는 이러한 노드를 설명합니다.

표 5. IBM SPSS Modeler Social Network Analysis 노드.

노드	아이콘	설명
그룹 분석		그룹 분석 노드는 고정 필드 텍스트 파일에서 호출 세부사항 레코드 데이터를 가져오고 기록으로 정의된 네트워크 내 노드 그룹을 식별하며 네트워크 내의 그룹 및 개인에 대한 핵심 성과 표시기를 생성합니다. 자세한 정보는 11 페이지의 『그룹 분석 개요』의 내용을 참조하십시오.
확산 분석		확산 분석 노드는 고정 필드 텍스트 파일에서 호출 세부사항 레코드 데이터를 가져오고 네트워크에 걸쳐 기록으로 정의된 효과를 전달하며 개인 노드의 영향에 대한 결과를 요약하는 핵심 성과 표시기를 생성합니다. 자세한 정보는 17 페이지의 『확산 분석 개요』의 내용을 참조하십시오.

노드 탭

IBM SPSS Modeler Social Network Analysis 노드는 분석을 정의하고 미리 보는 다음의 탭을 제공합니다.

- **데이터 탭.** 소셜 네트워크 정보가 들어 있는 파일 식별에 사용됩니다.
- **작성 옵션 탭.** 분석을 위한 설정 정의에 사용됩니다.
- **분석 탭.** 최적의 결과를 생산하기 위해 데이터 탭의 입력 설정을 수정하기 위한 지침을 제공하는 예비 출력 요약을 보는 데 사용됩니다.

또한 노드는 IBM SPSS Modeler의 소스 노드에 걸쳐 다음의 일반 탭을 제공합니다.

- **필터 탭.** 노드가 생산한 출력 필드를 제거하거나 이름을 바꾸는 데 사용됩니다. 이 탭은 필터 노드와 같은 기능을 제공합니다.
- **유형 탭.** 노드가 생산한 출력 필드에 측정 수준을 설정하는 데 사용됩니다. 이 탭은 유형 노드와 같은 기능을 제공합니다.
- **주석 탭.** 노드의 이름을 바꾸고 사용자 맞춤 도구팁을 제공하며 긴 주석을 저장하는 데 사용됩니다.

일반 탭에 대한 자세한 내용은 IBM SPSS Modeler 문서를 참조하십시오.

출력 미리보기

일반적으로 노드의 출력을 예측 모형에 대한 입력으로 사용합니다. 출력을 보려면 스트림에 테이블 노드를 추가할 수 있습니다. 하지만 데이터양이 보통 극단적으로 커서 테이블을 생성하는 데 시간이 많이 걸립니다. 게다가 테이블의 행 수는 테이블의 유용성을 제한합니다.

대안으로서, 만들어질 출력의 표본을 표시할 미리보기 테이블을 생성할 수 있습니다. 미리보기는 제한된 행 수를 위해 생성된 필드를 보여줍니다. 행의 수는 스트림 속성에서 정의됩니다. 노드 출력을 미리보려면 다음 단계를 수행합니다.

1. 노드를 엽니다.
2. 데이터 탭에서 데이터 설정을 지정합니다.
3. 작성 옵션 탭에서 분석 매개변수를 정의합니다.
4. **미리보기**를 클릭합니다.

결과를 표시하는 미리보기 창이 열립니다. 더불어 출력 미리보기가 결과의 요약 개요와 함께 노드의 분석 탭을 채웁니다.

데이터 분석

최적의 결과를 생산하는 분석 설정을 결정하는 것은 보통 반복적인 프로세스입니다. 설정을 정의하고 분석을 실행하며 결과를 검토합니다. 결과가 예상보다 유용하지 못할 경우 설정을 수정하여 분석을 다시 실행합니다.

노드의 입력 데이터를 분석하려면 다음 단계를 수행합니다.

1. 노드를 엽니다.
2. 데이터 탭에서 데이터 설정을 지정합니다.
3. 작성 옵션 탭에서 분석 매개변수를 정의합니다.
4. **데이터 분석**을 클릭합니다.

표시되어야 할 요약 통계를 작성 옵션이 나타내면 분석 탭에 결과가 보입니다.

분석을 다시 실행해야 할 경우 **분석 지우기**를 클릭하여 현재 결과를 삭제한 다음 **데이터 분석**을 클릭합니다.

응용 프로그램

IBM SPSS Modeler Social Network Analysis가 특히 유익한 특정 응용 프로그램은 다음을 포함합니다.

- **이탈 예측.** 그룹 특성은 이탈율에 영향을 줄 수 있습니다. 이탈 위험이 큰 그룹 내의 개인에 초점을 맞춰 이를 예방할 수 있습니다. 또한 이미 이탈한 사람을 통해 얻은 정보 흐름으로 이탈 위험이 있는 개인을 식별할 수 있습니다.
- **그룹 리더의 영향.** 그룹 리더는 다른 그룹 멤버에게 매우 영향력이 있습니다. 그룹 리더의 이탈을 막을 수 있다면 그룹 멤버의 이탈율이 감소될 수 있습니다. 또는 이탈할 경쟁자로부터 그룹 리더를 구하면 회사와 관련된 그룹 멤버의 이탈율은 감소하면서 해당 경쟁사와 관련된 그룹 멤버의 이탈율은 증가할 수 있습니다.
- **마케팅.** 그룹 리더는 새 상품 또는 서비스 제공을 시작하는 데 사용할 수 있습니다. 리더의 영향은 다른 멤버가 제공 품목을 구입할 가능성을 높일 수 있습니다. 확산 분석을 사용하여 그룹 리더에게 가장 큰 영향을 받는 개인을 식별하여 이들을 마케팅 대상으로 삼을 수 있습니다.

사용자의 스트림에서 소셜 네트워크 분석 결과를 어떻게 포함할 수 있는지에 대한 예를 제공하기 위해 두 개의 데모 스트림이 IBM SPSS Modeler Social Network Analysis와 함께 제공됩니다. 데이터 파일 및 샘플 스트림은 제품 설치 디렉토리 아래에 있는 *Demos* 폴더에 설치됩니다.

- *DA demo streams.str*은 데이터 분석의 예를 제공하여 이탈할 가능성이 가장 높은 최상의 300명을 식별합니다.
- *GA demo streams.str*은 KPI(key performance indicator) 사용 예를 제공하여 그룹 및 개인의 이탈을 예측하고 KPI 사용 예를 제공하여 마케팅 캠페인을 위해 특정 개인들을 대상으로 합니다.

데이터 구조

이익 네트워크의 개인에 대한 정보는 기업 전반에서 다양한 파일, 데이터베이스 및 시스템에 걸쳐 퍼질 수 있습니다. IBM SPSS Modeler Social Network Analysis를 사용하여 네트워크를 분석하려면 데이터 소스에서 관련 레코드 및 필드를 추출하고 노드에 입력하도록 형식을 구성해야 합니다.

분석적 노드는 단일 고정 너비 텍스트 파일에 저장된 호출 세부사항 레코드가 필요합니다. 파일의 각 행은 관계에 해당하며 다음 열에는 데이터가 체계화되어 있습니다.

- 관계를 시작한 개인에 대한 식별자.
- 관계의 대상인 개인에 대한 식별자.
- 관계에 대한 선택적 가중치.

모든 데이터는 정수로 제한되는 개인에 대한 식별자를 가진 숫자여야 합니다. 파일 내 데이터의 첫 행인 필드 이름 포함은 선택입니다. "호출 데이터 레코드의 예" 테이블은 데이터 구조를 보여줍니다.

헤더 레코드는 또한 숫자 또는 공백을 값으로 사용해야 합니다. 예를 들어, *EF BB BF* 값을 가진 헤더는 오류를 유발합니다.

표 6. 호출 데이터 레코드의 예.

소스	목적지	가중치
1000	5642	243
2190	8444	831
0299	9419	559

가중치는 네트워크의 다른 관계에 상대적인 관계의 중요성을 나타내기 위해 사용하려는 측도에 해당할 수 있습니다. 호출 데이터의 경우 일반 가중치는 호출 기간 또는 호출 빈도를 포함합니다. 이는 확산 분석의 경우 true이지만 집단 분석만 호출 빈도를 지원함을 참고하십시오.

분석을 호출 히스토리의 서브세트에 초점을 맞추려면 입력 파일을 만들 때 그 서브세트를 사용해야 합니다. 예를 들면 입력 텍스트 파일에 해당 데이터만 포함하여 개인의 지난 몇 달 간 또는 가장 최근 호출에 대한 분석을 제한할 수 있습니다.

스크립팅 속성

IBM SPSS Modeler Social Network Analysis에서 사용하는 스크립팅은 사용자 인터페이스의 프로세스 자동화를 위한 강력한 도구입니다. 스크립트는 마우스나 키보드로 수행하는 것과 같은 유형의 동작을 수행할 수 있으며, 스크립트를 사용하여 반복성이 높거나 수동으로 하려면 시간이 많이 걸리는 작업을 자동화할 수 있습니다. 스크립팅 사용에 대한 설명은 IBM SPSS Modeler에서 제공하는 *ScriptingAutomation.pdf* 안내서를 참조하십시오.

확산 노드 속성

다음 테이블에 확산 노드에 대한 스크립팅 속성이 나와 있습니다.

특성 이름	데이터 유형	특성 설명
input_data_file_name	*string*	
calling_field	*field*	
called_field	*field*	
frequency_weight_field	*field*	
read_field_names	*boolean*	
diffusion_list_file_name	*string*	
spreading_factor	*double*	기본값 = 50 최소값 = 1 최대값 = 99
max_number_iterations	*integer*	기본값 = 100 최소값 = 1
accuracy_threshold	*double*	기본값 = 0.01 최소값 = 0.001
calculate_statistics	*boolean*	

그룹 분석 노드 속성

다음 테이블에 그룹 분석 노드에 대한 스크립팅 속성이 나와 있습니다.

특성 이름	데이터 유형	특성 설명
input_data_file_name	*string*	
calling_field	*field*	
called_field	*field*	
frequency_weight_field	*field*	
read_field_names	*boolean*	
coverage_threshold	*double*	기본값 = 10 최소값 = 1 최대값 = 99
min_group_size	*integer*	기본값 = 2 최소값 = 2
max_group_size	*integer*	기본값 = 100 최소값 = 2
calculate_statistics	*boolean*	

제 2 장 그룹 분석

그룹 분석 개요

그룹 분석은 네트워크에 있는 개인의 상호작용 패턴을 사용하여 비슷한 개인의 그룹을 식별합니다. 이러한 그룹의 특징은 개별 그룹 멤버의 행동에 영향을 줍니다. 예를 들면, 상호 멤버 관계가 많고 강력한 리더가 있는 소그룹은 실제 이탈한 멤버가 없다 하더라도 이탈의 위험이 증가합니다. [3] 그룹과 개인 측정을 모두 통합하는 예측 모형은 개인 측정만 포함하는 모델보다 더 잘 수행합니다.

그룹 분석은 다음의 일반적 단계로 구성됩니다.

1. 사회 근접성을 반영하는 관계 강도를 결정합니다. 자세한 정보는『사회 유사성 결정』의 내용을 참조하십시오.
2. 크기 제한을 따르는 동시에 관계 강도에 기초하여 네트워크를 그룹으로 파티션합니다. 자세한 정보는 12 페이지의『그룹으로 파티션』의 내용을 참조하십시오.
3. 그룹 리더 식별을 포함하여 그룹과 개인을 모두 프로파일합니다. 자세한 정보는 13 페이지의『그룹 및 그룹 멤버 설명』의 내용을 참조하십시오.

사회 유사성 결정

그룹 멤버는 그룹에 있지 않은 개인보다 서로 더 비슷해야 합니다. 네트워크 분석에서 두 노드의 유사성은 노드의 관계에 따라 다릅니다. 네트워크에 있는 모든 노드에는 노드와 방향 관계가 되는 대상인 노드 집합이 있습니다. 통신 데이터의 경우 이 집합은 특정 개인이 연락한 모든 사람에 해당합니다. 두 개인이 같은 사람의 집합과 연락할 경우 이러한 개인은 서로 비슷한 것으로 간주됩니다. 관계 대상의 집합이 두 개인에게서 더 많이 겹칠수록 두 사람은 더 비슷합니다.

"10명의 노드 네트워크 예제" 그림에 나타난 네트워크를 고려해 보십시오.

3. Richter, Y., E. Yom-Tov, and N. Slonim. 2010. Predicting customer churn in mobile networks through analysis of social groups. In: *Proceedings of the 2010 SIAM international conference on data mining*. Columbus, OH: SDM 2010.

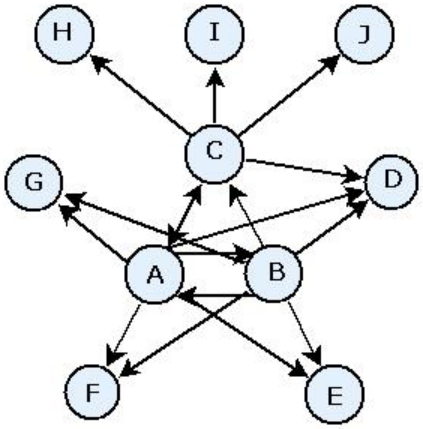


그림 2. 10명의 노드 네트워크 예제

"대상 노드" 테이블은 노드 A, B, C에서 시작한 관계의 대상 노드를 보여줍니다.

표 7. 대상 노드

소스 노드	대상 노드
A	B, C, D, E, F, G
B	A, C, D, E, F, G
C	A, D, H, I, J

노드 A와 B는 다섯 명의 공통 대상 노드가 있습니다. 반면 노드 A와 C는 한 명의 노드만 있습니다. 따라서 노드 A와 B는 노드 A와 C보다 사회적으로 더 비슷합니다.

호출 기간 또는 빈도와 같은 관계 가중치는 네트워크에서 노드의 유사성을 포착하지 않습니다. 유사성의 개념을 포함하기 위해 그룹 분석은 상호 정보[4]를 관계 가중치로 사용합니다. 이 통계는 두 노드가 같은 노드에 연결될 가능성을 반영합니다. 네트워크에 있는 관계가 사전 정의된 가중치(예: 통신 데이터를 위한 호출 빈도)을 가질 경우 상호 정보는 그에 따라 이러한 가중치를 통합합니다.

그룹으로 파티션

그룹은 다른 그룹 멤버와 높은 관계 가중치를 가진 개인으로 구성되며, 가중치는 관계에서 노드의 유사성을 측정합니다. 따라서 그룹 식별은 네트워크에 있는 더 약한 관계를 생략하는 것으로 시작합니다. *범위 임계값*는 유지할 가장 강한 관계의 비율을 정의하는 것으로 이 프로세스를 제어합니다. 예를 들어 범위 임계값이 0.4일 경우 관계의 가장 강한 40%가 그룹 식별에 사용되며 관계의 남은 60%는 생략됩니다.

남은 관계는 일부의 아주 작거나 매우 큰 그룹을 산출할 수 있으며, 그러한 그룹은 제한된 예측 효용을 가집니다. 이러한 그룹이 분석에 포함되는 것을 방지하기 위해 그룹의 최소 및 최대 *크기*를 지정할 수 있습니다. 최소 *크기*보다 작은 *크기*를 가진 그룹은 완전하게 생략됩니다. 그러나 최대 *크기*보다 큰 *크기*의 그룹은 허용 가능한 *크기* 범위 내에서 작은 그룹으로 나뉩니다. 약한 관계를 탈락시키고 *크기* 제한을 가한 남은 그룹을 *핵심 그룹*이라고 합니다.

---

4. Cover, T. M., and J. A. Thomas. 2006. *Elements of Information Theory, 2nd edition*. New York: John Wiley and Sons, Inc.

원래 네트워크에서 관계를 제거하면 일부 개인은 어떠한 핵심 그룹에도 속하지 못할 수 있습니다. 하지만 그러한 개인은 그룹에 포함되도록 보장하는 그룹의 멤버와 연결될 수 있습니다. 그룹의 핵심 멤버와 상대적으로 강한 관계가 많을 경우, 그룹 *크기* 제한을 위반하지 않는 한 개인은 그룹에 추가됩니다. 따라서 최종 그룹은 핵심 멤버의 집합에 핵심과의 연결성 때문에 추가된 멤버를 더해 구성됩니다.

그룹 및 그룹 멤버 설명

밀도, 내도 및 외도와 더불어 기타 통계는 그룹 역학을 설명합니다. 특히 권위 및 유포 스코어는 그룹 내에 있는 개인의 사회 상태에 대한 측도를 제공합니다. 그룹에 있는 각 개인의 역할은 그룹과 그룹 멤버의 행동을 예측하려 할 때 매우 중요합니다.

노드에 대한 *권위 스코어*는 그룹에 있는 다른 노드의 경향을 연결하기 위한 측도입니다. 여러 명의 개인이 특정인과 연락할 경우 권위 역할을 가진 그 특정인에게 정보나 의견을 구할 가능성이 큽니다. 그룹에 있는 노드의 권위 스코어는 그룹 네트워크를 통해 다시 시작하는 무작위 행보에 대한 정상 확률에 해당하며 범위는 0에서 1까지입니다. 권위 스코어가 1에 가까울수록 그룹 내의 노드가 가진 권위가 많습니다. 그룹에서 가장 높은 권위 스코어를 가진 노드를 그룹의 *권위 리더*라고 합니다. 그룹 내의 가장 큰 스코어를 가장 작은 스코어로 나누면 권위 리더에 대한 전체 강도의 측도가 됩니다.

반대로 노드의 *유포 스코어*는 그룹의 다른 노드에 연결하기 위한 노드의 경향을 측정합니다. 특정인이 그룹의 많은 다른 사람과 연락할 경우, 그 특정인은 전체 그룹의 견해에 상당한 영향을 미칠 수 있습니다. 그룹에 있는 노드의 유포 스코어는 거꾸로 그룹 네트워크를 통해 다시 시작하는 무작위 행보에 대한 정상 확률에 해당하며 범위는 0에서 1까지입니다. 유포 스코어가 1에 가까울수록 노드는 다른 그룹 멤버와 더 많이 연결합니다. 그룹에서 가장 높은 유포 스코어를 가진 노드를 그룹의 *유포 리더*라고 합니다. 그룹 내의 가장 큰 스코어를 가장 작은 스코어로 나누면 유포 리더에 대한 전체 강도의 측도가 됩니다.

그룹 분석 노드

소스 팔레트에서 사용 가능한 그룹 분석 노드는 네트워크에서 서로 사회적으로 비슷한 개인 집합을 식별하며 그룹에 있는 개인의 상대적 사회 상태를 결정합니다. 상대적으로 작은 그룹은 멤버 간의 연결이 많으며, 그룹 역학으로 인해 이탈의 위험이 큰 강력한 리더가 있습니다. 개인 이탈 스코어와 결합될 경우 개인 스코어만을 근거로 모델에 대한 이탈을 예측하는 능력을 향상시키는 그룹 이탈 스코어를 결정하기 위해 결과를 사용할 수 있습니다.

일반적으로 네트워크와 그룹 정의에 숨어 있는 사회적 상호작용의 특성을 구성하는 데이터양을 볼 때, 그룹 분석은 시간이 걸리는 프로세스입니다. 보통의 경우 데이터베이스나 예측 모형에 입력으로 사용되는 파일에 그룹 분석 결과를 저장할 수 있습니다. 모델은 훨씬 자주 새로 고칠 수 있는데 반해 그룹 결과는 월별과 같이 상대적으로 새로 고치는 빈도가 낮을 것입니다.

요구사항. 노드는 세 가지 필드를 사용하여 소셜 네트워크를 정의하는 고정 너비 텍스트 파일이 필요합니다. 한 필드는 각 방향 관계에 대한 소스를 식별하며, 한 필드는 각 방향 관계의 목적지를 정의하며, 나머지 필드는 각 관계에 대한 선택적 강도를 지정합니다. 네트워크의 모든 관계는 방향적이어야 합니다.

그룹 분석을 위한 데이터 지정

그룹 분석 소스 노드 창의 데이터 탭을 사용하여 네트워크 노드 관계에 들어 있는 입력 파일을 지정할 수 있습니다.

파일. 호출 세부사항 레코드가 들어 있는 파일 또는 폴더의 이름을 지정합니다. 이름을 입력하거나 생략표(...) 단추를 눌러 파일 시스템에서 이름을 선택할 수 있습니다. 이름을 선택하고 나면 경로가 나타나며 내용이 구분자와 함께 표시됩니다. 폴더를 지정할 경우 폴더에 포함된 모든 파일의 호출 세부사항 레코드는 분석을 위해 연결됩니다. 폴더의 모든 파일은 구조가 같아야 합니다.

파일에서 필드 이름을 읽습니다. 기본으로 선택된 이 옵션은 데이터 파일의 첫 행을 열에 대한 이름으로 취급합니다. 첫 행이 헤더가 아닐 경우 이 옵션을 선택 해제하여 자동으로 각 필드에 일반 이름(예: *Field1* 및 *Field2*)을 부여합니다.

네트워크 정의 설정

네트워크 정의 설정은 필드에 대한 역할을 정의합니다.

필드. 화살표 단추를 사용하여 이 목록에서 다양한 역할 필드로 항목을 수동으로 할당합니다. 아이콘은 각 역할 필드에 대한 유효한 측정 수준을 나타냅니다. **모두** 단추를 클릭하여 목록의 모든 필드를 선택하거나 개별 측정 수준 단추를 클릭하여 해당 측정 수준의 모든 필드를 선택합니다.

소스. 한 필드를 방향 관계의 원점으로 선택합니다.

대상. 한 필드를 방향 관계의 대상으로 선택합니다.

가중치. 선택적으로 네트워크에서 발생하는 관계의 횟수에 해당하는 한 필드를 선택합니다. 관계 강도를 결정할 때 분석은 그에 따라 레코드를 가중합니다.

그룹 분석을 위한 작성 옵션 설정

그룹 분석 소스 노드 창의 작성 옵션 탭을 사용하여 네트워크의 그룹을 식별하기 위한 옵션을 정의할 수 있습니다.

그룹 분석 설정

그룹 분석 설정은 형성된 그룹의 크기와 상대적 강도에 영향을 줍니다.

범위 임계값. 분석에 사용하기 위한 가장 강한 네트워크 관계의 비율을 정의합니다. 예를 들어 범위 임계값이 0.2일 경우 전체 가중치의 상위 20%에 드는 가중치를 가진 관계만 사용됨을 나타냅니다. 범위가 0에서 1인 이 매개변수를 통해 분석은 그룹 내에서 발생할 가장 강한 관계에 초점을 맞춥니다. 임계값이 높을수록 멤버 사이의 관계가 더 약한 그룹이 생깁니다. 자세한 정보는 12 페이지의 『그룹으로 파티션』의 내용을 참조하십시오.

최소 그룹 크기. 그룹 크기에 대한 하한을 지정합니다. 이 값보다 작은 그룹은 반환되지 않습니다.

최대 그룹 크기. 그룹 크기에 대한 상한값을 지정합니다. 이 값보다 큰 그룹은 작은 그룹으로 나뉩니다.

요약 통계 계산 및 표시. 선택했을 경우 노드는 분석에 대한 요약 통계를 계산하고 표시할 뿐만 아니라 핵심 성과 표시기 출력을 끌어냅니다. 이러한 통계의 계산은 극도로 큰 네트워크의 노드 성능에 부정적인 영향을 줄 수 있습니다.

그룹 분석 통계 보기

그룹 분석 소스 노드 창의 분석 탭은 네트워크에서 식별된 그룹의 요약 개요를 제공합니다. "그룹 분석 요약 통계" 테이블은 그룹에 대한 사용 가능한 요약 통계를 보여줍니다.

표 8. 그룹 분석 요약 통계.

통계	설명
그룹의 전체 노드	식별된 그룹에 포함된 노드 수
그룹의 전체 연결	식별된 그룹에 포함된 연결 수
전체 그룹 수	네트워크에서 식별된 그룹 수
평균 그룹 *크기*	그룹에 있는 노드의 평균 수
평균 그룹 밀도	그룹에 있는 노드 사이의 방향 연결의 평균 비율. 자세한 정보는 3 페이지의 『네트워크 밀도』의 내용을 참조하십시오.
핵심 멤버의 평균 비율	그룹의 핵심 노드인 그룹 내 노드의 평균 비율. 자세한 정보는 12 페이지의 『그룹으로 파티션』의 내용을 참조하십시오.
핵심 그룹의 평균 밀도	그룹에 있는 핵심 노드 사이의 방향 연결의 평균 비율.
평균 내도	들어오는 연결의 평균 수. 자세한 정보는 4 페이지의 『노드의 차수』의 내용을 참조하십시오.
평균 외도	나가는 연결의 평균 수. 자세한 정보는 4 페이지의 『노드의 차수』의 내용을 참조하십시오.

요약 통계 테이블에서 특정 통계를 선택하여 값의 분포, 표준 편차 및 통계에 대한 왜도를 봅니다.

이러한 통계를 검토하면 그룹 분석 설정을 정의하는 데 도움이 됩니다. 이상적으로 결과는 높은 밀도 값을 가진 상대적으로 작은 그룹을 보여주어야 합니다. 예를 들어 아주 큰 몇몇 그룹이 있을 경우 최대 그룹 *크기*를 줄여 작은 그룹으로 분할해 보십시오. 또는 그룹 밀도 값이 작을 것 같은 경우 범위 임계값을 줄여 더 강한 관계에 초점을 맞추십시오.

그룹 분석의 출력

그룹 분석 노드는 그룹 및 그룹의 개인을 설명하는 다양한 필드를 생성합니다. 기존 모델 및 이러한 핵심 성과 표시기를 가진 데이터를 증가시켜 모델이 생성하는 예측을 개선할 수 있습니다. 예를 들면 개인 이탈 성향 값을 업데이트하여 그룹 영향을 포함할 수 있습니다.

"그룹의 핵심성과지표(KPI)" 테이블은 분석에서 식별된 그룹의 핵심 성과 지표 필드를 보여줍니다.

표 9. 그룹의 핵심 성과 표시기.

필드	설명
GAG_GroupNumber	그룹에 대한 고유한 식별자
GAG_Size	그룹 내의 개인 수

표 9. 그룹의 핵심 성과 표시기 (계속).

필드	설명
GAG_Density	그룹 내 개인 간 방향 연결의 비율. 자세한 정보는 3 페이지의 『네트워크 밀도』의 내용을 참조하십시오.
GAG_KernelDensity	그룹 내 핵심 개인 간 방향 연결의 비율.
GAG_CoreNodesFraction	그룹의 핵심 개인인 그룹 내 개인의 비율. 자세한 정보는 12 페이지의 『그룹으로 파티션』의 내용을 참조하십시오.
GAG_MaxRankType1	그룹 멤버의 최대 권위 스코어. 자세한 정보는 13 페이지의 『그룹 및 그룹 멤버 설명』의 내용을 참조하십시오.
GAG_MinRankType1	그룹 멤버의 최소 권위 스코어
GAG_MaxMinRankRatioType1	가장 큰 권위 스코어 대 가장 작은 권위 스코어의 비율. 이 값은 그룹 리더의 권위 강도를 반영합니다.
GAG_MaxRankType2	그룹 멤버의 최대 유포 스코어. 자세한 정보는 13 페이지의 『그룹 및 그룹 멤버 설명』의 내용을 참조하십시오.
GAG_MinRankType2	그룹 멤버의 최소 유포 스코어
GAG_MaxMinRankRatioType2	가장 큰 유포 스코어 대 가장 작은 유포 스코어의 비율. 이 값은 그룹 리더의 유포 강도를 반영합니다.

"개인의 핵심성과지표(KPI)" 테이블은 네트워크에서 개인의 핵심 성과 지표 필드를 보여줍니다.

표 10. 개인의 핵심 성과 표시기.

필드	설명
GAI_NodeNumber	개인에 대한 고유한 식별자
GAI_CoreNode	개인이 그룹의 핵심 개인인지 아닌지에 대한 표시기. 자세한 정보는 12 페이지의 『그룹으로 파티션』의 내용을 참조하십시오.
GAI_RankType1	개인의 권위 스코어. 자세한 정보는 13 페이지의 『그룹 및 그룹 멤버 설명』의 내용을 참조하십시오.
GAI_RankOrderType1	권위 스코어에 근거한 그룹 내 순위
GAI_RankType2	개인의 유포 스코어. 자세한 정보는 13 페이지의 『그룹 및 그룹 멤버 설명』의 내용을 참조하십시오.
GAI_RankOrderType2	유포 스코어에 근거한 그룹 내 순위
GAI_InDegree	개인이 관계의 대상인 관계의 수. 자세한 정보는 4 페이지의 『노드의 차수』의 내용을 참조하십시오.
GAI_OutDegree	개인이 관계의 소스인 관계의 수. 자세한 정보는 4 페이지의 『노드의 차수』의 내용을 참조하십시오.
GAI_GroupLeaderType1	노드가 권위 리더이며, 리더의 리더십 스코어가 들어오는 연결에서 파생된 것인지에 대한 여부. 자세한 정보는 13 페이지의 『그룹 및 그룹 멤버 설명』의 내용을 참조하십시오.
GAI_GroupLeaderConfidenceType1	노드가 권위 리더라는 신뢰.
GAI_GroupLeaderType2	노드가 유포 리더이며, 리더의 리더십 스코어가 나가는 연결에서 파생된 것인지에 대한 여부. 자세한 정보는 13 페이지의 『그룹 및 그룹 멤버 설명』의 내용을 참조하십시오.
GAI_GroupLeaderConfidenceType2	노드가 유포 리더라는 신뢰.

제 3 장 확산 분석

확산 분석 개요

확산 분석은 소셜 네트워크의 다른 개인에 의해 영향을 가장 많이 받는 개인을 식별하여 그 효과를 *확산된 에너지*로 수량화합니다. 프로세스는 네트워크 노드에서 근접한 이웃 항목에 효과가 반복적으로 퍼지는 퍼짐 활성화 접근법을 사용하여 노드에서 노드[5]로 이동함에 따라 *크기*가 작아집니다. 에너지를 받으면 노드가 활성화되어 노드와 방향 관계가 되는 대상인 이웃 항목에 에너지의 일부를 전송합니다.

*퍼짐 요인*는 활성화된 노드가 전송한 에너지의 비율을 노드가 보유한 잔량과 함께 정의합니다. 이 에너지를 받는 모든 노드는 이웃 항목에 같은 비율을 전송하여 에너지 전송의 쇠퇴 프로세스를 만들어 냅니다. 퍼짐 요인이 크면 더 많은 에너지를 보낼 수 있으며 프로세스가 완전히 감소하기 전에 초기 활성화된 노드에서 에너지가 더 먼 거리의 노드에 닿을 수 있습니다. 작은 퍼짐 요인은 초기 노드에 상대적으로 가까이 남아 있는 전송된 에너지를 사용하여 빨리 감소하는 확산 프로세스를 만들어 냅니다.

활성화된 노드에 의해 확산 중인 에너지의 총량은 노드와 방향 관계가 되는 대상인 모든 노드에게 배포됩니다. 각 노드가 받는 양은 활성화된 노드와의 관계의 강도에 따라 다릅니다. 특정 노드에 보내진 비율은 활성화된 노드가 소스인 모든 관계에 대해 총 가중치로 나눈 관계 가중치와 같습니다. 따라서 관계 가중치가 높은 관계인 이웃 항목이 관계 가중치가 낮은 이웃 항목보다 많은 에너지를 받습니다.

다음 조건 중 하나가 발생하면 확산 프로세스는 중지됩니다.

- 활성화된 노드가 방향 관계의 소스이 아닐 경우
- 전송되고 있는 에너지의 양이 계속될 프로세스를 위해 전달되고 있는 에너지양의 한계인 *정확성 임계값*보다 낮을 경우
- 반복계산의 수가 특정 한계에 도달한 경우

확산이 완료되면 가장 확산된 에너지를 가진 노드가 프로세스를 시작한 효과에 가장 민감합니다. 예를 들어 이탈하는 노드에서 프로세스를 시작하는 경우 가장 높은 에너지를 가진 노드가 스스로 이탈할 위험이 가장 큽니다. 이러한 노드를 특별 주시하여 이탈을 예방할 수 있습니다.

확산 분석 예제

에 나타난 2 페이지의 『네트워크 표시』 네트워크를 고려해 보십시오. 네트워크는 서로 다양한 강도의 방향 관계를 가진 7명의 노드로 구성되어 있습니다.

5. Dasgupta, K., R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi. 2008. Social ties and their relevance to churn in mobile telecom networks. In: *Proceedings of the 11th international conference on extending database technology*. New York, NY: ADM.

"확산 프로세스의 예" 테이블은 퍼짐 요인 0.80을 사용한 네트워크의 확산 프로세스를 보여줍니다. 먼저 노드 A는 임의로 1.00의 값이 할당된 모든 에너지를 포함하고 있습니다. 1단계에서 이 노드가 활성화되어 자신이 20%를 보유하면서 관계의 대상인 세 이웃 항목에 자신이 가진 에너지의 80%를 퍼뜨립니다. 관계는 100의 전체 가중치를 가지고 있습니다. 노드 D와의 관계는 이 총 가중치의 반을 차지하므로 이 노드는 확산된 에너지의 반 또는 0.40을 받습니다. 노드 B와의 관계는 총 가중치의 20%를 차지하여 0.16의 확산된 에너지 값을 낳습니다. 노드 C는 노드 A에서 시작한 양의 30%인 0.24의 남은 에너지를 받습니다.

표 11. 예제 확산 프로세스.

단계	노드 A	노드 B	노드 C	노드 D	노드 E	노드 F	노드 G
0	1.00	0	0	0	0	0	0
1	0.20	0.16	0.24	0.40	0	0	0
2	0.20	0.16	0.05	0.08	0.08	0.24	0.19
3	0.20	0.16	0.05	0.08	0.08	0.08	0.38

1단계에서 받은 에너지로 노드 B, C, D는 2단계에서 활성화하여 대상 이웃 항목에 에너지를 확산합니다. 노드 B는 대상 노드가 없으므로 에너지를 확산하지 않습니다. 반면 노드 C는 대상이 있으므로 자신이 가진 에너지의 80%를 대상인 노드 G에게 확산합니다. 마지막으로 노드 D는 자신이 가진 에너지의 80%를 노드 E와 F에게 퍼뜨리며, F는 D와의 관계가 더 강력하므로 E가 받는 양의 세 배를 받습니다.

3단계에서 노드 E, F, G가 활성화됩니다. 노드 E와 G는 대상 노드가 없으므로 에너지를 확산하지 않습니다. 그러나 노드 F는 단일 대상 노드인 G에게 자신이 가진 에너지의 80%를 확산합니다. 이제 이 노드는 노드 C와 F에서 확산된 에너지로 구성된 0.38의 총 에너지 값을 가집니다.

이 지점에서 노드 G가 활성화되지만 확산할 대상 노드가 없으므로 프로세스가 종료됩니다. 노드 A에서 시작한 에너지는 네트워크에 걸쳐 퍼지며 노드 G가 가장 높은 양을 받습니다. 네트워크에 도입된 에너지가 이탈을 나타낼 경우 노드 G는 노드 A 이탈로 가장 큰 영향을 받습니다.

확산 분석 노드

소스 팔레트에서 사용 가능한 확산 분석 노드는 효과에 의해 가장 큰 영향을 받는 개인을 식별하기 위한 네트워크 관계를 사용하여 소셜 네트워크 전반에 걸쳐 지정된 개인의 집합으로부터 효과를 전달합니다. 예를 들어 효과가 이탈일 경우 노드는 네트워크의 다른 특정 개인의 이탈로 인해 가장 이탈할 것 같은 개인을 식별합니다. 기존 모델 및 노드 출력을 가진 데이터를 증가시켜 그러한 모델이 생성하는 예측을 개선할 수 있습니다. 예를 들면 개인 이탈 성향 값을 업데이트하여 확산 영향을 포함할 수 있습니다.

요구사항. 노드에는 분석할 데이터가 들어 있는 두 개의 고정 너비 텍스트 파일이 필요합니다. 첫 번째 파일은 세 개의 필드를 사용하여 소셜 네트워크를 정의합니다. 한 필드는 각 방향 관계에 대한 소스를 식별하며, 한 필드는 각 방향 관계의 목적지를 정의하며, 나머지 필드는 각 관계에 대한 선택적 강도를 지정합니다. 네트워크의 모든 관계는 방향적이어야 합니다. 두 번째 파일에는 효과를 시작하는 식별자의 목록이 들어 있습니다.

확산 분석을 위한 데이터 지정

확산 분석 소스 노드 창의 데이터 탭을 사용하여 네트워크 노드 관계에 들어 있는 입력 파일을 지정할 수 있습니다.

파일. 호출 세부사항 레코드가 들어 있는 파일 또는 폴더의 이름을 지정합니다. 이름을 입력하거나 생략표(...) 단추를 눌러 파일 시스템에서 이름을 선택할 수 있습니다. 이름을 선택하고 나면 경로가 나타나며 내용이 구분자와 함께 표시됩니다. 폴더를 지정할 경우 폴더에 포함된 모든 파일의 호출 세부사항 레코드는 분석을 위해 연결됩니다. 폴더의 모든 파일은 구조가 같아야 합니다.

파일에서 필드 이름을 읽습니다. 기본으로 선택된 이 옵션은 데이터 파일의 첫 행을 열에 대한 이름으로 취급합니다. 첫 행이 헤더가 아닐 경우 이 옵션을 선택 해제하여 자동으로 각 필드에 일반 이름(예: *Field1* 및 *Field2*)을 부여합니다.

네트워크 정의 설정

네트워크 정의 설정은 필드에 대한 역할을 정의합니다.

필드. 화살표 단추를 사용하여 이 목록에서 다양한 역할 필드로 항목을 수동으로 할당합니다. 아이콘은 각 역할 필드에 대한 유효한 측정 수준을 나타냅니다. **모두** 단추를 클릭하여 목록의 모든 필드를 선택하거나 개별 측정 수준 단추를 클릭하여 해당 측정 수준의 모든 필드를 선택합니다.

소스. 한 필드를 방향 관계의 원점으로 선택합니다.

대상. 한 필드를 방향 관계의 대상으로 선택합니다.

가중치. 관계 가중치를 나타내는 한 필드를 선택합니다. 호출 세부사항 레코드의 경우, 가중치는 호출 기간 또는 호출 빈도와 같이 관계 강도를 반영하는 호출 특성의 숫자에 해당할 수 있습니다.

확산 분석을 위한 작성 옵션 설정

확산 분석 소스 노드 창의 작성 옵션 탭을 사용하여 확산 프로세스의 옵션을 정의할 수 있습니다.

확산 시드 목록

확산 시드 목록은 확산을 시작할 에너지를 통해 네트워크 노드를 식별합니다. 예를 들어 이탈 분석에서 이 목록은 이미 이탈한 네트워크 노드를 식별합니다. 또는 이 목록은 노드가 실제로 이탈하는 네트워크에서 효과를 탐색하기 위해 이탈의 위험이 가장 큰 노드를 포함하고 있을 수도 있습니다.

파일. 노드 식별자가 들어 있는 텍스트 파일의 이름을 지정하며, 각 식별자는 파일의 선에 해당합니다. 파일 이름을 입력하거나 생략표(...) 단추를 눌러 파일을 선택할 수 있습니다.

확산 분석 설정

확산 분석 설정은 확산이 얼마나 공세적이며 프로세스가 언제 멈추는지 결정합니니다.

퍼짐 요인.확산 프로세스의 각 단계에서 노드가 전달하는 에너지의 비율을 정의합니다. 값이 작을수록 에너지를 확산 중인 노드에 가장 가까운 노드로 에너지를 제한합니다. 반대로 값이 클수록 거리가 먼 노드에 에너지가 영향을 미치도록 허락합니다.

최대반복계산수. 확산 반복계산수의 상한값을 지정합니다. 이 한계에 도달할 경우 확산 프로세스는 자동으로 종료됩니다.

정확성 임계값. 계속되는 확산 프로세스를 보장하는 노드 에너지의 가장 작은 변경을 지정합니다. 각 노드에 대한 에너지의 변경이 이 값보다 작을 경우 확산 프로세스는 종료됩니다.

요약 통계 계산 및 표시. 선택했을 경우 노드는 분석에 대한 요약 통계를 계산하고 표시할 뿐만 아니라 핵심 성과 표시기 출력을 끌어냅니다. 이러한 통계의 계산은 극도로 큰 네트워크의 노드 성능에 부정적인 영향을 줄 수 있습니다.

확산 분석 통계 보기

확산 분석 소스 노드 창의 분석 탭은 확산 결과의 요약 개요를 제공합니다. "확산 분석 요약 통계" 테이블은 사용 가능한 요약 통계를 보여줍니다.

표 12. 확산 분석 요약 통계.

통계	설명
네트워크의 총 노드	네트워크의 노드 수
네트워크의 총 링크	네트워크의 링크 수
네트워크의 총 확산 시드	확산 프로세스의 시드로 사용된 노드 수
평균 영향	개인과 연관된 확산된 에너지의 평균 양.
평균 내도	개인이 관계의 대상인 관계의 평균 수. 자세한 정보는 4 페이지의 『노드의 차수』의 내용을 참조하십시오.
평균 외도	개인이 관계의 소스인 관계의 평균 수. 자세한 정보는 4 페이지의 『노드의 차수』의 내용을 참조하십시오.

요약 통계 테이블에서 특정 통계를 선택하여 값의 분포, 표준 편차 및 통계에 대한 왜도를 봅니다.

확산 분석의 출력

확산 분석 노드는 네트워크의 개인을 설명하는 다양한 필드를 생성합니다. 기존 모델 및 이러한 핵심 성과 표시기를 가진 데이터를 증가시켜 모델이 생성하는 예측을 개선할 수 있습니다. 예를 들면 개인 이탈 성향 값을 업데이트하여 확산 영향을 포함할 수 있습니다.

"확산 분석 핵심성과지표(KPI)" 테이블은 분석에서 개인에 대한 핵심 성과 지표 필드를 보여줍니다.

표 13. 확산 분석 핵심 성과 표시기.

필드	설명
DA_NodeNumber	개인에 대한 고유한 식별자
DA_DiffusedEnergy	개인과 연관된 확산된 에너지의 양. 이탈 분석의 경우, 값이 높을수록 낮은 값보다 이탈하는 성향이 더 크다는 것을 나타냅니다.

표 13. 확산 분석 핵심 성과 표시기 (계속).

필드	설명
DA_InDegree	개인이 관계의 대상인 관계의 수. 자세한 정보는 4 페이지의 『노드의 차수』의 내용을 참조하십시오.
DA_OutDegree	개인이 관계의 소스인 관계의 수. 자세한 정보는 4 페이지의 『노드의 차수』의 내용을 참조하십시오.

주의사항

이 정보는 전 세계에 제공된 제품 및 서비스를 위해 개발되었습니다.

IBM은 다른 국가에서 이 책에 기술된 제품, 서비스 또는 기능을 제공하지 않을 수도 있습니다. 현재 사용할 수 있는 제품 및 서비스에 대한 정보는 한국 IBM 담당자에게 문의하십시오. 이 책에서 IBM 제품, 프로그램 또는 서비스를 언급했다고 해서 해당 IBM 제품, 프로그램 또는 서비스만을 사용할 수 있다는 것을 의미하지는 않습니다. IBM의 지적 재산권을 침해하지 않는 한, 기능상으로 동등한 제품, 프로그램 또는 서비스를 대신 사용할 수도 있습니다. 그러나 비IBM 제품, 프로그램 또는 서비스의 운영에 대한 평가 및 검증은 사용자의 책임입니다.

IBM은 이 책에서 다루고 있는 특정 내용에 대해 특허를 보유하고 있거나 현재 특허 출원 중일 수 있습니다. 이 책을 제공한다고 해서 특허에 대한 라이센스까지 부여하는 것은 아닙니다. 라이센스에 대한 의문사항은 다음으로 문의하십시오.

150-945
서울특별시 영등포구
국제금융로 10, 3IFC
한국 아이.비.엠 주식회사
대표전화서비스: 02-3781-7114

2바이트(DBCS) 정보에 관한 라이센스 문의는 한국 IBM에 문의하거나 다음 주소로 서면 문의하시기 바랍니다.

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

다음 단락은 현지법과 상충하는 영국이나 기타 국가에서는 적용되지 않습니다. IBM은 타인의 권리 비침해, 상품성 및 특정 목적에의 적합성에 대한 묵시적 보증을 포함하여(단, 이에 한하지 않음) 명시적 또는 묵시적인 일체의 보증 없이 이 책을 "현상태대로" 제공합니다. 일부 국가에서는 특정 거래에서 명시적 또는 묵시적 보증의 면책사항을 허용하지 않으므로, 이 사항이 적용되지 않을 수도 있습니다.

이 정보에는 기술적으로 부정확한 내용이나 인쇄상의 오류가 있을 수 있습니다. 이 정보는 주기적으로 변경되며, 변경된 사항은 최신판에 통합됩니다. IBM은 이 책에서 설명한 제품 및/또는 프로그램을 사전 통지 없이 언제든지 개선 및/또는 변경할 수 있습니다.

이 정보에서 언급되는 비IBM의 웹 사이트는 단지 편의상 제공된 것으로, 어떤 방식으로든 이들 웹 사이트를 옹호하고자 하는 것은 아닙니다. 해당 웹 사이트의 자료는 본 IBM 제품 자료의 일부가 아니므로 해당 웹 사이트 사용으로 인한 위험은 사용자 본인이 감수해야 합니다.

IBM은 귀하의 권리를 침해하지 않는 범위 내에서 적절하다고 생각하는 방식으로 귀하가 제공한 정보를 사용하거나 배포할 수 있습니다.

(i) 독립적으로 작성된 프로그램과 기타 프로그램(본 프로그램 포함) 간의 정보 교환 및 (ii) 교환된 정보의 상호 이용을 목적으로 본 프로그램에 관한 정보를 얻고자 하는 라이센스 사용자는 다음 주소로 문의하십시오.

150-945
서울특별시 영등포구
국제금융로 10, 3IFC
한국 아이.비.엠 주식회사
대표전화서비스: 02-3781-7114

이러한 정보는 해당 조건(예를 들면, 사용료 지불 등)하에서 사용될 수 있습니다.

이 정보에 기술된 라이센스가 부여된 프로그램 및 프로그램에 대해 사용 가능한 모든 라이센스가 부여된 자료는 IBM이 IBM 기본 계약, IBM 프로그램 라이센스 계약(IPLA) 또는 이와 동등한 계약에 따라 제공한 것입니다.

본 문서에 포함된 모든 성능 데이터는 제한된 환경에서 산출된 것입니다. 따라서 다른 운영 환경에서 얻어진 결과는 상당히 다를 수 있습니다. 일부 성능은 개발 단계의 시스템에서 측정되었을 수 있으므로 이러한 측정치가 일반적으로 사용되고 있는 시스템에서도 동일하게 나타날 것이라고는 보증할 수 없습니다. 또한 일부 성능은 추정을 통해 추측되었을 수도 있으므로 실제 결과는 다를 수 있습니다. 이 책의 사용자는 해당 데이터를 본인의 특정 환경에서 검증해야 합니다.

비IBM 제품에 관한 정보는 해당 제품의 공급업체, 공개 자료 또는 기타 범용 소스로부터 얻은 것입니다. IBM에서는 이러한 제품들을 테스트하지 않았으므로, 비IBM 제품과 관련된 성능의 정확성, 호환성 또는 기타 청구에 대해서는 확신할 수 없습니다. 비IBM 제품의 성능에 대한 의문사항은 해당 제품의 공급업체에 문의하십시오.

IBM이 제시하는 방향 또는 의도에 관한 모든 언급은 특별한 통지 없이 변경될 수 있습니다.

이 정보에는 일상의 비즈니스 운영에서 사용되는 자료 및 보고서에 대한 예제가 들어 있습니다. 이들 예제에는 개념을 가능한 완벽하게 설명하기 위하여 개인, 회사, 상표 및 제품의 이름이 사용될 수 있습니다. 이들 이름은 모두 가공의 것이며 실제 기업의 이름 및 주소와 유사하더라도 이는 전적으로 우연입니다.

이 정보를 소프트카피로 확인하는 경우에는 사진과 컬러 삽화가 제대로 나타나지 않을 수도 있습니다.

상표

IBM, IBM 로고 및 ibm.com은 전세계 여러 국가에 등록된 International Business Machines Corp.의 상표 또는 등록상표입니다. 기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다. 현재 IBM 상표 목록은 웹 (www.ibm.com/legal/copytrade.shtml)의 "저작권 및 상표 정보"에 있습니다.

Intel, Intel 로고, Intel Inside, Intel Inside 로고, Intel Centrino, Intel Centrino 로고, Celeron, Intel Xeon, Intel SpeedStep, Itanium 및 Pentium은 미국 또는 기타 국가에서 사용되는 Intel Corporation 또는 그 자회사의 상표 또는 등록상표입니다.

Linux는 미국 또는 기타 국가에서 사용되는 Linus Torvalds의 등록상표입니다.

Microsoft, Windows, Windows NT 및 Windows 로고는 미국 또는 기타 국가에서 사용되는 Microsoft Corporation의 상표입니다.

UNIX는 미국 및 기타 국가에서 사용되는 The Open Group의 등록상표입니다.

Java 및 모든 Java 기반 상표와 로고는 Oracle 및/또는 그 계열사의 상표 또는 등록상표입니다.

기타 제품 및 서비스 이름은 IBM 또는 타사의 상표입니다.

용어집

가

계량 관계(valued relationship) . 관계의 강도를 나타내는 연관된 가중치를 가진 두 노드 간의 연결입니다.

권위 리더(authority leader) . 그룹에서 권위 스코어가 가장 높은 개인입니다.

권위 스코어(authority score) . 그룹의 다른 멤버가 개인에게 얼마나 연결되어 있는지를 나타내는 측정입니다. 권위 스코어는 개인의 중요성을 그 개인에게서 끝나는 관계의 수와 연결합니다.

나

내도(indegree) . 방향 관계를 구성하는 네트워크에서 노드에 대한 명성 측정입니다. 노드가 대상인 관계의 총 숫자로 정의됩니다.

마

밀도 . 관측된 관계 수를 가능한 관계 수로 나눈 것으로 정의되는 네트워크 응집력의 측정입니다.

바

방향 관계(directional relationship) . 한쪽(소스)에서 시작하여 다른 쪽(목적지)에서 끝나는 두 노드 사이의 연결입니다.

범위 임계값(coverage threshold) . 그룹 식별에 사용된 가장 강한 관계의 비율입니다.

비방향 관계(nondirectional relationship) . 두 노드 사이의 연관입니다.

사

소시오그램(sociogram) . 개인 서로 간의 관계를 나타내는 연결 지점을 선으로 그린 소셜 네트워크의 시각적 표현입니다.

아

외도(outdegree) . 방향 관계를 구성하는 네트워크에서 노드에 대한 중심성 측정입니다. 노드가 소스인 관계의 총 숫자로 정의됩니다.

유포 리더(dissemination leader) . 그룹에서 유포 스코어가 가장 높은 개인입니다.

유포 스코어(dissemination score) . 개인이 그룹의 다른 멤버에게 얼마나 연결되어 있는지를 나타내는 측정입니다. 유포 스코어는 개인의 중요성을 그 개인에게서 시작하는 관계의 수와 연결합니다.

이분적 관계(dichotomous relationship) . 두 개의 값 중에서 하나만 가질 수 있는 두 노드 간의 연결로서, 연결 존재 또는 연결 부재를 나타냅니다.

자

정확성 임계값(accuracy threshold) . 계속되는 확산을 보장하는 노드 에너지의 가장 작은 변화에 해당하는 확산 프로세스의 중지 기준입니다.

차

차수(degree) . 네트워크에서 노드 활동 측정입니다. 노드와 관련된 관계의 총 숫자로 정의됩니다.

파

퍼짐 요인(spreading factor) . 노드가 확산 프로세스 단계에서 이웃 항목에게 전달하는 에너지의 비율입니다.

하

핵심 그룹(core group) . 네트워크에서 약한 관계를 생략하고 그룹 크기 제한을 적용한 후에 남아 있는 노드 그룹입니다.

확산된 에너지 . 전체 네트워크에 퍼진 네트워크의 특정 노드에서 시작하는 효과입니다. 초기 노드에서의 거리가 커지는 만큼 크기가 줄어듭니다.

색인

[가]

가중 필드 7, 14, 19
결과 지우기 6
계량 관계 2
관계
  계량 2
  방향 2
권위 리더 13, 15
권위 스코어 13, 15
그룹 분석 5
  그룹 크기 12, 14
  범위 임계값 12, 14
  요구 사항 13
그룹 크기 15
  한계 12, 14

[나]

내도 4, 15, 20

[다]

데이터 분석 6
데이터 탭 5
  그룹 분석 14
  확산 분석 19

[마]

목적지 필드 7
밀도 3, 15

[바]

방향 관계 2
범위 임계값 12, 14
분석 탭 5
  그룹 분석 15
  확산 분석 20
비방향 관계 2

[사]

소스 필드 7
속성
  스크립팅 8
스크립팅
  속성 8
스트림 실행
  IBM SPSS Modeler Server Social Network Analysis 사용 5
시드
  확산 분석용 19

[아]

외도 4, 15, 20
유포 리더 13, 15
유포 스코어 13, 15
유형 탭 5
이분적 관계 2

[자]

작성 옵션 탭 5
  그룹 분석 14
  확산 분석 19
정확성 임계값 17, 19
주석 탭 5

[차]

차수 4
출력 미리보기 6
출력 필터링 5
측정 수준 5

[카]

커널 밀도 15

[파]

퍼짐 요인 17, 19
평균 그룹 밀도 15
평균 그룹 크기 15
평균 내도 15, 20
평균 영향 20
평균 외도 15, 20
평균 핵심 그룹 밀도 15
평균 핵심 멤버 비율 15
필터 탭 5

[하]

핵심 그룹 12
핵심 노드 15
핵심 성과 표시기 15, 20
호출 세부사항 레코드 7
확산 분석 5, 17
  시드 19
  옵션 19
  요구 사항 18
  정확성 임계값 17, 19
  퍼짐 요인 17, 19
확산 시드 목록 19
확산된 에너지 20

G

GAG_GroupNumber 15