

SAS® Contextual Analysis

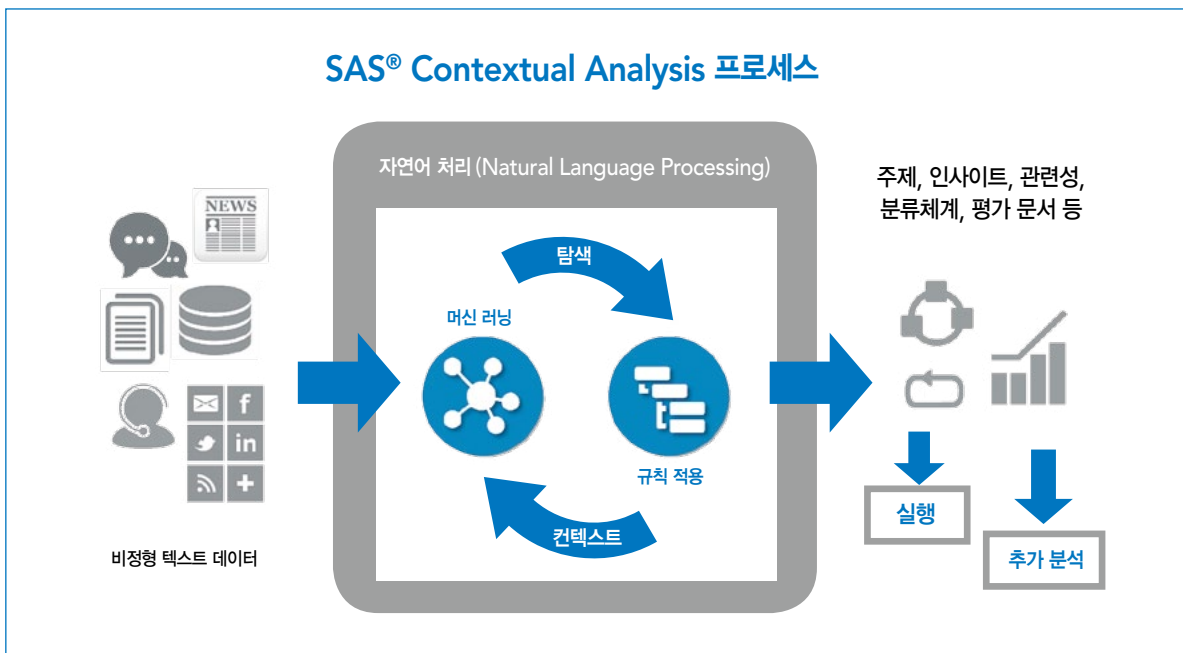
> fact sheet

비정형 텍스트의 체계적인 관리로 더욱 값진 인사이트 도출

SAS® Contextual Analysis 소개

SAS® Contextual Analysis는 비정형 데이터에서 중요한 비즈니스 인사이트를 확보할 수 있도록 도와주는 고급 언어 솔루션으로, 머신 러닝과 분야별 전문 지식을 결합해서 데이터에 감추어져 있는 트렌드와 주제를 찾아냅니다. 기존에는 비즈니스에 대한 심층 이해를 돕는 소중한 인사이트가 데이터 속에 묻혀 발견하기 어려웠을 것입니다.

SAS® Contextual Analysis는 매우 간편한 방법으로 사용할 수 있는 여러 가지 고급 기술을 제공합니다. 가령 분류체계 (taxonomies)를 구축하기 위해 학습 말뭉치 (training corpus)를 생성할 필요가 없으며, 대신 소프트웨어가 분류 체계를 찾아주고 자동으로 주제를 파악해서 분석 작업의 방향을 정해줍니다. 더욱이 이 솔루션은 뛰어난 유연성을 자랑합니다. 사용자는 시스템에서 생성한 분류 규칙에 변화를 주어 특정 영역에 더 초점을 맞추으로써 자신이 필요로 하는 정확한 인사이트를 도출하여 비즈니스 가치를 높일 수 있습니다.



SAS® Contextual Analysis 특징

- **단일 통합 시스템 환경에서 손쉽게 분석 작업 손쉽게 관리**
사용자는 SAS® Contextual Analysis를 곧바로 업무에 적용해서 가동할 수 있고, 필요에 따라 규칙을 맞춤 설정할 수 있습니다. 더욱이 이 모든 작업이 단일 통합 환경에서 이루어집니다. 탐색 작업은 처음부터 자동으로 수행되며, 사용자는 자연어 처리 기술을 이용해 자신만의 규칙 정의를 정할 수 있습니다.
- **체계적인 하이브리드 접근법으로 통제력 강화**
SAS® Contextual Analysis는 하이브리드 방식으로 사용자가 상황에 맞게 텍스트 분석을 맞춤화하고, 이를 위해 분야별 전문 지식을 자동 머신 러닝에 반영하여 인사이트를 도출합니다. 이 소프트웨어는 공통의 주제, 카테고리 분류 규칙, 문서 감성 등을 데이터에 기초하여 종합적으로 파악합니다. 사용자는 자신이 원할 때면 언제든지 특정 요구니즈에 부합하도록 머신 러닝 결과물을 검토하고 수정할 수 있습니다. 그 밖에도 사용자는 뛰어난 유연성과 분석 성능, 표준화된 분석 절차 그리고 신속한 결과 도출 등의 이점을 누릴 수 있습니다.
- **더욱 광범위한 데이터의 활용**
SAS® Contextual Analysis는 비정형 데이터에서 가치를 도출할 수 있게 해줄 뿐만 아니라, 사용자가 시각적으로 작업 결과물과 상호작용할 수 있도록 다양한 솔루션 환경에서 데이터를 지원합니다. 따라서 사용자는 시각적 분석을 구현하거나, 새로운 메타데이터를 생성해서 시각화 또는 고급 분석 등의 타 영역으로 응용 범위를 넓혀 나갈 수 있습니다.

SAS® Contextual Analysis 주요 기능

SAS® Contextual Analysis는 텍스트 데이터에서 새로 떠오르는 이슈나 미래의 트렌드 등을 쉽게 파악할 수 있게 해주며, 사용자는 이 정보를 활용해 비즈니스와 고객들에 대한 이해 증진을 도모하고 변화하는 시장 및 규제 등에 빠르게 대응할 수 있도록 합니다.

SAS® Contextual Analysis는 방대한 양의 문서데이터로 인해 업무를 제대로 처리하지 못하거나, 문서를 수작업으로 검토할 수 있는 시간이나 여력이 없는 조직들을 위해 설계된 솔루션입니다. 즉시 실무 실전에 적용해서 사용할 수 있도록 제작된 SAS® Contextual Analysis는 특히 IT와 비즈니스 사용자 모두에게 이상적인 맞춤형 솔루션입니다.

체계적인 텍스트 모델 개발과 적용을 위한 단일 시스템

언어학 모델 (linguistic models)의 생성/수정/적용 프로세스를 자연스럽게 진행되는 단일 환경으로 텍스트 분석 작업 과정과 문서 분류 작업을 더욱 간소화할 수 있습니다. 또한 시스템의 각종 설정을 단일 인터페이스 환경에서 쉽게 관리할 수 있으므로 사용자는 시스템에 대한 통제력을 강화할 수 있습니다.

- 포인트 앤 클릭을 지원하는 단일 GUI 인터페이스
 - 용어
 - 주제 탐색
 - 범주 생성
 - 컨셉 정의
 - 문서별 감성 분석
- 텍스트 모델의 효과적인 생성/수정/적용을 위한 체계적 텍스트 분석 절차 제공
- 프로젝트 초기에 기존 SAS® Content Categorization에서 생성된 분류체계 모델 불러오기 기능
- 사용자 지정 컨셉을 생성, 수정, 활성화(또는 비활성화)하고 동일한 사용자 환경에서 즉시 언어 규칙 정의의 유효성 검사의 수행
- 초기 프로젝트 정의를 위한 내장 프로젝트 생성 마법사 기능
- 텍스트 모델 생성 단계 전반에 적시적인 도움말 기능
- 텍스트 모델 생성 각 단계의 처리 상태를 상세히 보여주는 속성 패널 및 모델 개발 이슈 (예 : 주제 생성 시 필요한 문서 수 부족 등)의 진단을 도와주는 메시지 대화상자
- 복수의 텍스트 모델을 동시에 개발·실행할 수 있는 프로젝트 관리 옵션과 각 모델의 개발 상태를 알려주는 설명 기능
- 모든 프로젝트 속성에 대한 통합된 메타데이터 관리

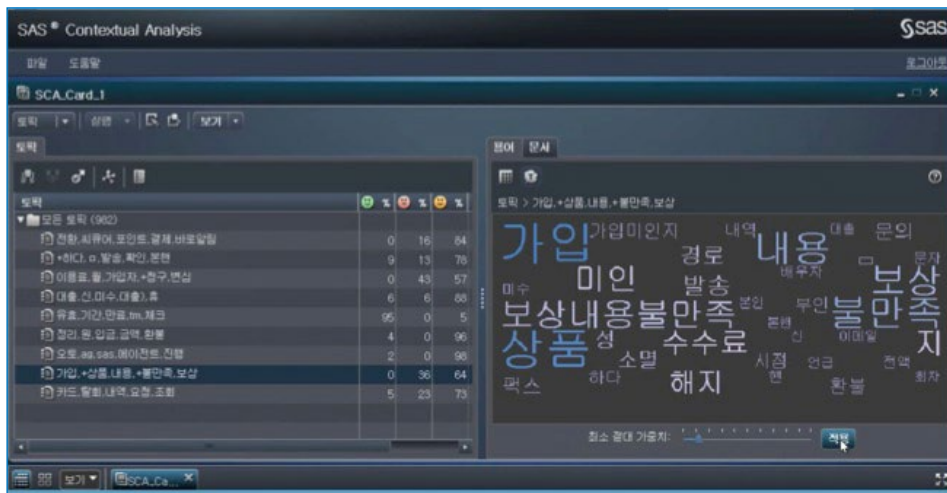


그림 1. 수많은 텍스트를 분석하여 특정 토픽을 추출하고 해당 토픽에 대한 대표 키워드들을 도출합니다.

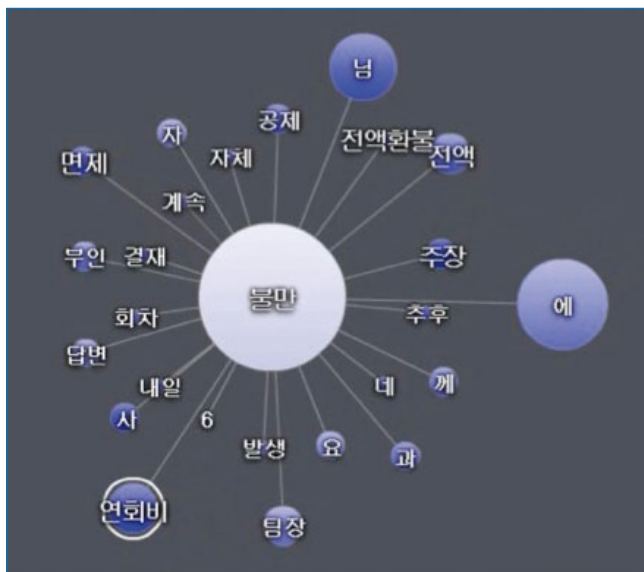


그림 2. 특정 키워드에 대하여 연관성이 깊은 주변 키워드 정보를 직관적으로 보여줍니다.

하이브리드 방식을 이용한 문서 분류

머신 러닝 기법을 이용해서 공통의 주제, 분류 규칙, 그리고 문서에 내포된 감성 등을 종합적으로 파악할 수 있고, 사용자의 분야별 전문 지식을 적용해서 맞춤형된 분석 작업을 수행할 수 있습니다. 이로써 사용자는 텍스트의 맥락을 더 깊이 파악할 수 있고, 궁극적으로 작업 결과를 더 빠르게 도출할 수 있습니다.

- 용어 식별, 어휘의 원형 추출 및 형태소 분석을 위한 자동화된 자연어 처리 (가용어 및 불용어 정의)
- Drag-and-drop으로 용어의 유지 및 제거를 손쉽게 변경
- 시스템에서 초기 분류 체계의 생성을 위한 머신 러닝 기반의 주제 탐색 기능
- 텍스트 문서를 선행 분류 체계 구축 이후 적용하여 분석 결과 생성 옵션 제공
- 주제에 포함된 용어를 시각화 환경에서 수정하고 유사한 주제를 병합하여 생성 결과를 개선하는 기능
- 초기 학습 데이터에 기반하여 자동으로 분류 규칙을 생성하고 주제를 분류 카테고리로 등록하는 기능
- 내장된 다양한 언어 연산자를 이용해 시스템에서 생성한 범주의 분류 규칙을 편집하는 기능
- 사전 정의된 17개의 공통 컨셉과 사용자 지정 컨셉을 생성해서 분석의 정확도를 높임
- 분석 결과의 시 컨셉을 적용하거나 적용하지 않는 기능
- 정의된 컨셉의 전부 또는 일부에 대한 구문 유효성 검사
- 주제 탐색 과정에서 생성된 주제 또는 범주로 분류된 문서 별 감성 자동 파악

자연어 처리

수작업으로 문서를 검토하다 보면 많은 시간이 낭비될 뿐만 아니라 작업 오류가 발생할 소지가 있습니다. 하지만 데이터 기반의 자연어 처리 기술을 이용하면 번거로운 수작업을 피할 수 있습니다. 가령 사용자가 입력 데이터만 지정하면, 자동으로 솔루션이 단어 빈도 수, 품사, 어간 등을 파악합니다.

- 소프트웨어가 제공하는 자연어 처리에 포함된 기능
 - 파싱
 - 토큰화
 - 품사 태깅
 - 형태소 분석
- 유지/제거할 용어의 리스트와 용어 빈도수 출력
- 유지/제거 용어 화면 간 Drag-and-drop 기능 지원
- 텍스트 처리 시 Start 및 Stop 리스트 적용 - 사용자 정의 Start/Stop 리스트 사용 가능
- 입력 데이터에 포함된 범주형 변수를 지정해서 자동적인 분류 범주 생성
- 내장된 연산자 구문을 사용하여 사용자 지정 컨셉 정의
- 텍스트 분석기는 입력 문서 없이 분류 체계를 개발 후 문서에 대해 분류 체계 실행 가능

초기 분류 체계 개발을 위한 주제 탐색

많은 문서들을 수작업으로 일일이 검토해서 핵심 주제를 정확히 찾아내기란 매우 어려운 일입니다. SAS® Contextual Analysis는 머신 러닝 및 통계 기법을 이용해 문서군에서 자동으로 핵심 주제를 찾아줍니다.

- 자동화된 머신 러닝 기반의 탐색 (machine discovery)은 연관 적합도 스코어 (associated relevance scores)에 근거하여 입력 문서에서 핵심 주제 식별
- 용어 클라우드, 컨셉 맵 (interactive term maps), 주제 세분화로 주제 내 용어 관계를 조사, 탐색하여 연관성 평가 및 탐색된 주제 정교화
- 해당 주제와 관련된 문서에 대한 감성의 수준을 시각적으로 표현
- 도출된 주제를 분리하거나 유사한 주제를 하나의 주제로 병합하는 기능
- 원하는 주제를 분류 규칙 정의를 위한 범주로 등록하고, 컨셉을 활용해 분류 정확도 향상

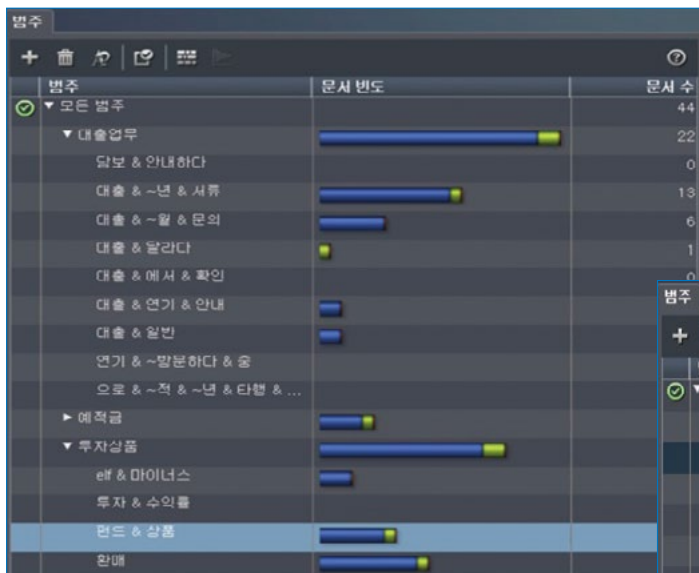
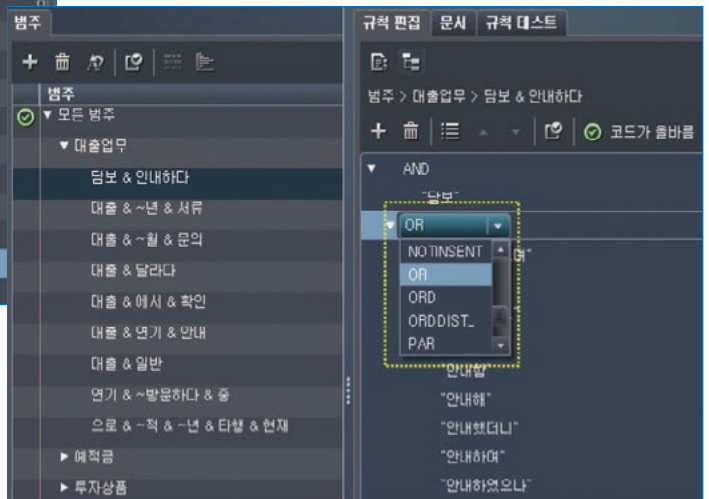


그림 4. 학습기반으로 생성된 룰을 하이브리드 방식으로 사용자의 의도에 맞게 수정하여 보다 정교한 모형을 만들 수 있습니다.

그림 3.

머신러닝 기반으로 분류 룰을 생성하고 분류 룰에 따라 문서를 자동 분류하며 선택된 범주에 따라 문서를 검증합니다.



설정 가능한 규칙 생성과 향상된 언어적 컨텍스트

SAS® Contextual Analysis는 특히 출원 중에 있는 독창적인 알고리즘을 이용해 분류 규칙을 자동으로 생성함으로써 논리연산자를 활용한 규칙 정의와 분류 체계 개발 작업을 간소화합니다. 이렇게 생성된 규칙은 정해진 컨셉트에 따라 의미론적으로 변경 및 개선 작업을 거치게 됩니다.

- 생성한 주제에 근거하여 자동화된 초기 범주 규칙 생성
- 이해하기 쉬운 논리연산자 규칙 정의를 통해 범주화 모델 (즉, 분류 체계) 생성
- 시스템에서 생성한 규칙의 적합성을 진양성 (true positive), 가양성 (false positive) 및 가음성 (false negative)으로 시각화하여 표현
- 범주의 감성이 문서 표시 화면에 시각적으로 묘사
- 초기 단계부터 규칙을 사용자 지정 범주로 편집 및 정의
- 분석가는 내장된 다양한 규칙 연산자를 사용하여 상세하게 규칙 모델 지정
- 입력 데이터에 사전 정의된 범주형 변수의 계층구조를 분류 체계로 생성
- 규칙을 배포하기 전에 입력 데이터에 대해 테스트 실시 가능
- 컨셉을 범주에 활용하여 보다 정확한 의미를 추출할 수 있도록 모델 정교화
- 사전 정의된 컨셉 사용 시 규칙 작성이 필요치 않고, 날짜, 장소, 시간 등이 표준으로 구성
- 또한 SAS Content Categorization Studio 프로젝트에서 작성된 기존의 컨셉트 불러오기 가능

SAS®와의 유연한 연계

텍스트 분석을 통해 핵심 비즈니스 인사이트를 찾아내는 것은 그 자체로도 큰 의미를 지닙니다. 하지만 이것은 심층 분석으로 나아가기 위한 과정의 일부이기도 합니다. SAS® Contextual Analysis는 다양한 텍스트 기반의 분석 결과물을 제공하므로, 사용자는 이를 통해 추가의 탐색 작업을 위한—가령 SAS Visual Analytics를 이용한—데이터를 준비할 수 있습니다.

- 처리 지연 시간 최소화: 초당 수백만 건의 이벤트를 처리할 수 있는 대용량 프로세싱
- 업데이트 기본 지원: 데이터가 스트림 처리되는 동안 삭제 / 삽입 가능
- 병렬 인-메모리 그리드 환경에서 연속 쿼리 지원 (유연한 스레드 프로세싱 모델에서 쿼리 실행)
- 고속 데이터 로드 및 데이터 유형 쿼리가 가능하도록 설계된 사용자 지정 어댑터
- 메모리 증가에 따른 선형 확장 기능: 이벤트 분석 루틴을 격리하도록 사용자 지정이 가능한 병렬 모델 프로세싱
- 상황에 따라 캐싱 스토어 정의 가능: 가용 분산 메모리의 한계를 초과하는 데이터 프로세싱 처리 및 가용 하드웨어 환경의 제약과 프로세싱 속도 허용 수준 간에 균형 유지

SAS® Contextual Analysis에 관한 더 자세한 내용은

www.sas.com/korea/contextual-analysis

페이지를 참조하시기 바랍니다.

시스템 사양

서버 및 호스트 플랫폼

- Linux x64 (64-bit): Novell SuSE 11 SP1; Red Hat Enterprise Linux 6.1; Oracle Linux 6.1
- Microsoft Windows on x64 (64-bit):
Desktop: Windows 7* x64 SP1; Windows 8** x64
Server: Windows Server 2008 x64 SP2 Family; Windows Server 2008 R2 SP1 Family; Windows Server 2012 Family

미드-티어

- HP/UX on Itanium
- IBM AIX on POWER
- Linux x64 (x86-64)
- Microsoft Windows x64 (x86-64)
- Solaris (SPARC and x64)

웹 브라우저

- Internet Explorer 9: Windows 7 (32-bit and x64 32-bit Web browsers)
- Internet Explorer 10: Windows 7 and Windows 8 (32-bit and x64 32-bit Web browsers)
- Firefox 6 and up: Windows 7 and Windows 8 (32-bit and x64 32-bit Web browsers); Linux x64: RHEL 6 and SLES 11 (32-bit Web browsers)
- Chrome 15 and up: Windows 7 and Windows 8 (32-bit and x64 32-bit Web browsers); Linux x64: RHEL 6.1 and SLES 11 SP 1 (32-bit Web browsers)

클라이언트

- Microsoft Windows (64-bit): Windows 7* x64 SP1; Windows 8** x64

* NOTE: Windows 7 supported editions are: Professional, Ultimate and Enterprise.

** NOTE: Supported editions include: Windows 8, Windows 8 Pro, Windows 8 Enterprise

한국패스소프트웨어(유) 서울특별시 서초구 서초대로 398 플래티넘타워2-3층 (우)06619

SAS 및 기타의 모든 SAS Institute Inc. 제품명 또는 서비스명은 미국 및 기타 국가에서 사용되는 SAS Institute Inc.의 등록상표 또는 상표입니다. © 기호는 미국 등록 상표임을 의미합니다. 기타의 브랜드 및 제품명은 각 해당 회사의 등록 상표입니다. Copyright © 2016, SAS Institute Inc. All rights reserved. 105943_S121718.0414

www.sas.com/korea

 **sas**
THE POWER TO KNOW®