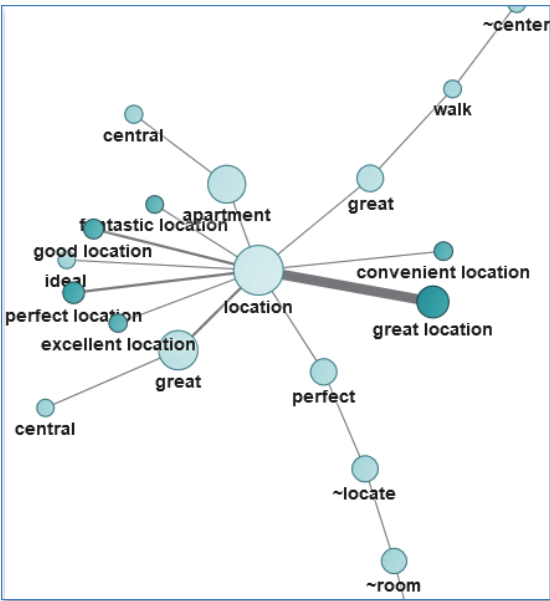


SAS® Visual Text Analytics

자연어 처리(NLP), 머신 러닝 및 언어 규칙을 결합하여
데이터에 숨어있는 인사이트 발견



SAS® Visual Text Analytics는 어떤 솔루션인가?

SAS Visual Text Analytics는 다양한 모델링 접근법을 제공하여 지도(Supervised) 및 비지도(Unsupervised) 머신 러닝, 언어 규칙, 범주화, 개체 추출, 감성 분석 및 주제 감지를 포함한 비정형 데이터의 가치를 극대화합니다.

SAS® Visual Text Analytics가 중요한 이유는?

텍스트 데이터는 산업 전반을 아우르며 점차 그 비중이 확대되고 있습니다. SAS Visual Text Analytics는 텍스트 데이터의 식별 및 범주화를 지원하는 종합적인 솔루션을 제공합니다. 이를 통해 기업은 방대한 양의 텍스트 데이터로부터 유용한 정보를 읽고 구성하고 추출하는 인적 활동을 조정할 수 있습니다. 사용자가 일련의 문서 집합을 자동으로 분석 및 범주화하는 모델을 구축하면 비정형 데이터가 유의미한 인사이트로 자동 변환되어 머신 러닝과 예측 모델에 투입됩니다.

SAS® Visual Text Analytics는 누구를 위한 솔루션인가?

SAS Visual Text Analytics는 대량의 비정형 데이터를 분석하여 새로운 인사이트를 도출해야 하는 비즈니스 분석가, 도메인 전문가, 리서치 분석가, 언어 전문가, 지식 근로자 및 데이터 사이언티스트를 위한 솔루션입니다.



텍스트 분석은 기록물의 관리 및 해석, 리스크 또는 사기의 검증, 문제의 조기 발견을 위한 고객 피드백 통합 등 갖가지 일상적인 비즈니스 문제의 해결을 지원합니다. 비정형 데이터가 확산됨에 따라 텍스트 분석 활동이 산업 전반으로 확대되고 있습니다.

SAS Visual Analytics는 사전 정의된 템플릿, 머신 러닝 기법, 자연어 처리(NLP; Natural Language Process)를 통해 대량의 비정형 데이터를 분석하여 심층적인 인사이트를 신속하게 도출합니다. 이는 현대적이고 유연한 프레임워크 내에서 텍스트 마이닝, 컨텍스트 추출, 범주화, 감성 분석 및 검색을 포함합니다. 엔드 투 엔드 시각 파이프라인을 통해 데이터 준비, 주제를 시각적으로 탐색, 개체 및 유의미한 정보 추출, 감성 분석, 텍스트 모델 등의 작업이 용이해지며 텍스트 모델을 구축하여 이를 기존 시스템 또는 프로세스 내에 배포할 수 있습니다.

주요 특징

- **떠오르는 트렌드를 찾아 새로운 활동 기회를 포착** 비정형 데이터를 유의미한 인사이트로 자동 변환합니다. 주제 전문성을 통해 강화할 수 있는 규칙 기반 접근 방식과 머신 러닝 기법을 결합하여 텍스트 모델의 정확도를 개선합니다.
- **자동화를 통한 신속한 결과 도출** 지능형 알고리즘과 NLP 기법이 텍스트 데이터의 관계 및 패턴을 자동으로 식별하므로 수동 분석이 필요 없습니다. 무형식 텍스트에서 중요 주제를 식별 및 추출하여 새로운 변수를 생산하므로 예측 모델, 리포트 및 검색 또는 필터링 애플리케이션의 개선 효과를 기대할 수 있습니다.
- **데이터에서 의사결정에 이르는 시간 단축** 데이터 소스에서 바로 의사 결정을 내릴 수 있습니다. 누군가가 코멘트를 남기거나 모바일 앱을 통해 클릭하는 즉시 해당 데이터가 분석되므로 데이터에서 의사결정까지의 과정을 단축시킬 수 있습니다. 또한, 인-메모리, 인-DB, 인-스트림 기술을 통해 정보를 입수한 후, 그에 따른 조치를 취할 때까지의 소요 시간을 단축할 수 있습니다.
- **오픈 환경에서 협업 및 정보 공유 강화** SAS Visual Text Analytics는 데이터 준비 및 시각적 탐색에서부터 분석 및 배포에 이르는 분석 라이프 사이클 전체를 지원하는 유연한 환경을 제공합니다. 다양한 분석 사례를 다루고 이를 검토하여 단일 전략을 제공합니다. 또한, 데이터를 준비하는 데이터 사이언티스트, 언어 규칙을 적용하는 도메인 전문가, 모델을 배포하는 IT 담당자들과 모든 관점에서 협업이 가능하며, 기존 시스템 및 개방형 소스 기술과 완벽하게 통합됩니다.

주요 기능

텍스트 기반 데이터가 수작업으로 검토 및 분석할 수 없을 정도로 많아 졌다면 인사이트를 위한 시스템 목록에 텍스트 분석을 추가할 때입니다. SAS Visual Text Analytics는 다음과 같은 기능을 결합합니다.

- 시각적 프로그래밍 플로우인 파이프라인 및 제어 가능한 노드
- 범주화 및 개체 추출을 위한 머신 러닝 및 규칙 기반의 언어적 방법
- 각 구성 요소에 대한 문서 차원에서 스코어링

텍스트 분석 아웃풋을 서비스 또는 품질 이슈에 대한 감지 및 추적, 고객 피드백 정량화, 리스크 평가, 운영 프로세스 개선, 예측 모델 강화 등을 포함한 다양한 비즈니스 활동에 적용할 수 있습니다.

머신 러닝과 규칙 기반 방법의 결합

여러 개의 문서에서 핵심 주제 및 화제를 정확히 찾아내는 일은 매우 어려운 작업입니다. SAS Visual Analytics는 머신 러닝 기법을 이용하여 문서 내 중요 단어들 중에서 화제를 자동으로 도출합니다. (화제는 주요 주제나 아이디어를 묘사하고 그 특징을 나타내는 용어들로 구성됩니다.) 텍스트 내 현재 화제를 탐색하고 시간에 따른 변화 양상을 확인할 수 있습니다. 이러한 인사이트를 바탕으로 즉각적인 조치를 취하거나 주제 전문성을 기반으로 인사이트를 상세히 조정할 수 있습니다.

발견된 주제에 대한 추가적인 조정이 필요하거나 사용자 정의가 선호되는 경우에는 정밀한 규칙 기반의 범주 및 개념을 생성할 수 있습니다.

머신 러닝 기술은 상당히 발전했지만 아직 특정 뉘앙스나 언어의 모호성에 따른 복잡성을 포착하지는 못합니다. 따라서 속어와 풍자를 파악하고 의도를 추론하기 위한 언어 규칙이 필요합니다. 고급 규칙 조정과 다양한 불리안(Boolean) 연산자 및 언어 한정자를 통해 사용자는 비정형 데이터를 자유롭게 분석할 수 있습니다.

- 비지도(Unsupervised) 머신 러닝을 통해 자동으로 화제 생성
- 자동 규칙 빌더로 지도(Supervised) 머신 러닝을 통한 토픽 범주화
- 규칙 기반 언어적 방법을 사용한 핵심 개념 추출
- 배포 전 입력 데이터 세트에서 규칙을 테스트

컨텍스트 추출

이름, 날짜, 측정값, 인과관계와 같은 핵심 데이터 요소가 자유형 텍스트 필드 내에 감춰져 있을 때는 필터링, 검색 또는 리포팅 등의 작업이 용이하지 않습니다. SAS Visual Text Analytics는 품사 태그, 정규식, 다양한 규칙 유형, 언어 한정자, 불리안 연산자를 비롯해 정확도와 컨텍스트의 특이성을 도출하는 종합적인 도구를 제공합니다.

사용자는 사전에 정의된 개념을 이용하여 비정형 데이터로부터 데이터 요소를 감지 및 추출하거나 사용자 정의와 개념을 생성할 수 있습니다. 이를 통해 정확한 컨텍스트를 확인하고 이러한 데이터 요소에서 핵심 부분만 추출함으로써 새로운 정형 데이터를 생성할 수 있습니다. 분석가들은 이를 통해 다운스트림 애플리케이션을 향상시키고 이를 자연어 이해(NLU) 및 AI 애플리케이션에 투입합니다.

- 사전 정의된 개념들을 이용하여 이름, 날짜, 통화, 측정값 등의 일반 개체 추출
- 키워드, 불리안 연산자, 정규식, 술어 로직 및 다양한 언어 연산자를 사용하여 사용자 정의 및 개념 생성
- 원하는 정보를 포함하는 필드 추가 가능. 관심 개체나 사실에 대한 별도의 기록 불필요
- 컨텍스트의 특이성이나 범위를 확보하기 위해 사전 정의 또는 사용자 정의된 개념을 범주화 규칙에서 참조

유연한 배포

SAS를 이용하면 API를 통해 배치(Batch), Hadoop 및 스트림 포맷으로 모델을 배포할 수 있습니다. 사용자 인터페이스를 거치지 않고도 텍스트 모델을 이용하여 데이터를 보강(Enrichment)할 수 있습니다. 데이터 저장위치와 가까운 곳에서 모델을 실행하여 데이터 이동을 줄이고 결과를 빠르게 도출하며 새로운 데이터에 대한 스코어링을 지원합니다.

- 개념, 감성, 주제 및 범주 노드가 외부 데이터 세트에 모델을 배포하는 스코어 코드 제공
- 분산 프로세싱을 위해 기본적으로 스레드 처리되는 스코어 코드를 통해 컴퓨팅 자원을 최대한 활용하여 대용량의 데이터 세트에서도 빠르게 결과를 도출

멀티 사용자 환경을 통한 팀워크 및 협업 강화

SAS 플랫폼은 소프트웨어 중 가장 심층적인 통합 분석 기능을 제공합니다. 이 기능은 협업과 정보 공유의 핵심 요소입니다. SAS 플랫폼의 일부인 SAS Visual Text Analytics는 Best Practice 파이프라인 및 방법을 공유할 수 있는 업무 공간을 제공하여 팀워크를 강화합니다.

여러 SAS 프로젝트 전반의 파이프라인에서 관련 용어 식별, 사용자 화제의 변경 또는 생성, 언어 규칙 생성 등의 광범위한 분석 활동을 Best Practice 노드로 공유할 수 있습니다.

SAS는 기존 시스템 외에도 Python, R, Java, Lua와 같은 개방형 소스 언어를 사용할 수 있습니다. 또한 RESTful API를 통해 SAS의 다양한 기능을 다른 애플리케이션에 쉽게 사용할 수 있습니다.

- 시각적 프로그래밍 플로우를 제공하는 그래픽 UI
- 다른 사용자와 프로젝트 공유
- 텍스트 분석 파이프라인을 통해 다섯 가지 노드(개념, 텍스트 파싱, 감성, 주제 및 범주) 이용
- 텍스트 분석 프로젝트의 전형적인 워크플로우를 보여주는 사전 정의된 기본 파이프라인
- 유연한 파이프라인을 통한 추가적인 노드 생성 또는 기본 파이프라인 수정
- 기존 시스템 및 개방형 소스 기술과의 원활한 통합
- SAS® Viya™ REST API를 통해 SAS Analytics의 기능을 다른 애플리케이션에 추가

- 텍스트 요약, 텍스트 데이터 분할, 텍스트 파싱 및 마이닝, 주제 모델링, 텍스트 규칙 개발 및 스코어링, 텍스트 규칙 탐색, 용어 매핑 및 주제 용어 매핑, 조건부 랜덤 필드, 검색을 위한 즉시 사용 가능한 분석 프로그래밍 인터페이스

자연어 처리

문서를 수작업으로 검토하면 시간이 많이 소요될 뿐만 아니라 오류가 발생할 수 있습니다. 자연어 처리(NLP)는 지루한 수동 분석을 줄여주고 텍스트 내 중요 패턴을 쉽게 식별 및 추출할 수 있도록 해줍니다.

NLP는 대체로 텍스트 분석 프로세스의 첫 단계에 속해 있습니다. NLP는 언어 분석을 수행하여 머신이 텍스트를 읽을 수 있도록 도와줍니다. SAS Visual Text Analytics는 NLP를 이용하여 텍스트를 분석하고 형식적 표현으로 변환하여 텍스트의 처리 및 이해를 지원합니다. 단어 및 문장 토큰화, 분할, 어간추출, 합성어 분해, 품사 태깅, 개체명 인식, 시맨틱 파싱(Semantic Parsing) 등을 수행하기 위해 서로 다른 NLP 라이브러리나 사용자 정의 코드를 억지로 맞출 필요가 없습니다. 따라서 텍스트 해석 방식에 대한 프로그래밍 시간을 줄이고 텍스트 데이터에서 가치를 생성하는데 집중할 수 있습니다.

자연어 이해(NLU; Natural Language Understanding)는 콘텐츠의 배경을 이해할 수 있도록 하는 자연어 처리(NLP)의 한 종류입니다. 개체 및 사실 추출, 범주화, 검색 및 요약과 같은 NLP 기능은 AI 애플리케이션 개발의 핵심이라고 할 수 있습니다.

- 자동화된 파싱, 토큰화, 품사 태깅 및 레마 처리(Lemmatization)
- 시작 및 정지 리스트 적용 가능
- 정밀도나 재현율, 추상화 향상을 위한 품사 태깅, 토큰화, 레마 처리를 활용하는 언어 규칙의 특수 태그, 한정자, 연산자 지원

자동화된 머신 생성 주제 감지

SAS Visual Text Analytics는 개방형 소스 커뮤니티에서 일반적으로 사용되는 특이값 분해(Singular Value Decomposition)와 잠재 디리클레 할당(Latent Dirichlet Allocation)이라는 두 가지 머신 러닝 기법을 사용하여 문서에서 화제를 자동으로 파악합니다.

이러한 방식을 이용하여 텍스트의 주제를 신속하게 파악하고 이러한 Natural Grouping을 구성하는 용어와 문서를 조사합니다. 미처 생각하지 못했던 주제들을 발견하고 이를 새로운 데이터와 비교하여 모델을 배포하거나, 주제 전문가가 직접 변경 또는 확장할 수 있는 로직으로 주제를 변환합니다.

- 자동 토픽 탐색 기능을 통해 공통 주제를 기반으로 주제가 없거나 하나 이상의 주제가 포함된 문서들을 그룹화
- 각 주제에 따른 문서의 적합성을 나타내는 연관성 스코어 및 임계값 초과 주제 멤버십을 나타내는 바이너리 플래그 생성

다양한 언어 지원

SAS Visual Text Analytics는 다수의 언어 전문가가 사전 및 언어 능력을 통해 다양한 언어를 지원하여 기업의 글로벌 과제 해결을 지원합니다. 언어별 품사 태깅 및 개체명의 표준화는 다양한 언어에 텍스트 분석을 구현하려는 기업에게 매우 중요한 기능입니다.

SAS Visual Text Analytics의 즉시 사용 가능한 분석 기능은 아라비아어, 중국어, 크로아티아어, 체코어, 덴마크어, 네덜란드어, 영어, 페르시아어, 핀란드어, 프랑스어, 독일어, 그리스어, 히브리어, 힌디어, 인도네시아어, 이탈리아어, 일본어, 한국어, 노르웨이어, 폴란드어, 포르투갈어, 러시아어, 슬로바키아어, 슬로베니아어, 스페인어, 스웨덴어, 타갈로그어, 터키어, 태국어 그리고 베트남어 등 30개 언어를 지원합니다. 이 언어 팩을 사용하면 분석을 수행하기 전에 번역할 필요 없이 자연어 분석이 가능합니다.

The image displays the SAS Visual Text Analytics interface. On the left, a table lists terms with columns for Term, Part of Speech, Documents, and Frequency. Below this, a 'Documents' section shows text snippets with associated scores and room types. On the right, a vertical flowchart illustrates the analysis pipeline: Data -> Concepts -> Text Parsing -> Sentiment -> Topics -> Categories. Each step in the flowchart is represented by a colored box with a checkmark, indicating successful completion.

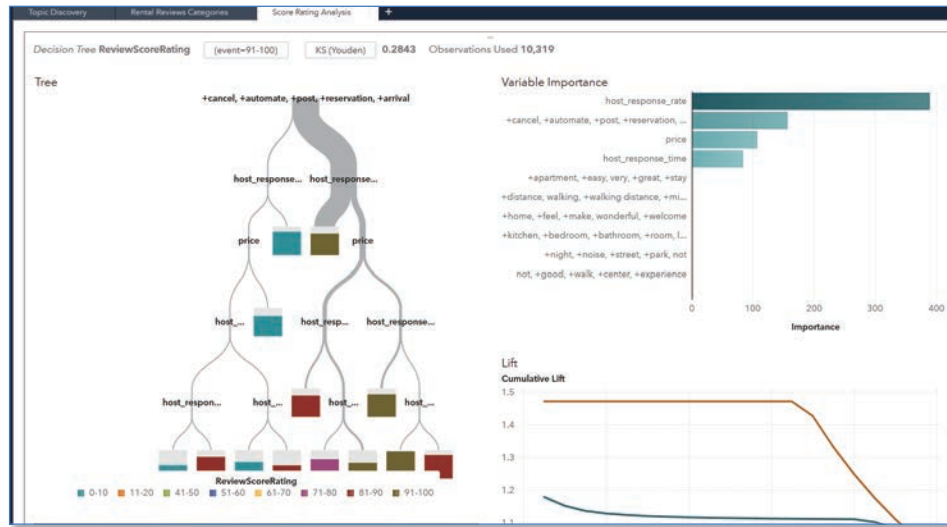
강력한 자연어 처리(NLP) 기능을 통해 텍스트 내 중요 패턴을 쉽게 식별 및 추출할 수 있습니다.

- 즉시 사용 가능한 30가지 언어에 대한 텍스트 분석
- 애플리케이션이 지원하는 각 언어에 대한 기본 정지 리스트
- 품사 정보 및 사전 기반 확장을 포함한 내장 어휘를 통해 표현형을 감지하여 원형(동사 활용, 복수 등)으로 분해

감성 분석

감성 분석은 텍스트에 표현된 저자의 어투나 태도(긍정, 부정 또는 중립)를 파악합니다. SAS Visual Text Analytics는 감성을 내포한 용어, 구문 및 문자열을 식별하고 분석하기 때문에 누군가의 경험에 감정적 컨텍스트가 연결되어 있을 경우 주제를 쉽게 파악할 수 있습니다.

이를 통해 단순히 특정 키워드에 대한 결과를 도출하는 것을 넘어, 사용자 고유의 비즈니스 애플리케이션과 관련성이 높은 트렌드를 추출할 수 있습니다. 지속적으로 감성을 추적하여 어떻게 변하는지를 확인합니다. 고객, 직원, 시민 등으로부터 얻은 피드백을 보다 효과적으로 파악 및 범주화하고, 이를 바탕으로 의사결정을 지원합니다.



텍스트 아웃풋의 자동 시각화를 통해 손쉽게 결과를 탐색합니다.

- 문서 및 화제에 관한 감성 인디케이터 디스플레이를 통해 문서에 기재된 감성을 시각적으로 표현
- 14개 언어에 대한 도메인의 영향을 받지 않는 기본 감성 분석 및 분류(아랍어, 중국어(간체), 중국어(번체), 네덜란드어, 영어, 페르시아어, 프랑스어, 독일어, 이탈리아어, 일본어, 한국어, 포르투갈어, 스페인어, 터키어)
- SAS Sentiment Analysis에 구축된 사용자 정의 감성 모델 가져오기 및 실행

더 자세한 내용은 sas.com/korea/vta에서 확인하실 수 있습니다.

