

# SVM 사용사례 조사

## SVM 이란?

서포트 벡터 머신 (SVM: Support Vector Machine) 알고리즘이란

데이터가 어느 카테고리에 속할지 판단하기 위해 가장 적절한 경계를 찾는 선형 모델

출처 :

서포트 벡터 머신, 가보자고! 🙌 : "디지털"한 일잘러 되는 비법  
뉴스레터 콘텐츠로 바로가기! 📧 오늘도 어김없이 돌아온 <노코드 머신러닝의 이해> 4번째 시간! 어려운 프로그래밍 대신 평소에 자주 사용하는 엑셀로 머신러닝을 이해해 볼 텐데요! 오늘은 데이터가 어느 카테  
📄 <https://ablearn.kr/newsletter/?idx=13552313&bmode=view>



▼ 좀 더 수학적으로 알아본다면 다음 참조

문과생도 이해하는 SVM(support vector machine)  
※본 글은 한양대학교 AI\_X\_DL 수업 기말과제로 작성된 글입니다.  
다. 유튜브 소개 영상 <https://www.youtube.com/watch?v=ymroOWbhlQM&t=22s> 멤버김민지 기계공학부  
📄 <https://binggre.tistory.com/1>



## SVM의 주요 특징

### 1. 마진 최대화 (Margin Maximization):

- SVM의 핵심 아이디어는 데이터 포인트들 간의 간격(마진)을 최대화하는 결정 경계(분류 초평면)를 찾는 것입니다.
- 마진은 결정 경계와 가장 가까운 데이터 포인트 사이의 거리입니다. 이 마진을 최대화하려면 결정 경계에 가장 가까운 데이터 포인트를 "서포트 벡터 (support

vector)"라고 부릅니다.

## 2. 서포트 벡터 (Support Vectors):

- 서포트 벡터는 결정 경계에 가장 가까운 데이터 포인트로, 다른 클래스에 속하는 데이터 포인트와 결정 경계 사이의 마진을 결정하는 데 중요한 역할을 합니다.
- SVM은 이러한 서포트 벡터에만 관심을 두고, 나머지 데이터 포인트는 무시합니다. 따라서 큰 데이터셋에서도 잘 동작할 수 있습니다.

## 3. 비선형 문제 해결 (Non-Linear Problems):

- SVM은 선형 및 비선형 분류 문제를 해결할 수 있습니다. 선형 분리 가능한 문제에서는 선형 결정 경계를 사용하고, 비선형 문제에서는 커널 트릭 (kernel trick)을 사용하여 데이터를 고차원 공간으로 매핑하고 선형 결정 경계를 찾습니다.

## 4. 커널 트릭 (Kernel Trick):

- 커널 트릭은 비선형 문제를 해결하기 위해 사용되는 기술로, 데이터를 고차원 공간으로 변환하여 선형 분리 가능한 문제로 변환합니다.
- 일반적으로 사용되는 커널 함수에는 선형 커널, 다항식 커널, 가우시안 RBF (Radial Basis Function) 커널 등이 있습니다.

## 5. C 파라미터와 규제 (C Parameter and Regularization):

- SVM은 모델의 복잡성을 제어하기 위해 C 파라미터를 사용합니다. C 파라미터는 오분류된 데이터 포인트와 마진 간의 균형을 조절하며, 작은 C 값은 모델을 단순화하고 큰 C 값은 모델을 복잡하게 만듭니다.

## 6. 멀티클래스 분류 (Multi-Class Classification):

- SVM은 이진 분류 (binary classification)를 기본으로 하지만, 멀티클래스 분류 문제를 해결하기 위해 여러 접근 방법을 사용할 수 있습니다. 일반적으로 일대다 (One-vs-Rest 또는 One-vs-All) 방법을 사용하여 다중 클래스를 처리합니다.

## 7. 장점과 단점:

- 장점: 효과적인 분류 성능, 서포트 벡터를 통한 과적합 방지, 비선형 분류 가능, 다양한 커널 함수 사용 가능
- 단점: 데이터 전처리 필요, 파라미터 조정이 필요, 대용량 데이터셋에 대한 효율성 문제

# 실제 사례

### <프로 서치>

- 프로서치는 (주)프로텐의 **검색엔진**으로 국가기록원 표준기록관리시스템 호환성 검증을 통과함
- 실시간/특정시간 색인작업 설정과 이관된 기록물 정보 수정/삭제 시 색인 정보 자동 삭제 기능을 적용함
- 100만건 이상의 대량 파일에 대한 빠른 색인 성능을 보여줌

### <프로서치에서의 SVM(Support Vector Machine) 활용>

- SVM 기반 형태소 분석기로 정확한 한글 검색을 제공함

### <SVM의 적용을 위한 방법>

#### 1. 데이터 수집

- '프로서치'의 성격(검색엔진)에 알맞은 다양한 데이터 수집(SNS, 웹 문서, 뉴스 등 대량의 텍스트 데이터)

- 형태소 분석의 정확도를 높이기 위해 데이터에 정확한 형태소 태깅이 필요함  
전처리

#### 2. 데이터 정제: 노이즈나 불필요한 정보를 제거하고, 텍스트를 일관된 형태로 변환

- Feature 추출: 형태소 분석의 성능을 높이기 위해 적절한 feature를 추출함(ex. 앞뒤 문자정보, n-gram, 형태소 패턴)

#### 3. 모델 훈련

#### 4. 평가 및 튜닝

출처 :

[시장동향] 빅데이터·인공지능·클라우드 통해 진화 중인 '검색엔진'

[컴퓨터월드] 우리는 정보 홍수 시대에 살고 있다. 정보가 기하급수적으로 늘어나면서 필요한 정보 또한 이에 비례해 늘어나고 있다. 그러나 필요한 정보를 찾아내는 것은 쉬운 일이 아니다. 빅데이터 분석을 위한 검색 기술

 <https://www.comworld.co.kr/news/articleView.html?idxno=50744>



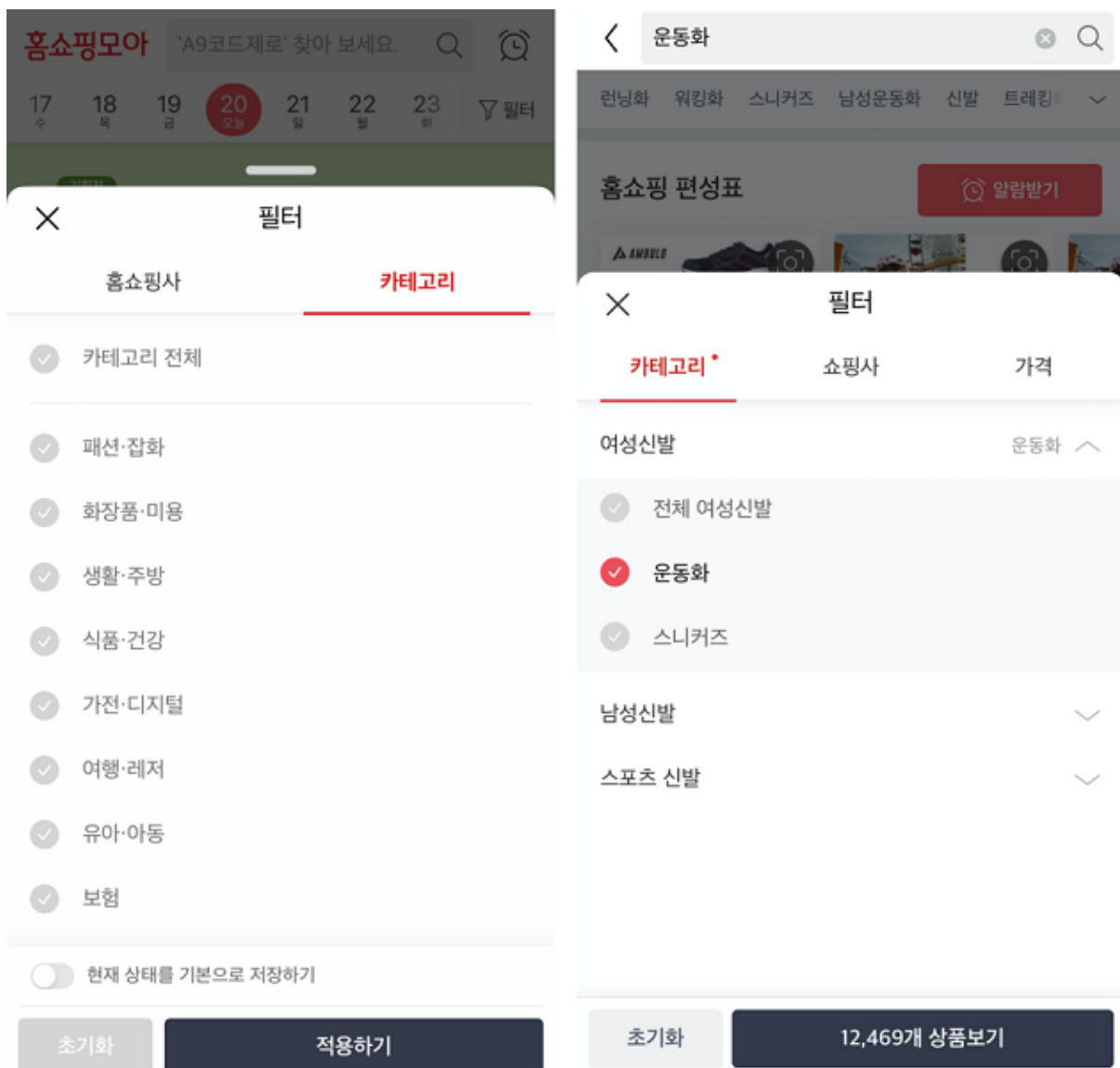
## <홈쇼핑모아 >

- 모바일 홈쇼핑 포털 앱

-1억 개 이상의 상품이 존재

-상품 카테고리 정보 활용해 사용자가 원하는 상품들만 확인할 수 있게 만들 :

카테고리 정보는 직접적으로 '편성표', '검색 결과' 페이지에서 필터 기능을 제공하여 직접적으로 필터 형태 노출 하는데 쓰일 뿐만 아니라 검색 과정에서 사용자가 입력한 질의어와 상품 사이의 연관도를 계산하는 과정, 사용자가 입력한 질의어의 의미를 분석하는 과정, 상품의 브랜드 정보를 추출하는 과정, 사용자가 좋아할 만한 상품을 추천해 주는 과정 등에 사용



## <홈쇼핑모아에서의 SVM(Support Vector Machine) 활용>

-카테고리 분류시 크게 상품명기반(텍스트정보), 이미지기반(이미지정보)이라는 두개의 정보가 활용되는데 딥러닝을 활용한 자연어처리 모델에 대한 많은 리서치가 진행되기 전에는 SVM(Support Vector Machine)을 초기 모델로 활용하여 카테고리 분류를 수행

### <SVM의 적용을 위한 방법>


- 1.데이터 수집
- 2.데이터 정제
  - 텍스트 정보는 버즈니에서 자체 개발한 쇼핑 데이터에 최적화된 상품명 토큰라이저를 포함해 공개돼 있는 여러 토큰라이저를 이용
  - 토큰라이저를 활용하여 생성된 텍스트 단어들을 TF-IDF 방법을 이용해 벡터화시킴
- 3.모델훈련
  - 벡터화된 상품명 정보를 SVM의 입력으로 하여 카테고리 분류를 수행
- 4.평가 및 튜닝

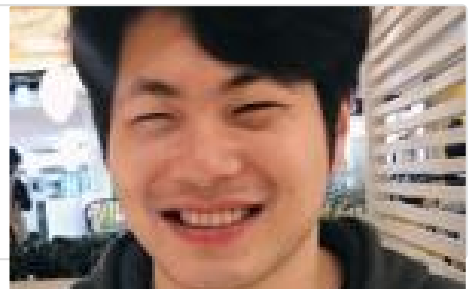
출처 :

[기고] 홈쇼핑모아의 상품 카테고리 자동 분류 시스템

[컴퓨터월드]

개요인터넷 쇼핑의 세계에는 셀 수 없을 만큼 많은 수의 상품이 존재한다. 때문에 인터넷 쇼핑을 운영하는 회사에서는 어떻게 하면 좀 더 호

 <https://www.comworld.co.kr/news/articleView.html?idxno=49901>



### <Saltlux의 INSIGHT STUDIO>

- 대량의 비정형 데이터로부터 필요한 정보들을 실시간으로 분석, 요약하여 제공

### <Insight studio에서의 SVM(Support Vector Machine) 활용>

-고품질의 형태소 분석 : Structural-SVM에 기반한 기계학습 알고리즘을 적용하여 다국어 문장을 기계가 이해할 수 있는 수준의 어절 단위로 나누는 기술로 고정밀 품사 태깅을 포함하는 형태소분석 기능을 내장하고 있음

### <SVM의 적용을 위한 방법>

1. 데이터 수집

-대량의 비정형 데이터 획득

## 2. 데이터 정제

-말뭉치에 의미역을 태깅하여 의미역 결정 학습데이터 생성

## 3. 모델 훈련

-Structural SVM 알고리즘을 이용

## 4. 평가 및 튜닝

출처 :

대한민국 인공지능 여기까지, 솔트룩스



챗GPT 같은 초거대 언어모델 생성, 생성 AI, 생성적 AI 기술, 대화형  
챗봇, 국내최다 AI특허, 인공지능, 인공지능·데이터과학 전문기업

◆ <https://www.saltlux.com/product/insight/>

[https://web.archive.org/web/20180424075949id\\_/http://kiise.or.kr/e\\_journal/2015/2/JOK/pdf/10.pdf](https://web.archive.org/web/20180424075949id_/http://kiise.or.kr/e_journal/2015/2/JOK/pdf/10.pdf)