

---

저자 (Authors)	임진욱, 유동균, 박혜진
출처 (Source)	<a href="#">대한산업공학회 추계학술대회 논문집</a> , 2009.10, 825-832(8 pages)
발행처 (Publisher)	<a href="#">대한산업공학회</a> Korean Institute Of Industrial Engineers
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE01987469">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE01987469</a>
APA Style	임진욱, 유동균, 박혜진 (2009). 프로야구 승/패 예측. 대한산업공학회 추계학술대회 논문집, 825-832
이용정보 (Accessed)	이화여자대학교 203.255.***.68 2020/08/29 10:54 (KST)

---

#### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

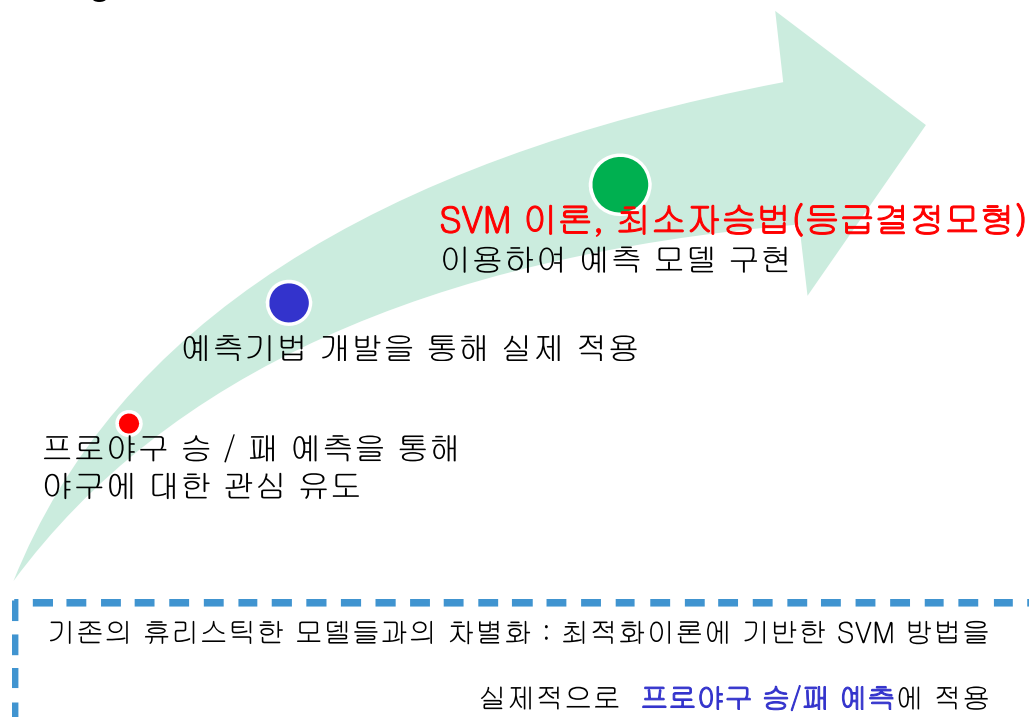
#### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

## 제 5회 대한산업공학회 대학생 프로젝트 경진대회 참가 지원서

작품 제목	프로야구 승 / 패 예측 : SVM과 최소자승법(등급결정모형) 적용				
대표저자	임진욱	공동저자	유동균, 박혜진		
소속	홍익대학교 정보·컴퓨터공학부(학과) 정보산업공학 전공				
대표저자 연락처	Tel.		핸드폰		E-mail jinugine@naver.com
내용 요약	1. 프로야구 경기 예측을 위해 2가지 모형을 적용하는 시스템 개발 ① 최적화 기법에 바탕을 둔 분류기법의 하나인 SVM모형 적용 ② 최소자승법에 의한 rating 결정 모형인 등급결정모형 적용 2. 향후 더 많은 데이터를 수집하여 예측률을 높여서 KBO홈페이지에 게시함으로서 대중의 관심을 유발시킬 수 있을 것으로 기대됨				
해당과목 구분	졸업 프로젝트, 과목명(시스템 분석, 시스템 설계)				
<b>지도교수 추천서</b>  본인은 2009년 대한산업공학 추계학술대회의 대학생 프로젝트 경진대회에 상기 지원자의 작품을 추천함과 더불어, 상기 작품이 “학생 주도”의 작업임을 확인합니다.  소속 : 홍익대학교    지도교수 이름 : 박구현    전화 : 02-320-1629 e-mail : <a href="mailto:khpark@hongik.ac.kr">khpark@hongik.ac.kr</a>					

## Project 목적



# Project 개요

## 1) SVM 모형

- 최적화 기법에 바탕을 둔 분류기법의 하나로 이는 프로야구 승패 예측에 적용함(각 팀 당 70경기에서 얻어진 데이터를 바탕으로 215경기 예측)

## 2) 등급결정 모형

- 최소자승법에 의한 rating결정 모형으로 이를 이용하여 프로야구 승패 예측에 적용함(106경기에서 얻어진 등급을 바탕으로 145경기 결과 예측)

# SVM(Support Vector machines)

## (1) SVM을 프로야구 승/패 예측에 적용한 이유

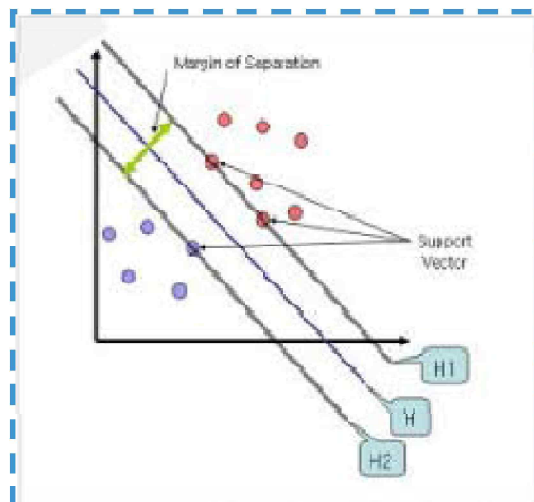
데이터 마이닝(data mining) 분류기법 중 최적화 기법을 사용한 방법

통계적 학습 이론 기반  
패턴 분류 알고리즘

구조적 리스크 최소화를 통해  
오류를 최소화

분리 초평면  
(Optimal Separating Hyperplane)을 통한  
분류기법

→프로야구 경기 결과 예측에 적용가능  
(기존 휴리스틱 방법, 통계적 예측 방법 등은 있음)



# SVM(Support Vector Machine)

## (2) 분리초평면

### 분리초평면

기존의 프로야구 경기 결과를 학습하여 분리초평면 식을 얻은후 이로부터 경기결과(승/패) 예측하고자 함

$X = (X_1, X_2, X_3, X_4, \dots, X_8)$  : 프로야구팀의 경기력을 나타내는 Data

$W = (W_1, W_2, W_3, W_4, \dots, W_8)$  : 대응되는 분리초평면 계수

$b$  = 분리초평면 상수

# Data 구성 (1)

## (1) 요소

- (예) 기아 타이거즈의 경우 (  $k = \text{'기아 타이거즈'}$  )

- 장타율 ( $X_{k1}$ ) = 안타수 / 타수  
 $\text{안타수} = \text{단타} + (2\text{루타} \times 2) + (3\text{루타} \times 3) + (\text{홈런} \times 4)$
- 타율 ( $X_{k2}$ ) = 안타수 / 타수
- 출루율 ( $X_{k3}$ ) = ( 안타수 + 베이스온볼스 ) /  
 ( 타수 + 베이스온볼스 + 희생플라이 )
- 방어율 ( $X_{k4}$ ) = ( 자책점 / 투구이닝수 )
- 사사구 ( $X_{k5}$ ) = (베이스온볼스 + 4구) / 투구이닝수
- 피안타율 ( $X_{k6}$ ) = 피안타수 / 타수
- 득점 ( $X_{k7}$ ) = 득점 / 타자이닝수
- 달삼진 ( $X_{k8}$ ) = 달삼진 / 투구이닝수

# 2008 프로야구 개요

## 2008 프로야구 팀 수 : 8개팀

KIA 타이거즈, 삼성 라이온즈, SK 와이번스, LG 트윈스,  
두산 베어스, 한화 이글스, 롯데 자이언츠, 히어로즈

각 팀 126경기 : 홈경기 63경기, 원정경기 63경기

총 504 경기 : 홈경기 504경기, 원정경기 504경기

## 데이터 구성 (2)

### 분리초평면

(홈기준)

데이터 = 홈팀 요소 값 - 원정 팀 요소 값

$$X_i = X_{ai} - X_{bi}$$

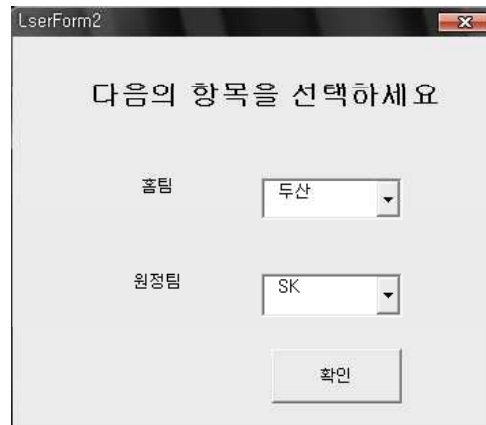
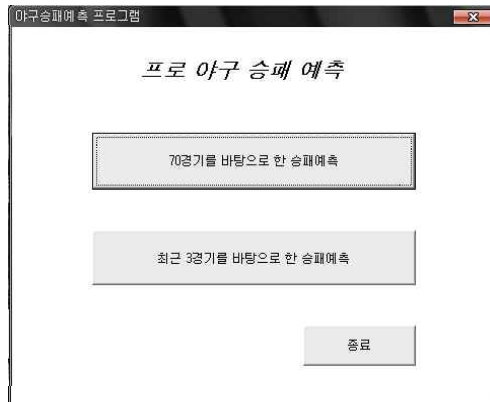
구분	장타 율	타율	출루 율	득 점	방어 율	사사 구	탈삼 진	피안 타
홈	Xa1	Xa2	Xa3	.....				Xa8
원정	Xb1	Xb2	Xb2					Xb8
(데이터) 결과	X1	X2	X3					X8
(변 수) 계수	W1	W2	W3					W8

분리 초평면 계수( $w_i$ )가 최적화모형에 의해 결정되면 분리초평면으로 구분되는  
한쪽구역( $(Y=1)$ 은 승), 다른 쪽 구역( $(Y=-1)$ 은 패)

$$y_i = 1 \text{ or } -1 \text{ (학습데이터 사용된 경기 결과)}, i = 1, \dots, 281$$



# 프로야구 승패예측 시스템 : SVM



## 결론

- 프로야구 경기 예측을 위해 2가지 모델을 적용하는 시스템을 개발하였음
  - 데이터마이닝 기법중 하나인 SVM 기법 적용: 59%적중률
  - 등급결정(최소자승법 이용) : 57.9%적중률
- 각 팀의 경기력을 8가지 요소로 수치화 하였음
  - 장타율, 타율, 홈런율, 득점, 방어율, 사사구, 탈삼진, 피안타
  - 281경기에 대한 실제 데이터
- 향후 개선점 및 기대효과
  - 더 많은 데이터를 수집하여 예측률을 높이는 작업이 필요
  - 경기 예측을 KBO홈페이지에 게시 함으로써 대중의 더욱 많은 관심 유발



## 참고문헌

한글 엑셀 97 사용자를 위한 VBA 매크로 길들이기 ,  
정성수, 21세기사, 1998

10일 만에 배우는 엑셀 2002 VBA, VB Tech Lab &  
Haruka Seto, (주)영진닷컴, 2002

엑셀활용 경영과학, 박구현 외, 교보문고, 2006

VBA for modelers : Developing Decision Support Systems with  
Microsoft Excel, S.C. Albright, Duxbury, 2006