

# Introduction to Probabilistic Machine Learning

Real-World Applications

Ralf Herbrich

# Overview

---

1. adPredictor: Bayesian Probit in e-Commerce
2. MatchBox: Bayesian Recommendation Systems
3. The Path of Go: Bayesian Pattern Ranking for Games

**Introduction to  
Probabilistic Machine  
Learning**

*Unit 12 – Real-World  
Applications*

# Overview

---

1. **adPredictor: Bayesian Probit in e-Commerce**
2. MatchBox: Bayesian Recommendation Systems
3. The Path of Go: Bayesian Pattern Ranking for Games

**Introduction to  
Probabilistic Machine  
Learning**

*Unit 12 – Real-World  
Applications*

# Predicting Clickthrough-Rate (CTR)

- **Display logic:**

$$b_1 \cdot p_1 \geq b_2 \cdot p_2 \geq \dots$$

- **Charge logic:**

$$c_i = b_{i+1} \cdot \frac{p_{i+1}}{p_i}$$

- **Advantages of improved probability estimates:**

- Increase user satisfaction by better targeting
- Fairer charges to advertisers
- Increase revenue by showing ads with high click-thru rate

The screenshot shows a Google search for 'potsdam vacation'. The search bar contains the query 'potsdam vacation'. Below the search bar, there are tabs for Images, News, Itinerary, Reviews, Maps, Videos, Books, Flights, and Finance. The results show 'About 31,000,000 results (0.82 seconds)'. The first section is 'Sponsored' and lists several ads. Each ad has a bid amount, a click-through rate (CTR), and a final charge. The ads are: 'Potsdam Holiday - Browse Our Top Holidays' (bid \$1.00, CTR \*10%, charge \$0.10), 'Hotels in Potsdam - Rundum sorglos buchen' (bid \$2.00, CTR \*4%, charge \$0.08), 'Kurzurlaub in Potsdam - INSELHOTEL Potsdam' (bid \$0.10, CTR \*50%, charge \$0.05), and 'CHECK24: Urlaub Potsdam buchen - Urlaub Potsdam zum...' (bid \$0.50, CTR \*4%, charge \$0.02). The second section is 'Germany Travel' and lists 'Potsdam: beautiful gardens and UNESCO World Heritage'.

search query

advertisements  
(paid-for content)

Introduction to  
Probabilistic Machine  
Learning

Unit 12 – Real-World  
Applications

# Bayesian Probit Regression with Factor Graphs

## ■ Learning algorithm speed requirement:

- 5,000-10,000 impression updates / sec = 100-200  $\mu$ s per impression update
- **Decision:** Use one-hot encoding for all raw input data values and features

## ■ Bayesian Probit model:

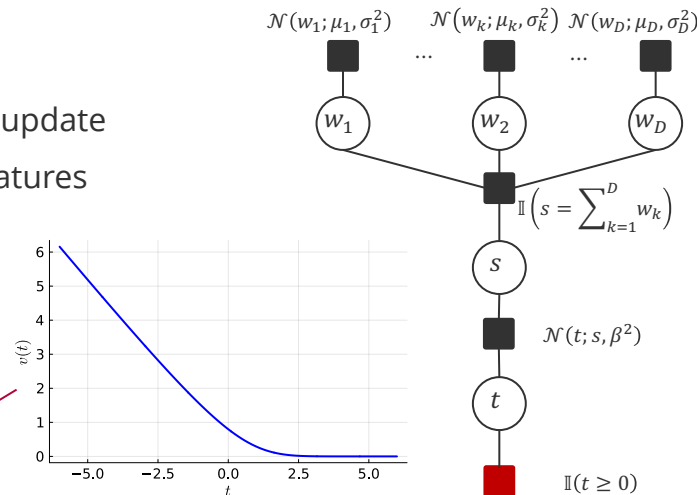
$$p(\mathbf{w}|y=+1) \propto \underbrace{\left[ \int_0^\infty \mathcal{N}\left(t; \sum_{k=1}^D w_k, \beta^2\right) dt \right]}_{\text{likelihood}} \cdot \underbrace{\left[ \prod_{k=1}^D \mathcal{N}(w_k; \mu_k, \sigma_k^2) \right]}_{\text{prior}}$$

**posterior**                      **likelihood**                      **prior**

## ■ Closed form using message passing:

$$\begin{aligned} \mu_k &\leftarrow \mu_k + \frac{\sigma_k^2}{s} \cdot v\left(\frac{\sum_{j=1}^D \mu_j}{s}\right) \\ \sigma_k^2 &\leftarrow \sigma_k^2 \cdot \left(1 - \frac{\sigma_k^2}{s^2} \cdot w\left(\frac{\sum_{j=1}^D \mu_j}{s}\right)\right) \\ s^2 &= \beta^2 + \sum_{j=1}^D \sigma_j^2 \end{aligned}$$

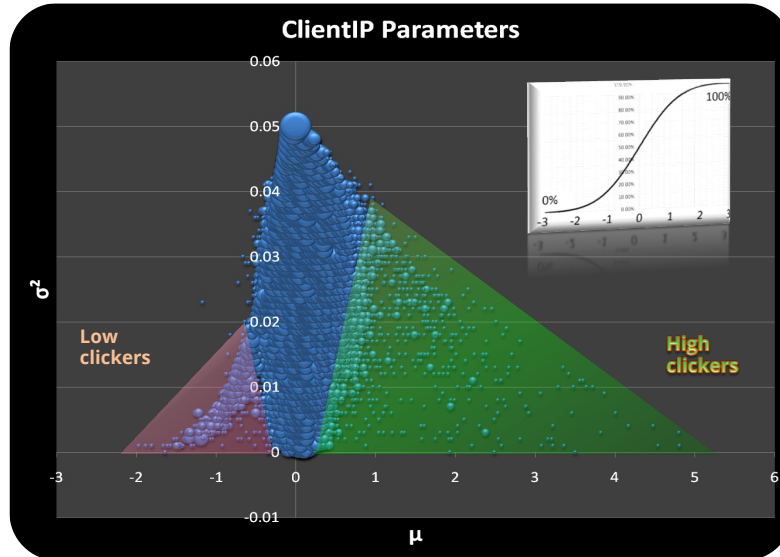
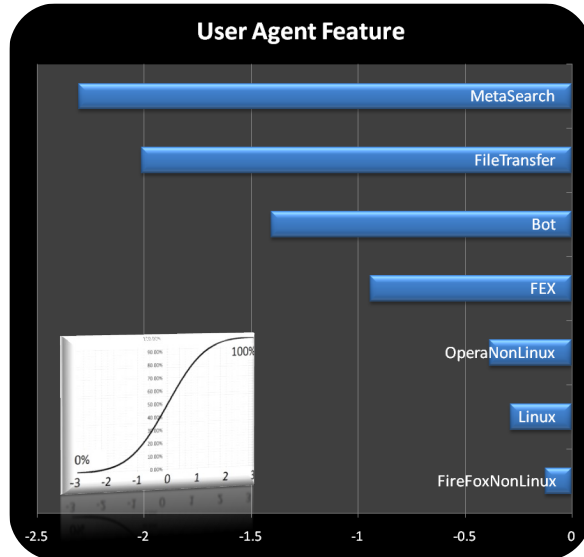
largest for parameter with largest uncertainty so far



Introduction to  
Probabilistic Machine  
Learning

§ 12 - Real-World  
Applications

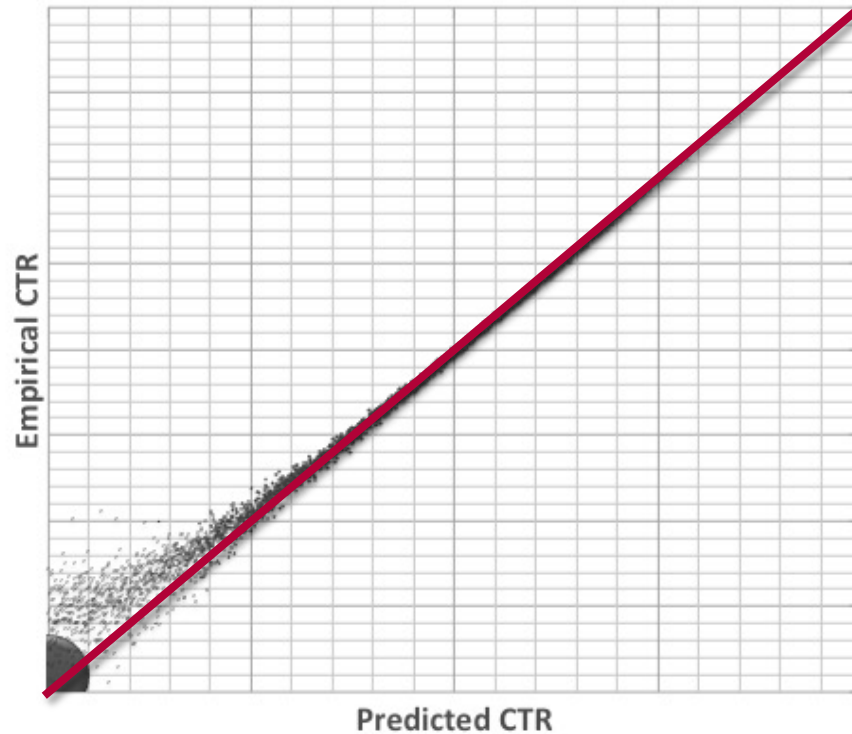
# Empirical Findings (bing.com)



## Introduction to Probabilistic Machine Learning

Unit 12 – Real-World  
Applications

# Accuracy



**Introduction to  
Probabilistic Machine  
Learning**

*Unit 12 – Real-World  
Applications*

# Overview

---

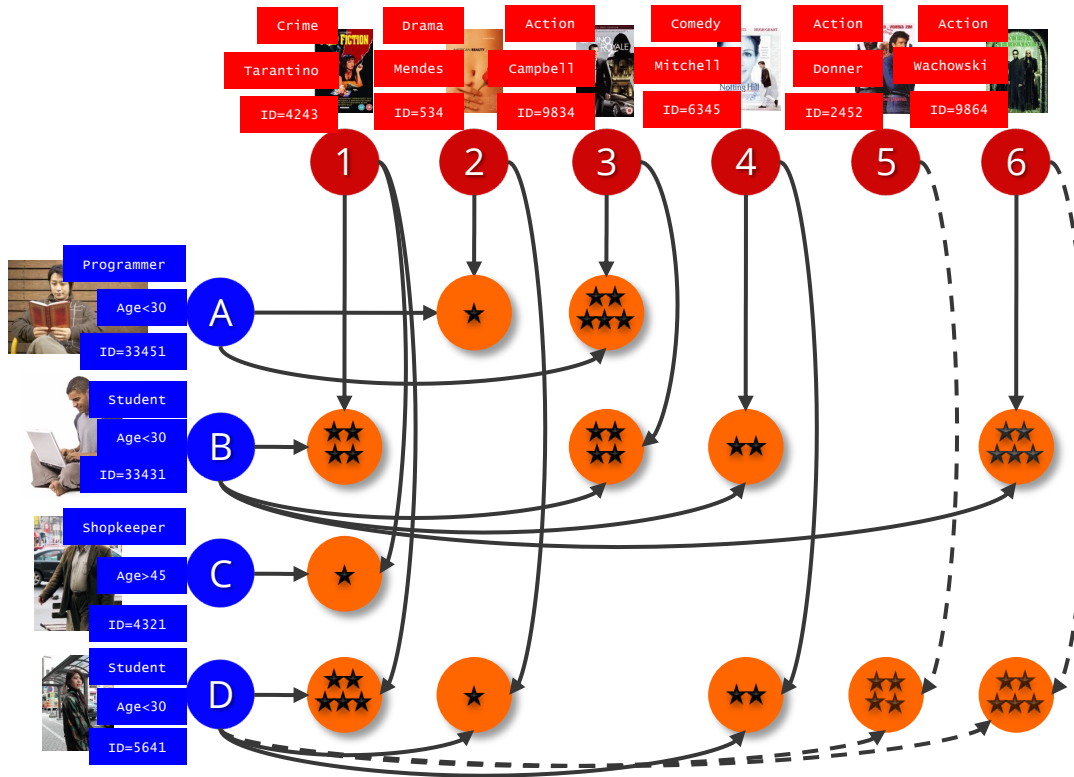
1. adPredictor: Bayesian Probit in e-Commerce
2. **MatchBox: Bayesian Recommendation Systems**
3. The Path of Go: Bayesian Pattern Ranking for Games

**Introduction to  
Probabilistic Machine  
Learning**

*Unit 12 – Real-World  
Applications*



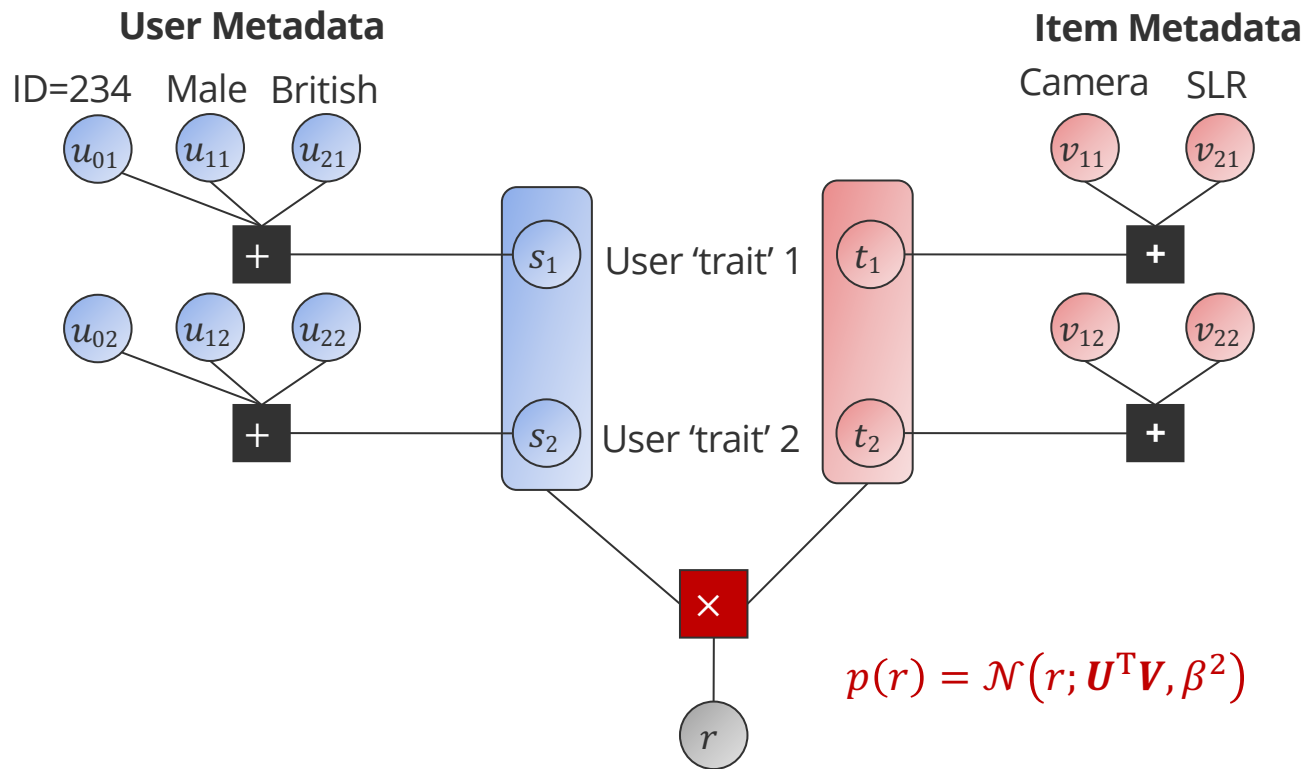
# MatchBox: The Recommendation Setting



**Introduction to  
Probabilistic Machine  
Learning**

*Unit 12 – Real-World  
Applications*

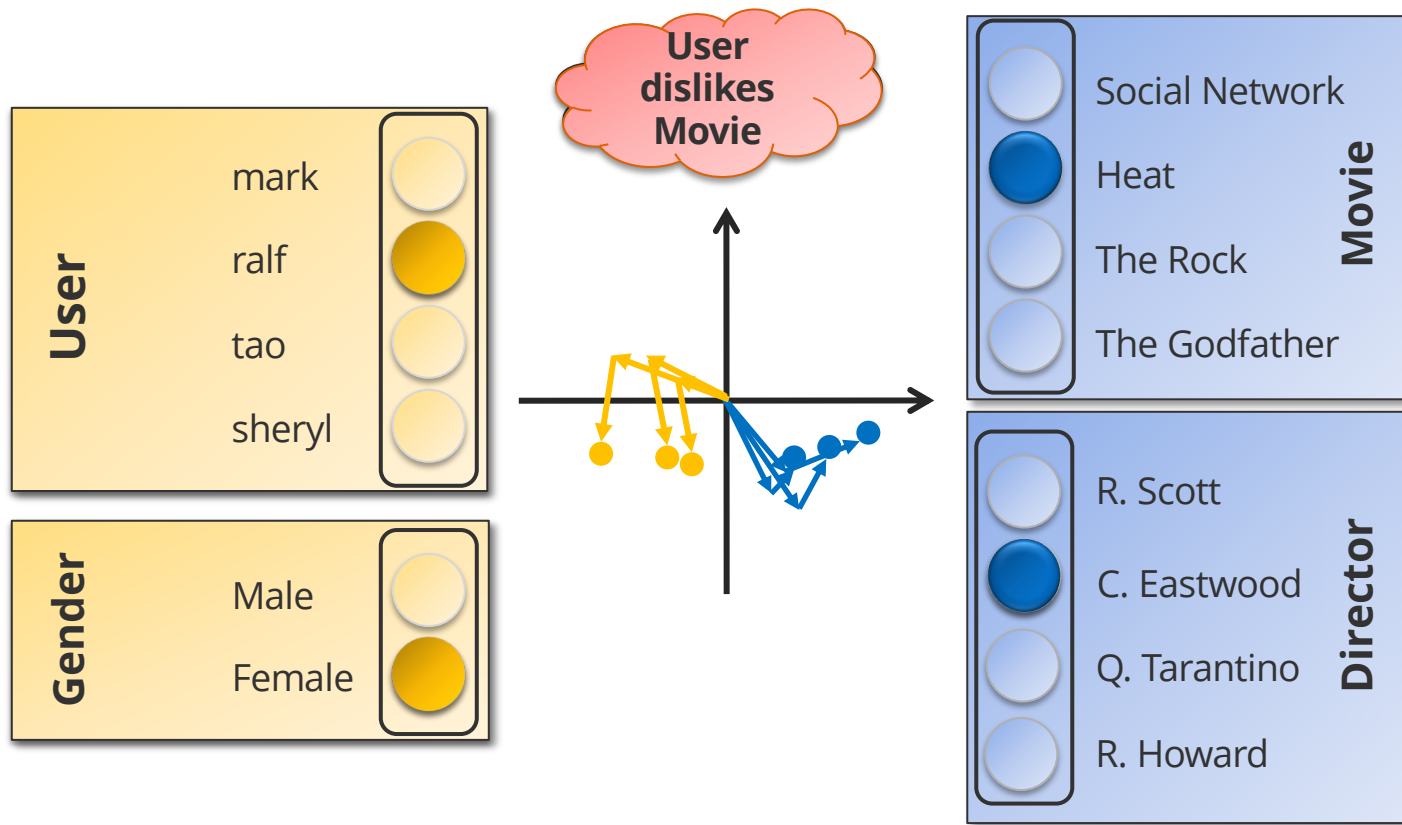
# MatchBox With Metadata



Introduction to  
Probabilistic Machine  
Learning

Unit 12 – Real-World  
Applications

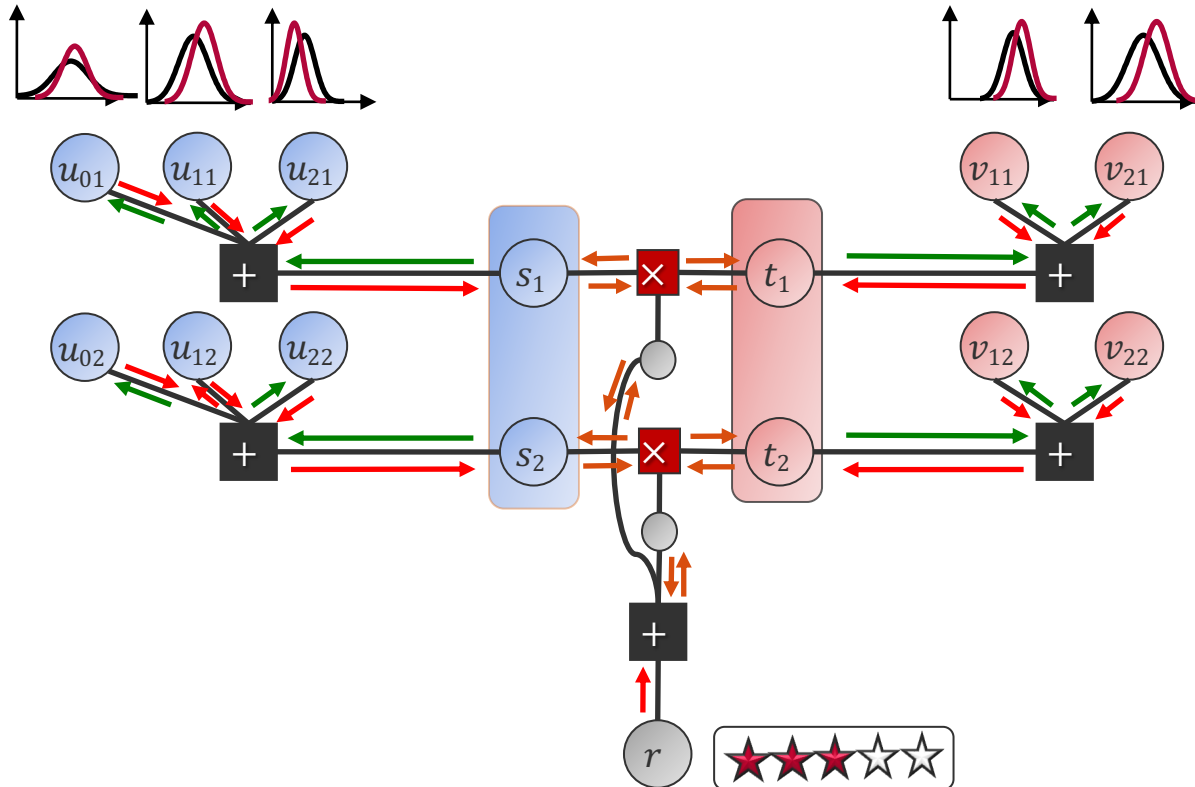
# Recommender System: MatchBox



**Introduction to  
Probabilistic Machine  
Learning**

*Unit 12 – Real-World  
Applications*

# Message Passing For Matchbox



**Introduction to  
Probabilistic Machine  
Learning**

*Unit 12 – Real-World  
Applications*

# MovieLens (1,000,000 Ratings)

6,040 users

User ID		
User Job		User Age
Other	Lawyer	<18
Academic	Programmer	18-25
Artist	Retired	25-34
Admin	Sales	35-44
Student	Scientist	45-49
Customer Service	Self-Employed	50-55
Health Care	Technician	>55
Managerial	Craftsman	User Gender
Farmer	Unemployed	Male
Homemaker	Writer	Female

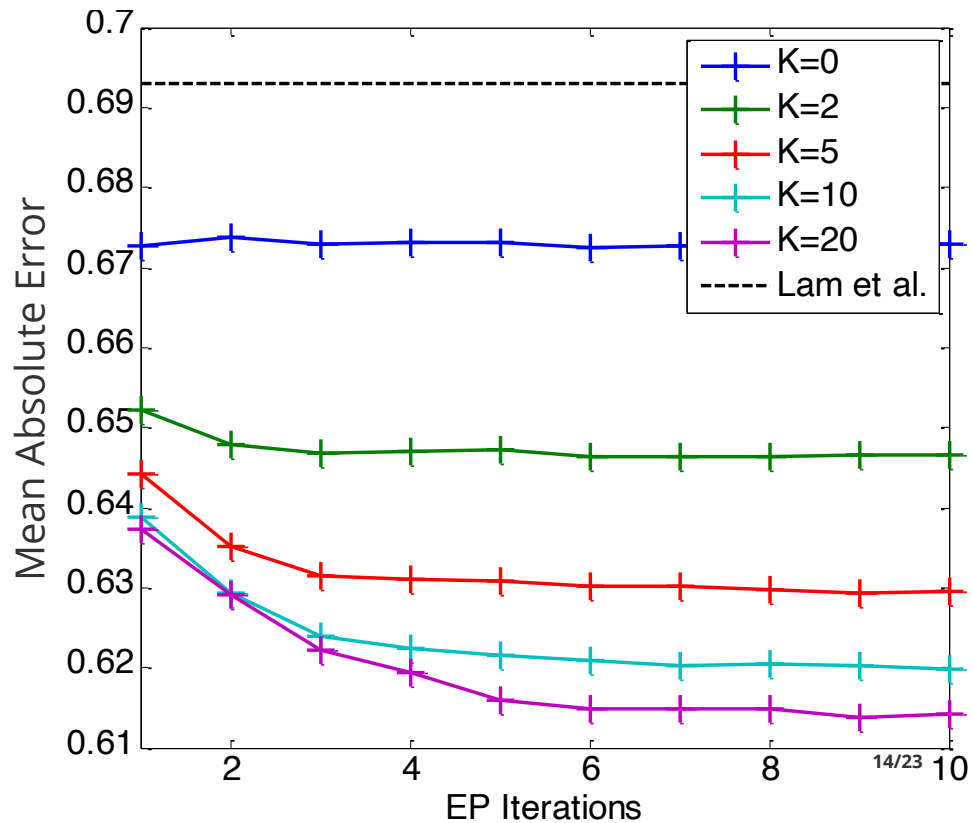
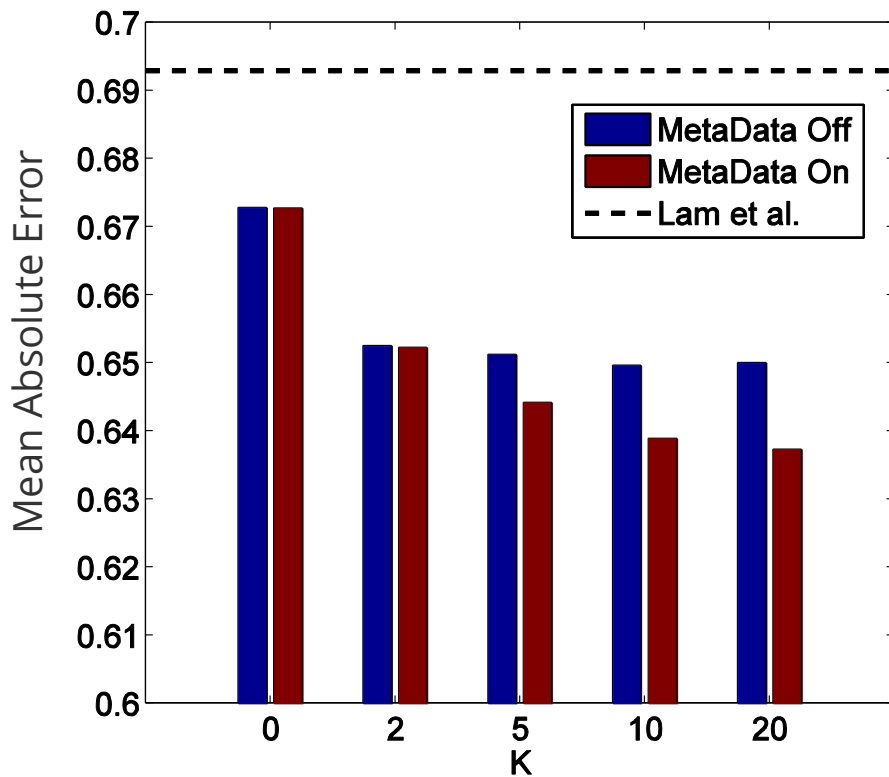
3,900 movies

Movie ID	
Movie Genre	
Action	Horror
Adventure	Musical
Animation	Mystery
Children's	Romance
Comedy	Thriller
Crime	Sci-Fi
Documentary	War
Drama	Western
Fantasy	Film Noir

**Introduction to  
Probabilistic Machine  
Learning**

*Unit 12 – Real-World  
Applications*

# MovieLens with Thresholds Model



# Overview

---

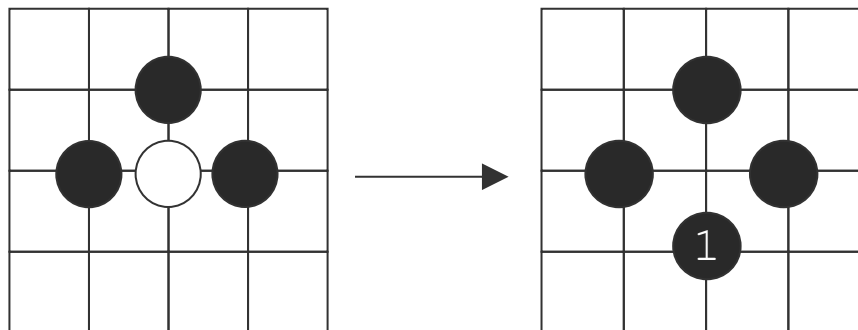
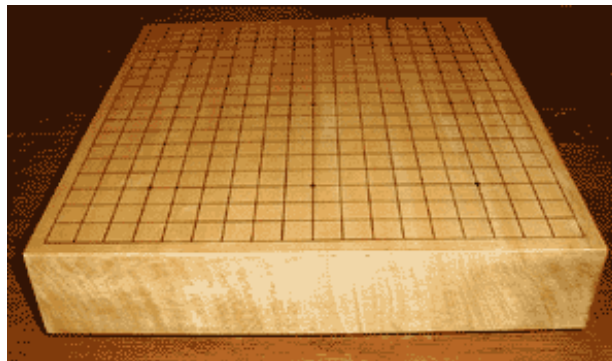
1. adPredictor: Bayesian Probit in e-Commerce
2. MatchBox: Bayesian Recommendation Systems
- 3. The Path of Go: Bayesian Pattern Ranking for Games**

**Introduction to  
Probabilistic Machine  
Learning**

*Unit 12 – Real-World  
Applications*

# The Game of Go

- Started about 4000 years ago in ancient China.
- About 20 million players worldwide.
- **2 Players:** Black and White.
- **Board:** 19×19 grid.
- **Rules:**
  - Turn: One stone placed on vertex.
  - Capture.
- **Aim:** Gather territory by surrounding it.



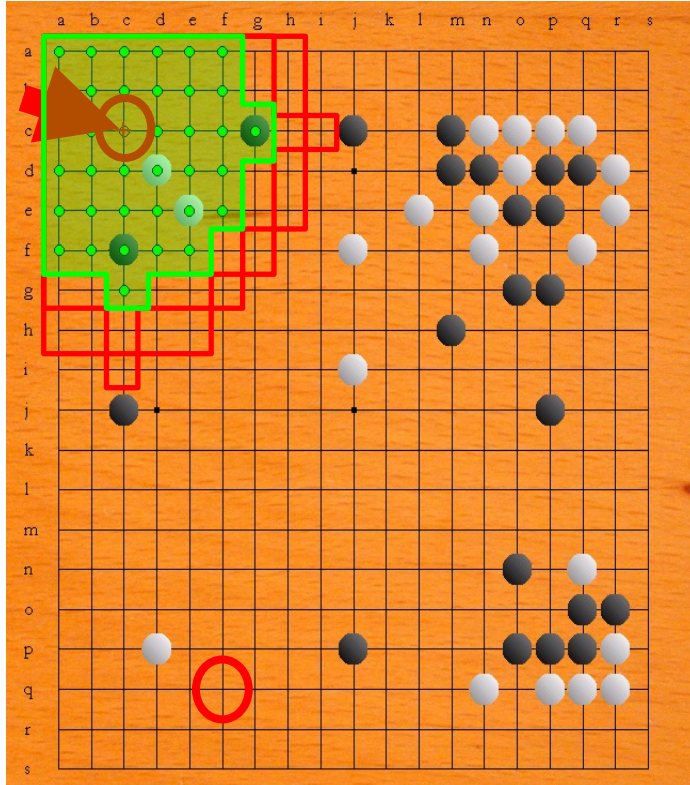
**Introduction to  
Probabilistic Machine  
Learning**

*Unit 12 – Real-World  
Applications*

A stone is **captured** by completely **surrounding** it.



# Moves Selection by Pattern Matching



Pattern Urgency Table

Not in database!

$\mu = 25, \sigma = 5.2$
⋮

Black to move

Introduction to  
Probabilistic Machine  
Learning

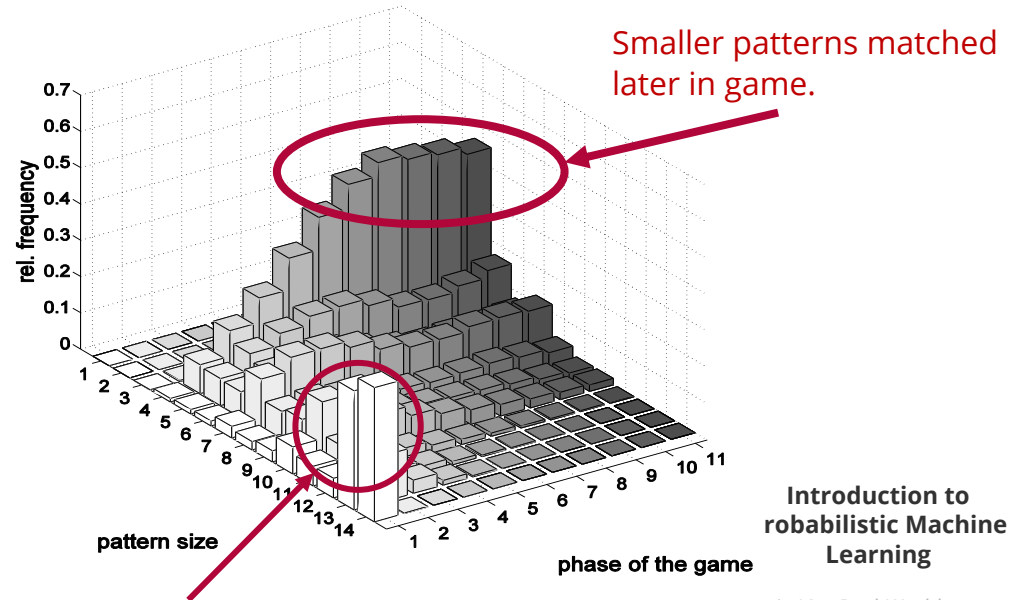
Unit 12 – Real-World  
Applications

## ■ Two processes:

1. Harvesting patterns (180,000 experts games → 600M possible patterns!)
2. Ranking patterns (requires model of moved selection)

## ■ Move selection model as partial ranking using *urgency*

- **Move made:** “Wins” over any other move available on the board indicating that it is most *urgent*.
- **Moves not made.** Nothing can be concluded about the ranking among the available but un-played moves.

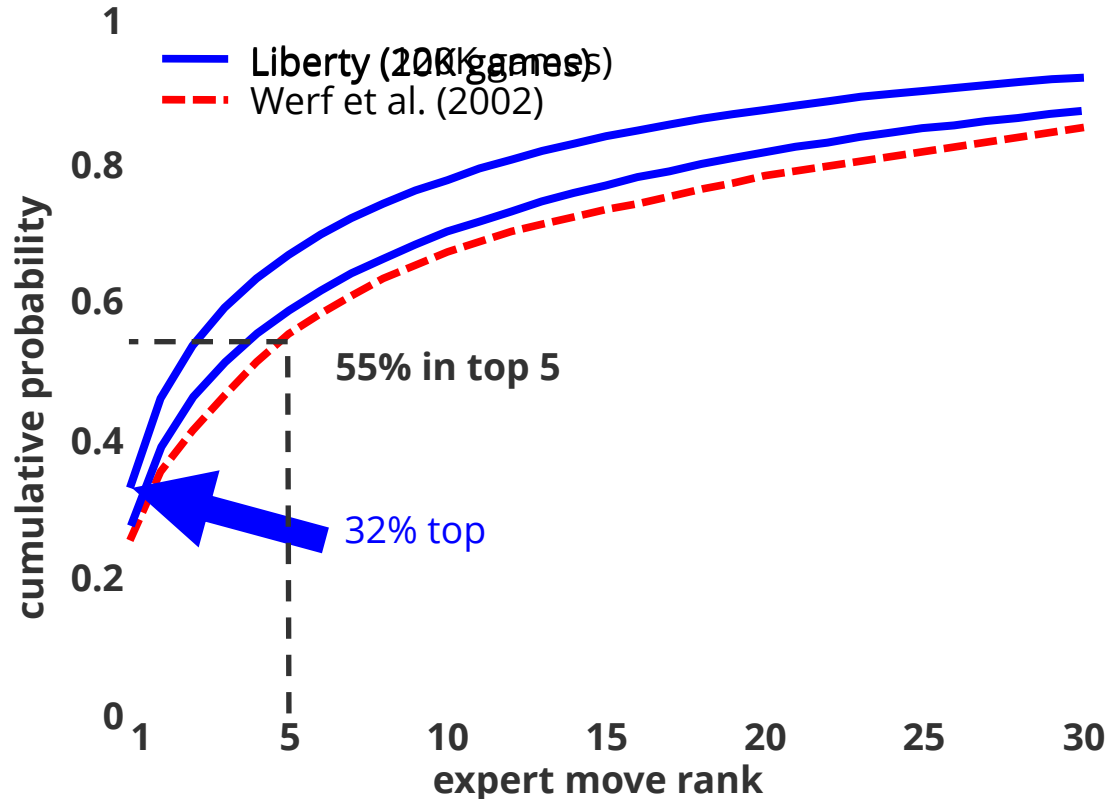


Introduction to  
robabilistic Machine  
Learning

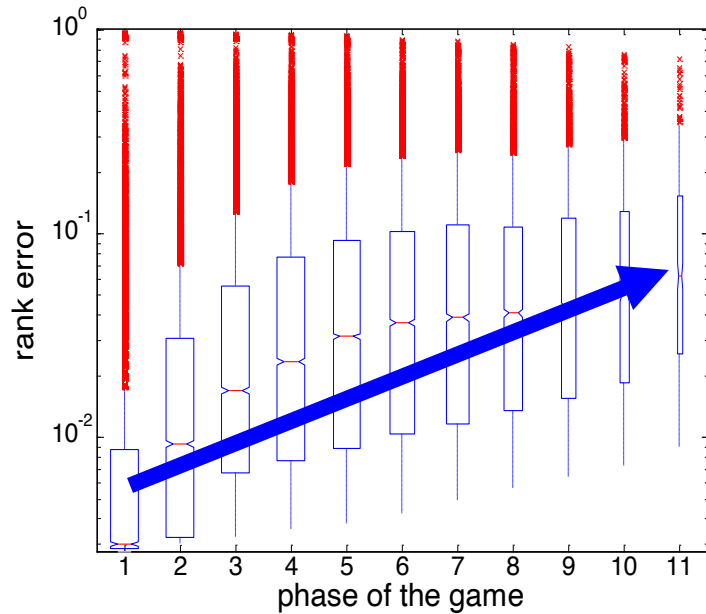
Unit 12 – Real-World  
Applications



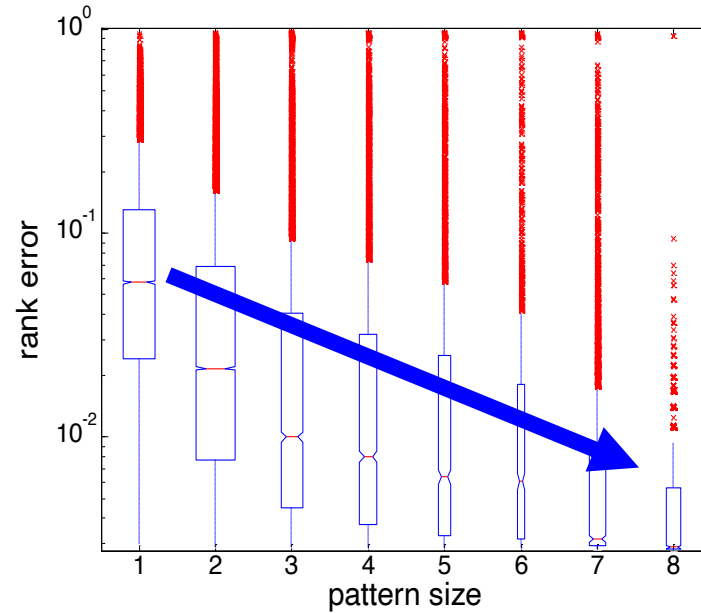
# Experimental Results



## Experimental Results (ctd.)



Error increases towards end of the game



Error decreases for increasing pattern size

**Introduction to  
Probabilistic Machine  
Learning**

*Unit 12 – Real-World  
Applications*

## ■ Real-World Applications

- Bayesian Probit regression using expectation propagation is used in over 200 production systems at Amazon today (and many more places around the world)
- Bayesian Probit regression is like a two-team game where one team is the “threshold-0 player”
- When using the Gaussian product factor for, we can even model bi-linear models used in recommendation → requires a different approximation due to multi-modality
- Partial rankings are another application of the pairwise difference factor
- Used today in AlphaGo!

**Introduction to  
Probabilistic Machine  
Learning**

*Unit 12 – Real-World  
Applications*

Thank You!