Learning and Inference with Statistical Relational Models

UNIST

Jaesik Choi

http://pail.unist.ac.kr/jaesik

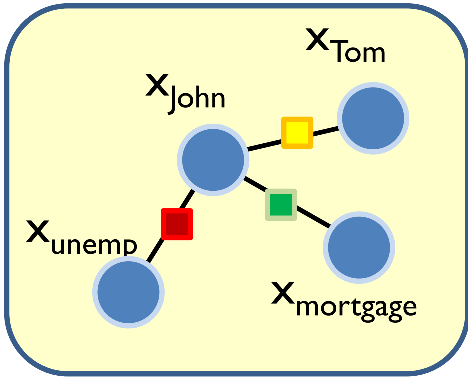
Contents

- **Statistical Relational Learning (확률관계형 학습)**
- Relational Kalman Filtering (관계형 칼만 필터)
- Probabilistic Artificial Intelligence Lab (확률 지능형 인공지능 연구실)

How to Estimate Future Events with Graphical Models?

- Choose a graphical model: e.g.,
  - Bayesian Networks,
  - Markov Random Fields,
  - Kalman Filter

- Collect observations:
  - Tom sold his home at $0.5 million.
  - The mortgage rates increased(↑) to average 5.5%.
  - The unemployment rate downed(↓) to 7%.



- Compute conditional probabilities by relationships:

P(value of John's home | observations)

Estimating Future Events with Large-Scale Graphical Models

- Estimating future events is essential in
  - Financial markets (housing)
  - Environment (extreme weather, groundwater)
  - Energy (smart grid)

housing

weather

energy

Challenges: Large-Scale Models

- Hard to handle large numbers of elements
  - US housing market: 75.56 million house units
  - Hurricane Sandy: spanning 1,100 miles (1,800 km)

- Computational Complexities
  - Kalman filter:
    $$O(n^3) = O(75.56^3 \cdot 10^6 \text{ trillion})$$
  - Dynamic Bayesian Networks and Markov Random Field:
    $$O(\exp^n) = O\left(\exp^{75.56 \text{ million}}\right)$$

Some Elements Share Relationships

- Elements share Relationships
  - If mortgage ↑ 1% → price of Tom's home ↓ 3%
  - If mortgage ↑ 1% → price of John's home ↓ 3%
  - If mortgage ↑ 1% → price of any home in the town ↓ 3%

- Relations over clusters
  - Town = {Tom, John, … }
  - Δ(price of *name*'s home)
    $$= -3\Delta\text{Mortgage} + \varepsilon$$
  - *name* $\in$ Town
  - $\varepsilon \sim N(0, \Sigma)$, Gaussian Noise

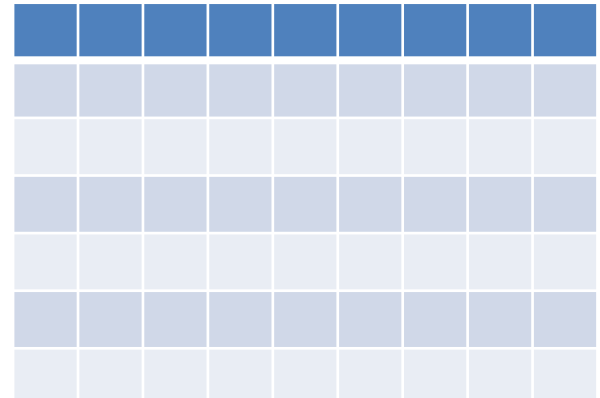
Machine Learning Models for Big Data

- Traditional **statistical** machine learning approaches assume
  - A random sample of homogeneous objects from single relation
  - Independent, identically distributed (IID)
- Traditional **relational** machine learning approaches assume:
  - Logical language for describing structure in sample
  - No noise and no uncertainty
- Real world data sets:
  - Multi-relational and heterogeneous
  - Noisy and uncertainty

[slide courtesy of Lise Getoor]

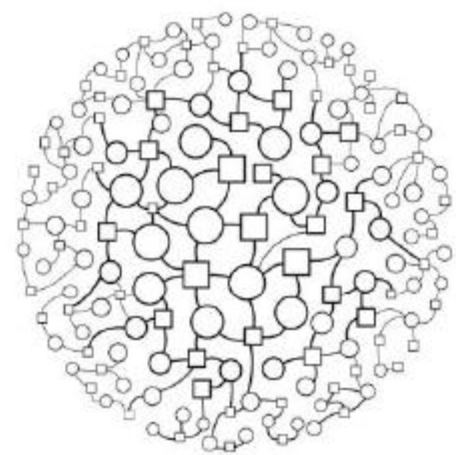
What is the difference?

Most of the data that is available in the newly emerging ear of big data does not look like this



Or even like this

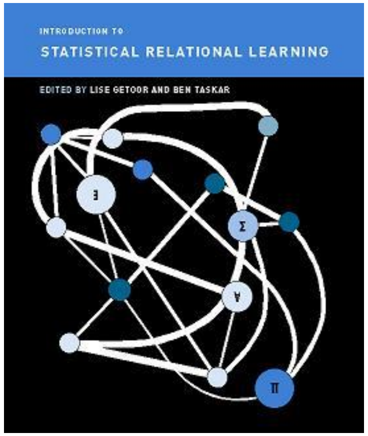
It looks more like this



[slide courtesy of Lise Getoor]

Statistical Relational Learning (SRL)
or Relational Graphical Models

- Collection of techniques which combine rich relational knowledge AI/DB representations with statistical models
  - First-order logic, SQL, graphs,
  - Graphical models, directed, undirected, mixed; relational decision trees, etc.
- Example:
  - Markov Logic Networks (Washington and Texas), Bayesian Logic Programs (Berkeley & MIT), Probabilistic Relational Models (Stanford), Factorie (UMass), **Relational Kalman Filtering (U of Illinois & UNIST)**, and …
- Key ideas
  - Relational feature construction
  - Collective reasoning
  - 'Lifted' representation, inference and learning



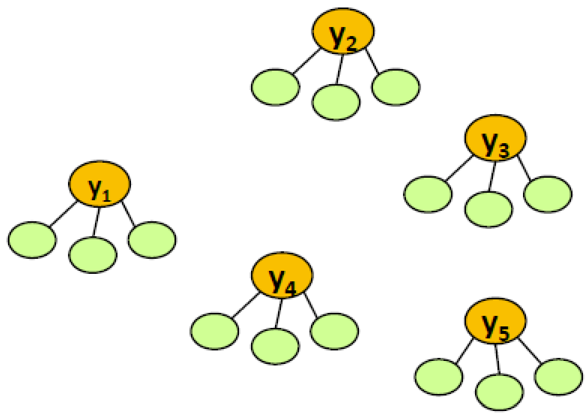
MIT Press

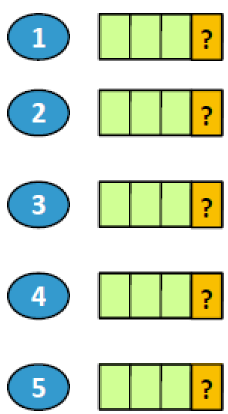
http://en.wikipedia.org/wiki/Statistical_relational_learning

[modified from a slide courtesy of Lise Getoor]

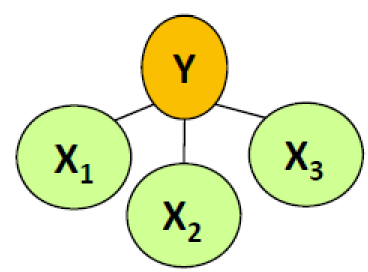
SRL in Classification (conventional) IID classification

Given:





Task: Predict Y given X_1, X_2, X_3

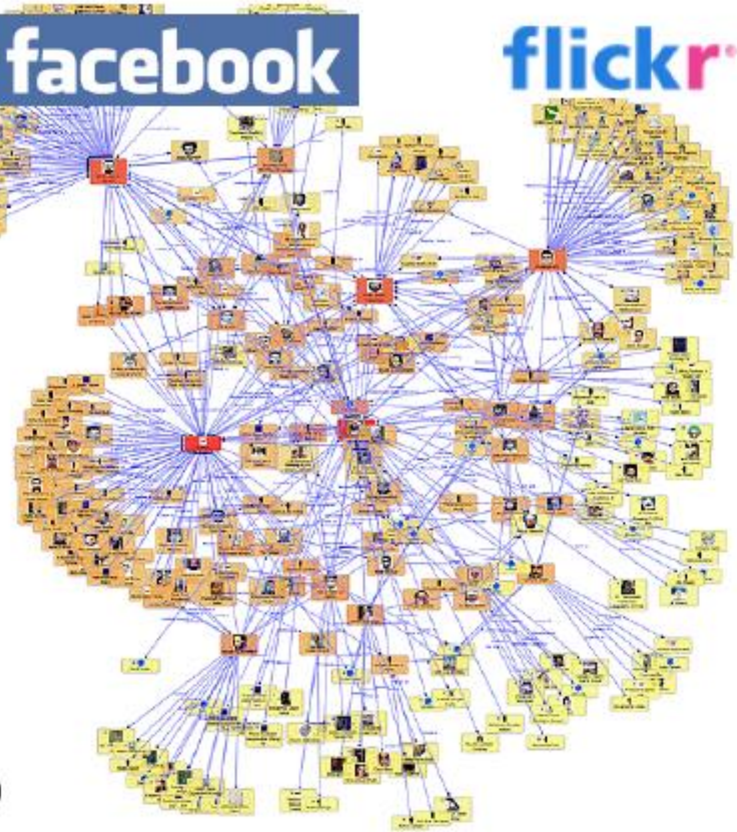


Classifiers: Use your favorite, logistic regression/SVM, neural net, naïve Bayes, decision trees, etc.

[slide courtesy of Lise Getoor]

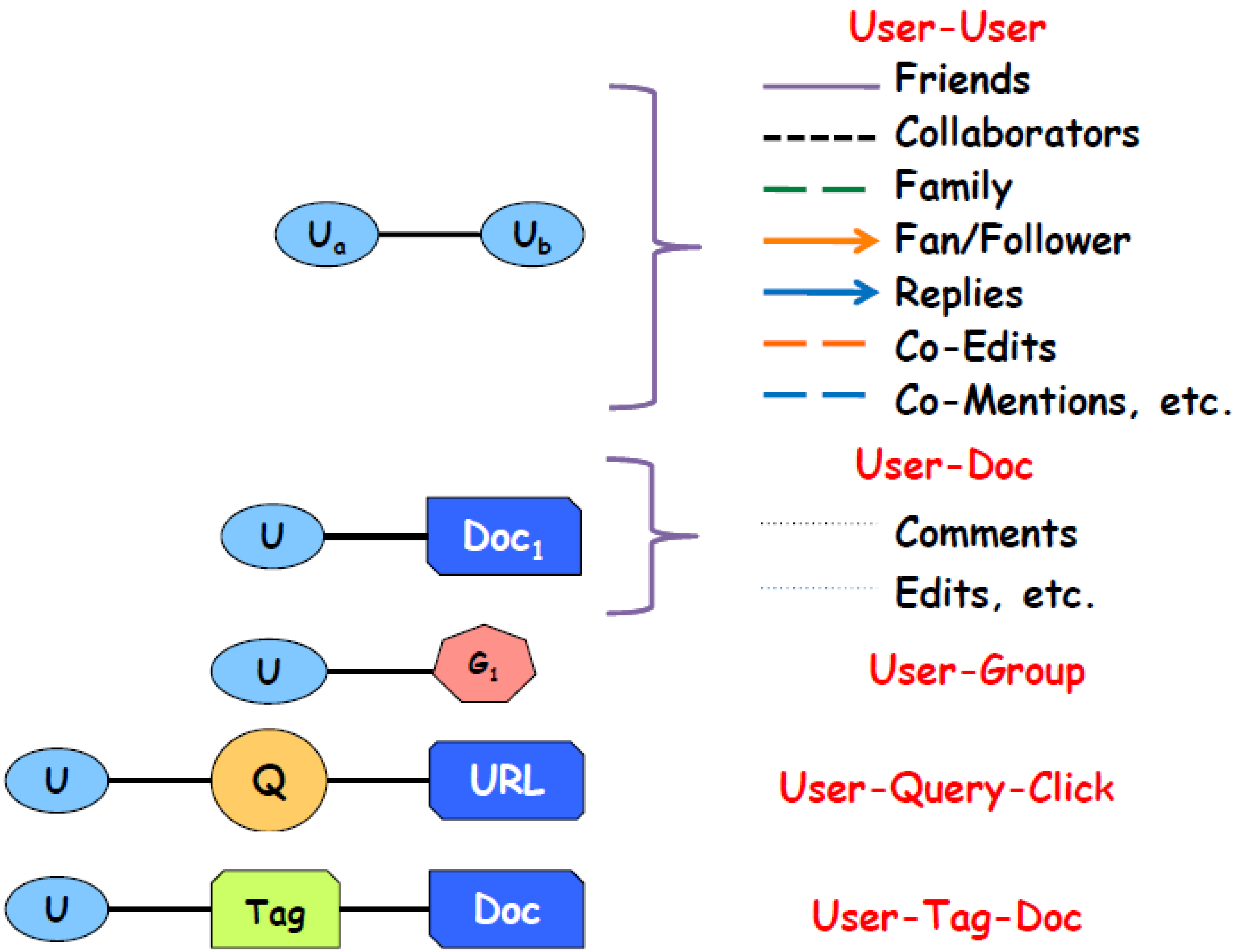
Example: Social Media

facebook
flickr
YouTube
LinkedIn
myspace.com
a place for friends..
BibSonomy ::
dogster
for the love of dog
last.fm
bebo
orkut
twitter



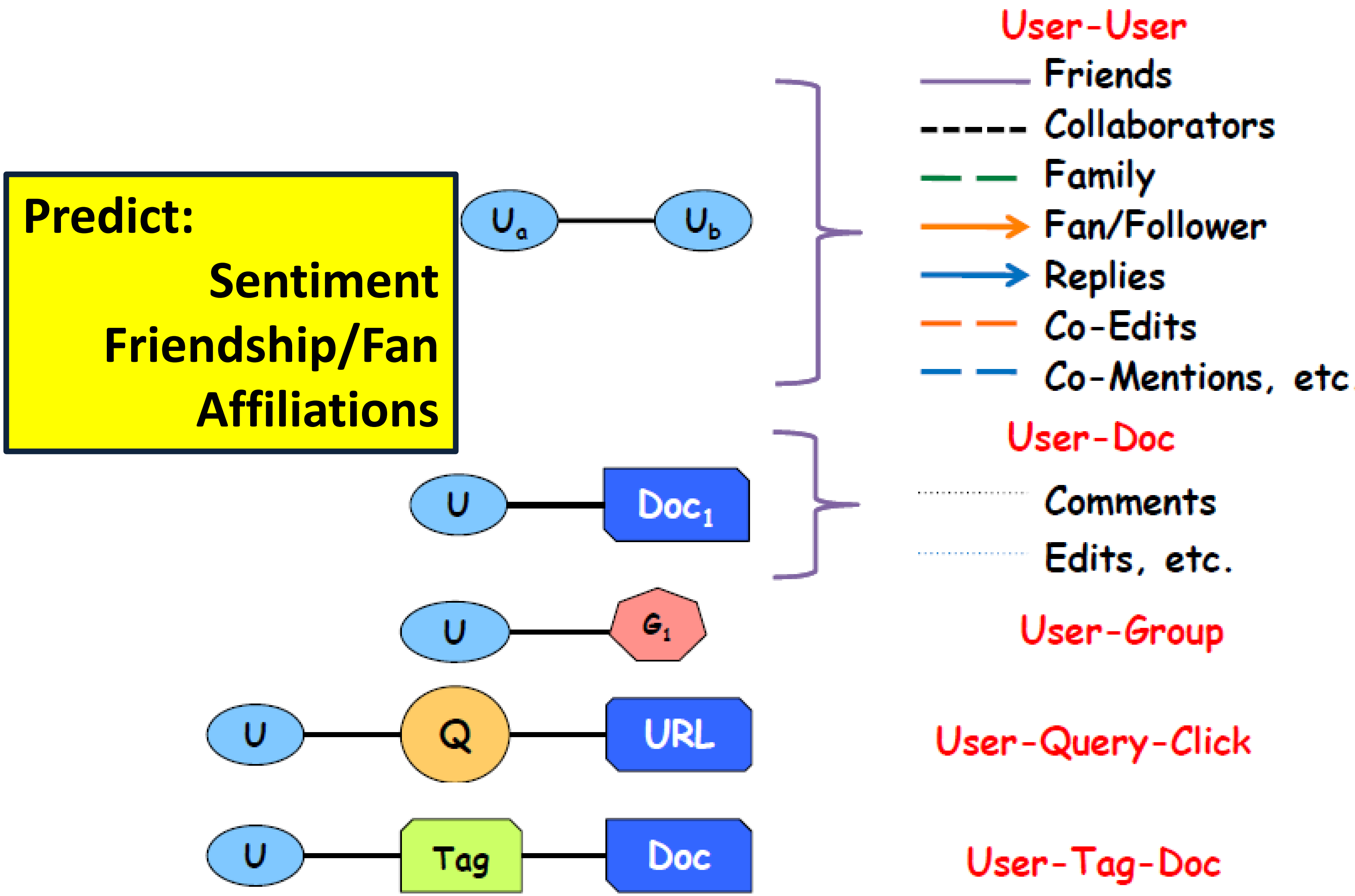
[slide courtesy of Lise Geetor]

Social Media Relationships



[slide courtesy of Lise Getoor]

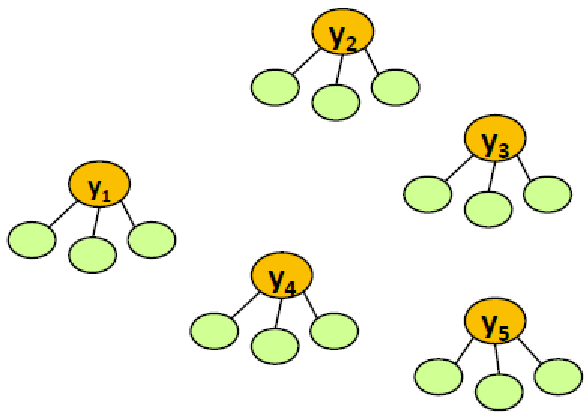
Social Media Relationships

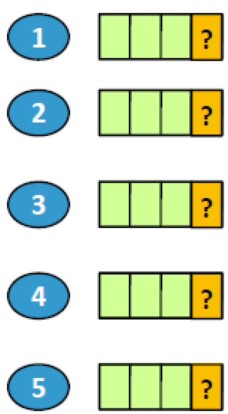


[slide courtesy of Lise Getoor]

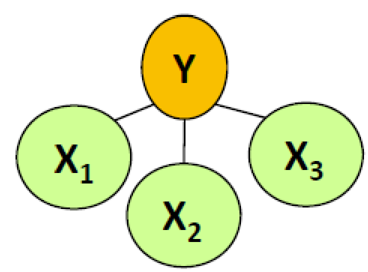
SRL in Classification (conventional) IID classification

Given:





Task: Predict Y given X_1, X_2, X_3

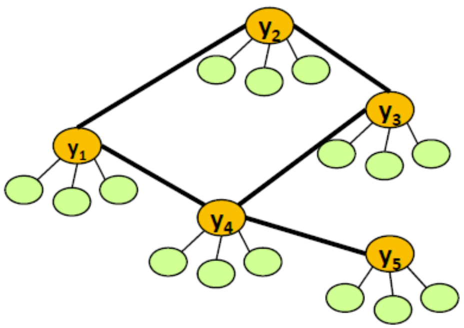


Classifiers: Use your favorite, logistic regression/SVM, neural net, naïve Bayes, decision trees, etc.

[slide courtesy of Lise Getoor]

SRL in Classification
Relational Classification

Given:

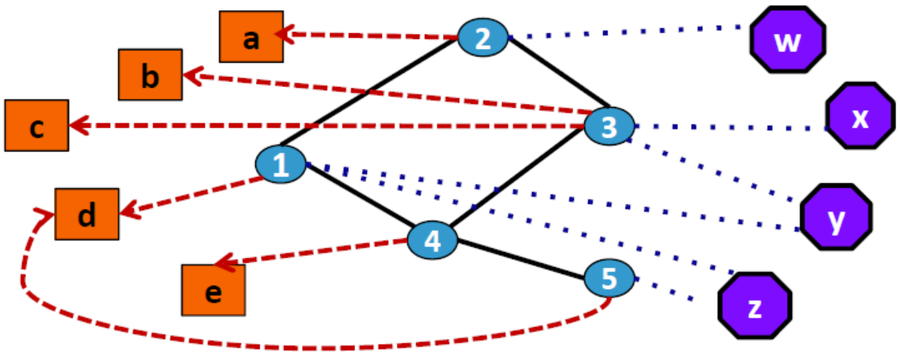


[slide courtesy of Lise Getoor]

SRL in Classification

Relational Classification – Attribute Prediction

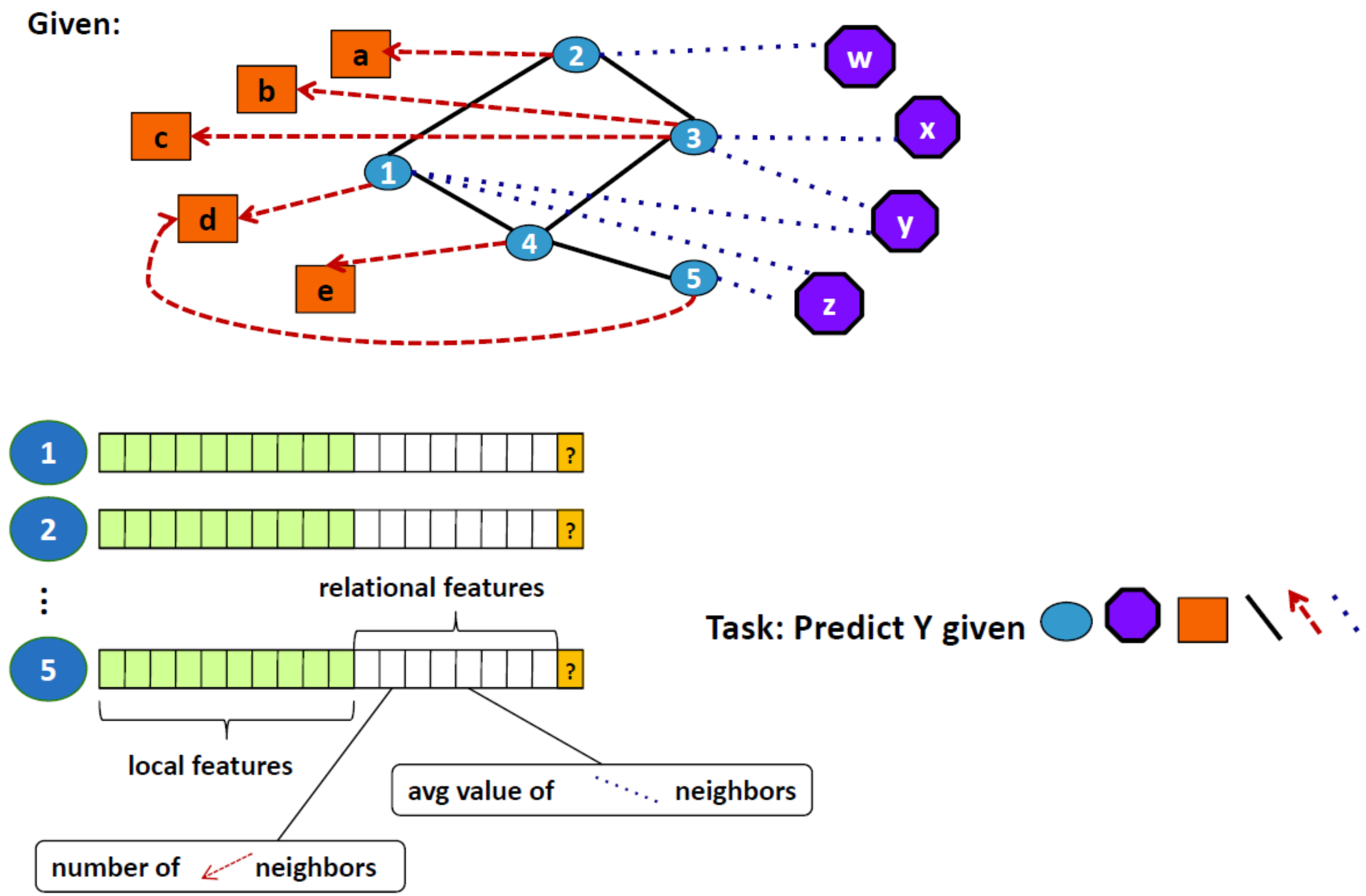
Given:



[slide courtesy of Lise Getoor]

SRL in Classification

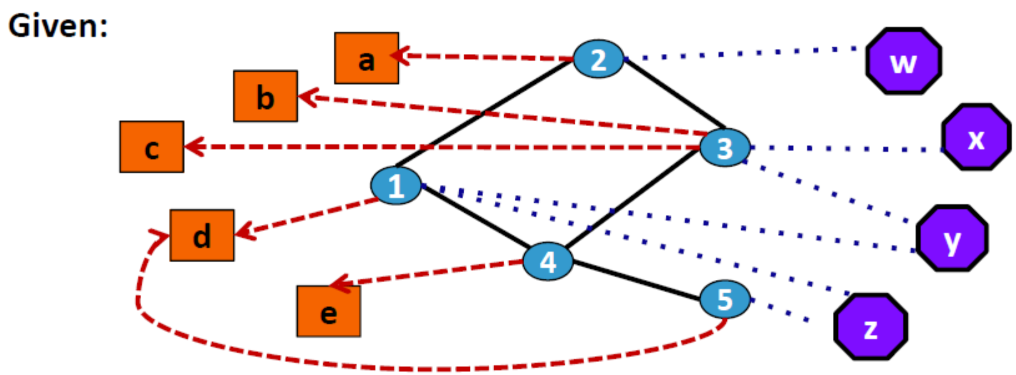
Relational Classification – Attribute Prediction

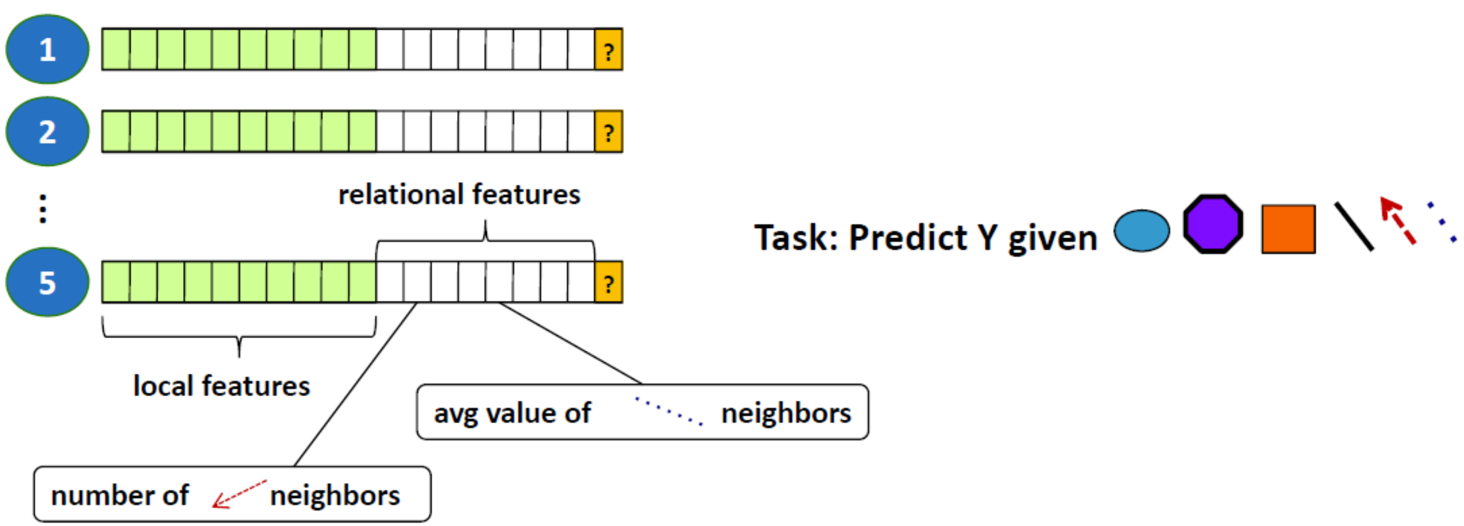


[slide courtesy of Lise Getoor]

SRL in Classification

Relational Classification – Attribute Prediction





[slide courtesy of Lise Getoor]

SRL in Classification

Predicting Ad Click-Through Rate

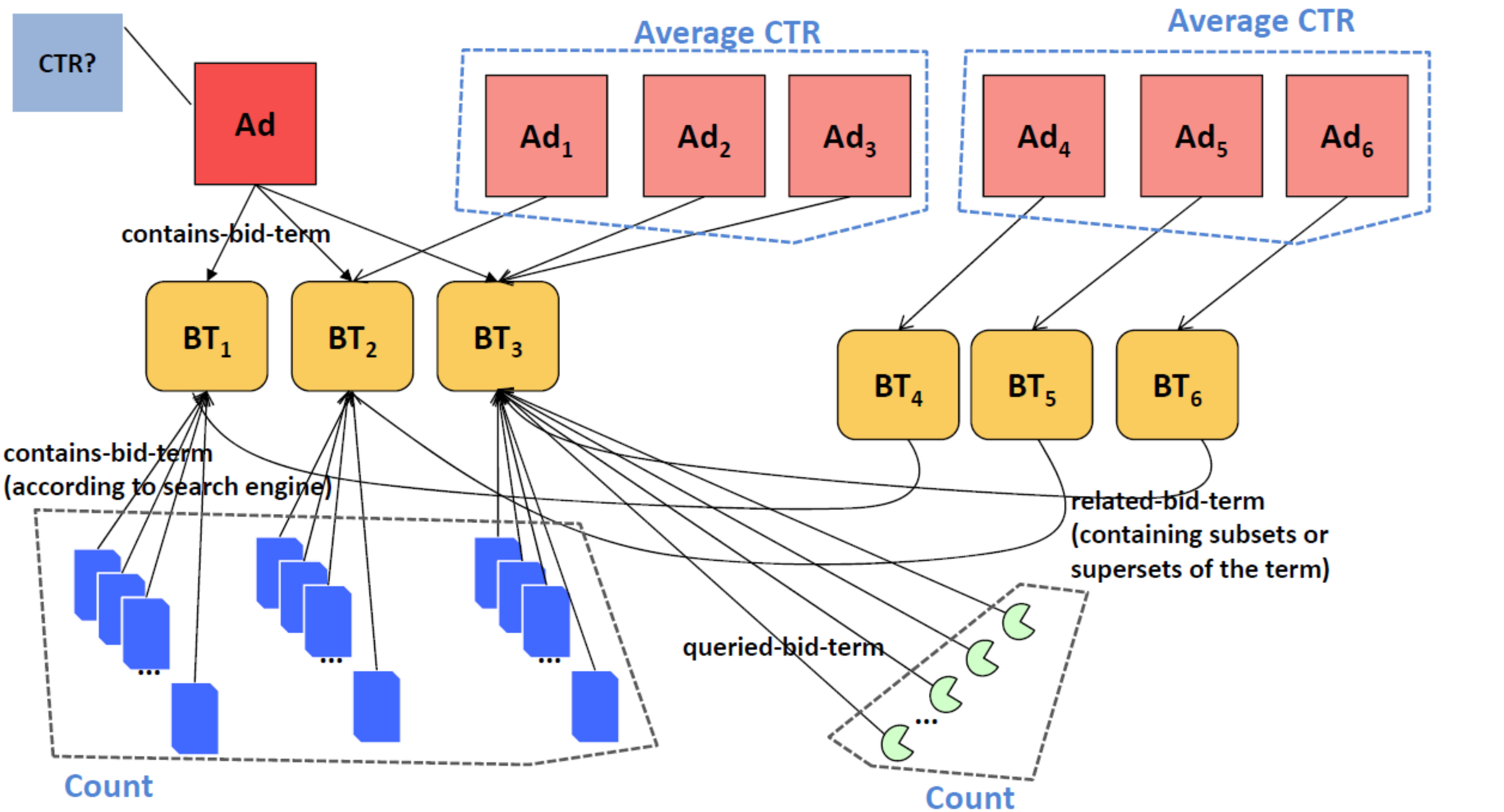
- Task: Predict the click-through rate (CTR) of an online ad, given that it is seen by the user, where the ad is described by
  - URL to which user is sent when clicking on ad
  - Bid terms used to determine when to display ad
  - Title and text of ad

- Based on approach by [Richardson et al., WWW07]

[slide courtesy of Lise Getoor]

SRL in Classification
Predicting Ad Click-Through Rate

- Relational Feature Used



[slide courtesy of Lise Getoor]

Contents

- Statistical Relational Learning
  (확률관계형 학습)
- **Relational Kalman Filtering**
  (관계형 칼만 필터)
- Probabilistic Artificial Intelligence Lab
  (확률 지능형 인공지능 연구실)

Kalman Filter

- Kalman Filter is an algorithm which produces estimates of unknown variables given a series of measurements (w/ noise) over time.

- Numerous applications in
  - Robot localization
  - Autopilot
  - Econometrics (time series)
  - Military: rocket and missile guidance
  - Weather forecasting
  - Speech enhancement
  - …

Stanley, Stanford University

Sebastian Thrun credited **Stanley's** victory to **Kalman filters**

Example – Kalman Filter for John's Home

- Input statements
  - **John**'s house price was **$0.39M at 2010**.
  - Each year, **John**'s house price **increases 5%**.
  - **John**'s house price is around the sold price.
  - **John**'s house is sold sporadically.
- Question: what is the price of John's house each year?

Example – Kalman Filter for John's Home

- Input statements
  - **John**'s house price was **$0.39M at 2010**.
  - Each year, **John**'s house price **increases 5%**.
  - **John**'s house price is around the sold price.
  - **John**'s house is sold sporadically.
- Question: what is the price of John's house each year?

$$\mathbf{x}^{11'}_{\mathbf{John}} = 1.05\mathbf{x}^{10'}_{\mathbf{John}} + \varepsilon_{tras}$$

Sold at $0.44M

$\varepsilon \sim \mathbf{N(0,\sigma^2)}$

$$\mathbf{x}^{10'}_{\mathbf{John}} = 0.39M + \varepsilon_{John}$$

$$\mathbf{x}^{11'}_{\mathbf{John}} = 0.44M + \varepsilon_{obs}$$

$x^{10'}_{John}$ $x^{11'}_{John}$ $x^{12'}_{John}$

2010 2011 2012 2013

Why Kalman Filter Takes $O(n^3)$ operations?

- Kalman Filtering steps

1. **Input: prior belief,** $X^t \sim N(\mu^t, \Sigma^t)$

   n variables, $X^t = \{x^t_{John}, x^t_{Tom}, x^t_{Ann}, \ldots\}$.

2. **Take the transition model:**

   $X^{t+1} = A_T X^t + \varepsilon_{trans}$ when $\varepsilon_{trans} = N(0, \Sigma_T)$.

3. **Updated covariance matrix:** $\Sigma^{t'} = A_T \Sigma^t A_T^T + \Sigma_T$.

4. **Take the observation model:**

   $X^{t+1} = A_O Obs^{t+1} + \varepsilon_{obs}$ when $\varepsilon_{obs} = N(0, \Sigma_O)$.

5. **Kalman gain:** $K = \Sigma^{t'} A_O^T (A_O \Sigma^{t'} A_O^T + \Sigma_O)^{-1}$.

6. **Output: update belief,** $X^{t+1} \sim N(\mu^{t+1}, \Sigma^{t+1})$

   New mean: $\mu^{t+1} = \mu^t + K(Obs^{t+1} - \mu^t)$

   New covariance: $\Sigma^{t+1} = (I - KA_O)\, \Sigma^{t'}$

Why Kalman Filter Takes $O(n^3)$ operations?

- Kalman Filtering steps
  1. **Input: prior belief,** $X^t \sim N(\mu^t, \Sigma^t)$

     n variables, $X^t = \{x^t_{John}, x^t_{Tom}, x^t_{Ann}, \ldots\}$.
  2. **Take the transition model:**

     $X^{t+1} = A_T X^t + \varepsilon_{trans}$ when $\varepsilon_{trans} = N(0, \Sigma_T)$.
  3. **Updated covariance matrix:** $\boxed{\Sigma^{t'} = A_T \Sigma^t A_T^T + \Sigma_T.}$
  4. **Take the observation model:**

     $X^{t+1} = A_O Obs^{t+1} + \varepsilon_{obs}$ when $\varepsilon_{obs} = N(0, \Sigma_O)$.
  5. **Kalman gain:** $K = \Sigma^{t'} A_O^T \boxed{(A_O \Sigma^{t'} A_O^T + \Sigma_O)^{-1}}.$
  6. **Output: update belief,** $X^{t+1} \sim N(\mu^{t+1}, \Sigma^{t+1})$

     New mean: $\mu^{t+1} = \mu^t + K(Obs^{t+1} - \mu^t)$

     New covariance: $\boxed{\Sigma^{t+1} = (I - KA_O)\, \Sigma^{t'}}$

Inversions and multiplications of the n by n matrix need $O(n^3)$ operations.

$$\Sigma^t = \begin{bmatrix} \sigma^2_{1,1} & \cdots & \sigma^2_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma^2_{n,1} & \cdots & \sigma^2_{n,n} \end{bmatrix}$$

Relational Kalman Filter

A set of element shares relationship!

- Input statements
  - **Town is a set of houses.**
  - **Town's** houses have initial prices **at 2010**.
  - Each year, **Town**'s house prices **increase 5%**.
  - **Town**'s house prices are around sold prices.
  - **Town**'s houses are sold sporadically.
- Question: what is the prices of Town's houses each year?

Relational Kalman Filter (IJCAI-11):
New Transition Models & Observation Models

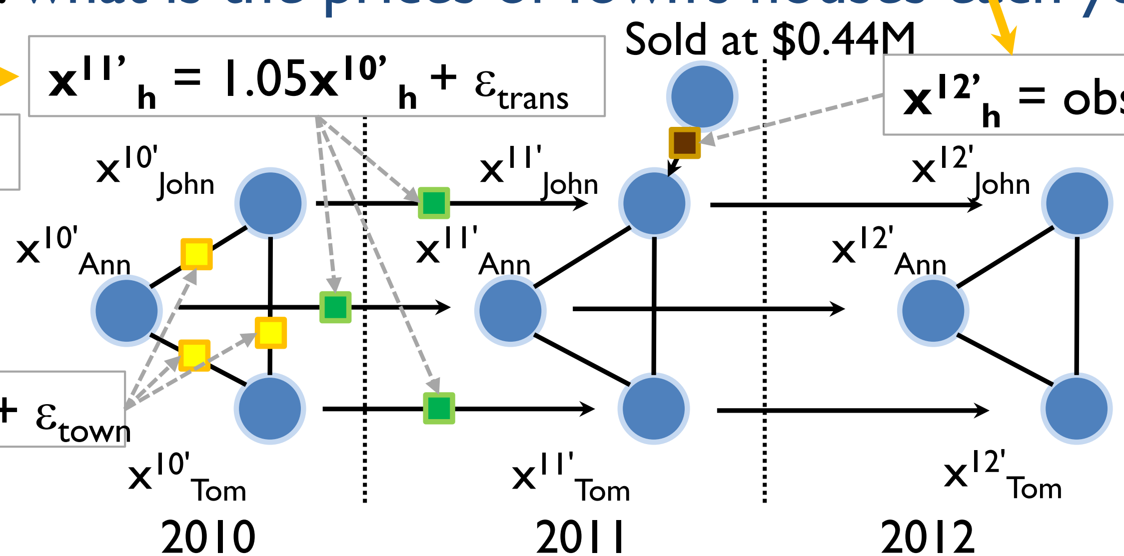
- Input statements
  - **Town is a set of houses.**
  - **Town's** houses have initial prices **at 2010**.
  - Each year, **Town**'s house prices **increase 5%**.
  - **Town**'s house prices are around sold prices.
  - **Town**'s houses are sold sporadically.
- Question: what is the prices of Town's houses each year?

$x^{11'}_{h} = 1.05x^{10'}_{h} + \varepsilon_{trans}$

$x^{12'}_{h} = obs^{12'}_{h} + \varepsilon_{obs}$

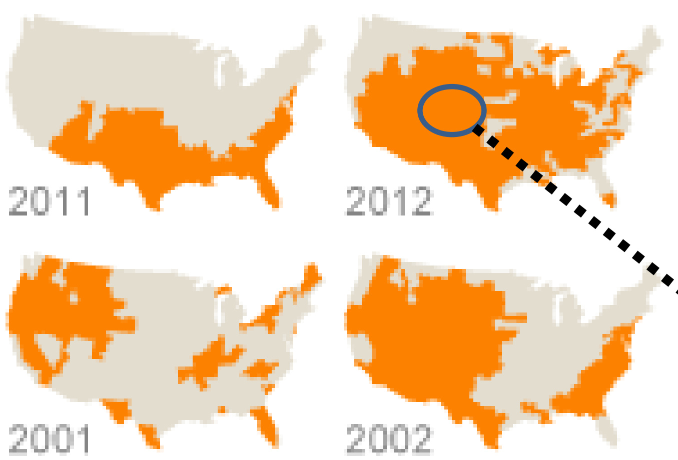
h, h' ∈ Town

$X^{10'}_{h} = x^{10'}_{h'} + \varepsilon_{town}$

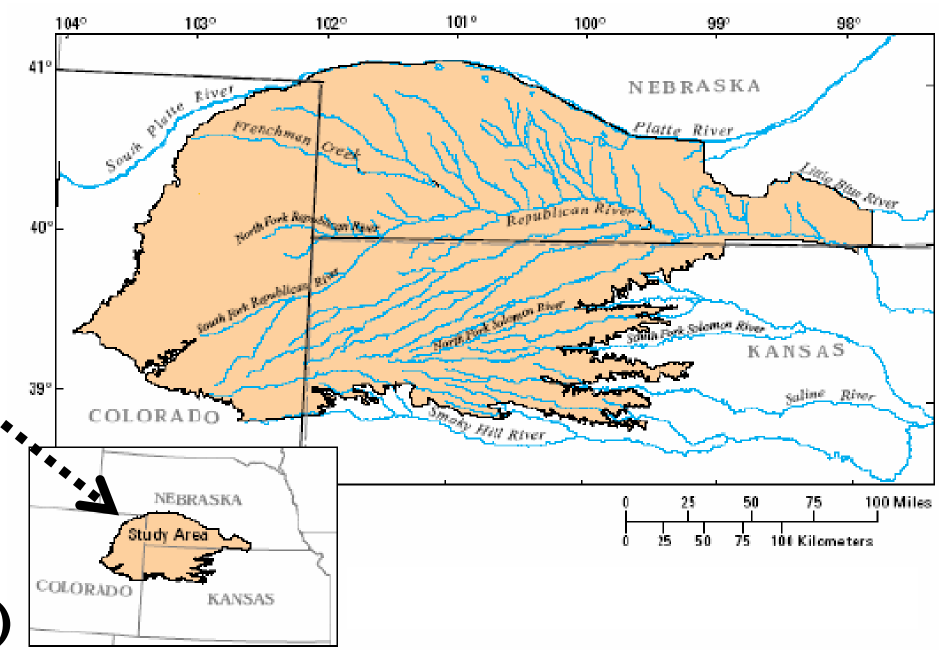


Experiments (Groundwater Models)

- Data is extracted in the largest aquifer (Ogallala) in US.
- Pumping (for farming) depletes many of water wells.
- Estimating level of groundwater is critical.

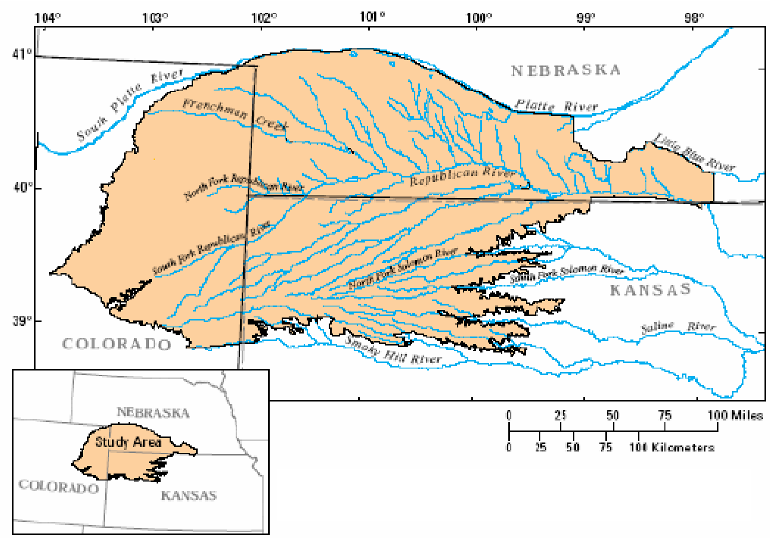


US drought maps (New York Times)

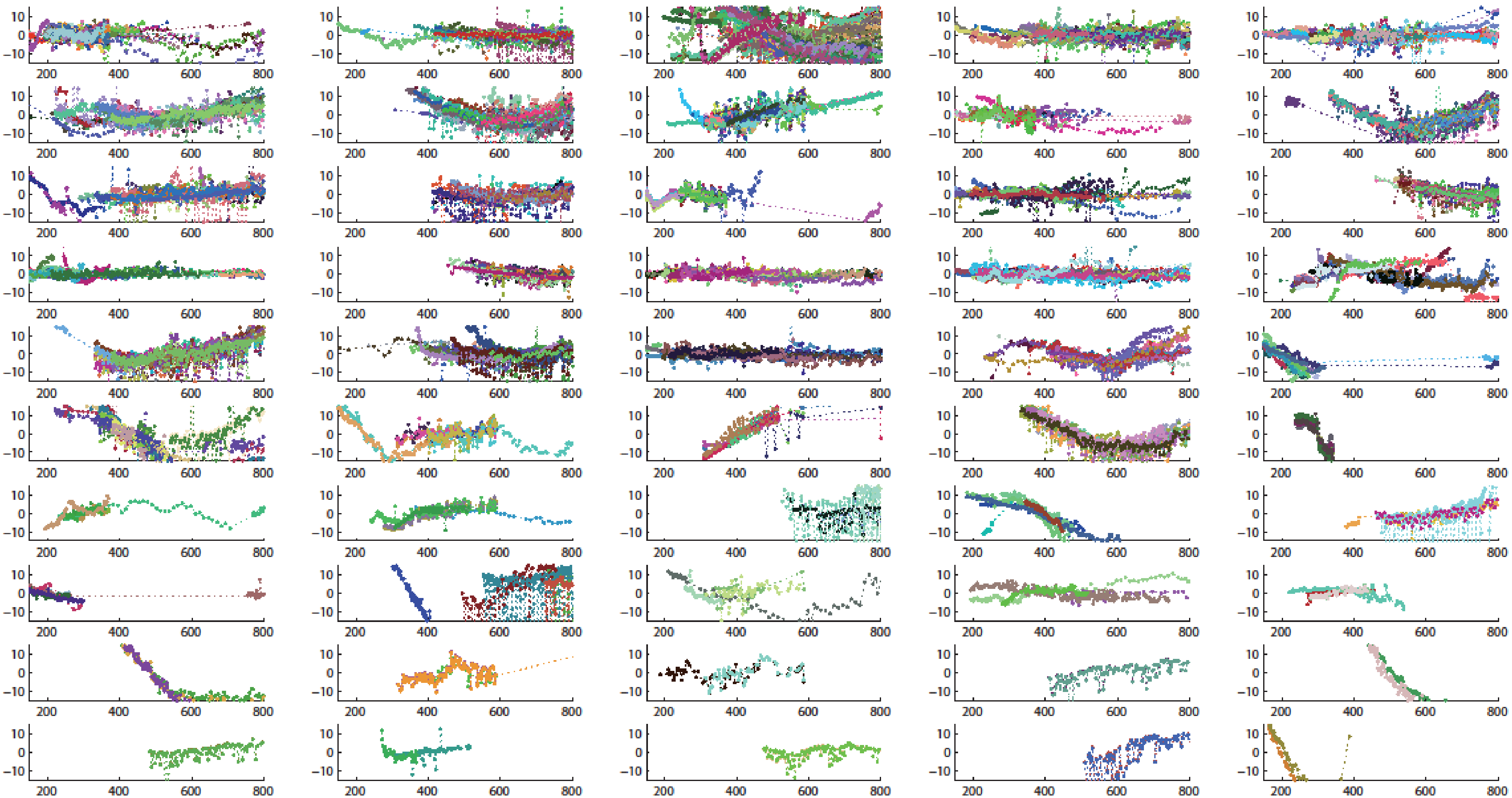


Experiments (Groundwater Models)

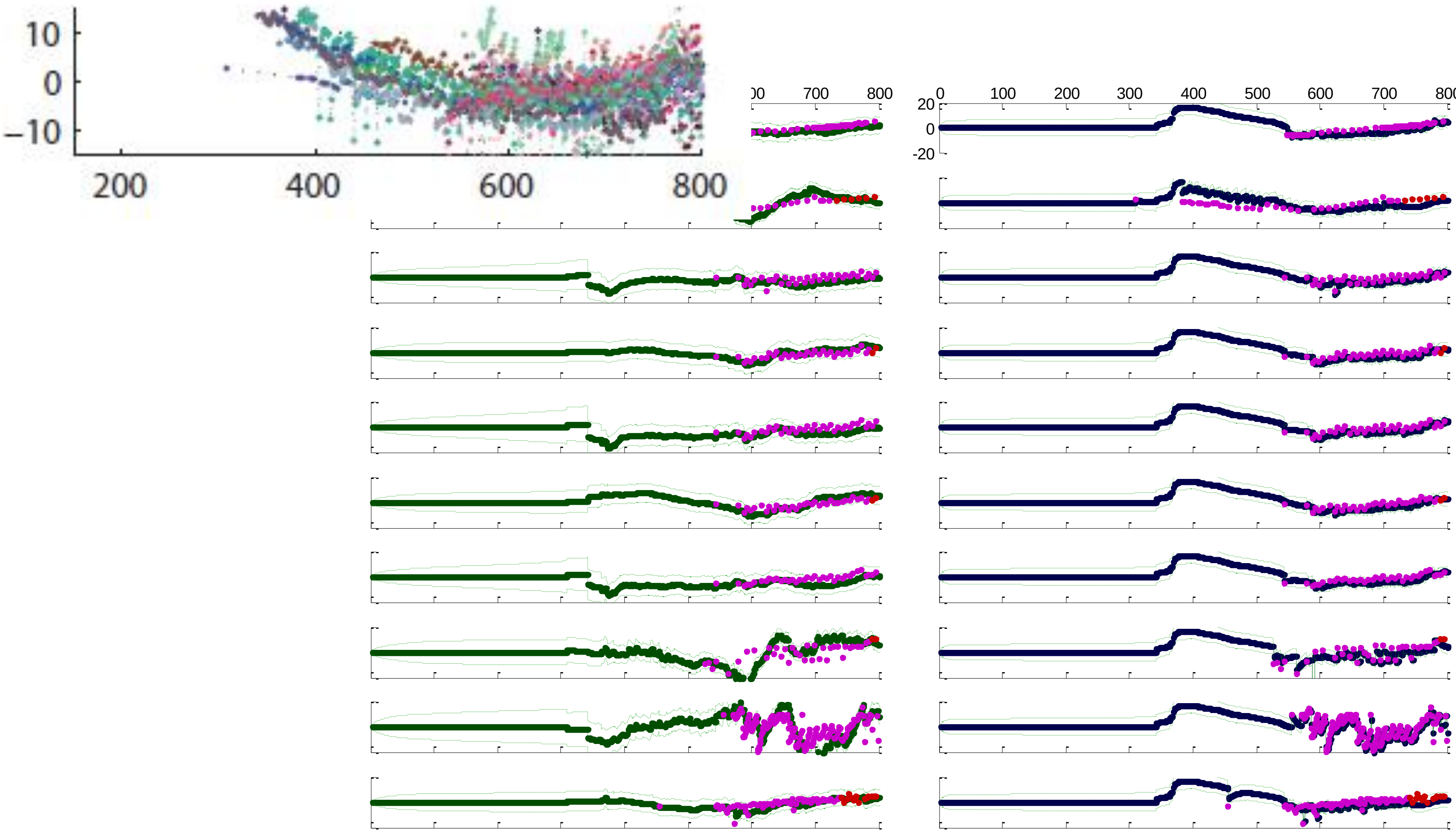
- Dataset
  - The model has measures (water levels) for 3078 water wells.
  - The measures span from 1918 to 2007 (about 900 months).
  - It has over 300,000 measurements.
- Cluster: 3078 wells into 10 groups.
- Train parameters using the auto regression (AR).
  - Vanilla Kalman filter
  - RKF



Extraction of Relational Information by Spectral Clustering



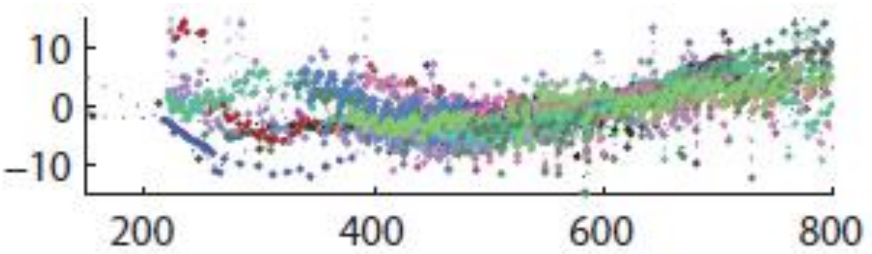
Extraction of Relational Models by Spectral Clustering

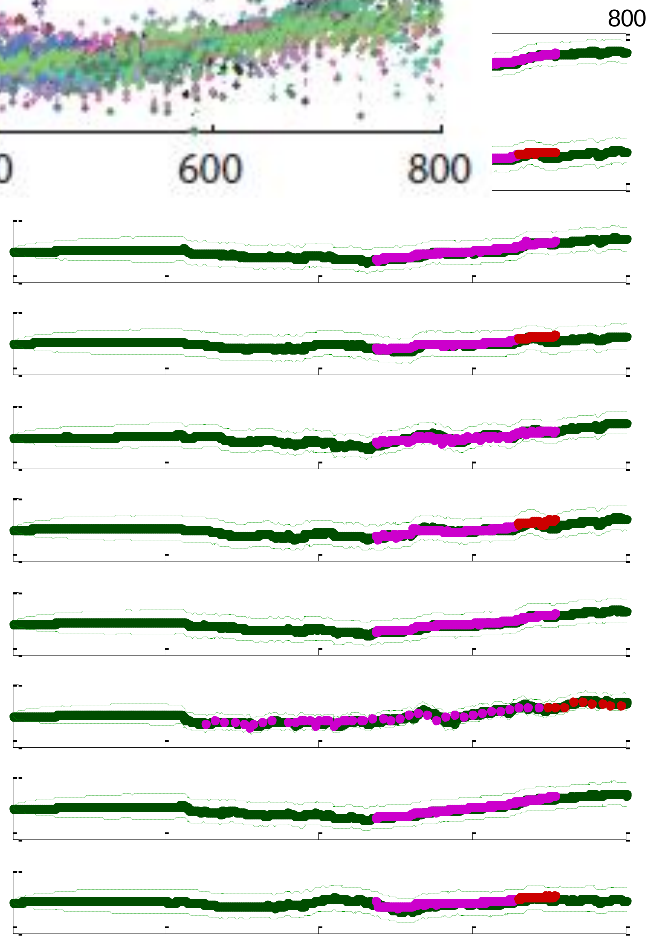


Vanilla KF

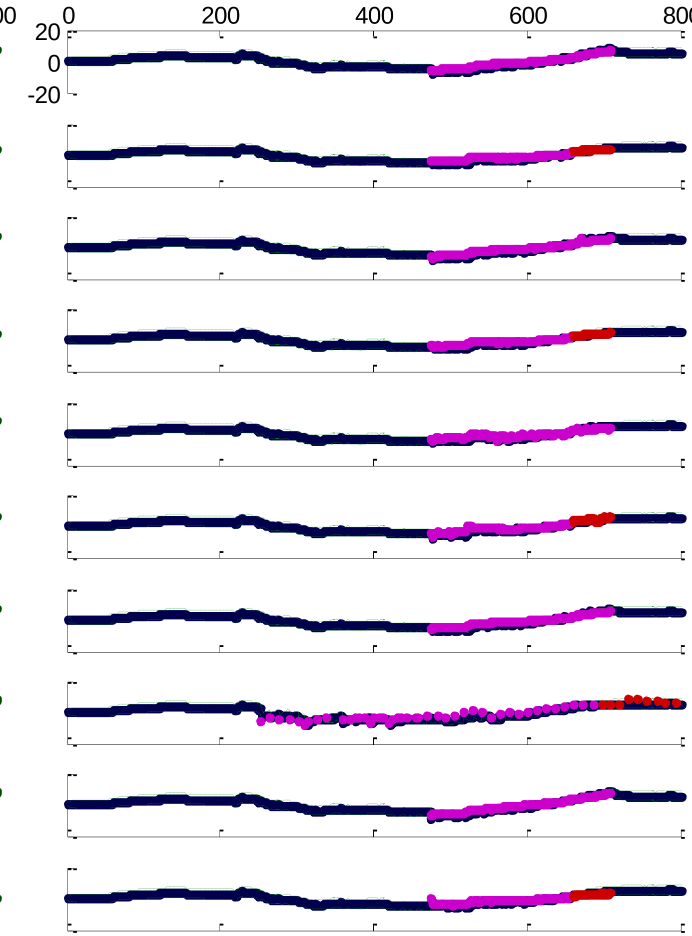
Relational KF

Extraction of Relational Models by Spectral Clustering





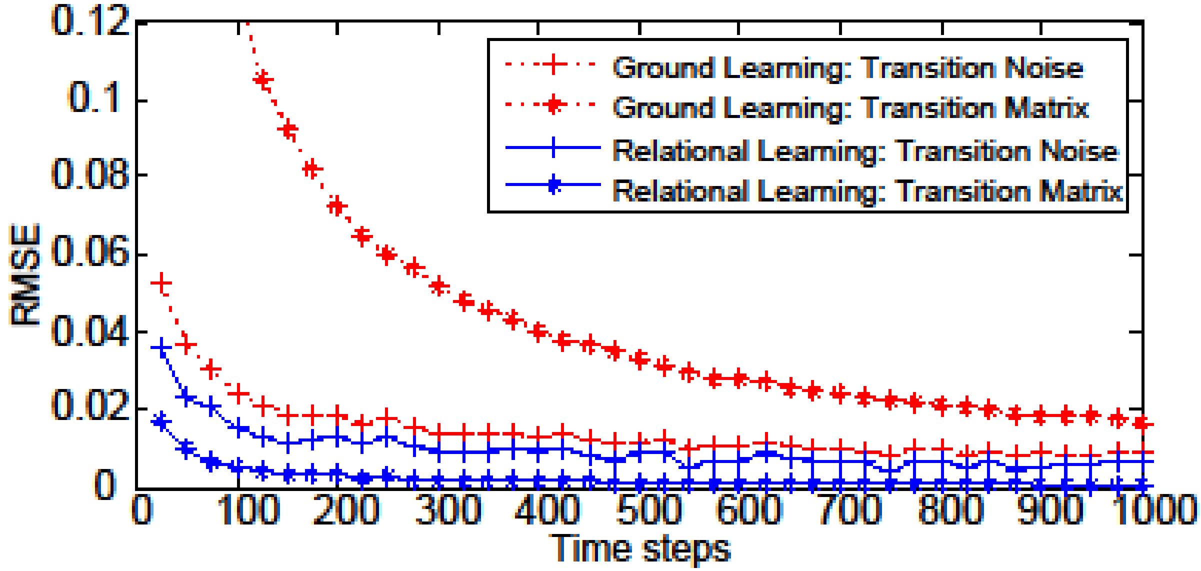
Vanilla KF



Relational KF

Learning and Prediction with RKF

- Parameter Learning in simulation

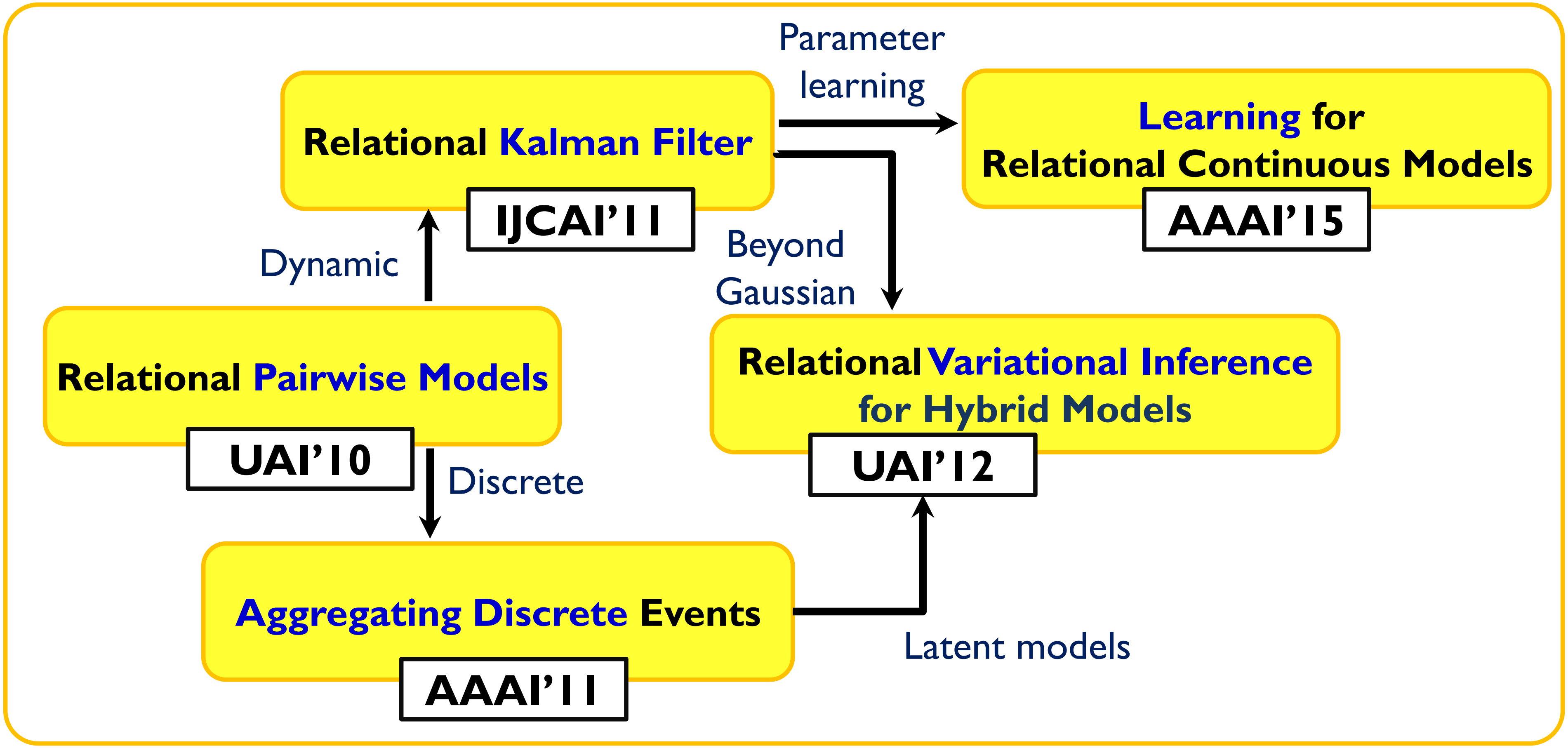


- Prediction accuracy on the RRCA model

	Vanilla KF	Relational KF
RMSE (Root Mean Square Error)	5.10	**4.36**
Negative Log of Probability -log(P(data|pred))	4.91	**3.88**

Statistical Relational Learning @PAI-Lab

Learning and inference with large-scale models



- **AAAI** and **IJCAI** are top #1 and #2 conferences in **Artificial Intelligence**
- **UAI** is top #3 conference in **Machine Learning**

(source: **academic**.research.**microsoft**.com/)

Contents

- Statistical Relational Learning
  (확률관계형 학습)
- Relational Kalman Filtering
  (관계형 칼만 필터)
- **Probabilistic Artificial Intelligence Lab**
  (확률 지능형 인공지능 연구실)

Research@PAI-Lab, UNIST

4. Machine Learning Based Intrusion Detections and Plant Diagnosis

2. State-of-the-art Human Action Classifier

3. Locally Exchangeable Measures & Compressive Sensing

Social Networks (w/ Linkedin)

Big Data Archiving (w/ Berkeley)

Fault/Attack Detection (w/ UIUC)

Financial Data

Computer Vision

1.The first O(n) Kalman Filter & Efficient Gaussian Process

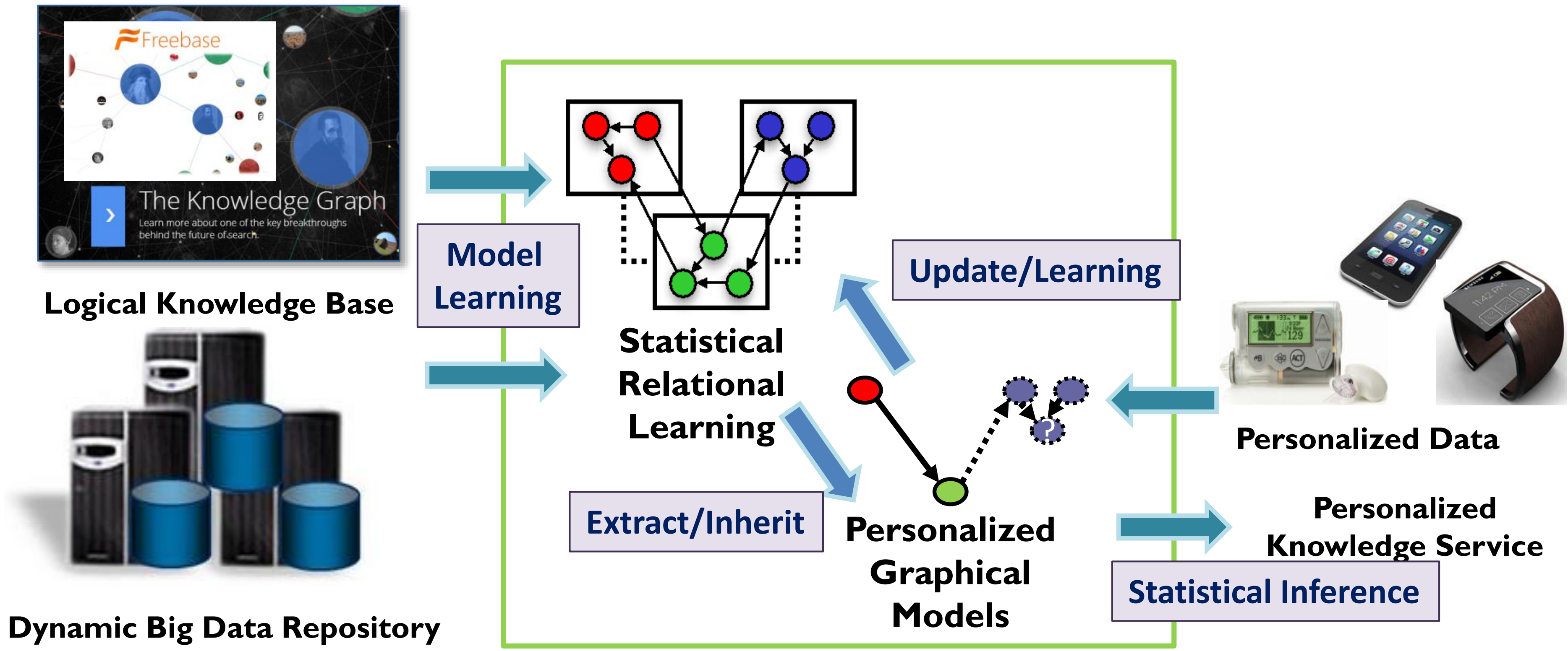
Relational Graphical Models

Dynamic Bayesians

Multivariate Kernel Methods

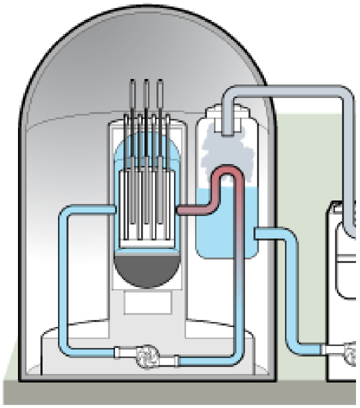
Machine Learning and Statistical Inference for Large-Scale Models

Smart Knowledge Service with Statistical Relational Learning



* Funded by Korean National Research Foundation Early Career Award May 2014 ~ April 2017

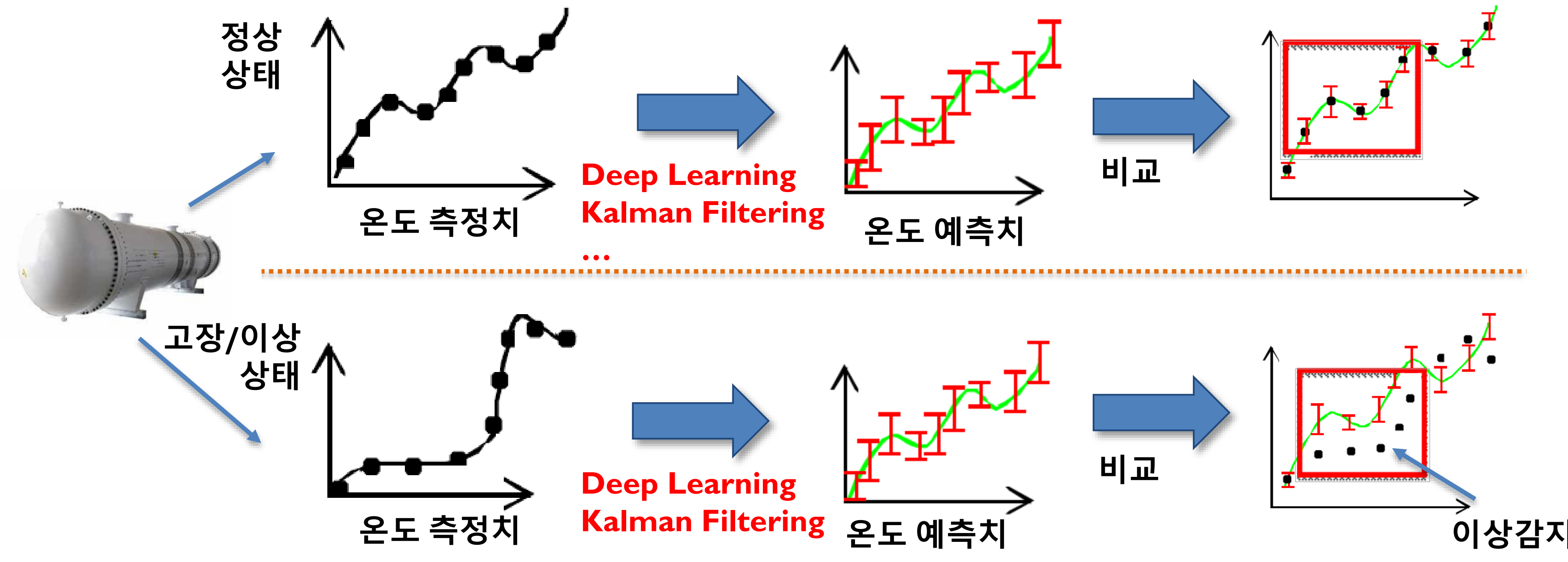
Fault Diagnosis for Nuclear Power Plants



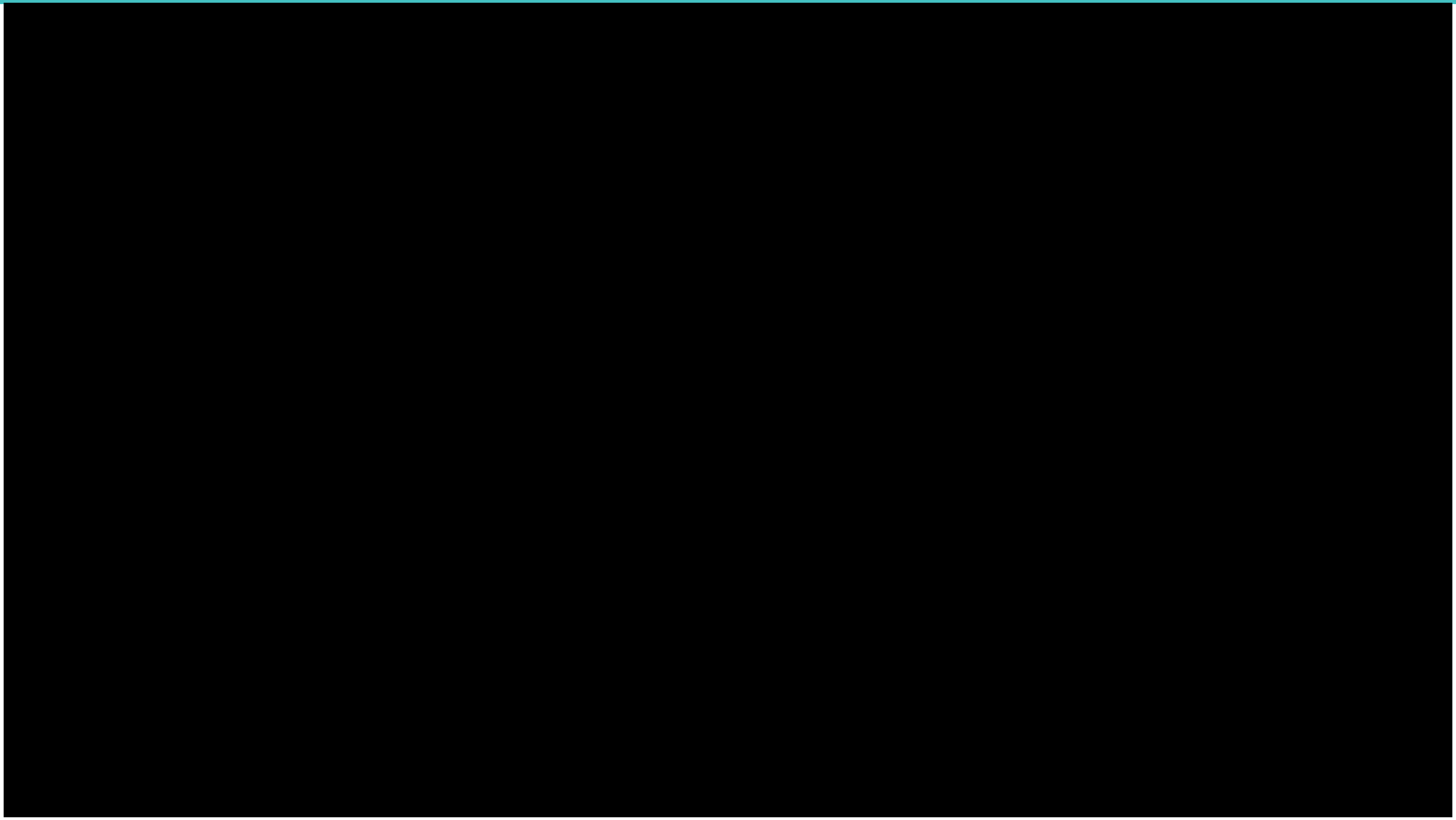
UNIST + Argonne NATIONAL LABORATORY = Safe Nuclear Energy

International Nuclear Energy Research Initiative (I-NERI)

한미국제공동연구(I-NERI)



Machine Learning Applications:
Human Face Detection in Google Glass



http://youtu.be/Z-siJX3LNO0

* Credit Chunggeol Ryu

Machine Learning Applications: Face Alignment and Facial Expression Recognition

http://youtu.be/5pe2IboFzjc
* Credit Kyungjoong Jeong

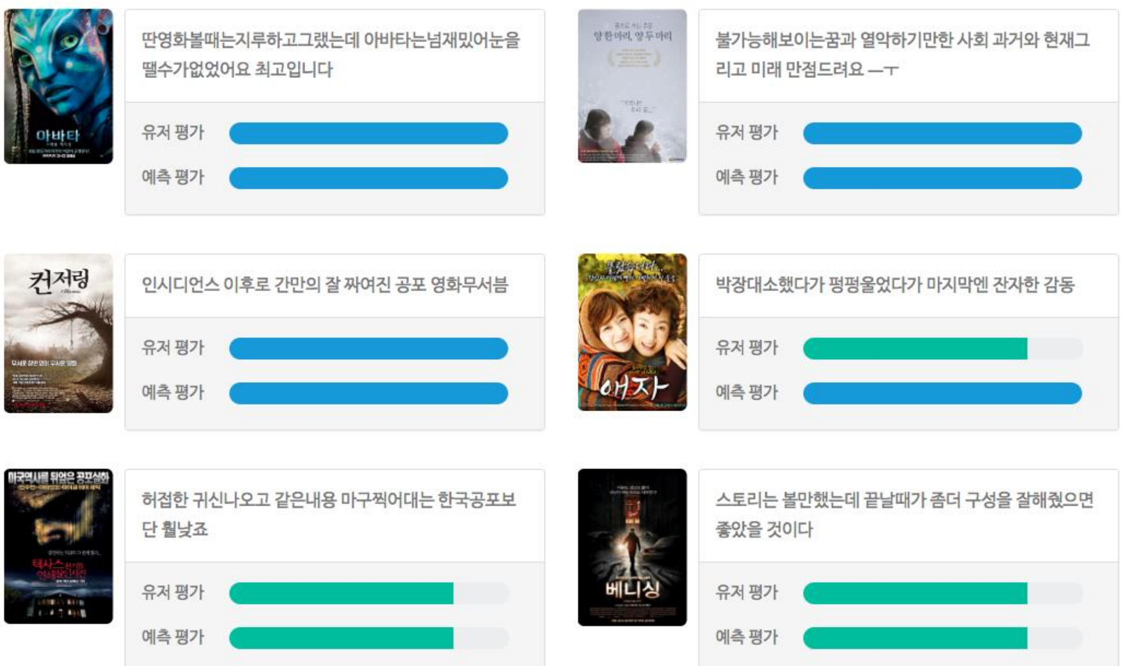
Machine Learning Applications:
Learning Manipulation Actions for Robot

http://youtu.be/SHBRF-sNpY8

* Credit Phuong Hoang

UNIST

Sentiment Analysis for Movie Review*



리뷰덕

세상의 모든 리뷰에서
감성을 찾다

Live demo: http://pail.unist.ac.kr:8080/

*by Taehoon Kim

UNIST

머신 러닝 알고리즘 소개글

TECH M

메인 | 최신뉴스 | 커버스토리 | 인뎁스스토리 | 피플&컴퍼니 | 이슈&트렌드

커버스토리 [인공지능과 딥러닝]

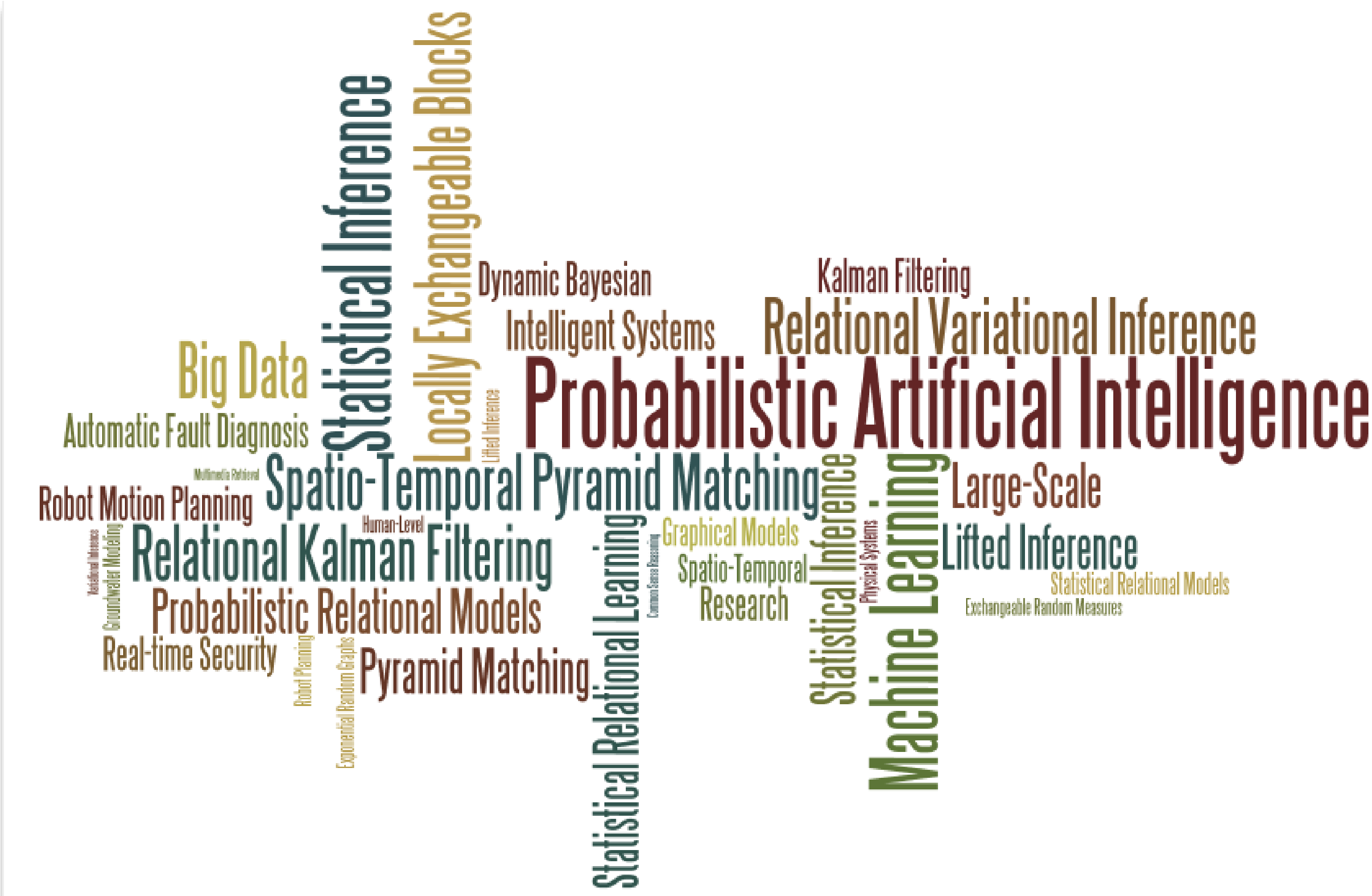
[딥러닝④] 진화하는 머신러닝 알고리즘 점점 더 ‘사람처럼’

다양한 머신러닝 알고리즘

최재식 울산과학기술대 교수 2015.02.23 17:59

http://techm.kr/home/bbs/board.php?bo_table=cover&wr_id=189&mg_id=

Thank you!



If you have any question, please send an e-mail to jaesik@unist.ac.kr