Bayesian Networks and Probabilistic Latent Semantic Analysis: AI Applications utilizing Big Data and its Future

Yoichi Motomura (y.motomura@aist.go.jp)

Principal researcher, Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST, Japan)
Probabilistic User Modeling

Human as elements in the system → uncertain, unstable, non-deterministic

Expand “system” to social domains → AI for human life, service and society

Bayesian Networks and Probabilistic Latent Semantic Analysis for Cyber Physical Systems and Applications
Bayesian network software developed in ETL/AIST (1998〜2009) (released as BayoNet → BayoLink from NTT data Mathematical System Inc.)

PLASMA (2010〜)
(Probabilistic Latent Structure Modeling API)

- Java API for probabilistic modeling
- Scala API
- Interpreter language using Java API
- Java/Web API
**PLASMA: as an AI Framework for real world problem solving**

Digitalized for Control and management

- Clustering
  - pLSA
  - Bayesian networks

- Structuring
  - 時
  - 場
  - 行
  - 何

User modeling

Probabilistic simulations

in real world applications

---

**Probabilistic Latent Semantic Analysis (PLSA)**

- Customer/Item segmentation from ID-POS data
- [Original ID-POS data (sep.2008-sep.2009)]
  - 700 millions transaction in supermarket chain 15 stores, 300,000 items.
  - 3981 customers who answered the questionnaires and 1000 popular items
  - Automatic clustering using PLSA (probabilistic latent semantic Analysis), then constructing Bayesian networks

- In real world applications, the best number of categories are automatically calculated using BIC.

- Items and Customers are classified into categories.
Re-usable computational models for real world applications

Big data in real world applications
behavior
Life logs
Interview/Questionnaires
User's Feedback data
Real services

Recognition model
PLSA, Naïve Bayes, text mining, etc...

Generative model
Recommendation Simurator, etc.

User model libraries (Bayesian networks)

Interest
like
scare
fun
want
comfortable

Applications
Ex. Probability of buying behavior controlling P(buy|condition)

Go out for shopping
Enter the store
Select the goods
Buy!

Attention
Interest
Desire
Action

商品情報
ノベルティ
話題
季節感

実物を見る
買い物の日
イベント
同行者として
なんとなく

実物を見る
フィッティング
素材感
時間がある
ディスプレイに
惹かれて

実感（似合う、機能）
揃える（色・デザイン）
安心（保有・確保）
接客に惹かれて
Probabilistic consumer models

- Behavior: ex. buying the product (yes or not)
  If the number of people is \((\text{buy}=10, \text{not buy}=90)\),
  \(P(\text{buy}=\text{yes}) = 10/(10+90) = 0.1\)

- Probability of behavior: \(P(\text{buy} | \text{condition})\)

- If the product is a cosmetic for women,
  \(P(\text{buy} | \text{female}) > P(\text{buy} | \text{male})\)

- If the day is weekend,
  \(P(\text{buy} | \text{female, weekend}) > P(\text{buy} | \text{weekday})\)

Information recommendation in supermarkets:
(BN model is constructed from ID-POS data)
Constructing models and probabilistic inference

【PLSA + Bayesian networks】
① PLSA works as dimension reduction for ID-POS
(Thousands customers and items are classified into 20 categories)

② Construct Bayesian networks for each latent class extracted by PLSA

Co-occurrence matrix

ID POS data
User1 buy Item1 and Item M,
User2 buy Item1 and Item2,
:
User N buy Item2

Co-occurrence matrix

<table>
<thead>
<tr>
<th></th>
<th>Item 1</th>
<th>Item 2</th>
<th>...</th>
<th>Item M</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>1</td>
</tr>
<tr>
<td>User 2</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>User N</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td>0</td>
</tr>
</tbody>
</table>

Co-occurrence: n(l,j)
Clustering and structuring consumer models
Using questionnaire data, lifestyles are classified into 6 types. Items are classified into 20 categories using PLSA. They can be estimated for all consumers from buying history.

Consumer model:
Bayesian network constructed from ID-POS data and questionnaire about lifestyle and personality
Represents joint probability distribution of
\( P(\text{ConsumerSegments, Personality, Lifestyle, ItemSegments, Time, Situations}) \)

(Using 4,200,000 records of ID-POS data, links are selected by AIC)
Bayesian network constructed with categories extracted by PLSA and questionnaires

Class06 Snack segment

Class07 Western breakfast meal segment
Example: Probabilistic inference using customer model (BN)

the consumer cluster buying quick/easy precooked foods in the evening
→ probability of answering "Not satisfying daily life" is High

Satisfying life = 'no'→Higher (Customer A)

Customer segment A: Buying precooked foods at evening

Customers who bought quick/easy precooked foods in the evening ⇒ answering Not satisfying daily life !

AI for service engineering

Living support applications

Data base

Preference, intent and behavior prediction

Information recommendation

User modeling

Big data

Probabilistic models

Artificial Intelligent Systems

PLSA and Bayesian networks (PLASMA)

Decision support systems

Phenomena modeling
AI applications for service systems
(Tablets’ applications for retail & restaurant service)

Questionnaires
Recommendation
Campaign promotion

POSEIDON application (AI for service)

used in Japanese restaurants
Product promotion

In store digital signage systems
AI applications in public space (social events)
(for supporting services and getting user’s feedback)

Digital signage
Motion detection using Kinect
for health care purpose
Science museum in Odaiba

AI for human life
(AI living lab)

Located in Real Town
PLASMA
IoT devices
Everyday Life Computing Package
Real Person Lives
Sensor Network
Life-log DB
Object DB
Bayesian network constructing from life log data

Predict/simulate human behavior by BN

P(watching TV| night, at living) = 0.9
P(watching TV| night, at kitchen) = 0.3
P(watching TV| morning, at living) = 0.6
P(watching TV| morning, at kitchen) = 0.1
Bayes estimation for pattern recognition

Bayesian network for Behavior prediction

velocity $x, y$: $V_{x,y}(t)$
- height: $Z(t)$
- Previous behavior: $C(t-1)$
Time study in medical services

• We developed a system consisting of 7’ Android tablet devices, bar-code scanner, and micro-server (WiFi, linux).
• Bar code scanners send the nurse and the patient ids to the tablets, and tablets sending the data to the nearest server with the location id.

Variance analysis using Bayesian networks for productivity improvement

Discretize the variance of time

Constructing Bayesian networks

Analysis

Understanding the causes of variances, evaluation, and simulation by Bayesian networks

Dependent variables: time variance
Independent variables: conditions
AI for service (supporting system in the hospital)

Proposed assistant system

Hospital room
Observer (intern)

Nurse (in charge)

In the ward
Assistant nurse

Tablet PC

In the Nurse station

Hospital

Data correcting and Analyzing server PC

Proposed service process optimization

Workflow/Environment

optimization

Education

To be

As is

Nurses, care giver and etc.
patients

Daily service process

Visualization and Data mining

Time, variance and conditions

Assist and navigation

Big data

Proposed system

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management

AI for service (supporting system in the hospital)

Daily monitoring
Reporting

Nursing director
Head nurse
Chief nurse

Management support
optimization

Guidelines

Management
Discussion

- Probabilistic modeling is core of AI technology
- Gap between real world problems and toy problems
- We need better evaluation and creation methods

→ Competition with practical regulation (start from 2017 Nov.)