AAAI2019 Tutorial

Behavior Analytics: Methods and Applications

Professor Longbing Cao

Data Science Lab

Advanced Analytics Institute, University of Technology Sydney, Australia

Web: www.datasciences.org | **Email**: Longbing.Cao@uts.edu.au

This tutorial covers both classic sequence-based behavior analysis and more recent research on behavior modeling, learning, and management by using statistical methods, deep learning etc.

Behaviors consist of transaction-based behaviors, sequential activities, actions, interactions, and visual behaviors of humans or non-human subjects.

Key terms

- Behavior analytics
- Behavior computing
- Behavior informatics
- Behavior learning
- Behavior modeling
- Behavior representation
- Behavior simulation
- Behavior imitation
- Behavior impact
- Behavior utility
- Behavior management

- Action
- Activity
- Behavior
- Event
- Interaction
- Occurring behavior
- Non-occurring behavior
- Individual behavior
- Group/collective behavior
- Positive behavior
- Negative behavior

Outline

1. Overview of Behavior Informatics
2. What is Behavior?
3. High Impact Behavior Sequence Analysis
4. Impact-oriented Combined Behavior Analysis
5. High Utility Behavior Analysis
6. Nonoccurring Behavior Analysis
7. Coupled/Group/Collective Behavior Analysis

Outline

8 Statistical Modeling of Coupled Behaviors

9 Probabilistic Modeling of Sparse Rating Behaviors

10 Understanding Behavior Drivers: Choice and Attraction

11 Behavior Analysis with Recurrent Networks

12 Behavior Analysis in Visual Data

13 Behavior Learning from Demonstrations

14 Challenges and Prospects

References Download

- www.behaviorinformatics.org

- http://www.datasciences.org/

Behavior Informatics: Overview

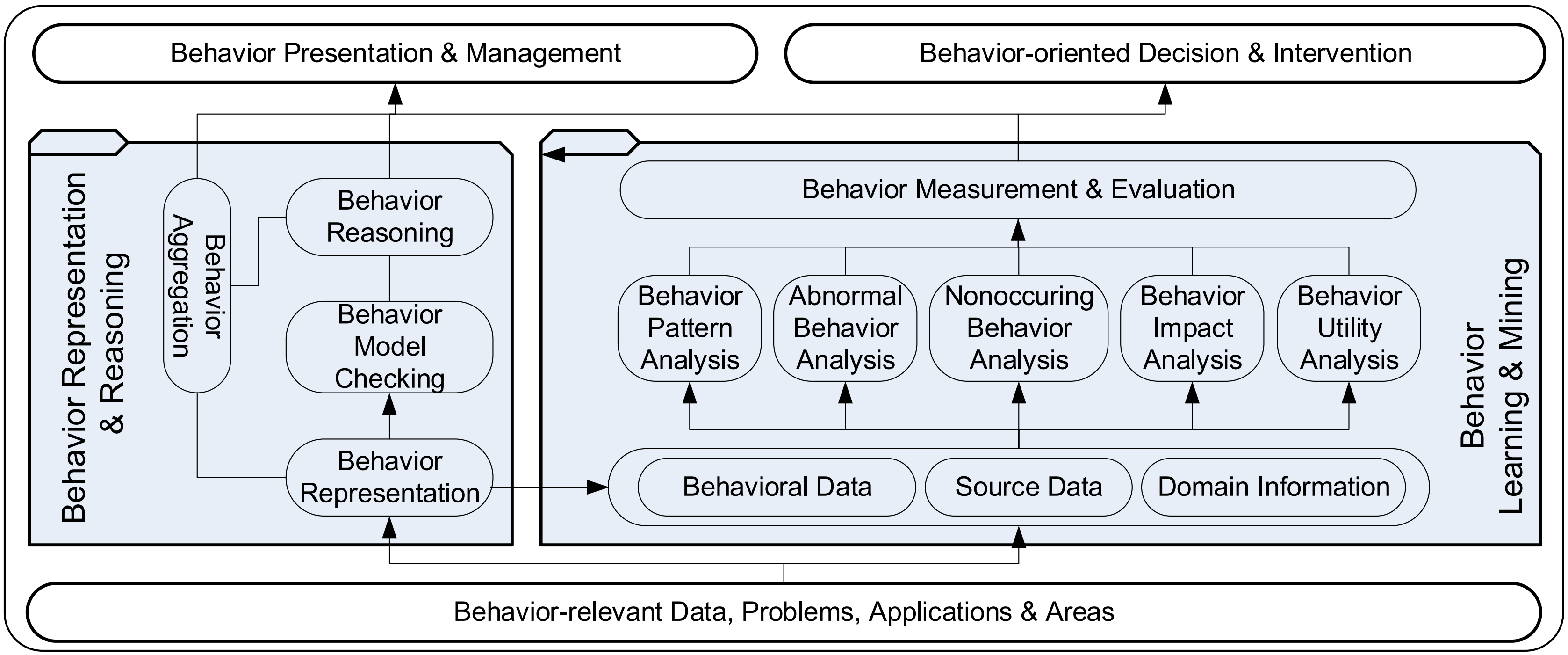
Longbing Cao, In-depth Behavior Understanding and Use: the Behavior Informatics Approach, *Information Science*, 180(17); 3067-3085, 2010.

Longbing Cao and Philip S Yu (eds) *Behavior Computing*, 21-35, Springer, 2012

IJCAI'2013 tutorial: behavior informatics

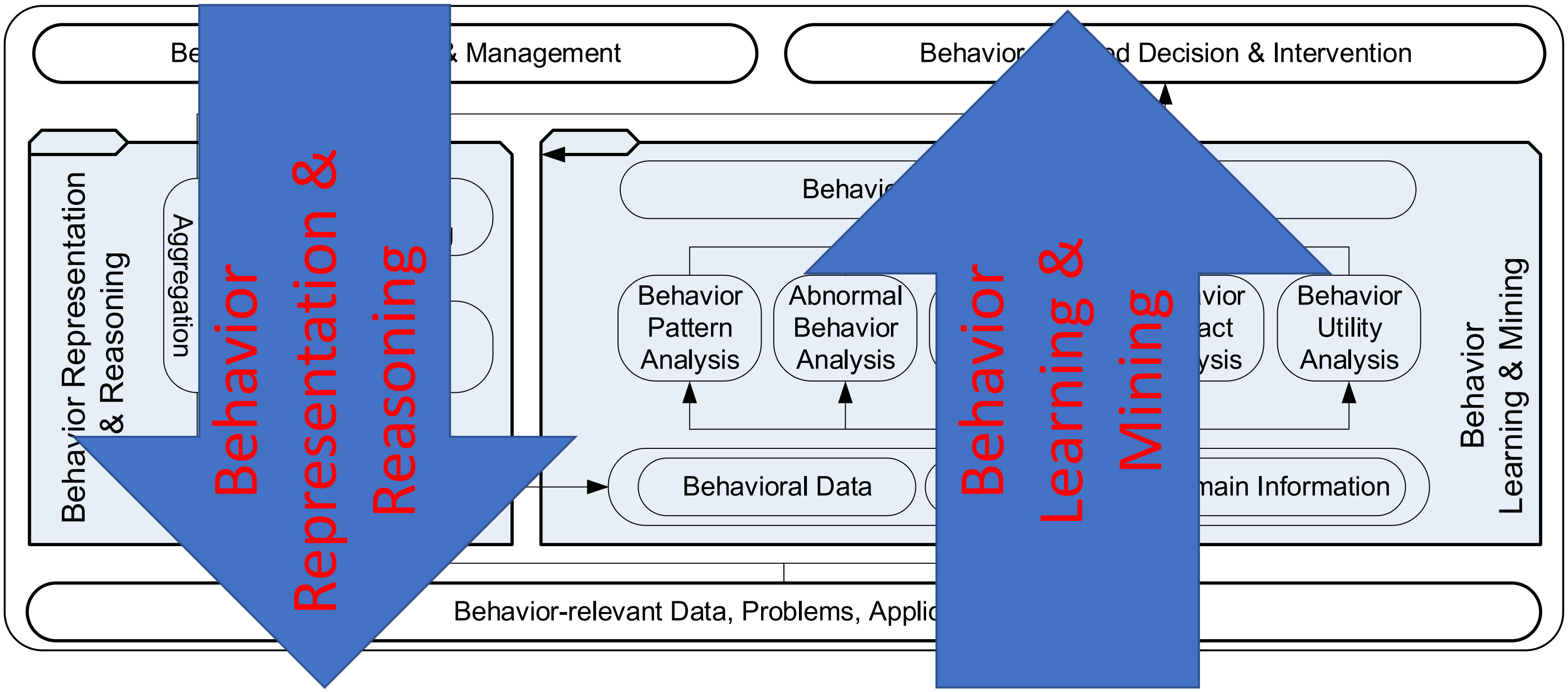
BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Behavior informatics – Concept Map



http://www.behaviorinformatics.org/

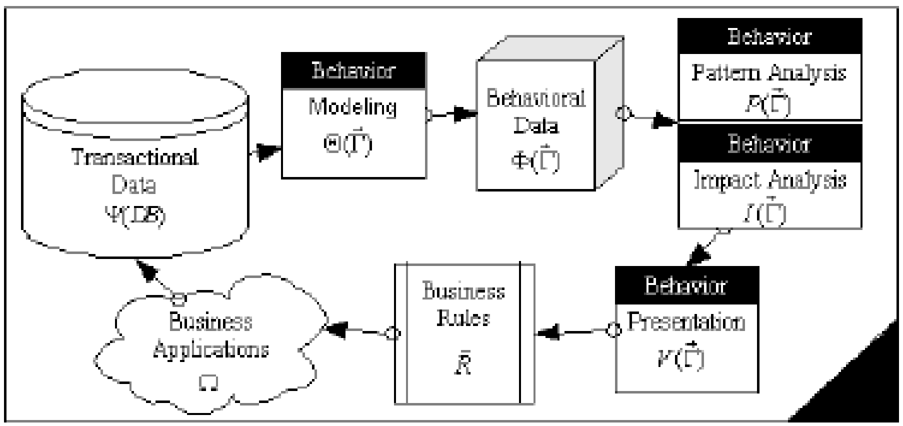
Behavior informatics – Concept Map



http://www.behaviorinformatics.org/

Behavior analysis process

- Behavior-centric modeling, analysis, and intervention



$$BIA : \Psi(DB) \xrightarrow{\Theta(\vec{\Gamma})} \vec{\Gamma} \xrightarrow{\Omega, e, c, t_i()} \widetilde{P} \xrightarrow{\Lambda, e, c, b_i()} \widetilde{R}$$

BIA PROCESS: The Process of Behavior Informatics and Analytics

INPUT: original dataset Ψ;

OUTPUT: behavior patterns $\widetilde{P}$ and operationalizable business rules $\widetilde{R}$;

Step 1: Behavior modeling $\Theta(\vec{\Gamma})$;
 Given dataset Ψ;
 Develop behavior modeling method θ ($\theta \in \Theta$) with technical interestingness $t_i()$;
 Employ method θ on the dataset Ψ;
 Construct behavior vector set $\vec{\Gamma}$;

Step 2: Converting to behavioral data $\Phi(\vec{\Gamma})$;
 Given behavior modeling method θ;
 FOR $j = 1$ to $(count(\Psi))$
  Deploy behavior modeling method θ on dataset Ψ;
  Construct behavior vector $\vec{\gamma}$;
 ENDFOR
 Construct behavior dataset $\Phi(\vec{\Gamma})$;

Step 3: Analyzing behavioral patterns $P\vec{\Gamma}$;
 Given behavior data $(\Phi(\vec{\Gamma})$;
 Design pattern mining method $\omega \in \Omega$;
 Employ the method ω on dataset $\Phi\vec{\Gamma}$;
 Extract behavior pattern set $\widetilde{P}$;

Step 4: Converting behavior patterns $\widetilde{P}$ to operationalizable business rules $\widetilde{R}$;
 Given behavior pattern set $\widetilde{P}$;
 Develop behavior modeling method Λ;
 Involve business interestingness $b_i()$ and constraints c in the environment e;
 Generate business rules $\widetilde{R}$;

Why Behavior Analytics Is Important?

Longbing Cao, Philip S Yu (Eds). Behavior Computing: Modeling, Analysis, Mining and Decision, Springer, 2012.

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

External Market Behavior

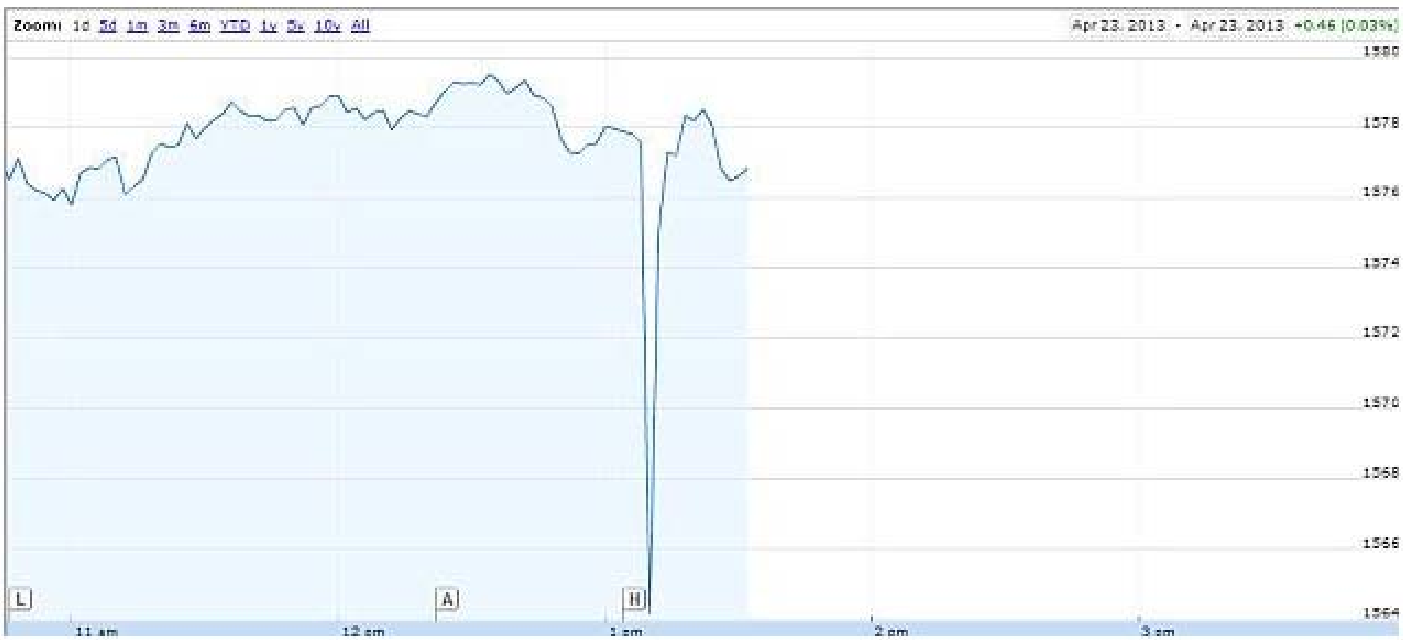
Plunge: Immediately after the false Twitter report the DOW Jones Industrial average plunged 100 points before going back up again following reassurances from both the AP and the White House

The Cause: Another Negative Behavior

Hackers: A group claiming responsibility for the hacking posted this apparent screen grab taken from their computer showing their infiltration

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

The Effect of Negative Behavior



Billions gone: Similarly seen the S&P 500 dropped about 1 percent wiping out $136.5 billion before almost as quickly recovering from their losses amid the scare

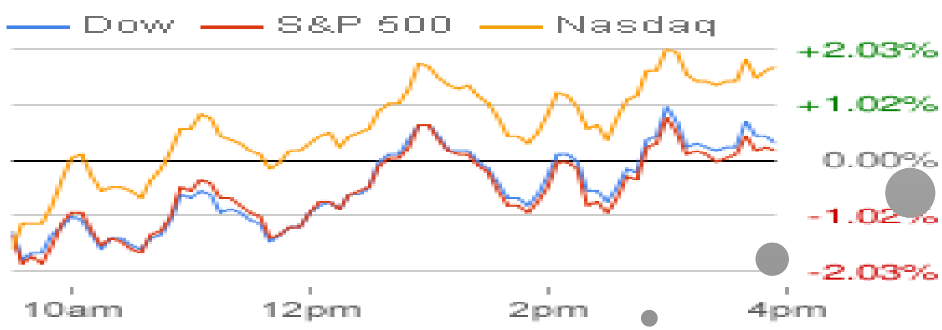
Argument 1: Behavior is ubiquitous

- Behavior is an important analysis object in
  - Consumer analysis
  - Marketing strategy design
  - Business intelligence
  - Customer relationship management
  - Social computing
  - Intrusion detection
  - Fraud detection
  - Event analysis
  - Risk analysis
  - Group decision-making, etc.

➢ Customer behavior analysis
➢ Consumer behavior and market strategy
➢ Web usage and user preference analysis
➢ Exceptional behavior analysis of terrorist and criminals
➢ Trading pattern analysis of investors in capital markets

Argument 2: Major work focuses on Behavior exterior-driven analysis

- Example 1: Price movement as market behavior

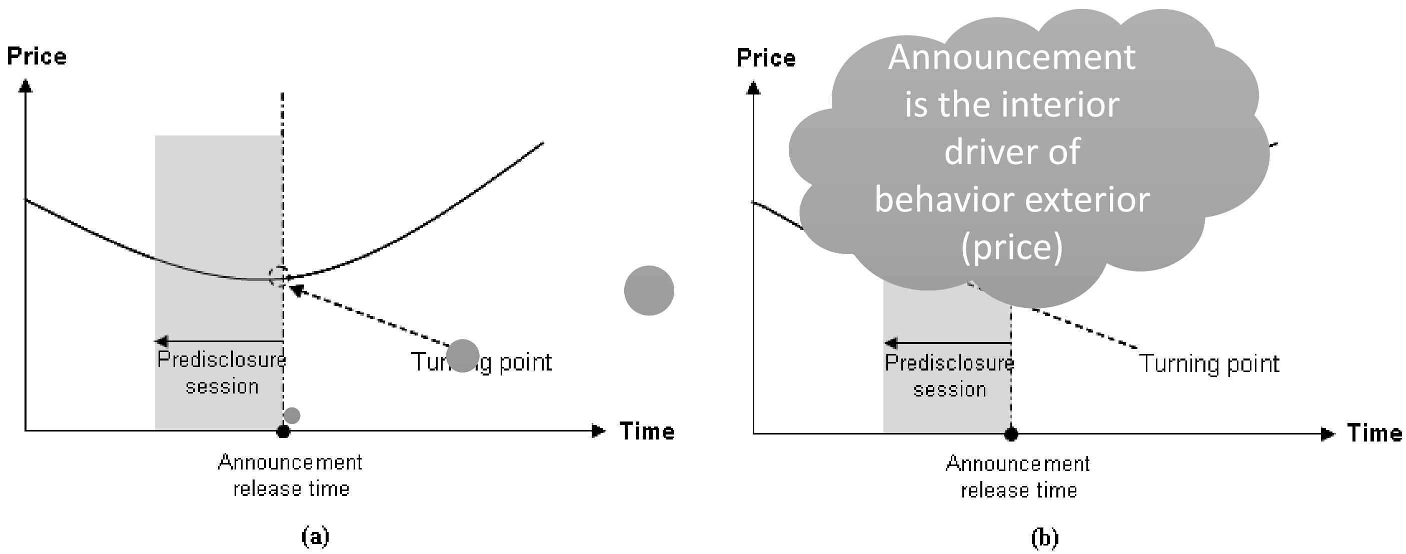


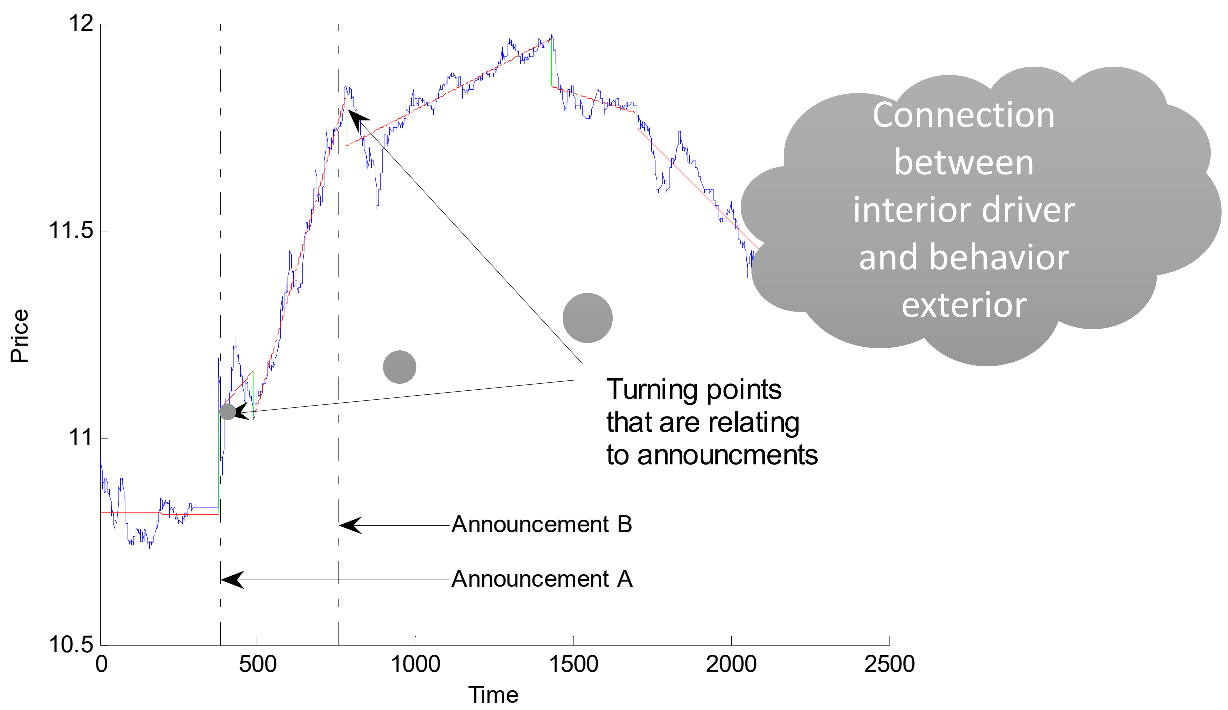
Dow	8,629.68	+64.59 (0.75%)
S&P 500	879.73	+6.14 (0.70%)
Nasdaq	1,540.72	+32.84 (2.18%)
10y bond	2.57%	+0.01 (0.39%)

Price/index movement is the behavior exterior

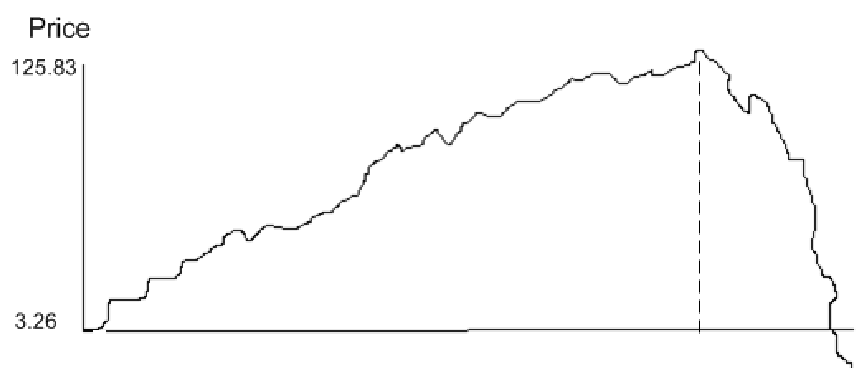
Argument 3: Behavior interior-driven analysis can make difference

- Example 2: Announcement as market behavior driver

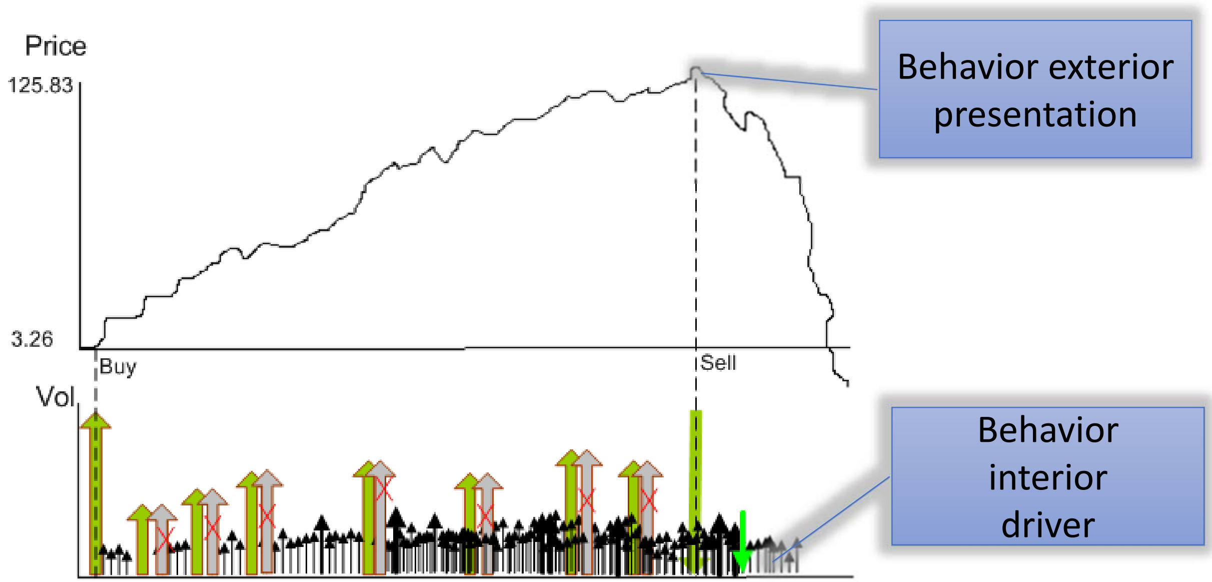


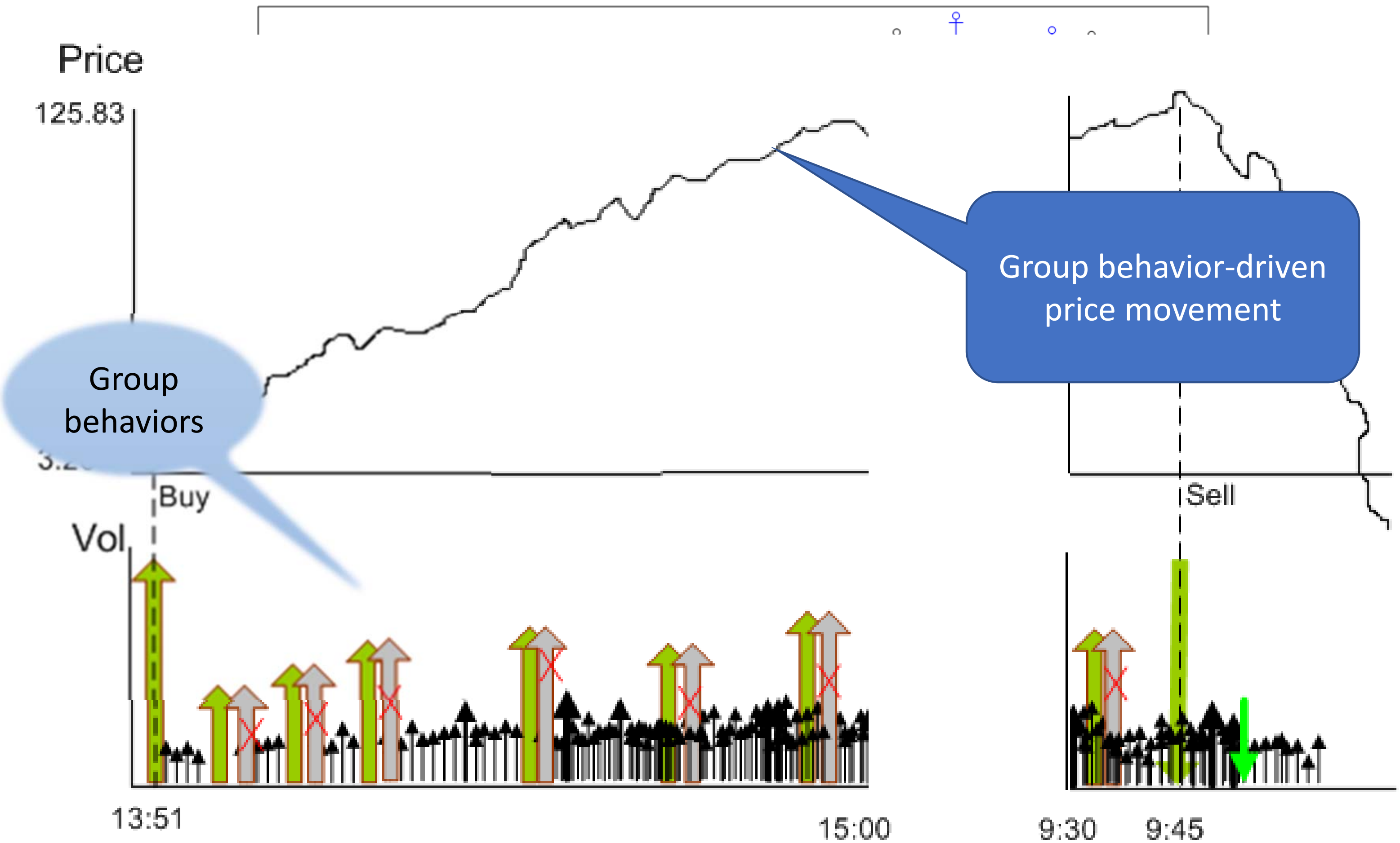

12
11.5
11
10.5
Price
0
500
1000
1500
2000
2500
Time
Connection between interior driver and behavior exterior
Turning points that are relating to announcments
Announcement B
Announcement A

- Why does this stock go so crazily?



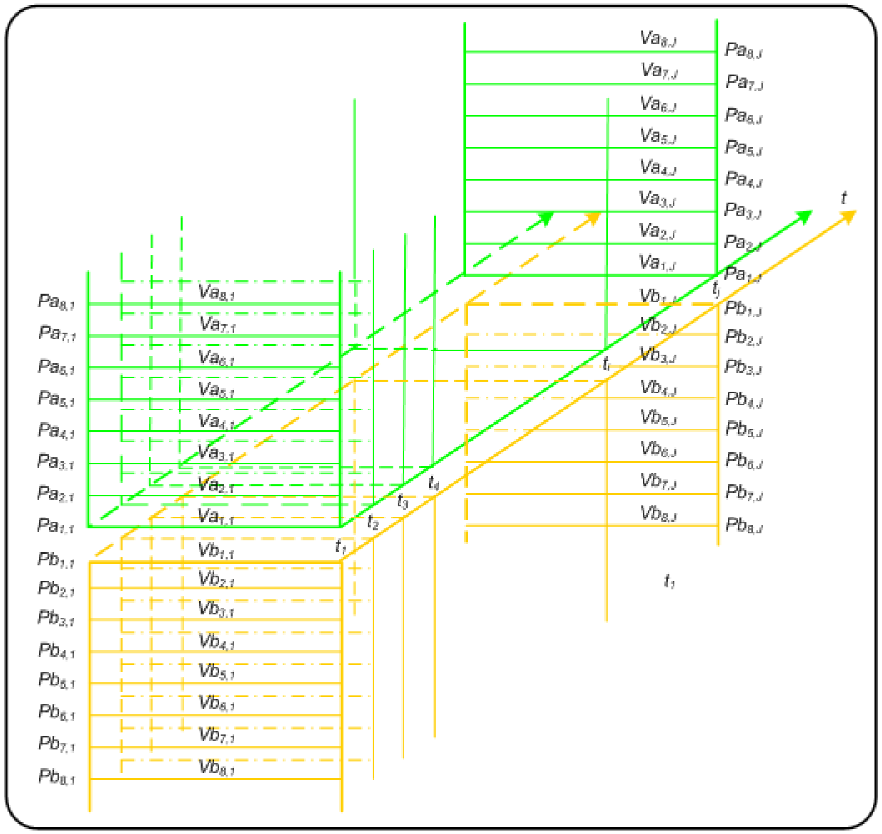
- Short-term manipulation behavior as cause




Price
125.83
Group behaviors
Buy
Vol
Group behavior-driven price movement
Sell
13:51
15:00
9:30
9:45

Argument 4: Need to consider behavior context

- Microstructure data



Observation: Traditional analysis on behavior

- Empirical, qualitative, psychological, social etc
- Behavior-oriented analysis was usually conducted on customer demographic and transactional data directly
  - Telecom churn analysis, customer demographic data and service usage data are analyzed to classify customers into loyal and non-loyal groups based on the dynamics of usage change
  - Outlier mining of trading behavior, price movement is usually focused to detect abnormal behavior

so-called behavior-oriented analysis *is actually not on customer behavior-oriented elements, rather on straightforward customer demographic data and business usage related appearance data (transactions)*

Problems with traditional behavior analysis

- Customer demographic and transactional data is not organized in terms of behavior but entity relationships
- Human behavior is *implicit* in normal transactional data: *behavior implication*
    - cannot support in-depth analysis on *behavior interior:* focus on *behavior exterior*
    - Cannot scrutinize behavioral actor's belief, desire, intention and impact on business appearance and problems

Such behavior implication indicates the limitation or even ineffectiveness of supporting behavior-oriented analysis on transactional data directly.

Genuine behavior analysis does matter

- Behavior plays the role as internal driving forces or causes for business appearance and problems
- Complement traditional pattern analysis solely relying on demographic and transactional data
- Disclose extra information and relationship between behavior and target business problem-solving

A multiple-dimensional viewpoint and solution may exist that can uncover problem-solving evidence from not only demographic and transactional but behavioral (including intentional, social, interactive and impact aspects) perspectives

Support genuine behavior analysis

- Make behavior 'explicit' by squeezing out behavior elements hidden in transactional data
- *A conversion from transactional space to behavior feature space is necessary*

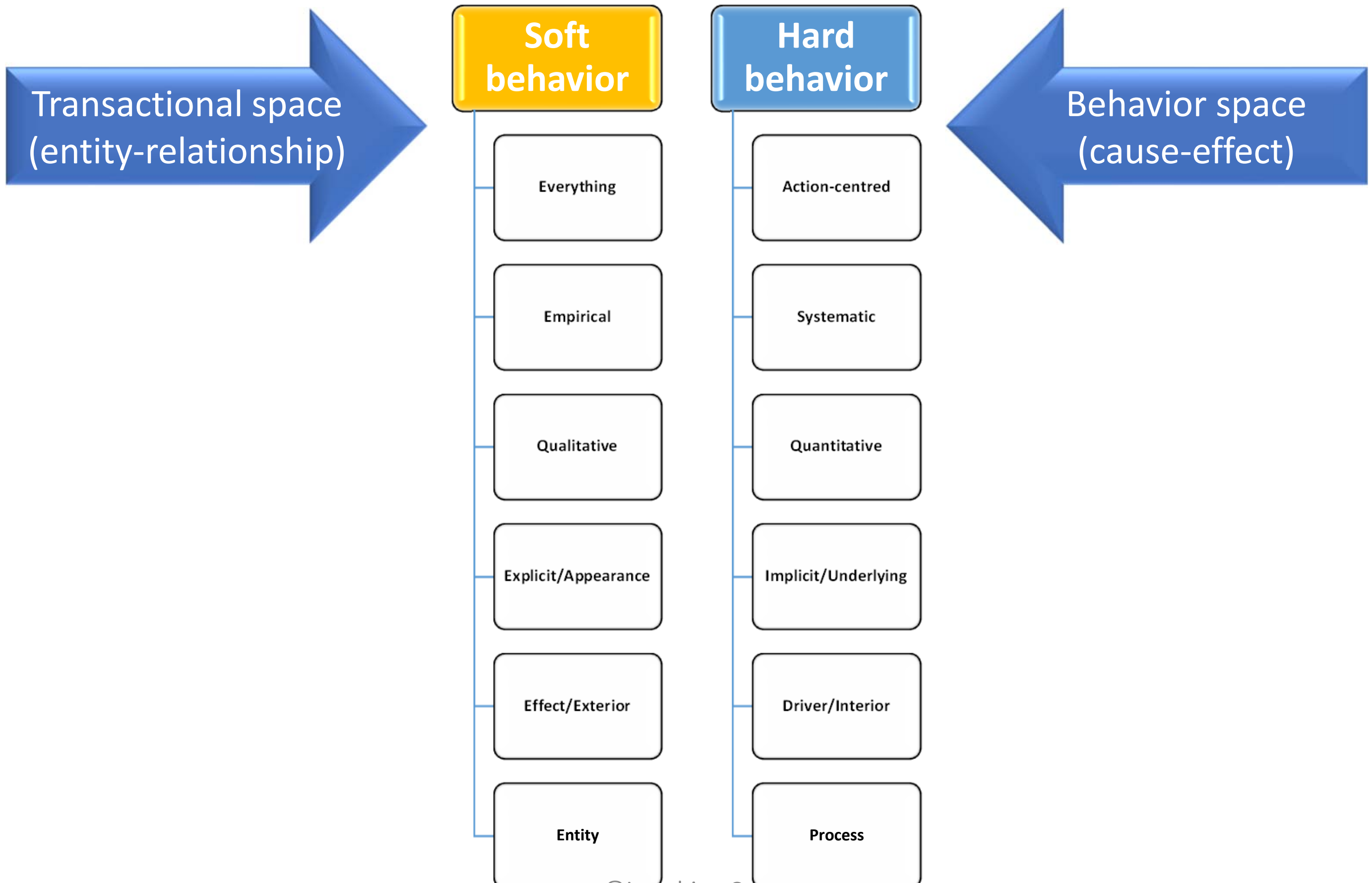
- Constructing behavioral data corresponding to the physical world
    - *behavior modeling and mapping*
    - organized in terms of behavior, behavior relationship and impact
- Developing behavior analytics theories and tools

Explicitly and more effectively analyze behavior patterns and behavior impacts than on transactional data

What is Behavior?

Longbing Cao, In-depth Behavior Understanding and Use: the Behavior Informatics Approach, *Information Science*, 180(17); 3067-3085, 2010.

Behavior: soft vs. hard



What is behavior?

- An abstract behavior model
  - Demographics and circumstances of behavioral subjects and objects
  - Associates of a behavior may form into certain behavior sequences or network;
  - Social behavioral network consists of sequences of behaviors that are organized in terms of certain social relationships or norms.
  - Impact, costs, risk and trust of behavior/behavior network

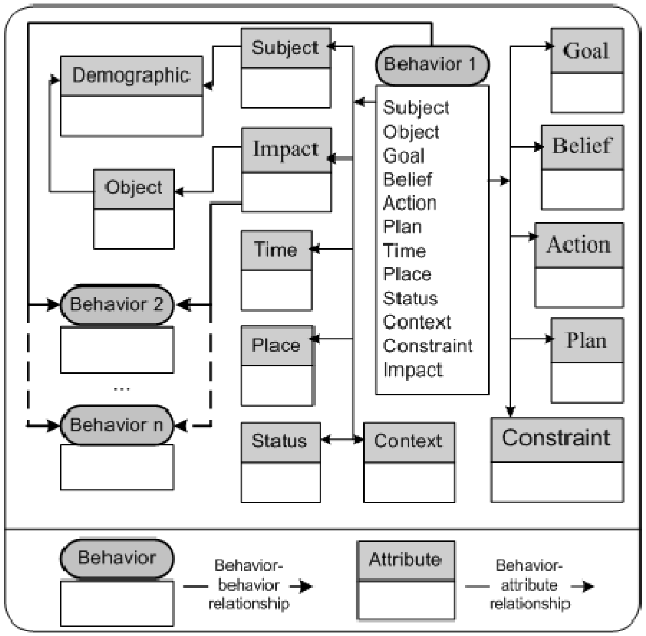


Figure 1. An Abstract Behavioral Model

Abstract Behavior Model

Definition 1. *A* ***behavior*** *($\mathbb{B}$) is described as a four-ingredient tuple* $\mathbb{B} = (\mathscr{E}, \mathscr{O}, \mathscr{C}, \mathscr{R})$,

- *Actor* $\mathscr{E} = \langle \mathcal{SE}, \mathcal{OE} \rangle$ *is the entity that issues a behavior (subject, $\mathcal{SE}$) or on which a behavior is imposed (object, $\mathcal{OE}$).*
- *Operation* $\mathscr{O} = \langle \mathcal{OA}, \mathcal{SA} \rangle$ *is what an actor conducts in order to achieve certain goals; both objective ($\mathcal{OA}$) and subjective ($\mathcal{SA}$) attributes are associated with an operation. Objective attributes may include time, place, status and restraint; while subjective aspects may refer to action and its actor's belief and goal etc of the behavior and the behavior impact on business.*
- *Context* $\mathscr{C}$ *is the environment in which a behavior takes place.*
- *Relationship* $\mathscr{R} = \langle \theta(\cdot), \eta(\cdot) \rangle$ *is a tuple which reveals complex interactions within an actor's behaviors (named intra-coupled behaviors, represented by function $\theta(\cdot)$) and that between multiple behaviors of different actors (inter-coupled behaviors by relationship function $\eta(\cdot)$).*

- Behavior instance: behavior vector

$$\vec{\gamma} = \{s, o, e, g, b, a, l, f, c, t, w, u, m\}$$

  - basic properties
  - social and organizational factors
- Vector-based behavior sequences

$$\vec{\Gamma} = \{\vec{\gamma}_1, \vec{\gamma}_2, ..., \vec{\gamma}_n\}$$

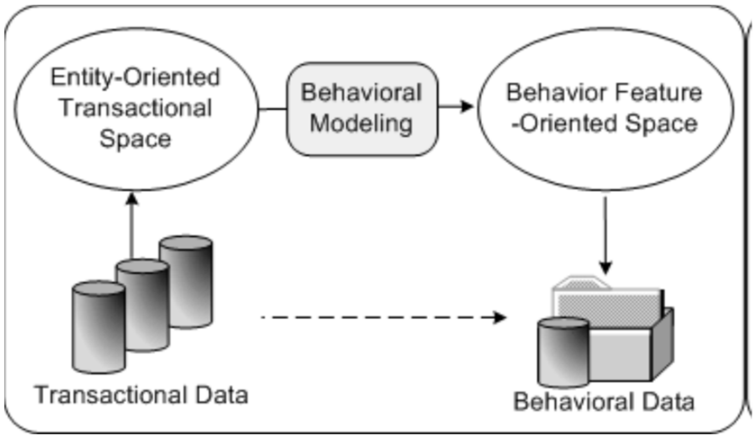
- Vector-oriented patterns

- Vector-oriented behavior pattern analysis
  - Behavior performer:
    - Subject (*s*), action (*a*), time (*t*), place (*w*)
  - Social information:
    - Object (*o*), context (*e*), constraints (*c*), associations (*m*)
  - Intentional information:
    - Subject's: goal (*g*), belief (*b*), plan (*l*)
  - Behavior performance:
    - Impact (*f*), status (*u*)

➢ *New methods for vector-based behavior pattern analysis*

Behavioral data

- Behavioral elements hidden or dispersed in transactional data
- *behavioral feature space*



- ➢ Behavioral data modeling
- ➢ Behavioral feature space
- ➢ Mapping from transactional to behavioral data
- ➢ Behavioral data processing
- ➢ Behavioral data transformation

Impact-oriented Sequential Patterns

Longbing Cao. Zhao Y., Zhang, C. Mining Impact-Targeted Activity Patterns in Imbalanced Data, *IEEE Trans. on Knowledge and Data Engineering*, 20(8): 1053-1066, 2008.

Longbing Cao, Huaifeng Zhang, Yanchang Zhao, Dan Luo, Chengqi Zhang. Combined Mining: Discovering Informative Knowledge in Complex Data, IEEE Trans. SMC Part B, 41(3): 699 - 712, 2011.

Basic paradigms

- Nonimpact-oriented combined patterns

$$\mathcal{P}_n : R_l(X_1 \wedge \cdots \wedge X_i) \rightarrow I_m \quad (3)$$

$$\mathcal{P} := \mathcal{G}(P_1 \wedge \cdots \wedge P_n) \rightarrow \mathcal{I} \quad (4)$$

- Impact-oriented combined patterns

$$P_n : \{R_l(X_1 \wedge \cdots \wedge X_i) \rightarrow I_m\} \rightarrow T_1 \quad (5)$$

$$\mathcal{P} := \mathcal{G}(P_1, \cdots, P_n) \quad (6)$$

Impact definition

- *Impact* is defined as a feasible detrimental outcome of an activity (action) or a sequence of activities (e.g., launch or operation of a spacecraft) subject to hazard(s)

(1) *magnitude (or severity)* of the adverse consequence(s) that can potentially result from the given activity, action or behavior, and

(2) *likelihood* of occurrence of the given adverse consequence(s).

Impact

- Business impact of behavior
  - Consequence:
    - Fraud
    - Debt
    - Exception ...
  - Magnitude:
    - Positive/negative
    - Multi-level
    - Ratio
    - Probabilistic

Impact modeling

- Impact measuring
  - Cost
  - Cost-sensitive
  - Profit
  - Cost-benefit
  - Risk score
  - ...
- Impact evolution
  - Positive → Negative
  - Negative → Positive

- *qualitative risk assessment:*
  - severity and likelihood are both expressed qualitatively (e.g., high, medium, or low)
- *quantitative risk assessment/probabilistic risk assessment:*
  - Consequences are expressed numerically
  - Their likelihoods of occurrence are expressed as *probabilities or frequencies*

Probabilistic Risk Assessment

- Causes/Initiators:
  - What can go wrong with the studied technological entity, or what are the *initiators or initiating events (undesirable starting events) that lead to adverse* consequence(s)?
- Effects/Consequences:
  - What and how severe are the potential detriments, or the adverse *consequences that the technological entity may be eventually subjected to as* a result of the occurrence of the initiator?
- Functions(cause, effect):
  - How likely to occur are these undesirable consequences, or what are their *probabilities or frequencies?*

- Risk of a pattern, e.g.,

$$Risk(P \rightarrow T) = \frac{Cost(P \rightarrow T)}{TotalCost(P)}$$

$$AvgCost(P \rightarrow T) = \frac{Cost(P \rightarrow T)}{Cnt(P \rightarrow T)}$$

Impact-Targeted Activity Mining

- Frequent impact-oriented activity patterns
- Frequent impact-contrasted activity patterns
- Sequential impact-reversed activity patterns

Here:
Impact → Debt, Fraud, Risk …

$\bar{T}$

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Impact-Oriented Activity Patterns

$\{P \dashrightarrow T\}$ or $\{P \dashrightarrow \bar{T}\}$ $\quad (P \dashrightarrow \bar{T}$, or $\bar{P} \dashrightarrow \bar{T})$

- *frequent positive impact-oriented (T) activity patterns*
  - $P \dashrightarrow T$, or $\bar{P} \dashrightarrow T$
- *frequent negative impact-oriented ( ) activity patterns*
  - $\bar{T}$, or $P \dashrightarrow \bar{T}$

    $\bar{P} \dashrightarrow \bar{T}$

P is an activity sequence, (P = {a_i, a_{i+1}, … }, i=0, 1, …).

Impact-contrasted Activity Patterns

- **Pattern:** *P* is of high significance in positive impact dataset, and of low significance in negative impact dataset, or vice versa.

$$\{P \to T,\ P \to \overline{T}\} \qquad \{P \to \overline{T},\ P \to T$$

- *Positive impact-contrasted pattern*

$$P_{T\overline{T}}: \{P \to T,\ P \to \overline{T}\}$$

- *Negative impact-contrasted pattern*

$$P_{\overline{T}T}: \{P \to \overline{T},\ P \to T\}$$

Impact-reversed Activity Patterns

$\{P \dashrightarrow T\}\ \{PQ \dashrightarrow \overline{T}\}\quad \{P \dashrightarrow \overline{T}\}\ \{PQ \dashrightarrow T\}$

- *Sequential impact-reversed activity pattern pair*
  - *underlying pattern*:
  - *derivative pattern:*

$$\begin{tikzcd}
\{P \dashrightarrow T\} \arrow[d] & \{P \dashrightarrow \overline{T}\} \arrow[d] \\
\{PQ \dashrightarrow \overline{T}\} & \{PQ \dashrightarrow T\}
\end{tikzcd}$$

Case Study

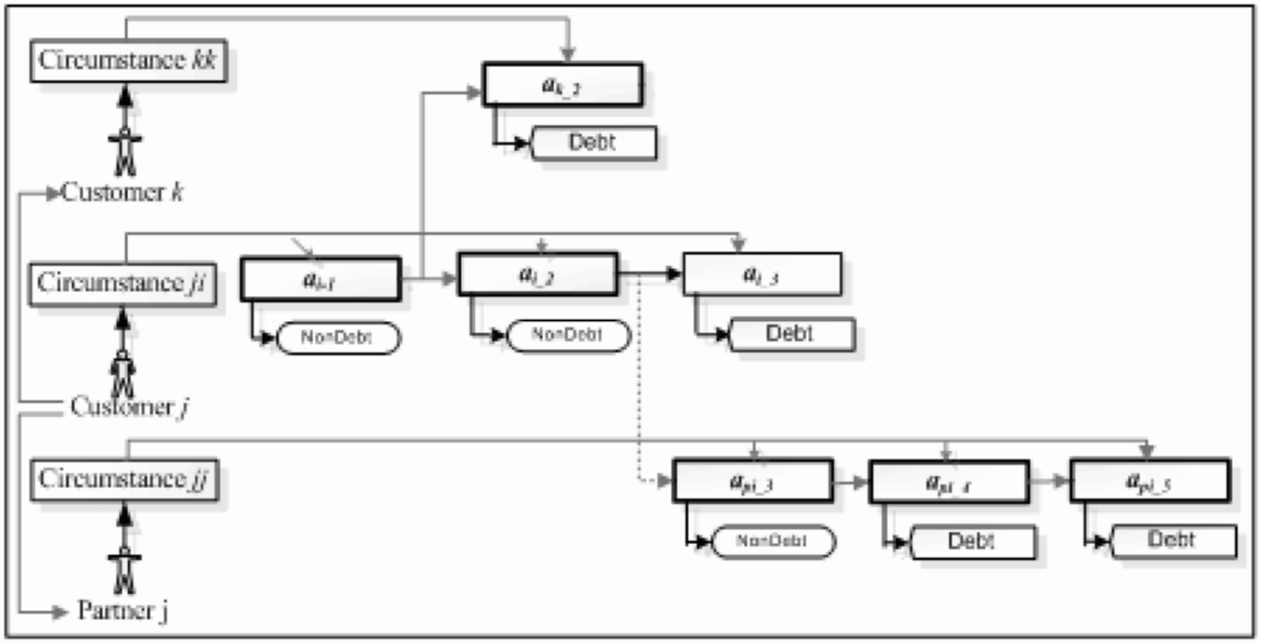
- Mining Combined Patterns and Patterns Clusters for Debt Recovery in social security

Business Problem

- To profile customers according to their capacity to pay off their debts in shortened timeframes.
- To target those customers with recovery and amount options suitable to their own circumstances, and increase the frequency and level of repayment.

Impact-oriented group behaviors

- Coupled group behaviors with impact



Raw Data

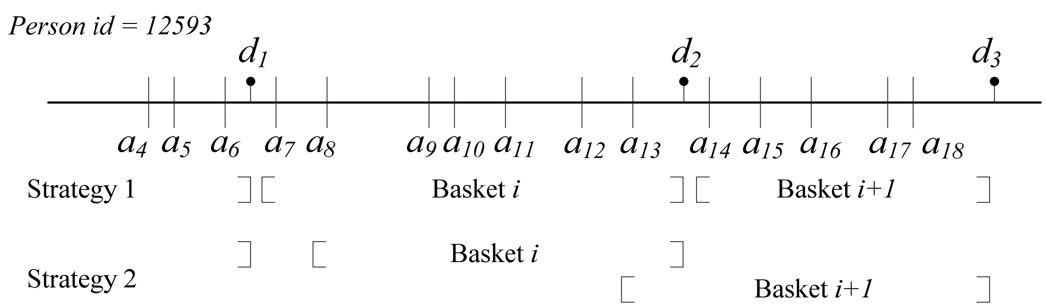
- Data:
  - No. of activity transactions: 15,932,832
  - No. of customers: 495,891
  - No. of debts: 30,546

Data

- Customer demographic data
  - Customer ID, gender, age, marital status, number of children, declared wages, location, benefit type, …
- Debt data
  - Debt amount, debt start/end date, …
- Repayment data (transactional)
  - Repayment method, amount, time, date, …
- Class ID: Quick/Moderate/Slow Payer

Constructing Activity Baskets and Sequences

- Positive-impact activity sequences: the activities before a debt are put in a basket. E.g., {*a8*, *a9*, *a10*, *a11*, *a12*, *a13*, ***d2***}, {*a13*, *a14*, *a15*, *a16*, *a17*, *a18*, ***d3***}



- Negative-impact activity sequences
  A virtual activity "NDT" is created for those customers have never had a debt.

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Examples of Debt/Non-Debt Activity Sequences

Table 1. Example of an activity sequence associated with a debt from target dataset a15, a9, a18, a19, a16, a9, DET

ACTIVITY CODE	START DATE	TIME
a_{15}	15/02/2006	13:34:05
a_9	16/02/2006	16:26:16
a_{18}	16/02/2006	16:26:17
a_{19}	20/02/2006	16:12:35
a_{16}	28/02/2006	11:27:50
a_9	1/03/2006	13:50:03
Debt	1/03/2006	23:59:59

Table 2. Example of an activity sequence related to non-debt from non-target dataset a14, a16, a1, a20, a14, a21, a22, NDT

ACTIVITY CODE	START DATE	TIME
a_{14}	6/02/2006	2:19:37
a_{16}	6/02/2006	10:21:50
a_1	7/02/2006	3:51:07
a_{20}	7/02/2006	4:44:48
a_{14}	7/02/2006	9:48:59
a_{21}	8/02/2006	10:03:13
a_{22}	15/02/2006	13:55:39
No-Debt	15/02/2006	23:59:59

Impact-contrasted sequential patterns

$P_{T\bar{T}}: \{P \rightarrow T,\ P \rightarrow \bar{T}\}$ $\qquad$ $P_{\bar{T}T}: \{P \rightarrow \bar{T},\ P \rightarrow T\}$

TABLE 8
Common Frequent Sequential Patterns in Separate Data Sets

Patterns (P)	$Supp_{D_T}(P)$	$Supp_{D_{\bar{T}}}(P)$	$Cd_{T,\bar{T}}(P)$	$Cdr_{T,\bar{T}}(P)$	$Cd_{\bar{T},T}(P)$	$Cdr_{\bar{T},T}(P)$	$AvgAmt$ (cents)	$AvgDur$ (days)	$risk_{amt}$	$risk_{dur}$
a_5	0.382	0.178	0.204	2.15	-0.204	0.47	18290	6.2	0.363	0.334
a_7	0.312	0.154	0.157	2.02	-0.157	0.50	19090	6.8	0.310	0.300
a_6	0.367	0.257	0.110	1.43	-0.110	0.70	18947	7.2	0.362	0.370
a_{14}	0.903	0.684	0.219	1.32	-0.219	0.76	19251	6.6	0.905	0.840
a_{15}	0.746	0.567	0.179	1.32	-0.179	0.76	19192	7.4	0.745	0.775
a_{16}	0.604	0.597	0.007	1.01	-0.007	0.99	17434	8.7	0.548	0.738
a_{14}, a_{15}	0.605	0.374	0.231	1.62	-0.231	0.62	19235	7.0	0.606	0.594
a_{15}, a_{15}	0.539	0.373	0.167	1.45	-0.167	0.69	18918	7.7	0.531	0.584
a_{16}, a_{14}	0.479	0.402	0.076	1.19	-0.076	0.84	16726	8.1	0.417	0.549
a_{14}, a_{16}	0.441	0.393	0.049	1.12	-0.049	0.89	17013	8.5	0.391	0.532
a_{16}, a_{16}	0.367	0.410	-0.043	0.90	0.043	1.12	14627	9.6	0.279	0.496
a_{14}, a_{14}, a_{15}	0.477	0.257	0.220	1.85	-0.220	0.54	19087	6.5	0.474	0.437
a_{14}, a_{15}, a_{14}	0.435	0.255	0.179	1.70	-0.179	0.59	18279	6.7	0.413	0.412
a_{16}, a_{14}, a_{14}	0.361	0.267	0.093	1.35	-0.093	0.74	16092	7.6	0.302	0.387
a_{16}, a_{14}, a_{16}	0.265	0.255	0.010	1.04	-0.010	0.96	14262	9.3	0.197	0.346

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Impact-reversed sequential patterns

$\{P \dashrightarrow T\}, \{PQ \dashrightarrow \bar{T}\}$ $\{P \dashrightarrow \bar{T}\}$ $\{PQ \dashrightarrow T\}$

TABLE 9
Impact-Reversed Sequential Activity Patterns in Separate Data Sets

Underlying sequence (P)	Impact 1	Derivative activity Q	Impact 2	Cir	Cps	Local support of $P \to$ Impact 1	Local support of $PQ \to$ Impact 2
a_{14}	$\bar{T}$	a_4	T	2.5	0.013	0.684	0.428
a_{16}	$\bar{T}$	a_4	T	2.2	0.005	0.597	0.147
a_{14}	$\bar{T}$	a_5	T	2.0	0.007	0.684	0.292
a_{16}	$\bar{T}$	a_7	T	1.8	0.004	0.597	0.156
a_{14}	$\bar{T}$	a_7	T	1.7	0.005	0.684	0.243
a_{15}	$\bar{T}$	a_5	T	1.7	0.007	0.567	0.262
a_{14}, a_{14}	$\bar{T}$	a_4	T	2.3	0.016	0.474	0.367
a_{16}, a_{14}	$\bar{T}$	a_5	T	2.0	0.006	0.402	0.133
a_{14}, a_{16}	$\bar{T}$	a_5	T	2.0	0.005	0.393	0.118
a_{16}, a_{15}	$\bar{T}$	a_5	T	1.8	0.006	0.339	0.128
a_{15}, a_{14}	$\bar{T}$	a_5	T	1.7	0.007	0.381	0.179
a_{16}, a_{14}	$\bar{T}$	a_7	T	1.6	0.004	0.402	0.108
a_{14}, a_{16}, a_{14}	$\bar{T}$	a_{15}	T	1.2	0.005	0.248	0.188
a_{16}, a_{14}, a_{14}	$\bar{T}$	a_{15}	T	1.2	0.005	0.267	0.220

Impact-oriented Combined Sequential Patterns

Longbing Cao. Combined Mining: Analyzing Object and Pattern Relations for Discovering and Constructing Complex but Actionable Patterns, WIREs Data Mining and Knowledge Discovery.

Longbing Cao, Huaifeng Zhang, Yanchang Zhao, Dan Luo, Chengqi Zhang. Combined Mining: Discovering Informative Knowledge in Complex Data, IEEE Trans. SMC Part B, 41(3): 699 - 712, 2011.

Yanchang Zhao, Huaifeng Zhang, **Longbing Cao,** Chengqi Zhang. Combined Pattern Mining: from Learned Rules to Actionable Knowledge, LNCS 5360/2008, 393-403, 2008.

Number of constituent atoms

- Pair patterns

$$\mathcal{P} ::= \mathcal{G}(P_1, P_2)$$

- Cluster patterns

$$\mathcal{P} ::= \mathcal{G}(P_1, \ldots, P_n)(n > 2)$$

Structural relations

- Peer-to-peer patterns

$$\mathcal{P} ::= P_1 \cup P_2$$

- Master-slave patterns

$$\{\mathcal{P} ::= P_1 \cup P_2, P_2 = f(P_1)\}$$

- Hierarchical patterns

$$\{\mathcal{P} ::= P_i \cup P_i' \cup P_j \cup P_j', P_j = \mathcal{G}(P_i), \ldots, P_j' = \mathcal{G}'(P_i)'\}$$

Time frame

- Independent patterns

$$\{P_1 : P_2\}$$

- Sequential patterns

$$\{P_1 ; P_2\}$$

- Hybrid patterns

$$\{P_1 \otimes P_2 \cdots \otimes P_n; \otimes \in \{:, \|, ;\}\}$$

Combined Pattern Pairs

- A combined rule pair is composed of two contrasting rules.
- Eg,. for customers with the same characteristics U, different policies/campaigns, V_1 and V_2, can result in different outcomes, T_1 and T_2.

DEFINITION COMBINED PATTERN PAIRS. *For impact-oriented combined patterns, a Combined Pattern Pair (CPP) is in the form of*

$$\mathcal{P}: \begin{cases} X_1 \to T_1 \\ X_2 \to T_2 \end{cases},$$

where 1) $X_1 \cap X_2 = X_{\mathrm{p}}$ *and* X_{p} *is called the* prefix *of pair* $\mathcal{P}$; $X_{1,e} = X_1 \setminus X_{\mathrm{p}}$ *and* $X_{2,e} = X_2 \setminus X_{\mathrm{p}}$; *2)* X_1 *and* X_2 *are different itemsets; and 3)* T_1 *and* T_2 *are contrary to each other, or* T_1 *and* T_2 *are same but there is a big difference in the interestingness (say confidences* $conf$*) of the two patterns.*

Interestingness of Pattern Pairs

$$I_{\text{pair}}(\mathcal{P}) = \begin{cases} |Conf(P_1) - Conf(P_2)|, & \text{if } T_1 = T_2; \\ \sqrt{Conf(P_1)\ Conf(P_2)}, & \text{if } T_1 \text{ and } T_2 \text{ are contrary}; \\ 0, & \text{otherwise}; \end{cases}$$

Combined Pattern Clusters

- Based on a combined rule pair, related combined rules can be organized into a cluster to supplement more information to the rule pair.
- The rules in cluster C have the same U but different V , which makes them associated with various results T.

DEFINITION COMBINED PATTERN CLUSTERS. *Assume there are k local patterns $X_i \to T_i, (i = 1, \ldots, k)$, $k \geq 3$ and $X_1 \cap X_2 \cap \cdots \cap X_k = X_{\mathrm{p}}$, a combined pattern cluster (CPC) is in the form of*

$$\mathcal{C}: \begin{cases} X_1 \to T_1 \\ \cdots \\ X_k \to T_k \end{cases},$$

where X_{p} is the prefix *of cluster $\mathcal{C}$.*

Interestingness of Pattern Clusters

$$I_{\text{cluster}}(\mathcal{C}) = \max_{P_i, P_j \in \mathcal{C}, i \neq j} I_{\text{pair}}(P_i, P_j)$$

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Interestingness of Rule Pair/Cluster

- dist(): the dissimilarity between the descendants of R_1 and R_2
- The interestingness of combined rule pair/cluster is decided by both the interestingness of rules and the most contrasting rules within the pair/cluster.
- A cluster made of contrasting confident rules is interesting, because it explains why different results occur and what can be done to produce an expected result or avoid an undesirable consequence.

$$I_{\text{pair}}(\mathcal{P}) = Lift_V(R_1)\; Lift_V(R_2)\; dist(T_1, T_2)$$

$$I_{\text{cluster}}(\mathcal{C}) = \max_{i \neq j, R_i, R_j \in \mathcal{C}, T_i \neq T_j} I_{\text{pair}}(R_i, R_j)$$

Rule Pair vs Rule Cluster

- From P, we can see that V_1 is a preferable policy for customers with characteristics U.
- If, for some reason, policy V_1 is inapplicable to the specific customer group, P is no longer actionable.
- Rule cluster C suggests that another policy V_3 can be employed to retain those customers.

$$\mathcal{P}:\begin{cases} U \wedge V_1 \rightarrow stay \\ U \wedge V_2 \rightarrow churn \end{cases}, \qquad \mathcal{C}:\begin{cases} U \wedge V_1 \rightarrow stay \\ U \wedge V_2 \rightarrow churn \\ U \wedge V_3 \rightarrow stay \end{cases}.$$

Extended Combined Pattern Pairs

DEFINITION EXTENDED COMBINED PATTERN PAIRS. *An Extended Combined Pattern Pair (ECPP) is a special combined pattern pair as follows*

$$\mathcal{E}: \begin{cases} X_{\mathrm{p}} \rightarrow T_1 \\ X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T_2 \end{cases},$$

where $X_{\mathrm{p}} \neq \emptyset$, $X_{\mathrm{e}} \neq \emptyset$ *and* $X_{\mathrm{p}} \cap X_{\mathrm{e}} = \emptyset$.

Contribution

DEFINITION CONTRIBUTION. *For a multi-feature combined pattern* $P : X \rightarrow T$, *where* $X = X_{\mathrm{p}} \wedge X_{\mathrm{e}}$, *the* contribution *of* X_{e} *to the occurrence of outcome* T *in rule* P *is*

$$Cont_{\mathrm{e}}(P) = \frac{Lift(X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T)}{Lift(X_{\mathrm{p}} \rightarrow T)} = \frac{Conf(X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T)}{Conf(X_{\mathrm{p}} \rightarrow T)}$$

$Cont_{\mathrm{e}}(P)$ is the lift of X_{e} with X_{p} as a precondition, which shows how much X_{e} contributes to the rule. *Contribution* can be taken as the increase of *lift* by appending additional items X_{e} to a rule. Its value falls in $[0, +\infty)$. A *contribution* greater than one means that the additional items in the rule contribute to the occurrence of the outcome, and a *contribution* less than one suggests that it incurs a reverse effect.

Conditional P-S ratio

DEFINITION *A metric for measuring the difference led by the occurrence of* X_e *in the above scenario is* Conditional Piatetsky-Shapiro's *(P-S)* ratio Cps*, which is defined as follows.*

$$Cps(X_e \to T|X_p) = Prob(X_e \to T|X_p) - Prob(X_e|X_p) \times Prob(T|X_p)$$

$$= \frac{Prob(X_p \wedge X_e \to T)}{Prob(X_p)} - \frac{Prob(X_p \wedge X_e)}{Prob(X_p)} \times \frac{Prob(X_p \to T)}{Prob(X_p)}$$

Interestingness of Combined Pattern

$$I_{\mathrm{rule}}(X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T) = \frac{Cont_{\mathrm{e}}(X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T)}{Lift(X_{\mathrm{e}} \rightarrow T)}$$

I_{rule} indicates whether the *contribution* of X_{p} (or X_{e}) to the occurrence of T increases with X_{e} (or X_{p}) as a precondition. Therefore, "$I_{\mathrm{rule}} < 1$" suggests that $X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T$ is less interesting than $X_{\mathrm{p}} \rightarrow T$ and $X_{\mathrm{e}} \rightarrow T$. The value of I_{rule} falls in $[0, +\infty)$. When $I_{\mathrm{rule}} > 1$, the higher I_{rule} is, the more interesting the rule is.

Extended Combined Pattern Clusters

DEFINITION EXTENDED COMBINED PATTERN SEQUENCES. *An Extended Combined Pattern Sequence (ECPC), or called Incremental Combined Pattern Sequence (ICPS), is a special combined pattern cluster with additional items appending to the adjacent local patterns incrementally.*

$$\mathcal{S}: \begin{cases} X_{\mathrm{p}} \rightarrow T_1 \\ X_{\mathrm{p}} \wedge X_{\mathrm{e},1} \rightarrow T_2 \\ X_{\mathrm{p}} \wedge X_{\mathrm{e},1} \wedge X_{\mathrm{e},2} \rightarrow T_3 \\ \cdots \\ X_{\mathrm{p}} \wedge X_{\mathrm{e},1} \wedge X_{\mathrm{e},2} \wedge \cdots \wedge X_{\mathrm{e,k-1}} \rightarrow T_k \end{cases},$$

where $\forall i,\ 1 \leq i \leq k-1,\ X_{i+1} \cap X_i = X_i$ *and* $X_{i+1} \setminus X_i = X_{e,i} \neq \emptyset$, *i.e.,* X_{i+1} *is an* increment *of* X_i. *The above cluster of rules actually makes a sequence of rules, which can show the impact of the increment of patterns on the outcomes.*

Impact

DEFINITION IMPACT. *The* impact *of* X_e *on the outcome in the rule is*

$$impact_e(P) = \begin{cases} cont_e(P) - 1 & : \textit{if } cont_e(P) \geq 1, \\ \frac{1}{cont_e(P)} - 1 & : otherwise. \end{cases}$$

Intervention Strategy 1

- Type A: Demographics differentiated combined pattern
  - Customers with the same actions but different demographics
    → different classes/business impact

$$\text{Type A:} \begin{cases} A_1 + D_1 & \rightarrow \text{quick payer} \\ A_1 + D_2 & \rightarrow \text{moderate payer} \\ A_1 + D_3 & \rightarrow \text{slow payer} \end{cases}$$

Intervention Strategy 2

- Type B: Action differentiated combined pattern
  - Customers with the same demographics but taking different actions
    → different classes/business impact

$$
\text{Type B:} \left\{ \begin{array}{lcl} A_1 + D_1 & \rightarrow & \text{quick payer} \\ A_2 + D_1 & \rightarrow & \text{moderate payer} \\ A_3 + D_1 & \rightarrow & \text{slow payer} \end{array} \right.
$$

Business Impact

- Able to move customers from one class to another class
- Useful for designing business policy

	Behavior 1	Behavior 2
Demographic 1	Slow	Fast
Demographic 2	Fast	Slow

Results (2)

Traditional Association Rules

V		T	$Conf(\%)$	$Count$	$Lift$
Arrangement	Repayment	Class			
irregular	cash or post office	A	82.4	4088	1.8
withholding	cash or post office	A	87.6	13354	1.9
withholding & irregular	cash or post office	A	72.4	894	1.6
withholding & irregular	cash or post office & withholding	B	60.4	1422	1.7

An Example of Combined Patterns

Rules	X_p	X_e		T	Cnt	$Conf$	I_r	$Lift$	$Cont_p$	$Cont_e$	$Lift$ of	$Lift$ of
	Demographics	Arrangements	Repayments	Class		(%)					$X_p \to T$	$X_e \to T$
P_1	age:65+	withholding & irregular	withholding	C	50	63.3	2.91	3.40	2.47	4.01	0.85	1.38
P_2	income:0 & remote:Y & marrital:sep & gender:F	withholding	cash or post & withholding	B	20	69.0	1.47	1.95	1.34	2.15	0.91	1.46
P_3	income:0 & age:65+	withholding	cash or post & withholding	A	1123	62.3	1.38	1.35	1.72	1.09	1.24	0.79
P_4	income:0 & gender:F & benefit:P	withholding	cash or post	A	469	93.8	1.36	2.04	1.07	2.59	0.79	1.90

Results (3)

An Example of Combined Pattern Clusters

Clusters	Rules	X_p demographics	X_e arrangements	X_e repayments	T	Cnt	$Conf$ (%)	I_r	I_c	$Lift$	$Cont_p$	$Cont_e$	$Lift$ of $X_p \rightarrow T$	$Lift$ of $X_e \rightarrow T$
$\mathcal{P}_1$	P_5	marital:sin	irregular	cash or post	A	400	83.0	1.12	0.67	1.80	1.01	2.00	0.90	1.79
	P_6	&gender:F	withhold	cash or post	A	520	78.4	1.00		1.70	0.89	1.89	0.90	1.90
	P_7	&benefit:N	withhold & irregular	cash or post & withhold	B	119	80.4	1.21		2.28	1.33	2.06	1.10	1.71
	P_8		withhold	cash or post & withhold	B	643	61.2	1.07		1.73	1.19	1.57	1.10	1.46
	P_9		withhold & vol. deduct	withhold & direct debit	B	237	60.6	0.97		1.72	1.07	1.55	1.10	1.60
	P_{10}		cash	agent	C	33	60.0	1.12		3.23	1.18	3.07	1.05	2.74
$\mathcal{P}_2$	P_{11}	age:65+	withhold	cash or post	A	1980	93.3	0.86	0.59	2.02	1.06	1.63	1.24	1.90
	P_{12}		irregular	cash or post	A	462	88.7	0.87		1.92	1.08	1.55	1.24	1.79
	P_{13}		withhold & irregular	cash or post	A	132	85.7	0.96		1.86	1.18	1.50	1.24	1.57
	P_{14}		withhold & irregular	withhold	C	50	63.3	2.91		3.40	2.47	4.01	0.85	1.38

Business Rule

BUSINESS RULES: Customer Demographic-Arrangement-Repayment combination business rules

For All *customer* i *(*$i \in I$ *is the number of valid customers)*
Condition:
satisfies *S/he is a debtor aged 65 or plus*;
relates
S/he is under arrangement of 'withholding' and 'irregularly',
and
His/her favorite Repayment method is 'withholding';
Operation:
Alert = "*S/he has 'High' risk of paying off debt in a very long timeframe.*"
Action = "*Try other arrangements and repayments in* R_2, *such as trying to persuade her/him to repay under 'irregular' arrangement with 'cash or post'.*"
End-All

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Combined pattern presentation

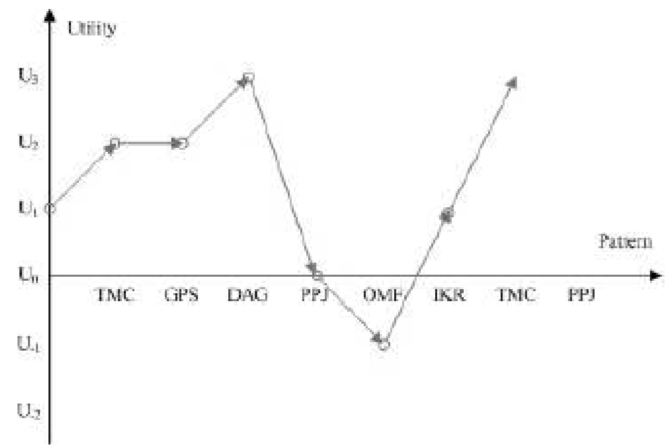


Figure 2: Pattern Evolution Chart

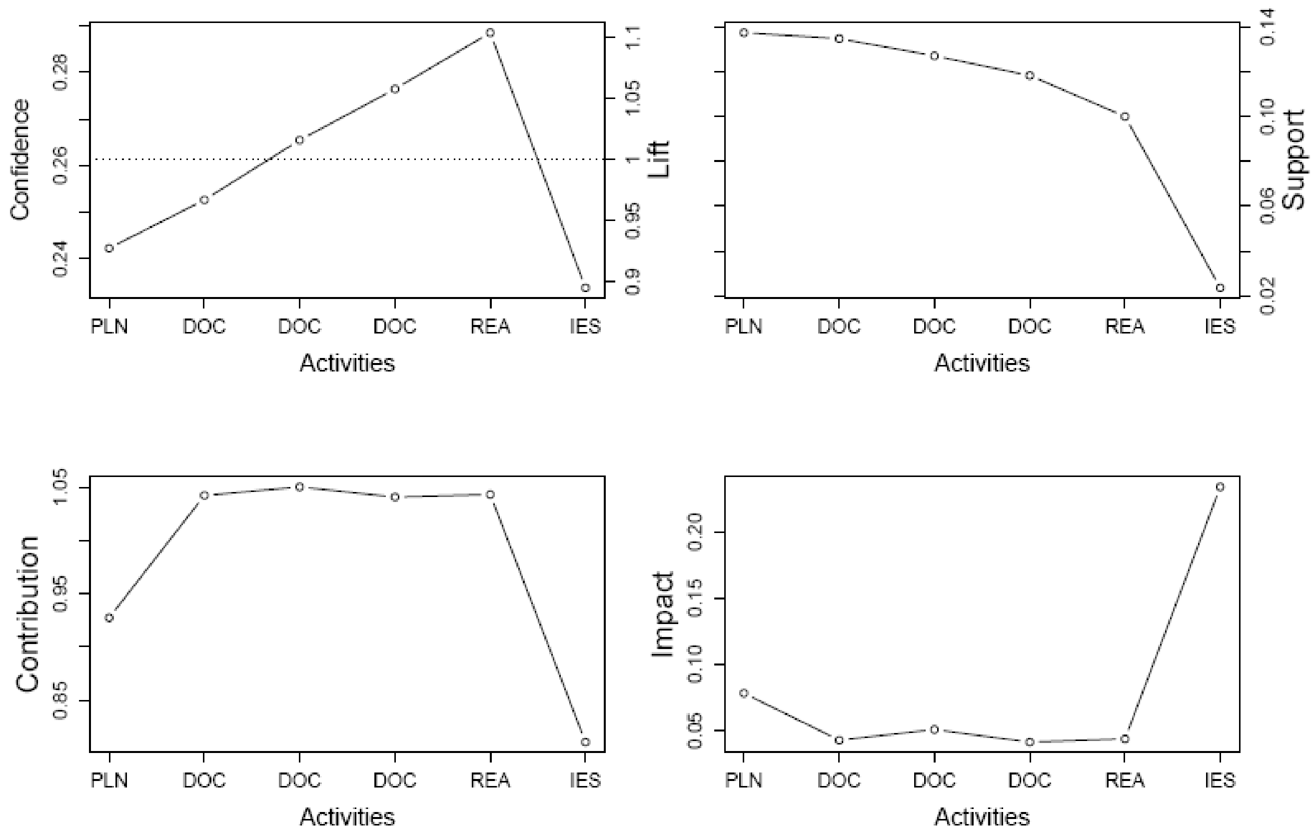
$$\begin{cases} TMC \rightarrow U_1 \\ TMC, GPS \rightarrow U_2 \\ TMC, GPS, DAG \rightarrow U_2 \\ TMC, GPS, DAG, PPJ \rightarrow U_3 \\ TMC, GPS, DAG, PPJ, OMF \rightarrow U_0 \\ TMC, GPS, DAG, PPJ, OMF, IKR \rightarrow U_{-1} \\ TMC, GPS, DAG, PPJ, OMF, IKR, TMC \rightarrow U_1 \\ TMC, GPS, DAG, PPJ, OMF, IKR, TMC, PPJ \rightarrow U_3 \end{cases}, \qquad (6)$$

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

An Example of Extended Combined Pattern Cluster

$$\begin{cases} PLN \rightarrow T \\ PLN, DOC \rightarrow T \\ PLN, DOC, DOC \rightarrow T \\ PLN, DOC, DOC, DOC \rightarrow T \\ PLN, DOC, DOC, DOC, REA \rightarrow T \\ PLN, DOC, DOC, DOC, REA, IES \rightarrow T \end{cases}$$

An Example of Extended Combined Pattern Cluster



High Utility Sequence Analysis

Yin, Junfu, Zhigang Zheng, and Longbing Cao. "USpan: an efficient algorithm for mining high utility sequential patterns." In *KDD*, pp. 660-668. ACM, 2012

Jingyu Shao, Junfu Yin, Wei Liu, Longbing Cao. Mining actionable combined patterns of high utility and frequency. DSAA 2015: 1-10

Introduction

- **Sequential pattern mining**
  - Very essential for handling order-based critical business problems.
  - Interesting and significant sequential patterns are generally selected by frequency.
- **Insufficient of frequency/support framework**
  - They do not show the business value and impact.
  - Some truly interesting sequences may be filtered because of their low frequencies.
- Example: Retail business

Introduction

Table 1: Quality Table

Items	*a*	*b*	*c*	*d*	*e*	*f*
Quality	2	5	4	3	1	1

Table 2: Quantitative Sequence Database

SID	Quantitative Sequence
1	< **(*e*, 5)** [(*c*, 2)(*f*, 1)] **(*b*, 2)** >
2	< [(*a*, 2)(*e*, 6)] [(*a*, 1)(*b*, 1)(*c*, 2)] [(*a*, 2)(*d*, 3)(*e*, 3)] >
3	< **(*c*, 1)** [(*a*, 6)(*d*, 3)(*e*, 2)] >
4	< [(*b*, 2)(*e*, 2)] [(*a*, 7)(*d*, 3)] [(*a*, 4)(*b*, 1)(*e*, 2)] >
5	< [(*b*, 2)(*e*, 3)] [(*a*, 6)(*e*, 3)] [(*a*, 2)(*b*, 1)] >

In sequence s_2, there are three transactions:

[(*a*, 2)(*e*, 6)],
[(*a*, 1)(*b*, 1)(*c*, 2)] and
[(*a*, 2)(*d*, 3)(*e*, 3)].

Transaction [(*a*, 2)(*e*, 6)] means the customer buys two items, namely *a* and *e*. (*a*, 2) means the quanity of item *a* is 2.

The square brackets omitted when there is only one item in the transaction. For example: (*e*, 5), (*b*, 2) in s_1 and (*c*, 1) in s_3.

Introduction

Table 1: Quality Table

Items	*a*	*b*	*c*	*d*	*e*	*f*
Quality	**2**	5	4	3	**1**	1

Table 2: Quantitative Sequence Database

SID	Quantitative Sequence
1	< (*e*, 5) [(*c*, 2)(*f*, 1)] (*b*, 2) >
2	< [(*a*, 2)**(*e*, 6)**] [**(*a*, 1)**(*b*, 1)(*c*, 2)] [**(*a*, 2)**(*d*, 3)(*e*, 3)] >
3	< (*c*, 1) [(*a*, 6)(*d*, 3)(*e*, 2)] >
4	< [(*b*, 2)**(*e*, 2)**] [**(*a*, 7)**(*d*, 3)] [**(*a*, 4)**(*b*, 1)(*e*, 2)] >
5	< [(*b*, 2)**(*e*, 3)**] [**(*a*, 6)**(*e*, 3)] [**(*a*, 2)**(*b*, 1)] >

The utility of <*e*> in (*e*, 6) is 6×1 = 6

The utility of <*ea*> in s_2 is
{ ((6×1) + (1×2)), ((6×1) + (1×2)) }
= {8, 10}

The utility of <*ea*> is the database is
{{}, {8, 10}, {}, {16, 10}, {15, 7}}.

Add the highest utility in each sequence to represent the utility of <*ea*>:
10 + 16 + 15 = 41

If the minimum utility threshold ξ = 40 then <*ea*> is a high utility pattern.

Introduction

1. We define the problem of mining high utility sequential patterns systematically.

2. USpan as a novel algorithm for mining high utility sequential patterns.

3. Two pruning strategies, namely width and depth pruning, are proposed to reduce the search space substantially.

Related Work

- **High utility pattern mining**
  - Two-Phase Algorithm (Liu et al., UBDM' 2005)
  - IHUP Algorithm (Ahmed et al., IEEE Trans. TKDE' 2009)
  - UP-Growth (Tseng et al., SIGKDD' 2010)

- **High utility sequential pattern mining**
  - UMSP (Shie et al., DASFAA' 2011) Designed for mining high utility mobile sequential patterns.

  - UWAS-tree / IUWAS-tree (Ahmed et al., SNPD' 2010) Designed for mining the high utility weblog data. IUWAS-tree is for incremental environment.

  - UI / US (Ahmed et al., ETRI Journal' 2010) Uses two measurements of utilities of sequences. No generic framework is proposed.

Problem Statement: Containing

Table 1: Quality Table

Items	*a*	*b*	*c*	*d*	*e*	*f*
Quality	2	5	4	3	1	1

Table 2: Quantitative Sequence Database

SID	Quantitative Sequence
1	< (*e*, 5) [(*c*, 2)(*f*, 1)] (*b*, 2) >
2	< [(*a*, 2)(*e*, 6)] [(*a*, 1)(*b*, 1)(*c*, 2)] [(*a*, 2)(*d*, 3)(*e*, 3)] >
3	< (*c*, 1) [(*a*, 6)(*d*, 3)(*e*, 2)] >
4	< [(*b*, 2)(*e*, 2)] [(*a*, 7)(*d*, 3)] [(*a*, 4)(*b*, 1)(*e*, 2)] >
5	< [(*b*, 2)(*e*, 3)] [(*a*, 6)(*e*, 3)] [(*a*, 2)(*b*, 1)] >

(*a*, 2): Q-item
[(*a*, 2)(*e*, 6)]: Q-itemset
s_1 - s_5: Q-sequence

- Q-itemset containing

[(*a*, 4)(*b*, 1)(*e*, 2)] contains q-itemsets (*a*, 4), [(*a*, 4)(*e*, 2)] and [(*a*, 4)(*b*, 1)(*e*, 2)] but not [(*a*, 2)(*e*, 2)] and [(*a*, 4)(*c*, 1)].

- Q-sequence containing

<[(*b*, 2)(*e*, 3)][(*a*, 6)(*e*, 3)][(*a*, 2)(*b*, 1)]> contains q-sequences
<(*b*, 2)>, <[(*b*, 2)(*e*, 3)]> and
<[(*b*, 2)][(*e*, 3)](*a*, 2)>
but not [(*a*, 2)(*e*, 2)] and [(*a*, 4)(*c*, 1)].

Problem Statement: Matching

Table 1: Quality Table

Items	*a*	*b*	*c*	*d*	*e*	*f*
Quality	2	5	4	3	1	1

Table 2: Quantitative Sequence Database

SID	Quantitative Sequence
1	< (*e*, 5) [(*c*, 2)(*f*, 1)] (*b*, 2) >
2	< [(*a*, 2)**(*e*, 6)**] [**(*a*, 1)**(*b*, 1)(*c*, 2)] [**(*a*, 2)**(*d*, 3)(*e*, 3)] >
3	< (*c*, 1) [(*a*, 6)(*d*, 3)(*e*, 2)] >
4	< [(*b*, 2)**(*e*, 2)**] [**(*a*, 7)**(*d*, 3)] [**(*a*, 4)**(*b*, 1)(*e*, 2)] >
5	< [(*b*, 2)**(*e*, 3)**] [**(*a*, 6)**(*e*, 3)] [**(*a*, 2)**(*b*, 1)] >

Sequence <*ea*> **matches:**

<(*e*, 6)(*a*, 1)> and <(*e*, 6)(*a*, 2)> in s_2 ;
<(*e*, 2)(*a*, 7)> and <(*e*, 2)(*a*, 4)> in s_4 ;
<(*e*, 3)(*a*, 6)> and <(*e*, 3)(*a*, 2)> in s_5 ;

Denote as <(*e*, 6)(*a*, 1)> ~ <*ea*>

Problem Statement: Utilities

The Sequence Utility Framework

The q-item utility:

$$u(i,q) = f_{u_i}(p(i),q)$$

The q-itemset utility:

$$u(l) = f_{u_{is}}(\bigcup_{j=1}^{n} u(i_j,q_j))$$

The q-sequence utility:

$$u(s) = f_{u_s}(\bigcup_{j=1}^{m} u(l_j))$$

The q-sequence database utility:

$$u(S) = f_{u_{db}}(\bigcup_{j=1}^{r} u(s_j))$$

The sequence utility in a q-sequence:

$$v(t,s) = \bigcup_{s' \sim t \cap s' \subseteq s} u(s')$$

The sequence utility in a database:

$$v(t) = \bigcup_{s \in S} v(t,s)$$

For example:

v(<*ea*>, s_4) = {u(<(*e*, 2)(*a*, 7)>), u(<(*e*, 2)(*a*, 4)>)}
v(<*ea*>) = {v(<*ea*>, s_2), v(<*ea*>, s_4), v(<*ea*>, s_5)}

Problem Statement: Utilities

High Utility Sequential Pattern Mining

The q-item utility:

$$f_{u_i}(p(i), q) = p(i) \times q$$

The q-itemset utility:

$$f_{u_{is}}(\bigcup_{j=1}^{n} u(i_j)) = \sum_{j=1}^{n} u(i_j, q_j)$$

The q-sequence utility:

$$f_{u_s}(\bigcup_{j=1}^{m} u(l_j)) = \sum_{j=1}^{m} u(l_j)$$

The q-sequence database utility:

$$f_{u_{db}}(\bigcup_{j=1}^{r} u(s_j)) = \sum_{j=1}^{r} u(s_j)$$

The sequence utility in a database:

$$v(t) = u_{max}(t) = \sum \max\{u(s')|s'\sim t \cap s' \subseteq s \cap s \in S\}$$

For example:

V(<*ea*>, s_4) = {16, 10}

V(<*ea*>) = { {8, 10}, {16, 10}, {15, 7} }

Sequence t is a high utility sequential pattern if and only if $u_{max} \geq \xi$ where ξ is a user-specified minimum utility.

Target: Extracting all high utility sequential patterns in *S* satisfying ξ.

USpan Algorithm

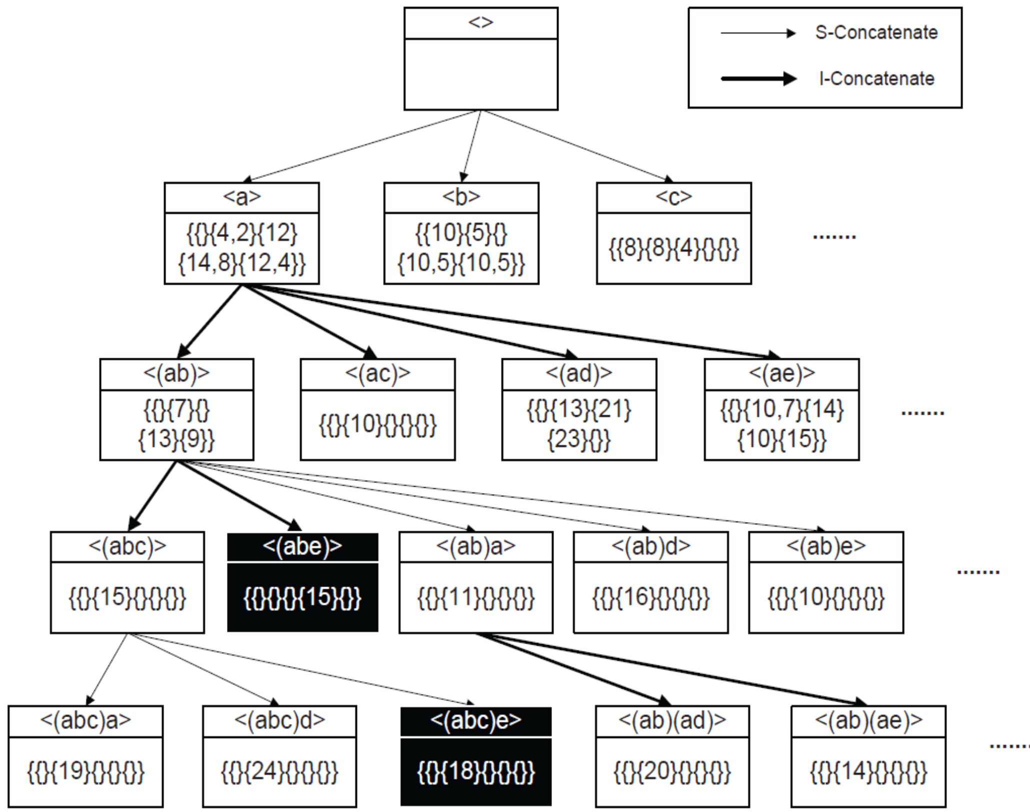
Challenges of mining for high utility patterns

u_{max} ($<a>$)= 4 + 12 + 14 + 12 = 42
u_{max} ($<ab>$)= 7 + 13 + 9 = 29
u_{max} ($<abc>$)= 15
u_{max} ($<(abc)a>$)= 19

No Downward Closure Property

USpan Algorithm

Lexicographic Q-sequence Tree



USpan Algorithm

Table 1: Quality Table

Items	*a*	*b*	*c*	*d*	*e*	*f*
Quality	2	5	4	3	1	1

Table 2: Quantitative Sequence Database

SID	Quantitative Sequence
1	< (*e*, 5) [(*c*, 2)(*f*, 1)] (*b*, 2) >
2	< [(*a*, 2)(*e*, 6)] [(*a*, 1)(*b*, 1)(*c*, 2)] [(*a*, 2)(*d*, 3)(*e*, 3)] >
3	< (*c*, 1) [(*a*, 6)(*d*, 3)(*e*, 2)] >
4	< [(*b*, 2)(*e*, 2)] [(*a*, 7)(*d*, 3)] [(*a*, 4)(*b*, 1)(*e*, 2)] >
5	< [(*b*, 2)(*e*, 3)] [(*a*, 6)(*e*, 3)] [(*a*, 2)(*b*, 1)] >

Items	Itemset 1	Itemset 2	Itemset 3
a		14	8
b	10		5
d		9	
e	2		2

v(<*b*>) = {10, 5}

Items	*I 1*	*I 2*	*I 3*
a		14	8
b	**10●**		**5■**
d		9	
e	2		2

v(<(*be*)>) = {10 + 2, 5 + 2} = {12, 7}

Items	*I 1*	*I 2*	*I 3*
a		14	8
b	10		5
d		9	
e	2●		2■

v(<(*be*)*a*>) = {12 + 14, 12 + 8} = {26, 20}

Items	*I 1*	*I 2*	*I 3*
a		14●	8■
b	10		5
d		9	
e	2		2

v(<(*be*)(*ad*)>) = {26 + 9} = {35}

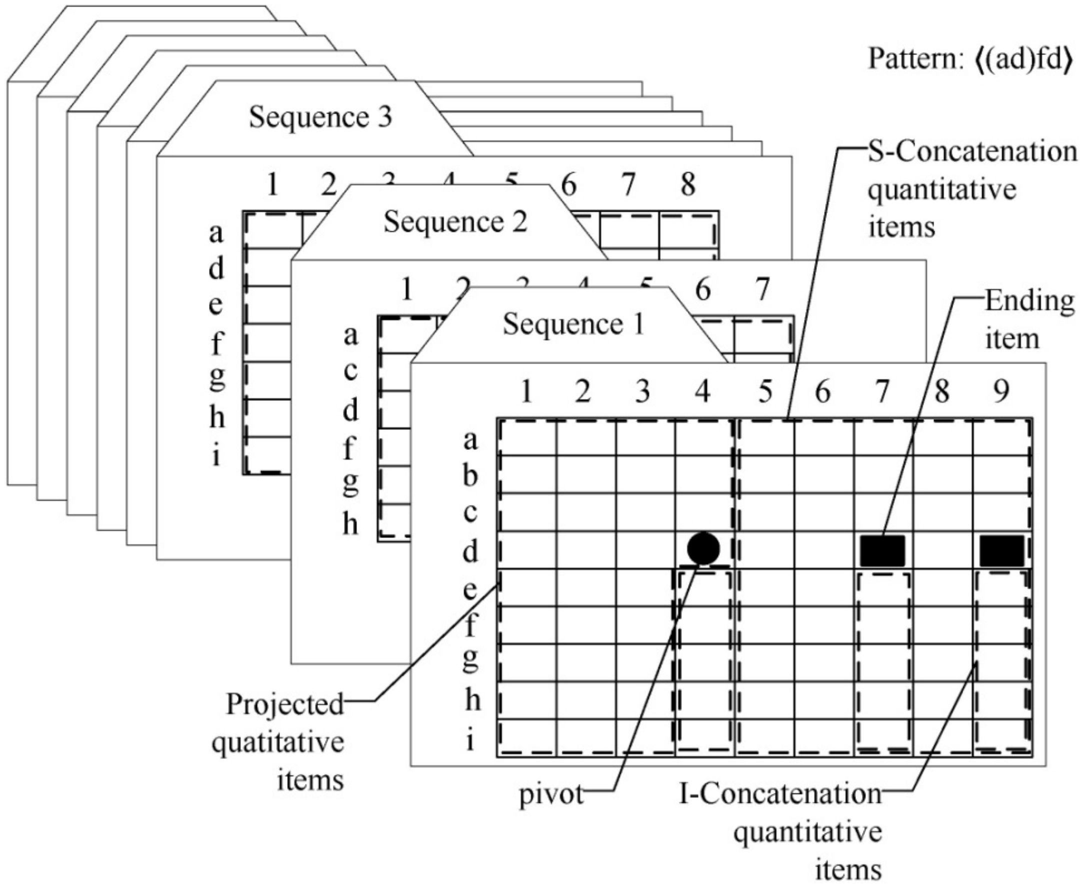
Items	*I 1*	*I 2*	*I 3*
a		14	8
b	10		5
d		9●	
e	2		2

v(<(*be*)(*ad*)*a*>) = {35 + 8}

Items	*I 1*	*I 2*	*I 3*
a		14	8●
b	10		5
d		9	
e	2		2

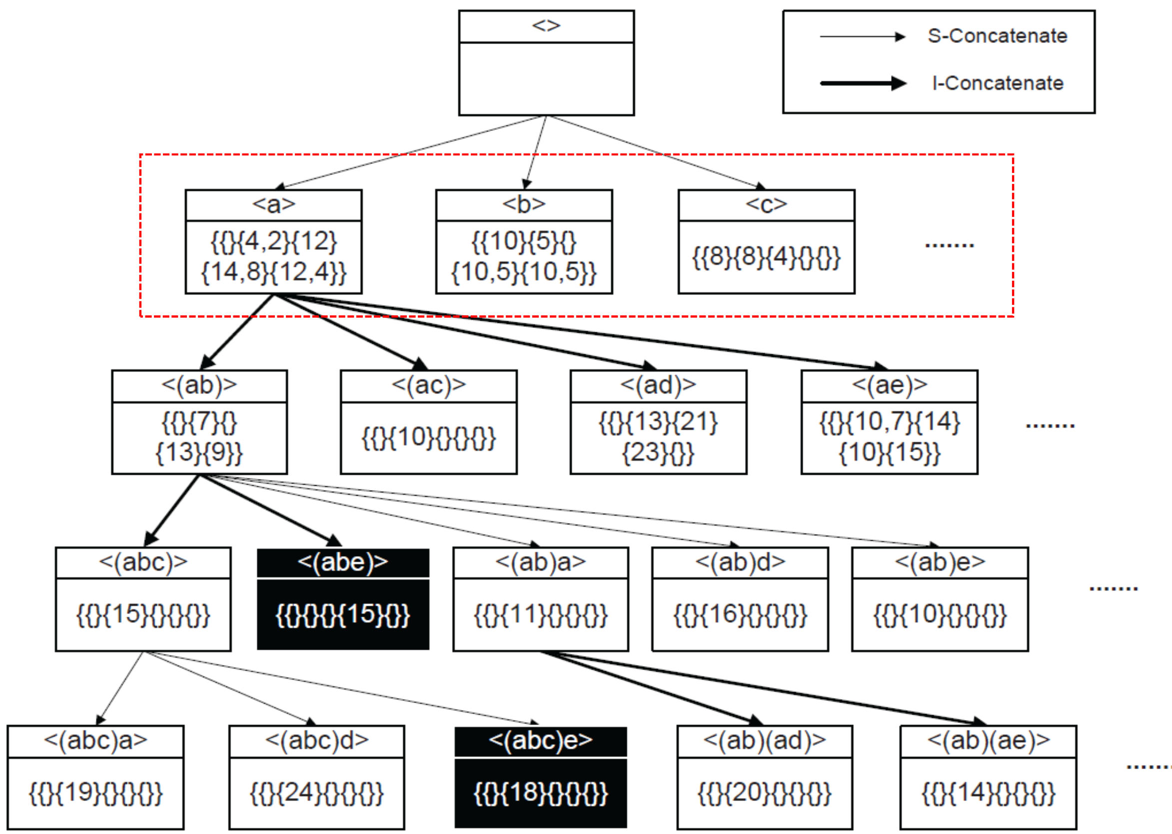
USpan Algorithm: Concatenation

Data Representation



USpan Algorithm: Width Pruning

What is Width Pruning



USpan Algorithm: Width Pruning

What to Width Prune

Table 1: Quality Table

Items	*a*	*b*	*c*	*d*	*e*	*f*
Quality	2	5	4	3	1	1

Table 2: Quantitative Sequence Database

SID	Quantitative Sequence	SU
1	< (*e*, 5) [(*c*, 2)(*f*, 1)] (*b*, 2) >	24
2	**< [(*a*, 2)(*e*, 6)] [(*a*, 1)(*b*, 1)(*c*, 2)] [(*a*, 2)(*d*, 3)(*e*, 3)] >**	**41**
3	< (*c*, 1) [(*a*, 6)(*d*, 3)(*e*, 2)] >	27
4	**< [(*b*, 2)(*e*, 2)] [(*a*, 7)(*d*, 3)] [(*a*, 4)(*b*, 1)(*e*, 2)] >**	**50**
5	**< [(*b*, 2)(*e*, 3)] [(*a*, 6)(*e*, 3)] [(*a*, 2)(*b*, 1)] >**	**42**

<f> should be **width-pruned**

$$\text{SWU}(<ea>) = u(s_2) + u(s_4) + u(s_5) = 41 + 50 + 24 = 115$$

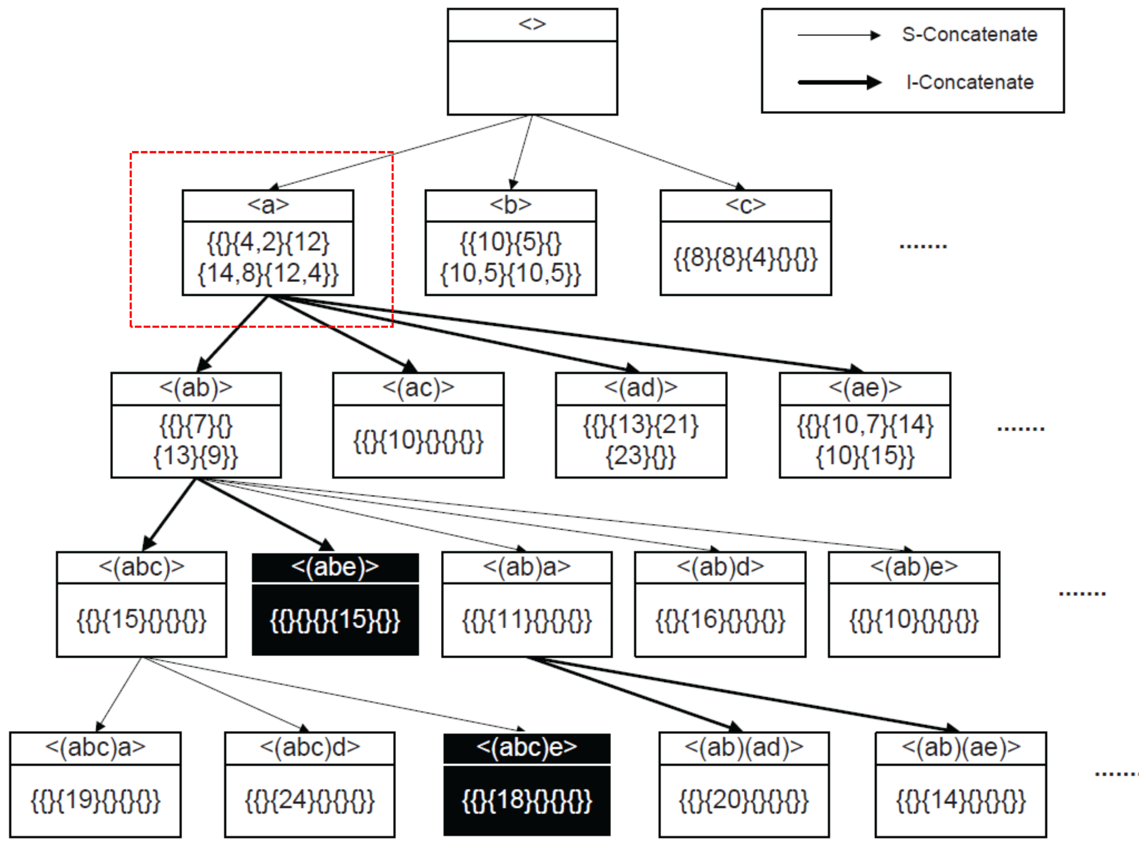
SID	Quantitative Sequence	SU
1	**< (*e*, 5) [(*c*, 2)(*f*, 1)] (*b*, 2) >**	**24**
2	< [(*a*, 2)(*e*, 6)] [(*a*, 1)(*b*, 1)(*c*, 2)] [(*a*, 2)(*d*, 3)(*e*, 3)] >	41
3	< (*c*, 1) [(*a*, 6)(*d*, 3)(*e*, 2)] >	27
4	< [(*b*, 2)(*e*, 2)] [(*a*, 7)(*d*, 3)] [(*a*, 4)(*b*, 1)(*e*, 2)] >	50
5	< [(*b*, 2)(*e*, 3)] [(*a*, 6)(*e*, 3)] [(*a*, 2)(*b*, 1)] >	42

$$\text{SWU}(<f>) = u(s_1) = 24$$

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

USpan Algorithm: Depth Pruning

What is Depth Pruning



USpan Algorithm: Depth Pruning

What to Depth Prune

Table 1: Quality Table

Items	*a*	*b*	*c*	*d*	*e*	*f*
Quality	2	5	4	3	1	1

Table 2: Quantitative Sequence Database

SID	Quantitative Sequence	SU
1	< (*e*, 5) [(*c*, 2)(*f*, 1)] (*b*, 2) >	24
2	< [(*a*, 2)**(*e*, 6)**] [**(*a*, 1)(*b*, 1)(*c*, 2)**] [**(*a*, 2)(*d*, 3)(*e*, 3)**] >	41
3	< (*c*, 1) [(*a*, 6)(*d*, 3)(*e*, 2)] >	27
4	< [(*b*, 2)**(*e*, 2)**] [**(*a*, 7)(*d*, 3)**] [**(*a*, 4)(*b*, 1)(*e*, 2)**] >	50
5	< [(*b*, 2)**(*e*, 3)**] [**(*a*, 6)(*e*, 3)**] [**(*a*, 2)(*b*, 1)**] >	42

SID	Quantitative Sequence	SU
1	< (*e*, 5) [(*c*, 2)(*f*, 1)] (*b*, 2) >	24
2	< [(*a*, 2)(*e*, 6)] [(*a*, 1)(*b*, 1)(*c*, 2)] [(*a*, 2)(*d*, 3)(*e*, 3)] >	41
3	< (*c*, 1) [(*a*, 6)(*d*, 3)(*e*, 2)] >	27
4	< [(*b*, 2)(*e*, 2)] [(*a*, 7)(*d*, 3)] [(*a*, 4)(*b*, 1)(*e*, 2)] >	50
5	< [(*b*, 2)**(*e*, 3)**] [**(*a*, 6)(*e*, 3)**] [**(*a*, 2)(*b*, 1)**] >	42

<e(ae)> should be **depth-pruned**

$$u_{rest}(<ea>) = (8+29) + (16+24) + (15+17) = 37 + 40 + 32 = 109$$

$$u_{rest}(<e(ae)>) = (18 + 9) = 27$$

Experiments

Datasets

Synthetic Datasets

Parameters	DS1	DS2
that the average number of elements	10	8
the average number of items in an element	2.5	2.5
the average length of a maximal pattern	4	6
the average number of items per element	2.5	2.5
Number of sequences	10k	10k
Number of items	1k	10k

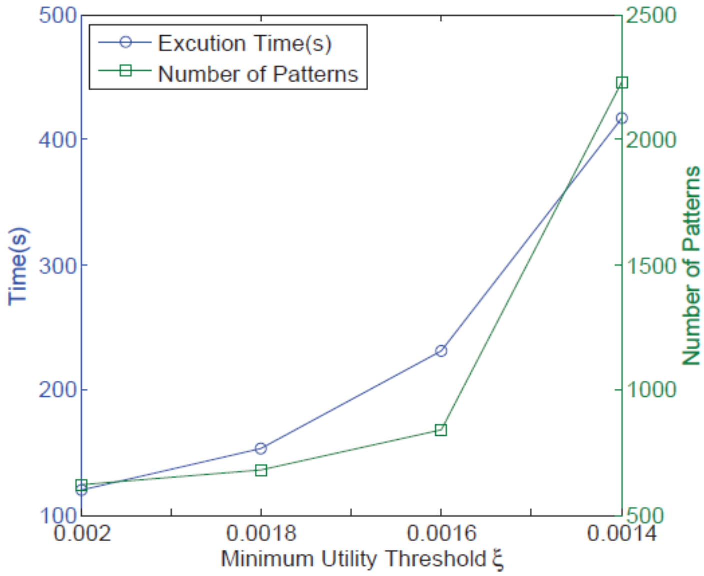
Real Datasets

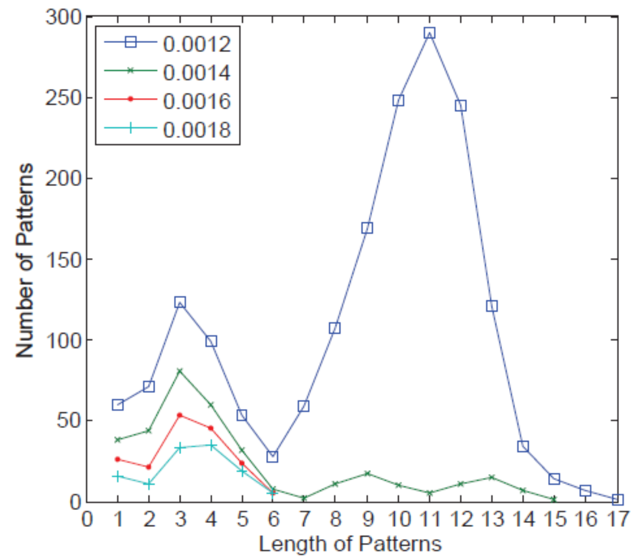
DS3 is a dataset consisting of online shopping transactions which contains 350,241 transactions and 59,477 customers.

DS4 is a real dataset that includes mobile communication transactions. The dataset is a 100,000 mobile call history from a specific day. There are 67,420 customers in the dataset.

Experiments

Performance and distributions (DS2)

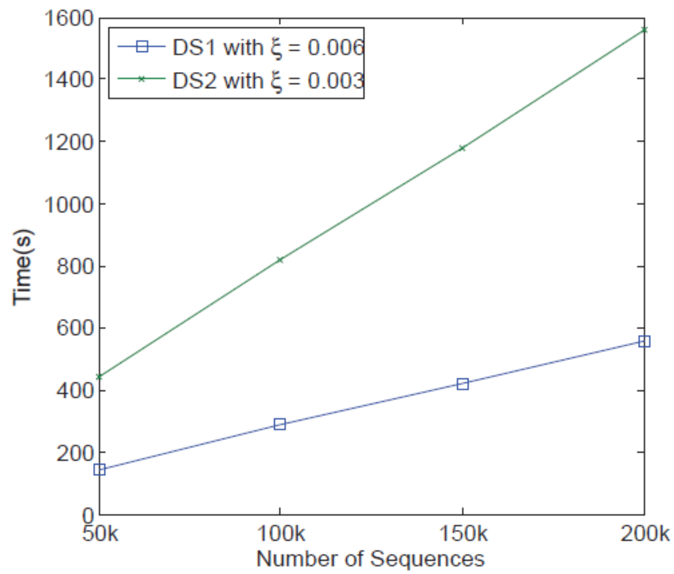


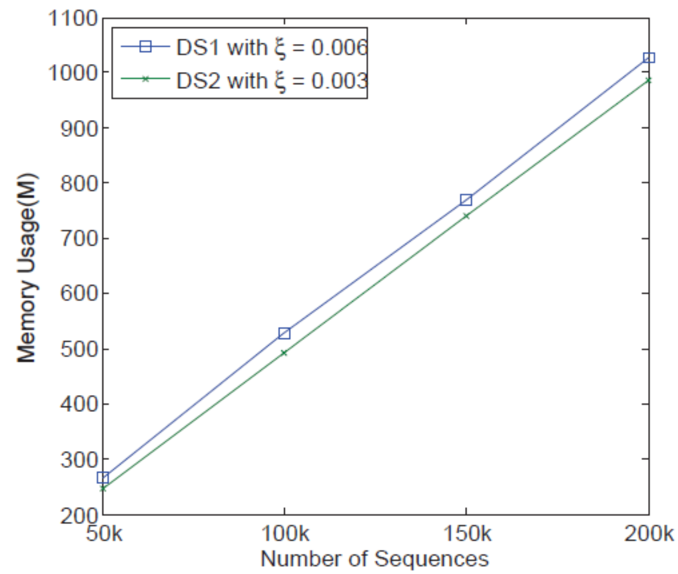


- The running time and the number of patterns grow exponentially with respect to ξ.
- The high utility sequential patterns are mid-long patterns.

Experiments

Scalability Test (DS1 & DS2)

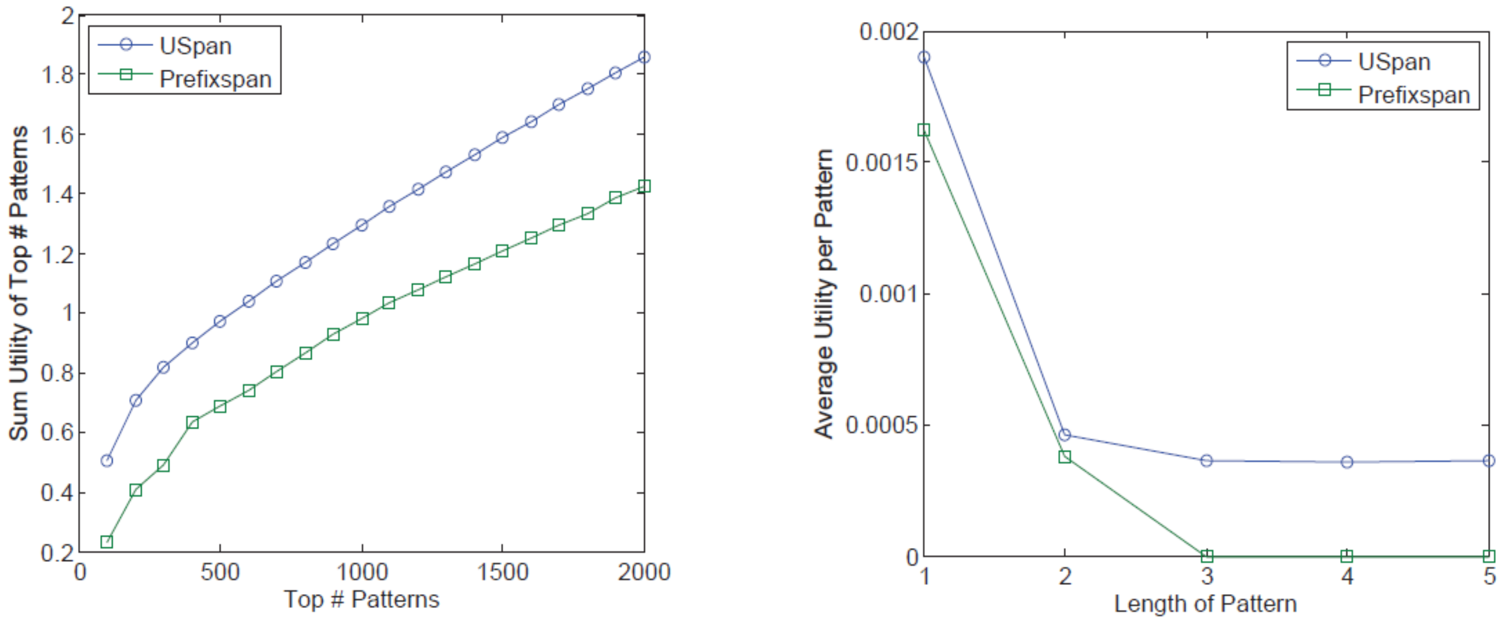




- Both the time and memory usage grow linearly with respect to the size of the DB.

Experiments

High Utility Sequential Pattern vs. Frequent Sequential Patterns (DS3)



- USpan out performs Prefixspan with respect to the utilities of the patterns.

Conclusions

1. We define the problem of mining high utility sequential patterns.
2. We propose the USpan to efficiently mine for mining high utility sequential patterns.
3. Two pruning strategies are proposed to substantially reduce the search space.
4. Experiments on both synthetic and real datasets show that USpan can discover the high utility sequential patterns efficiently.

Non-occurring Behavior Analysis /Negative Sequence Analysis

Longbing Cao, Philip S. Yu, Vipin Kumar. Nonoccurring Behavior Analytics: A New Area. IEEE Intelligent Systems 30(6): 4-11 (2015).

Longbing Cao, Xiangjun Dong and Zhigang Zheng. e-NSP: Efficient Negative Sequential Pattern Mining, Artificial Intelligence, 235: 156-182, 2016

Zhigang Zheng, Yanchang Zhao, Ziye Zuo, Longbing Cao, Huaifeng Zhang, Yanchang Zhao, Chengqi Zhang. An Efficient GA-Based Algorithm for Mining Negative Sequential Patterns, PAKDD 2010, 262-273

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Problem description

- What is negative sequential patterns?
- *Focus on negative relationship between itemsets*
- *Absent items are taken into consideration*

- Example:

  $p_1 = <a\ b\ c\ d>$ *vs* $p_2 = <a\ b\ \neg c\ e>$
- *Each item, a, b, c, d* and *e*, stands for a claim item of insurance.
- *p1: an insurant usually claims for a, b, c and d in a claim.*
- *p2: does NOT claim c after a and b, then claim item e instead of d.*

PSP & NSP

PSP: Positive Sequential Pattern

- **Only contain occurring itemsets**

  E.g. p1=<a b c X>.

  Existing Methods:

  AprioriAll, GSP, FreeSpan, PrefixSpan, SPADE , SPAM

NSP: Negative Sequential Pattern

- **Also contain non-occurring itemsets**

  E.g. p1=<a b ¬c X>.

  Limited research:

  Neg_GSP, PNSP

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Challenges for NSP

- *Apriori principle doesn't work for some situations*

- *Huge search space*
  - 10 distinct items
  - 3-item PSC: 10^3
  - 3-item NSC: 20^3

Difficulties in Mining NSP

- **High Computational Complexity.**

Additionally scanning database after identifying PSP.

- **Large NSC Search Space.**

k-size NSC by conducting a joining operation on (k-1)-size NSP. (NSC : Negative Sequential Candidates)

- **No Unified Definition about Negative Containment.**

How a data sequence contains a negative sequence?
<a> contains < a¬a >?   <a> contains < ¬a a¬a >?

Non-occurrence behaviour analysis

(Negative sequence analysis)

Table 1. Supports, Confidences and Lifts of Four Types of Sequential Rules

	Rules	Support	Confidence	Lift
I	$A \rightarrow B$	$P(AB)$	$\frac{P(AB)}{P(A)}$	$\frac{P(AB)}{P(A)P(B)}$
II	$A \rightarrow \neg B$	$P(A)-P(AB)$	$\frac{P(A)-P(AB)}{P(A)}$	$\frac{P(A)-P(AB)}{P(A)(1-P(B))}$
III	$\neg A \rightarrow B$	$P(B)-P(A\&B)$	$\frac{P(B)-P(A\&B)}{1-P(A)}$	$\frac{P(B)-P(A\&B)}{P(B)(1-P(A))}$
IV	$\neg A \rightarrow \neg B$	$1-P(A)-P(B)+P(A\&B)$	$\frac{1-P(A)-P(B)+P(A\&B)}{1-P(A)}$	$\frac{1-P(A)-P(B)+P(A\&B)}{(1-P(A))(1-P(B))}$

Table 4. Selected Positive and Negative Sequential Rules

Type	Rule	Support	Confidence	Lift
I	REA ADV ADV→DEB	0.103	0.53	2.02
	DOC DOC REA REA ANO→DEB	0.101	0.33	1.28
	RPR ANO→DEB	0.111	0.33	1.25
	RPR STM STM RPR→DEB	0.137	0.32	1.22
	MCV→DEB	0.104	0.31	1.19
	ANO→DEB	0.139	0.31	1.19
	STM PYI→DEB	0.106	0.30	1.16
II	STM PYR RPR REA RPT→ ¬DEB	0.166	0.86	1.16
	MND→ ¬DEB	0.116	0.85	1.15
	STM PYR RPR DOC RPT→ ¬DEB	0.120	0.84	1.14
	STM PYR RPR REA PLN→ ¬DEB	0.132	0.84	1.14
	REA PYR RPR RPT→ ¬DEB	0.176	0.84	1.14
	REA DOC REA CPI→ ¬DEB	0.083	0.83	1.12
	REA CRT DLY→ ¬DEB	0.091	0.83	1.12
	REA CPI→ ¬DEB	0.109	0.83	1.12
III	¬{PYR RPR REA STM}→DEB	0.169	0.33	1.26
	¬{PYR CCO}→DEB	0.165	0.32	1.24
	¬{STM RPR REA RPT}→DEB	0.184	0.29	1.13
	¬{RPT RPR REA RPT}→DEB	0.213	0.29	1.12
	¬{CCO RPT}→DEB	0.171	0.29	1.11
	¬{CCO PLN}→DEB	0.187	0.28	1.09
	¬{PLN RPT}→DEB	0.212	0.28	1.08
IV	¬{ADV REA ADV}→ ¬DEB	0.648	0.80	1.08
	¬{STM EAN}→ ¬DEB	0.651	0.79	1.07
	¬{REA EAN}→ ¬DEB	0.650	0.79	1.07
	¬{DOC FRV}→ ¬DEB	0.677	0.78	1.06
	¬{DOC DOC STM EAN}→ ¬DEB	0.673	0.78	1.06
	¬{CCO EAN}→ ¬DEB	0.681	0.78	1.05

Genetic-Algorithm based NSP approach: GA-NSP

- Find good (frequent) genes with good performance (supp), and optimize genes (FP) through crossover and mutation, m*generations
- Improve gene quality (making more and more frequent)

Strengths:

- Treat candidates unequally
- Very low support threshold
- Find long-NSP at the beginning

Zhigang Zheng, Yanchang Zhao, Ziye Zuo, Longbing Cao, Huaifeng Zhang, Yanchang Zhao, Chengqi Zhang. An Efficient GA-Based Algorithm for Mining Negative Sequential Patterns, PAKDD 2010, 262-273

GA-NSP

- *New generations: good genes (freq patterns) through crossover and mutation operations.*
- *Population evolution control: fitness and dynamic fitness.*
- *Performance improvement: pruning method (check constraints of NSP)*

Problem Statement

- Sequence (general)

  $s = \langle e_1\ e_2 \ldots e_n \rangle$

  i.e. $\langle a\ b\ (c,d)\ e \rangle$, $\langle a\ \neg b\ c\ e \rangle$

- Positive/Negative Sequence

  $s_p = \langle e_1\ e_2 \ldots e_n \rangle$, *all elements are positive*

  $s_n = \langle e_1\ e_2 \ldots e_n \rangle$, *at least one element is negative*

- Negative Sequential Pattern
  - *Its support is greater than minimum support threshold.*
  - *Two or more continuous negative elements are not accepted.*
  - *For each negative item, its corresponding positive item is required to be frequent.*
  - *Items in an element should be all positive or all negative. i.e.* $\langle a\ (a,\neg b)\ c \rangle$ *is not allowed.*

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

- **Negative Matching**

Negative Matching. A negative sequence s_n=<e_1 e_2 ... e_k> *matches* a data sequence s=<d_1 d_2 ... d_m>, iff:

1) s contains the max positive subsequence of s_n
2) for each negative element $e_i(1 \leq i \leq k)$, there exist integers p, q, $r(1 \leq p \leq q \leq r \leq m)$ such that: $\exists e_{i-1} \subseteq d_p \wedge e_{i+1} \subseteq d_r$, and for $\forall d_q$, $e_i \not\subseteq d_q$

	Sequence	Matching	Data Sequence
S_1	<b ¬c a>	No	<b f d c a>
S_2	<b ¬c d a>	Yes	<b f d c a>

GA-NSP Algorithm

- Encoding

Sequence		Chromosome		
		$gene_1$	$gene_2$	$gene_3$
$<a\ b\ \neg(c,d)>$	$\Rightarrow$	$+a$	$+b$	$\neg(c,d)$

- Crossover

$parent1$	$b\ \neg c \updownarrow a$	$\Rightarrow$	$child1$	$b\ \neg c\ \mathbf{e}$
$parent2$	$\mathbf{d} \updownarrow \mathbf{e}$	$\Rightarrow$	$child2$	$\mathbf{d}\ a$

$parent1$	$b\ \neg c\ a \updownarrow$	$\Rightarrow$	$child1$	$b\ \neg c\ a\ \mathbf{d}\ \mathbf{e}$
$parent2$	$\updownarrow \mathbf{d}\ \mathbf{e}$	$\Rightarrow$	$child2$	$\mathbf{d}\ \mathbf{e}\ b\ \neg c\ a$

- Mutation

  Select a random position and then replace all genes after that position with 1-item patterns

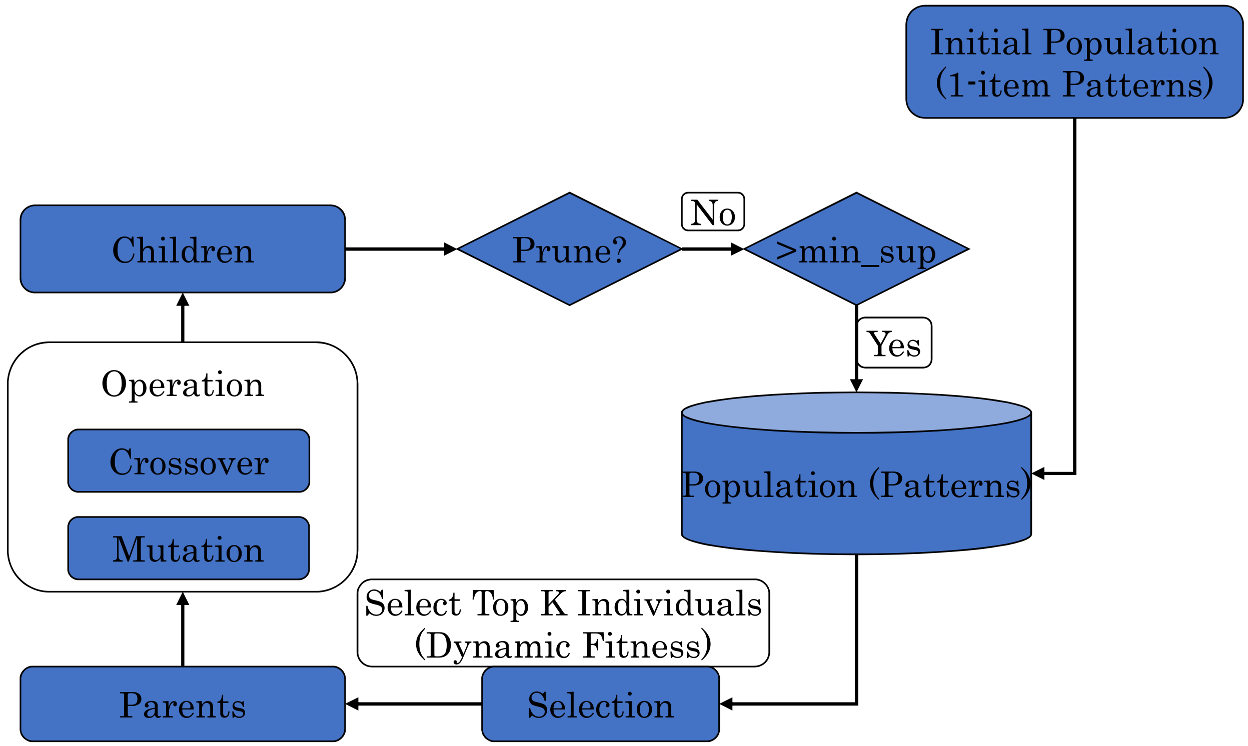
- Fitness & Dynamic Fitness

$$ind.fitness = (ind.support - min\_sup) \times DatasetSize. \quad (1)$$

$$ind.dfitness = \begin{cases} ind.fitness, & \text{initial set} \\ ind.dfitness \times (1 - \underline{DecayRate}), & \text{if } ind \text{ is selected} \end{cases} \quad (2)$$

- Selection

Selection(pop){ //Subfunction for selecting top K individuals from population
  for (each ind with top K dfitness in pop){
    $popK$.add(ind);
    $ind.dfitness = ind.dfitness * (1\text{-}decay\_rate)$;
    if ($ind.dfitness < 0.01$) $ind.dfitness = 0$;
  }
  return $popK$;
}



$$ind.fitness = (ind.support - min\_sup) \times DatasetSize. \quad (1)$$

$$ind.dfitness = \begin{cases} ind.fitness, & \text{initial set} \\ ind.dfitness \times (1 - DecayRate), & \text{if } ind \text{ is selected} \end{cases} \quad (2)$$

- GA-NSP Pseudocode

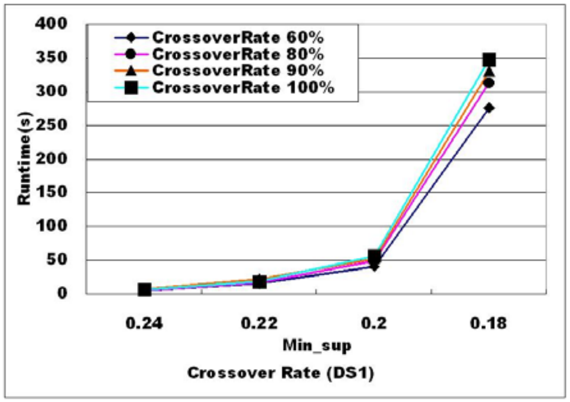
```
RunGA(min_sup, decay_rate, crossover_rate, mutation_rate){
  pop = initialPopulation();
  for (each individual ind in pop){
    ind.fitness = calculateFitness(ind);
    ind.dfitness = ind.fitness
    pop.sum_dfitness = pop.sum_dfitness + ind.dfitness
  }
  while ( pop.sum_dfitness > 0 ){
    popK = Selection(pop);
    if (Random()<crossover_rate) Crossover(popK);
    if (Random()<mutation_rate) Mutation(popK);
    for (each individual ind in popK)
      if (Prune(ind) ! =true && ind.sup >= min_sup)    pop.add(ind);
  }
  return pop;
}
```

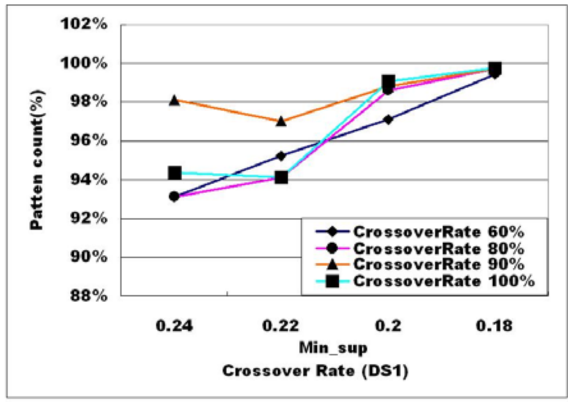
Experiments Result .1

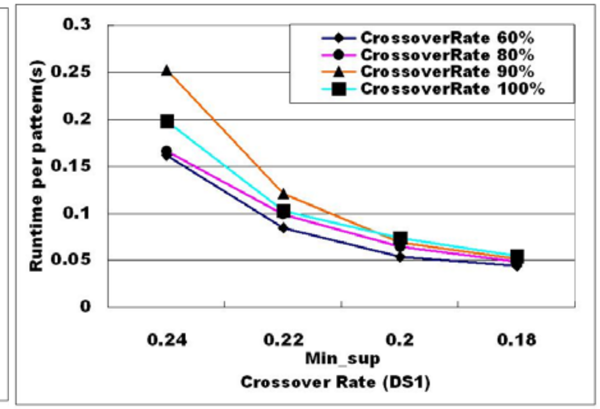
- Datasets

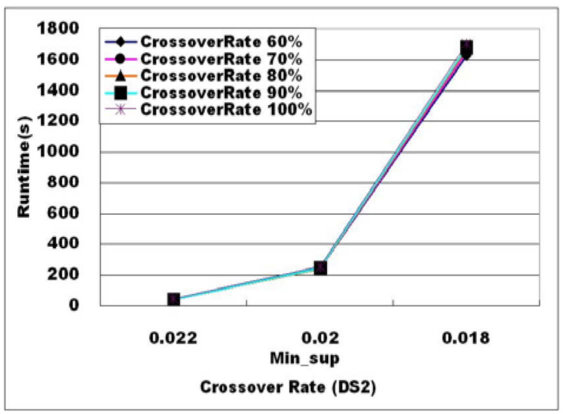
- *Dataset*1(*DS*1) is C8.T8.S4.I8.DB10k.N1k, which means the average number of elements in a sequence is 8, the average number of items in an element is 8, the average length of a maximal pattern consists of 4 elements and each element is composed of 8 items average. The data set contains 10k sequences, the number of items is 1000.
- *Dataset*2(*DS*2) is C10.T2.5.S4.I2.5.DB100k.N10k.
- *Dataset*3(*DS*3) is C20.T4.S6.I8.DB10k.N2k.
- *Dataset*4(*DS*4) is real application data for insurance claims.

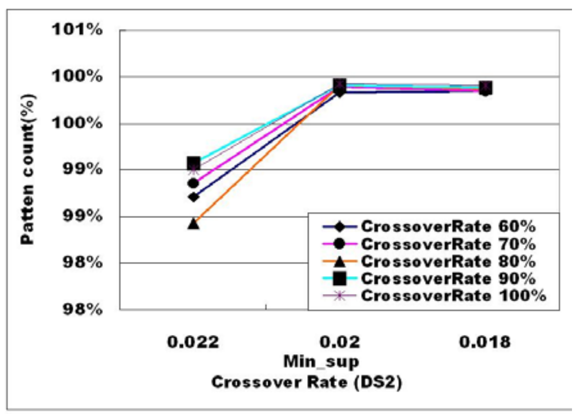
- Crossover Rate

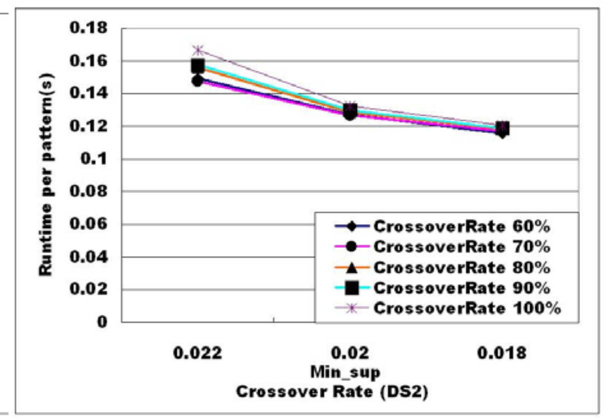




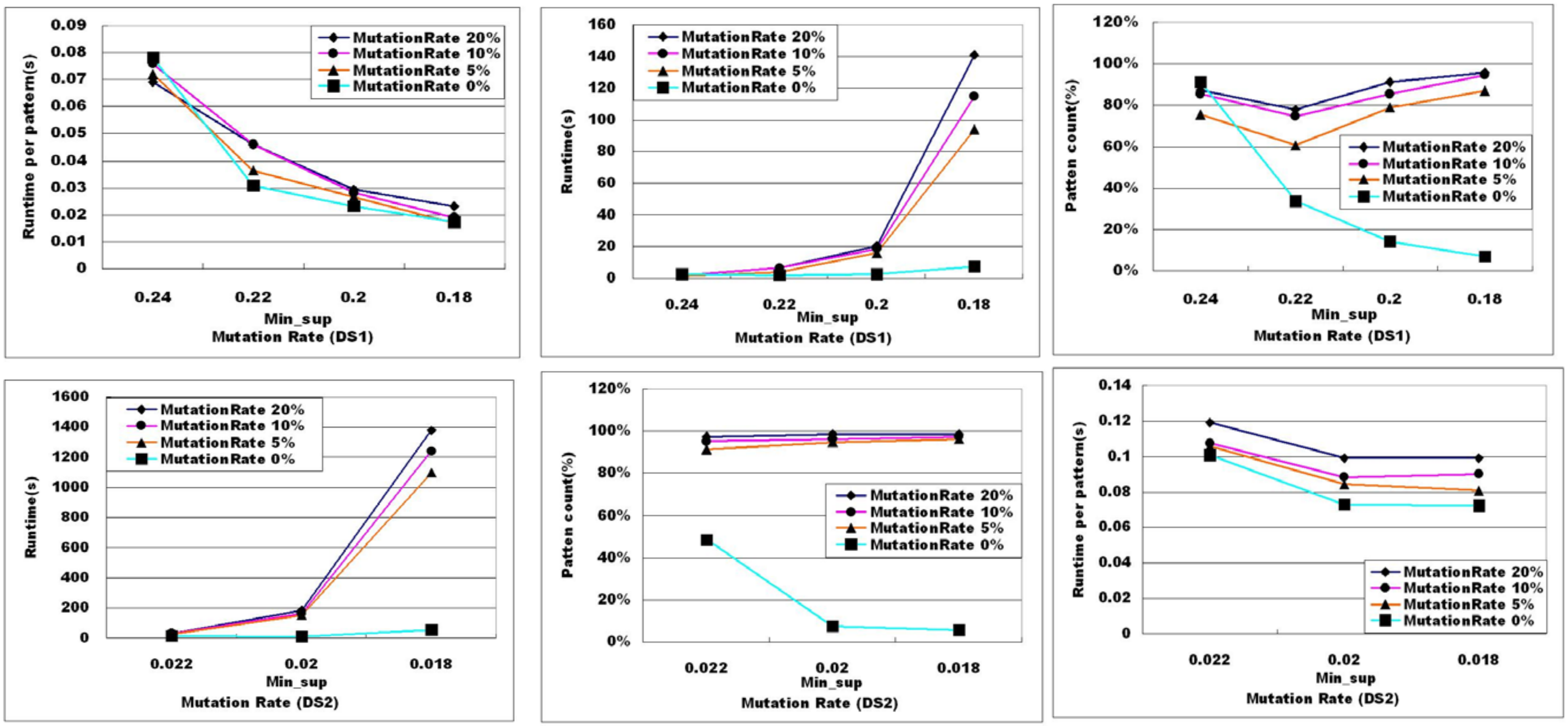




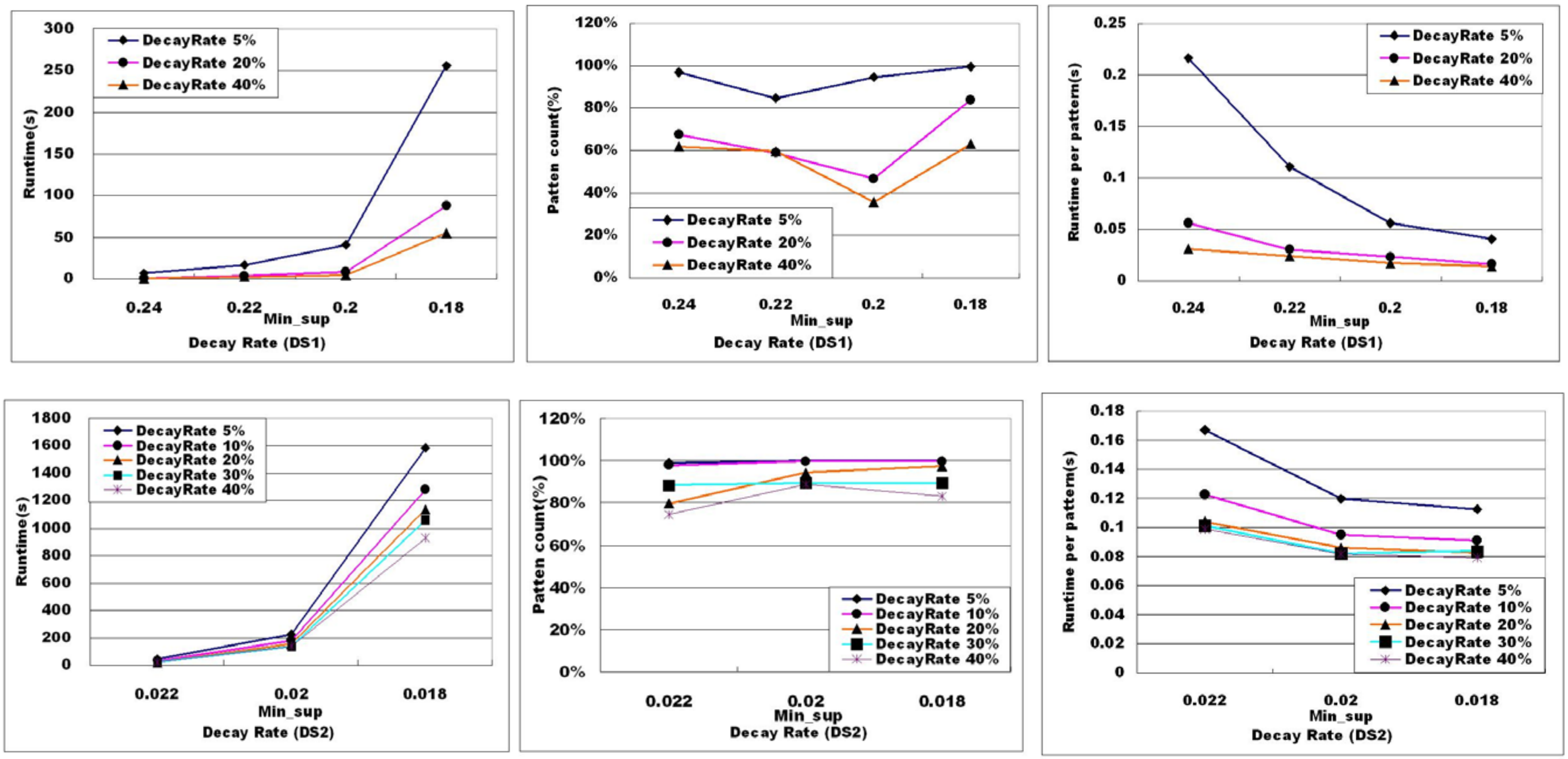




- Mutation Rate

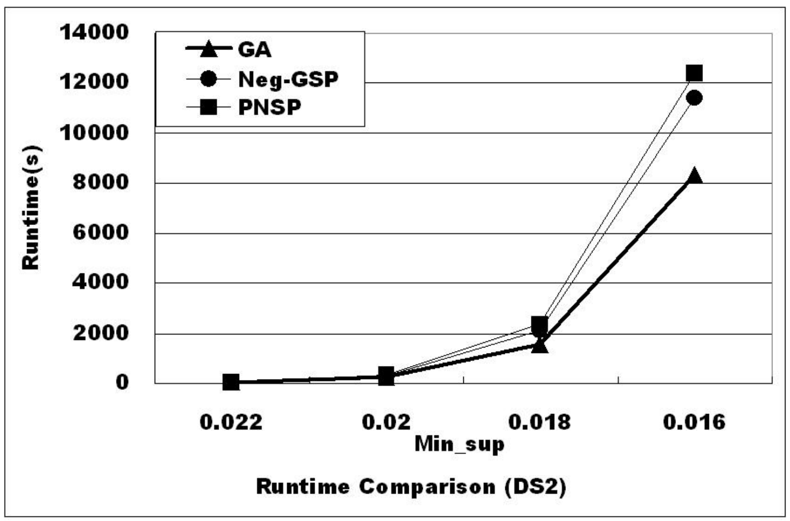


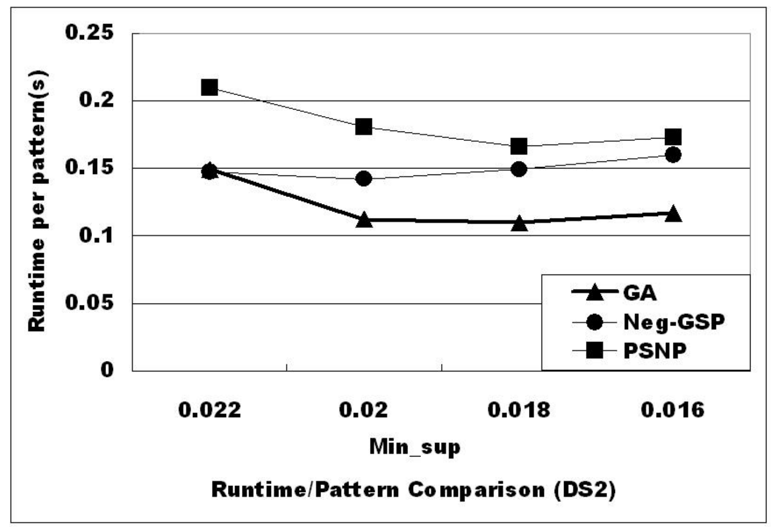
- Decay Rate



Decay Rate (DS1)
Decay Rate (DS1)
Decay Rate (DS1)
Decay Rate (DS2)
Decay Rate (DS2)
Decay Rate (DS2)
Runtime(s)
Patten count(%)
Runtime per pattern(s)
Min_sup
DecayRate 5%
DecayRate 10%
DecayRate 20%
DecayRate 30%
DecayRate 40%

- Comparison with PNSP, Neg-GSP





Classification of both positive and negative behavior patterns

Huaifeng Zhang, Yanchang Zhao, Longbing Cao, Chengqi Zhang and Hans Bohlscheid. Customer Activity Sequence Classification for Debt Prevention in Social Security, *Journal of Computer Science and Technology*, 24(6): 1000-1009 (2009).

Yanchang Zhao, Huaifeng Zhang, Shanshan Wu, Jian Pei,Longbing Cao, Chengqi Zhang and Hans Bohlscheid. Debt Detection in Social Security by Sequence Classification Using Both Positive and Negative Patterns, *ECML/PKDD2009*, 648-663.

Sequence classification

Let $\mathcal{T}$ be a finite set of *class labels*. A *sequential classifier* is a function

$$\mathcal{F} : \mathcal{S} \rightarrow \mathcal{T}. \quad (1)$$

In sequence classification, the classifier $\mathcal{F}$ is built on the base of frequent *classifiable sequential patterns* $\mathcal{P}$.

Definition 3.1 (Classifiable Sequential Pattern). *Classifiable Sequential Patterns (CSP) are frequent sequential patterns for the sequential classifier in the form of $p_a \Rightarrow \tau$, where p_a is a frequent pattern in the sequence database $\mathcal{S}$.*

Based on the mined classifiable sequential patterns, a sequential classifier can be formulised as

$$\mathcal{F} : s \xrightarrow{\mathcal{P}} \tau. \quad (2)$$

- Class correlation ratio

$$CCR(p_a \rightarrow \tau) = \frac{\hat{corr}(p_a \rightarrow \tau)}{\hat{corr}(p_a \rightarrow \neg\tau)} = \frac{a \cdot (c+d)}{c \cdot (a+b)},$$

$$\hat{corr}(p_a \rightarrow \tau) = \frac{sup(p_a \cup \tau)}{sup(p_a) \cdot sup(\tau)} = \frac{a \cdot n}{(a+c) \cdot (a+b)}.$$

Table 2. Feature-Class Contingency Table

	p_a	$\neg p_a$	$\sum$
τ	a	b	$a+b$
$\neg\tau$	c	d	$c+d$
$\sum$	$a+c$	$b+d$	$n=a+b+c+d$

Table 4. Selected Positive and Negative Sequential Rules

Type	Rule	Support	Confidence	Lift
I	REA ADV ADV→DEB	0.103	0.53	2.02
	DOC DOC REA REA ANO→DEB	0.101	0.33	1.28
	RPR ANO→DEB	0.111	0.33	1.25
	RPR STM STM RPR→DEB	0.137	0.32	1.22
	MCV→DEB	0.104	0.31	1.19
	ANO→DEB	0.139	0.31	1.19
	STM PYI→DEB	0.106	0.30	1.16
II	STM PYR RPR REA RPT→ ¬DEB	0.166	0.86	1.16
	MND→ ¬DEB	0.116	0.85	1.15
	STM PYR RPR DOC RPT→ ¬DEB	0.120	0.84	1.14
	STM PYR RPR REA PLN→ ¬DEB	0.132	0.84	1.14
	REA PYR RPR RPT→ ¬DEB	0.176	0.84	1.14
	REA DOC REA CPI→ ¬DEB	0.083	0.83	1.12
	REA CRT DLY→ ¬DEB	0.091	0.83	1.12
	REA CPI→ ¬DEB	0.109	0.83	1.12
III	¬{PYR RPR REA STM}→DEB	0.169	0.33	1.26
	¬{PYR CCO}→DEB	0.165	0.32	1.24
	¬{STM RPR REA RPT}→DEB	0.184	0.29	1.13
	¬{RPT RPR REA RPT}→DEB	0.213	0.29	1.12
	¬{CCO RPT}→DEB	0.171	0.29	1.11
	¬{CCO PLN}→DEB	0.187	0.28	1.09
	¬{PLN RPT}→DEB	0.212	0.28	1.08
IV	¬{ADV REA ADV}→ ¬DEB	0.648	0.80	1.08
	¬{STM EAN}→ ¬DEB	0.651	0.79	1.07
	¬{REA EAN}→ ¬DEB	0.650	0.79	1.07
	¬{DOC FRV}→ ¬DEB	0.677	0.78	1.06
	¬{DOC DOC STM EAN}→ ¬DEB	0.673	0.78	1.06
	¬{CCO EAN}→ ¬DEB	0.681	0.78	1.05

Table 5. The Number of Patterns in PS10 and PS05

	PS10 ($min\_sup = 0.1$)		PS05 ($min\_sup = 0.05$)	
	Number	Percent(%)	Number	Percent(%)
Type I	93,382	12.05	127,174	3.93
Type II	45,821	5.91	942,498	29.14
Type III	79,481	10.25	1,317,588	40.74
Type IV	556,491	71.79	846,611	26.18
Total	775,175	100	3,233,871	100

Table 6. Classification Results with Pattern Set PS05-4K

Pattern Number		40	60	80	100	150	200	300
Neg&Pos	Recall	.438	.416	**.286**	**.281**	**.422**	.492	.659
	Precision	.340	.352	.505	**.520**	**.503**	.474	.433
	Accuracy	.655	.670	**.757**	**.761**	**.757**	.742	.705
	Specificity	.726	.752	.909	.916	.865	.823	.720
Positive	Recall	.130	.124	.141	.135	.151	.400	.605
	Precision	.533	.523	.546	.472	.491	.490	.483
	Accuracy	.760	.758	.749	.752	.754	.752	.745
	Specificity	.963	.963	.946	.951	.949	.865	.790

e-NSP: Efficient negative sequential pattern mining

Cao, Longbing, Xiangjun Dong, and Zhigang Zheng. "e-NSP: Efficient negative sequential pattern mining." *Artificial Intelligence* 235 (2016): 156-182.

Dong, Xiangjun, Zhigang Zheng, Longbing Cao, Yanchang Zhao, Chengqi Zhang, Jinjiu Li, Wei Wei, and Yuming Ou. "e-NSP: efficient negative sequential pattern mining based on identified positive patterns without database rescanning." In *CIKM*, pp. 825-830. ACM, 2011.

Some Definitions

- **Negative Item/Element:**
  Non-occurring item / element
- **Negative Sequence**
  A sequence includes at least one negative item
- **Positive-partner of a Negative Element /Sequence**
  p(¬e)= e.
  p(<a¬(ab) c>) =<a(ab) c>.
- **Max Positive Sub-sequence**
  MPS(<a¬(ab) c>) = <ac>.

Constraints to Negative Sequence

Constraint 1. Frequency Constraint

This paper only focuses on the negative sequences ns whose positive partner is frequent, i.e., sup(p(ns))>=min_ sup.

Constraint 2. Format Constraint

Continuous negative elements in a NSC are not allowed.

< ¬(ab) c ¬ d> ✔

< ¬(ab) ¬ c d> ✗

Constraint 3. Element Negative Constraint

The minimum negative unit in a NSC is an element.

< ¬(ab) c d> ✔

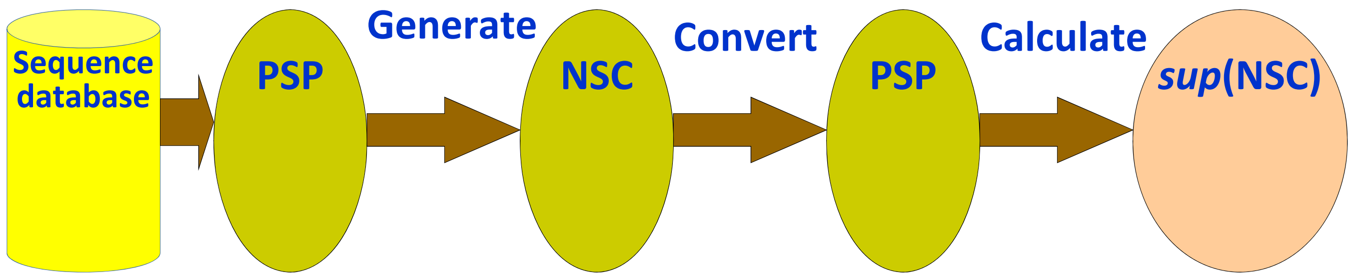
<(¬ab) c d> ✗

What does This Paper Do

E-NSP: Only use corresponding PSP information to calculate the support of negative sequence, without additional database scanning.

- A definition about negative containment.
- Three constraints for negative sequence
- A smart method to generate negative sequence candidate (NSC).
- A conversion strategy to convert negative containment problems to positive containment problems.
- A method to calculate the support of NSC.

The framework of E-NSP



1. **Mine all PSP by traditional PSP mining algorithms;**
2. **Generate NSC based on these PSP;**
3. **Convert these NSC to corresponding PSP;**
4. **Get supports of NSC by calculating support of corresponding PSP.**

Negative Containment Definition

Definition 4. Negative Containment Definition

Let ds=<d_1 d_2 ... d_t> be a data sequence, ns=<s_1 s_2 ... s_m> be an *m-size* and *n-neg-size* negative sequence, (1) if $m>2t+1$, then *ds does not contain ns*; (2) if $m=1$ and $n=1$, then *ds contains ns* when $p(ns) \nsubseteq ds$; (3) otherwise, *ds contains ns* if, $\forall (s_i, id(s_i)) \in EidS_{ns}^{-}$ $(1 \leqslant i \leqslant m)$, one of the following three holds:

(a) $(lsb=1)$ or $(lsb>1) \wedge p(s_1) \nsubseteq <d_1 \ldots d_{lsb-1}>$, when $i=1$,

(b) $(fse=t)$ or $(0<fse<t) \wedge p(s_m) \nsubseteq <d_{fse+1} \ldots d_t>$, when $i=m$,

(c) $(fse>0 \wedge lsb=fse+1)$ or $(fse>0 \wedge lsb>fse+1) \wedge p(s_i) \nsubseteq <d_{fse+1} \ldots d_{lsb-1}>$, when $1<i<m$,

where $fse=FSE(MPS(<s_1\ s_2 \ldots s_{i-1}>), ds)$, $lsb=LSB(MPS(<s_{i+1} \ldots s_m>), ds)$.

Negative Containment Definition

$$
\begin{array}{ccc}
\textbf{ns=<}\ ns_{left}, & \neg e, & ns_{right}\ > \\
MPS(ns_{left}) & e & MPS(ns_{right}) \\
\subseteq & \not\subset & \subseteq \\
\text{ds=<}s_1,\ldots\ldots,s_i, & s_{i+1},\ldots s_{j-1}, & s_j,\ldots\ldots\ldots s_t>
\end{array}
$$

ds **contains** ***ns*** **if** $<s_1,\ldots,s_i>$ **contain** ***MPS*(*ns*$_{left}$)** **,** $<s_j,\ldots s_t>$ contain ***MPS*(*ns*$_{right}$)** **, and** $< s_{i+1},\ldots s_{j-1}, >$ **doesn't contain** $<e>$**.** **(To EACH negative element ¬e in *ns*)**

©Longbing.Cao

Example: Negative Containment Definition

ns=<*a*¬*bb*(*cde*)>. *ds*=<*a*(*bc*)*d*(*cde*)>.

< *a* ¬*b* *b*(*cde*)>

$\subseteq$ $\not\subset$ $\subseteq$

ds= <*a* (*bc*)*d*(*cde*)>.

***ds* contains *ns*.**

Definitions

1-neg-size Maximum Sub-sequence **is a sequence that includes *MPS*(*ns*) and one negative element e in original sequence order.**

1-neg-size maximum sub-sequence set **is a set that includes all 1-neg-size maximum sub-sequences of *ns*, denoted as $1\text{-}negMSS_{ns}$.**

Example ***ns*=<*a*¬*bc*¬*d*>,**

***1-negMSSns* ={<*a*¬*bc*>, <*ac*¬*d*>}**

Negative Conversion Strategy

Given a data sequence ds=<d_1 d_2 ... d_t>, and ns=<s_1 s_2 ... s_m>, which is an *m-size* and *n-neg-size* negative sequence, the negative containment definition can be converted as follows: data sequence ds contains negative sequence ns if and only if the two conditions hold: (1) $MPS(ns) \subseteq ds$; and (2) $\forall$ *1-negMS* $\in$ *1-negMSS*$_{ns}$, *p(1-negMS)* $\not\subseteq$ ds.

Example ***ns*** **=<*a*¬*bb*¬*a*(*cde*)>, *ds*=<*a*(*bc*)*d*(*cde*)>.**

1-*negMSSns*={ <*a*¬*bb*(*cde*)> , <*ab*¬*a*(*cde*)> }

(1)MPS(ns)=< *ab*(*cde*)>$\subseteq$ds;

(2)p(<*a*¬*bb*(*cde*)>)= <*abb*(*cde*)> $\not\subset$ ds;

p(<*ab*¬*a*(*cde*)>)= <*aba*(*cde*)> $\not\subset$ ds;

ds* contains *ns

Negative Conversion Strategy

problem
whether a data sequence contains a negative sequence

→

problem
whether the data sequence does not contain its corresponding positive sequences

Now we can calculate the support of NSC only using the NSC's corresponding PSP.

Calculate the Support of NS

$$sup(ns)=|\{ ns \}|=|\{MPS(ns)\} - \bigcup_{i=1}^{n} \{p(1\text{-}negMS_i)\}| \quad (1)$$

Because $\bigcup_{i=1}^{n} \{p(1\text{-}negMS_i)\} \subseteq \{MPS(ns)\}$**, equation 1 can be rewritten as:**

$$sup(ns)= |\{MPS(ns)\}| - |\bigcup_{i=1}^{n} \{p(1\text{-}negMS_i)\}|$$

$$= sup(MPS(ns)) - |\bigcup_{i=1}^{n} \{p(1\text{-}negMS_i)\}| \quad (2)$$

Example 10 $sup(<a\neg bc\neg de>)=sup(<ace>\})-|\{<abce>\} \cup \{<acde>\}|;$

$sup(<\neg aa\neg a>)=sup(<a>)-|\{<aa>\}\cup\{<aa>\}|=sup(<a>)-sup(<aa>).$

***If** ns* **only contains a negative element, the support of** *ns* **is:**

$$sup(ns) = sup(MPS(ns)) - sup(p(ns)) \quad (3)$$

Example 11 $sup(<a\neg bce>) = sup(<ace>) - sup(<abce>)$

Specially, for negative sequence $<\neg e>$**,**

$$sup(<\neg e>) =|D| - sup(<e>). \quad (4)$$

Calculate the Support of NS

$$sup(ns) = |\{MPS(ns)\}| - |\cup_{i=1}^{n}\{p(1\text{-}negMS_i)\}|$$
$$= \underline{sup(MPS(ns))} - |\underline{\cup_{i=1}^{n}\{p(1 - negMS_i)\}}| \qquad (2)$$

Known

PSP	Support	{sid}
<a>	4	-
<b>	3	-
<c>	2	-
<a a>	3	{20,30,40}
<a b>	3	{10,20,30}
<a c>	2	{10,30}
<b c>	2	{10,30}
<(ab)>	2	-
<a b c>	2	{10,30}
<a (ab)>	2	{20,30}

Calculate the union set of {*p*(1-*negMSi*)}. (*p*(1-*negMSi*) are frequent.)

Negative Sequential Candidates Generation

Definition . e-NSP Candidate Generation

For a *k*-size PSP, its NSC are generated by changing any *m* non-contiguous element(s) to its (their) negative one(s), *m=1,2, …,*$\lceil k/2 \rceil$, where $\lceil k/2 \rceil$ is a minimum integer that is not less than *k/*2.

Example. **s= *<(ab) c d>* include:**

m=1, <¬(ab) c d>,<(ab) ¬cd>,<(ab) c¬d>;

m=2, <¬(ab) c ¬d>.

An Example

Table 1: Example Data Set

Sid	Data Sequence
10	$<a\ b\ c>$
20	$<a\ (ab)>$
30	$<(ae)\ (ab)\ c>$
40	$<a\ a>$
50	$<d>$

Table 2: Example Result - Positive Patterns

PSP	Support	{sid}
$<a>$	4	-
$<b>$	3	-
$<c>$	2	-
$<a\ a>$	3	{20,30,40}
$<a\ b>$	3	{10,20,30}
$<a\ c>$	2	{10,30}
$<b\ c>$	2	{10,30}
$<(ab)>$	2	-
$<a\ b\ c>$	2	{10,30}
$<a\ (ab)>$	2	{20,30}

An Example

Table 3: Example Result - NSC and Support (min_sup=2)

PSP	NSC	Related PSP	Sup
$<a>$	$<\neg a>$	$<a>$	1
$<b>$	$<\neg b>$	$<b>$	**2**
$<c>$	$<\neg c>$	$<c>$	**3**
$<a\ a>$	$<\neg a\ a>$	$<a>, <a\ a>$	1
	$<a\ \neg a>$	$<a>, <a\ a>$	1
$<a\ b>$	$<\neg a\ b>$	$<b>, <a\ b>$	0
	$<a\ \neg b>$	$<a>, <a\ b>$	1
$<a\ c>$	$<\neg a\ c>$	$<c>, <a\ c>$	0
	$<a\ \neg c>$	$<a>, <a\ c>$	**2**
$<b\ c>$	$<\neg b\ c>$	$<c>, <b\ c>$	0
	$<b\ \neg c>$	$<b>, <b\ c>$	1
$<(ab)>$	$<\neg (ab)>$	$<(ab)>$	**3**
$<a\ (ab)>$	$<\neg a\ (ab)>$	$<(ab)>, <a\ (ab)>$	0
	$<a\ \neg (ab)>$	$<a>, <a\ (ab)>$	**2**
$<a\ b\ c>$	$<\neg a\ b\ c>$	$<b\ c>, <a\ b\ c>$	0
	$<a\ \neg b\ c>$	$<a\ c>, <a\ b\ c>$	0
	$<a\ b\ \neg c>$	$<a\ b>, <a\ b\ c>$	1
	$<\neg a\ b\ \neg c>$	$<b>, <a\ b>, <b\ c>$	0

Experiment and Evaluation

Data Sets

Four source datasets including both real data and synthetic datasets generated by IBM data generator. Partition these datasets to 14 datasets according to different data factors.

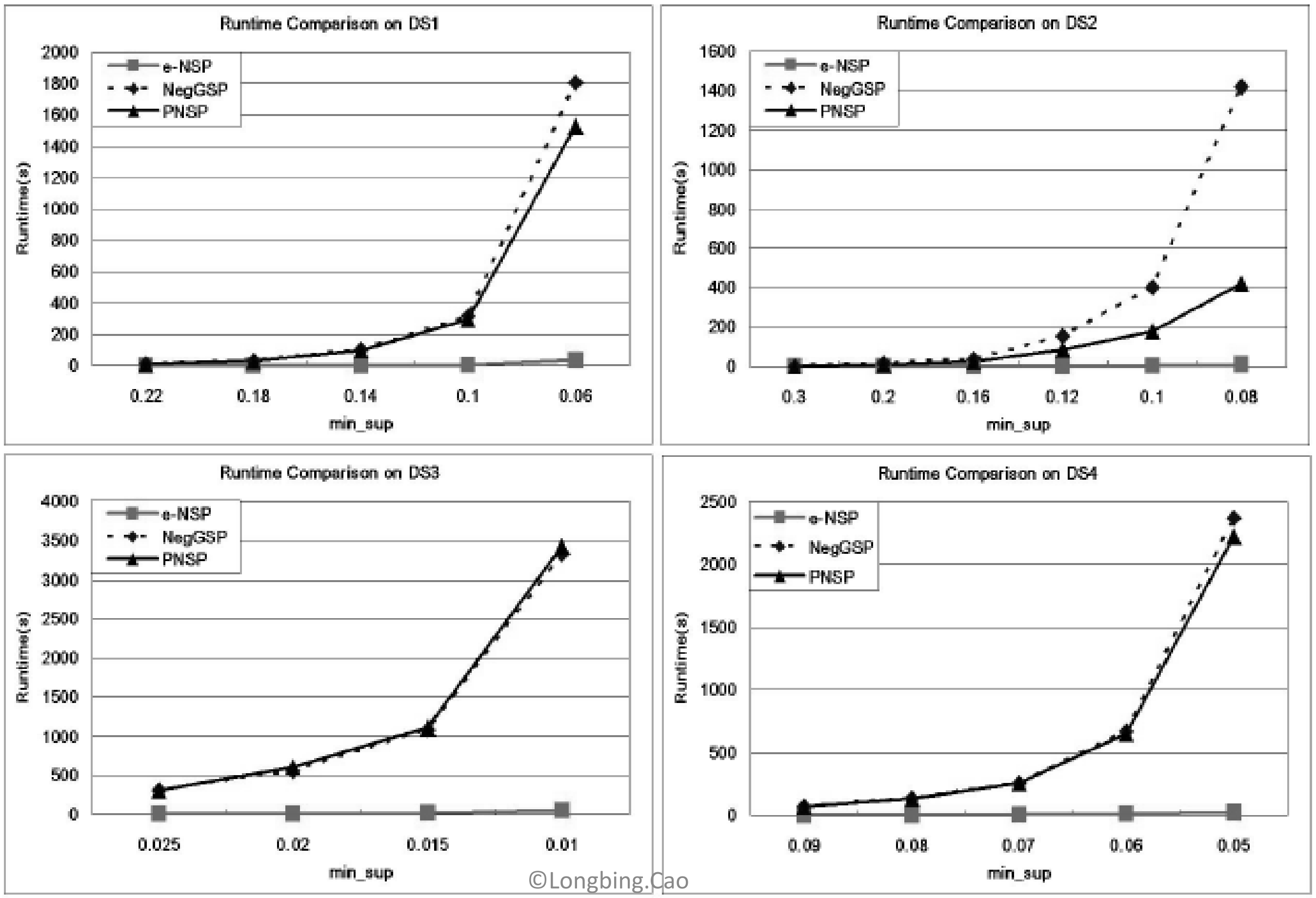
An Example

Table 4: Dataset Characteristics Analysis Result

ID	Dataset Characteristics	min sup	NGSP (t_1,s)	PNSP (t_2,s)	eNSP (t_3,s)	t_3/t_2
DS1	C8T4S6I6.DB10k.N100	0.04 0.06 0.08	1451.7 241.4 78.9	638.2 163.1 61.9	14.94 4.16 1.53	2.3% 2.5% 2.5%
DS1.1	**C4**T4S6I6.DB10k.N100	0.01 0.015 0.02	517.5 130.4 48.0	208.4 64.5 28.4	1.08 0.33 0.16	0.5% 0.5% 0.5%
DS1.2	**C12**T4S6I6.DB10k.N100	0.14 0.16 0.18	229.0 127.6 73.8	191.9 109.5 66.9	7.99 4.49 2.53	4.2% 4.1% 3.8%
DS1.3	C8**T8**S6I6.DB10k.N100	0.22 0.24 0.26	130.8 83.7 55.9	118.5 76.5 52.8	5.22 3.19 2.14	4.4% 4.2% 4.1%
DS1.4	C8**T12**S6I6.DB10k.N100	0.3 0.4 0.5	1205.2 133.2 23.6	969.3 123.5 23.0	57.55 6.75 1.06	5.9% 5.5% 4.6%
DS1.5	C8T4**S12**I6.DB10k.N100	0.04 0.06 0.08	1130.0 187.0 61.2	478.6 124.7 47.5	12.22 3.39 1.23	2.6% 2.7% 2.6%
DS1.6	C8T4**S18**I6.DB10k.N100	0.04 0.06 0.08	297.1 64.2 23.5	157.4 45.5 19.0	3.47 0.97 0.36	2.2% 2.1% 1.9%
DS1.7	C8T4S6**I10**.DB10k.N100	0.06 0.07 0.08	690.2 334.7 188.1	395.1 227.5 138.0	7.33 4.23 2.63	1.9% 1.9% 1.9%
DS1.8	C8T4S6**I14**.DB10k.N100	0.08 0.1 0.12	983.9 320.5 141.8	630.8 248.9 112.7	8.88 3.63 1.61	1.4% 1.5% 1.4%
DS1.9	C8T4S6I6.DB10k.**N200**	0.03 0.04 0.05	378.2 101.8 39.5	98.4 43.1 23.3	0.59 0.17 0.06	0.6% 0.4% 0.3%
DS1.10	C8T4S6I6.DB10k.**N400**	0.015 0.02 0.025	823.0 197.3 99.8	97.4 42.0 20.6	0.08 0.03 0.02	0.1% 0.1% 0.1%

Experiment and Evaluation

Computational Cost



Conclusions

- **We have proposed a simple but very efficient NSP mining algorithm: e-NSP. E-NSP includes:**

■ A formal definition, negative containment, to define how a data sequence contains a negative sequence.

■ A negative conversion strategy to convert negative containing problems to positive containing problems.

■ A method to calculate the supports of NSC only using the corresponding PSP.

■ A simple but efficient approach to generate NSC.

■ The experimental results and comparisons on 14 datasets from different data characteristics perspectives have clearly shown that e-NSP is much more efficient than existing approaches.

Coupled/Group/Collective Behavior Analysis

Physical world

Intelligent Transport Systems

Virtual world

Problem-solving world

What is Coupled Behavior?

Longbing Cao. Coupling Learning of Complex Interactions, Information Processing and Management, 51(2): 167-186 (2015)

Cao, L., Ou, Y., Yu, P.S. Coupled Behavior Analysis with Applications, *IEEE Transactions on Knowledge and Data Engineering*, 24 (8): 1378-1392, 2012.

Relationship crossing behaviors

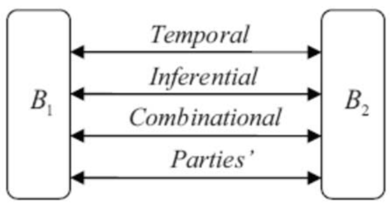


Figure 6: Relationships between Multiple Behaviors

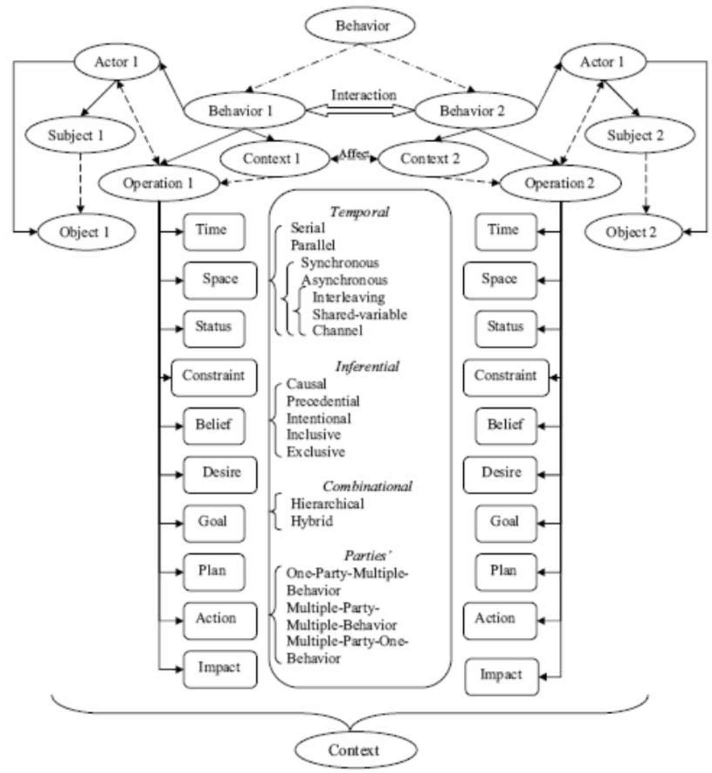


Figure 7: Relationships between behaviors

Behavior Coupling Types

- Logic/semantic relation based behavior coupling
- Statistical/Probabilistic relation based behavior coupling

Behavior Feature Matrix

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

I actors (customers): $\{\mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_I\}$

J_i behaviors for an actor $\mathscr{E}_i$: $\{\mathbb{B}_{i1}, \mathbb{B}_{i2}, \ldots, \mathbb{B}_{iJ_i}\}$

Behavior $\mathbb{B}_{ij}$: $\overrightarrow{\mathbb{B}}_{ij} = ([p_{ij}]_1, [p_{ij}]_2, \cdots, [p_{ij}]_K)$

Behavior Feature Matrix :

$$FM(\mathbb{B}) = \begin{pmatrix} \mathbb{B}_{11} & \mathbb{B}_{12} & \ldots & \mathbb{B}_{1J_{max}} \\ \mathbb{B}_{21} & \mathbb{B}_{22} & \ldots & \mathbb{B}_{2J_{max}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{B}_{I1} & \mathbb{B}_{I2} & \ldots & \mathbb{B}_{IJ_{max}} \end{pmatrix}$$

An Example of Stock Market

Transactional Data → **Behavior Feature Matrix**

	Investor	Time	Direction	Price	Volume
B1	(1)	09:59:52	Sell	12.0	155
B2	(2)	10:00:35	Buy	11.8	2000
B3	(3)	10:00:56	Buy	11.8	150
B4	(2)	10:01:23	Sell	11.9	200
B5	(1)	10:01:38	Buy	11.8	200
B6	(4)	10:01:47	Buy	11.9	200
B7	(5)	10:02:02	Buy	11.9	250
B8	(2)	10:02:04	Sell	11.9	500

$$FM(\mathbb{B}) = \begin{pmatrix} B_1 & B_5 & \emptyset \\ B_2 & B_4 & B_8 \\ B_3 & \emptyset & \emptyset \\ B_6 & \emptyset & \emptyset \\ B_7 & \emptyset & \emptyset \end{pmatrix}$$

Behavior Intra-relationship

Definition 2. *(Intra-Coupled Behaviors) Actor* $\mathscr{E}_i$*'s behaviors* $\mathbb{B}_{ij}$ $(1 \leq j \leq J_{max})$ *are intra-coupled in terms of coupling function* $\theta_j(\cdot)$,

$$\mathbb{B}_{i\cdot}^{\theta} ::= \mathbb{B}_{i\cdot}(\mathscr{E}, \mathscr{O}, \mathscr{C}, \theta) | \sum_{j=1}^{J_{max}} \theta_j(\cdot) \odot \mathbb{B}_{ij} \tag{1}$$

$$|\theta_j(\cdot)| \geq \theta_0 \tag{2}$$

where θ_0 *is the intra-coupling threshold,* $\sum_{j=1}^{J_{max}} \odot$ *means the subsequent behavior of* $\mathbb{B}_i$ *is* $\mathbb{B}_{ij}$ *intra-coupled with* $\theta_j(\cdot)$*, and so on, with nondeterminism.*

$$FM(\mathbb{B}) = \begin{pmatrix} \mathbb{B}_{11} & \mathbb{B}_{12} & \dots & \mathbb{B}_{1J_{max}} \\ \mathbb{B}_{21} & \mathbb{B}_{22} & \dots & \mathbb{B}_{2J_{max}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{B}_{I1} & \mathbb{B}_{I2} & \dots & \mathbb{B}_{IJ_{max}} \end{pmatrix}$$

Behavior Inter-relationship

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Definition 3. *(Inter-Coupled Behaviors) Actor $\mathscr{E}_i$'s behaviors $\mathbb{B}_{ij}$ $(1 \le i \le I)$ are inter-coupled with each other in terms of coupling function $\eta_i(\cdot)$,*

$$\mathbb{B}^{\eta}_{\cdot j} ::= \mathbb{B}_{\cdot j}(\mathscr{E}, \mathscr{O}, \mathscr{C}, \eta) | \sum_{i=1}^{I} \eta_i(\cdot) \odot \mathbb{B}_{ij} \qquad (3)$$

$$|\eta_i(\cdot)| \ge \eta_0 \qquad (4)$$

where η_0 is the inter-coupling threshold, $\sum_i^I \odot$ means the subsequent behavior of $\mathbb{B}_i$ is $\mathbb{B}_{ij}$ inter-coupled with $\eta_i(\cdot)$, and so on, with nondeterminism.

$$FM(\mathbb{B}) = \begin{pmatrix} \mathbb{B}_{11} & \mathbb{B}_{12} & \dots & \mathbb{B}_{1J_{max}} \\ \mathbb{B}_{21} & \mathbb{B}_{22} & \dots & \mathbb{B}_{2J_{max}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{B}_{I1} & \mathbb{B}_{I2} & \dots & \mathbb{B}_{IJ_{max}} \end{pmatrix}$$

Behavior Relationship

Definition 4 *(Coupled Behaviors) Coupled behaviors* $\mathbb{B}_c$ *refer to behaviors* $\mathbb{B}_{i_1 j_1}$ *and* $\mathbb{B}_{i_2 j_2}$ *that are coupled in terms of relationships* $f(\theta(\cdot), \eta(\cdot))$, *where* $(i_1 \neq i_2) \vee (j_1 \neq j_2) \wedge (1 \leq i_1, i_2 \leq I) \wedge (1 \leq j_1, j_2 \leq J_{max})$

$$\mathbb{B}_c = (\mathbb{B}^{\theta}_{i_1 j_1})^{\eta} * (\mathbb{B}^{\theta}_{i_2 j_2})^{\eta} ::= \mathbb{B}_{ij}(\mathscr{E}, \mathscr{O}, \mathscr{C}, \mathscr{R}) | \sum_{i_1, i_2 = 1}^{I} \sum_{j_1, j_2 = 1}^{J_{max}} f(\theta_{j_1 j_2}(\cdot), \eta_{i_1 i_2}(\cdot)) \odot (\mathbb{B}_{i_1 j_1} \mathbb{B}_{i_2 j_2}) \tag{5}$$

$$FM(\mathbb{B}) = \begin{pmatrix} \mathbb{B}_{11} & \mathbb{B}_{12} & \cdots & \mathbb{B}_{1J_{max}} \\ \mathbb{B}_{21} & \mathbb{B}_{22} & \cdots & \mathbb{B}_{2J_{max}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{B}_{I1} & \mathbb{B}_{I2} & \cdots & \mathbb{B}_{IJ_{max}} \end{pmatrix}$$

Behavior Behavior Analysis

Theorem 1. *(Coupled Behavior Analysis (CBA)) The analysis of coupled behaviors* (CBA Problem *for short) is to build the objective function* $g(\cdot)$ *under the condition that behaviors are coupled with each other by coupling function* $f(\cdot)$*, and satisfy the following conditions.*

$$f(\cdot) ::= f(\theta(\cdot), \eta(\cdot)), \tag{9}$$

$$g(\cdot)|(f(\cdot) \geq f_0) \geq g_0 \tag{10}$$

Logic/Semantic Relation-based Group Behavior Analysis

Longbing Cao. Combined Mining: Analyzing Object and Pattern Relations for Discovering and Constructing Complex but Actionable Patterns, *WIREs Data Mining and Knowledge Discovery*.

Longbing Cao. Zhao Y., Zhang, C. Mining Impact-Targeted Activity Patterns in Imbalanced Data, *IEEE Trans. on Knowledge and Data Engineering*, 20(8): 1053-1066, 2008.

Coupling relationships

- From temporal aspect
  - Serial Coupling: $TS_1; TS_2; \cdots ; TS_n$
  - Interleaving Coupling: $TS_1 : TS_2 : \cdots : TS_n$
  - Shared-variable Coupling: $TS_1 ||| TS_2 ||| \cdots ||| TS_n$
  - Channel System Coupling: $TS_1 \mid TS_2 \mid \cdots \mid TS_n$
  - Synchronous Coupling: $TS_1 \parallel TS_2 \parallel \cdots \parallel TS_n$
- From inferential aspect
  - Causal Coupling: $TS_1 \rightarrow TS_2$
  - Precedential Coupling: $TS_1 \Rightarrow TS_2$
  - Intentional Coupling: $TS_1 \twoheadrightarrow TS_2$
  - Inclusive Coupling: $TS_1 \mapsto TS_2$
  - Exclusive Coupling: $TS_1 \oplus TS_2$
- From combinational aspect
  - Hierarchical Coupling: $f(g(TS_1, TS_2, \cdots, TS_n))$
  - Hybrid Coupling: $f(TS_1).g(TS_2)$, $f(TS_1)^*$, $(TS_1)^\omega$
  - One-Party-Multiple-Behavior Coupling: $f(TS_1, TS_2, \cdots, TS_n)^{[A_1]}$
  - Multiple-Party-One-Behavior Coupling: $f(TS_1)^{[A_1 A_2 \cdots, A_n]}$
  - Multiple-Party-Multiple-Behavior Coupling: $f(TS_1, TS_2, \cdots, TS_n)^{[A_1 A_2 \cdots A_n]}$

Basic Behavior Patterns

- Tracing: Different actions with sequential order.

$$\{a_1, a_2, \cdots, a_n\}$$

- Consequence: Different actions have causalities in occurrence.

$$\{a_i \rightarrow a_j\}$$

- Synchronization: Different actions occur at the same time.

$$\{a_1 \leftrightarrow, \cdots, \leftrightarrow a_n\}$$

- Combination: Different actions occur in concurrency.

$$\{a_1 \| a_2 \|, \cdots, \| a_n\}$$

- Exclusion: Different actions occur mutually exclusively.

$$\{a_1 \oplus a_2 \oplus, \cdots, \oplus a_n\}$$

- Precedence: Different actions have required precedence

$$\{a_i \Rightarrow a_j\}$$

And more to be explored…

- *Sequential Combination* $\longrightarrow A \times B \times C \times \cdots$
- *Parallel Combination* $\longrightarrow A \otimes B \otimes C \otimes \cdots$
- *Nested Combination*
- *Fuzzy or probabilistic Combination*

Logic Coupling Based Combined Pattern Pairs

DEFINITION EXTENDED COMBINED PATTERN PAIRS. *An Extended Combined Pattern Pair (ECPP) is a special combined pattern pair as follows*

Group 1 behavior

$$\mathcal{E}: \begin{cases} X_{\mathrm{p}} \rightarrow T_1 \\ X_{\mathrm{p}} \wedge X_{\mathrm{e}} \rightarrow T_2 \end{cases},$$

Group 2 behavior

where $X_{\mathrm{p}} \neq \emptyset$, $X_{\mathrm{e}} \neq \emptyset$ *and* $X_{\mathrm{p}} \cap X_{\mathrm{e}} = \emptyset$.

Logic Coupling Based Combined Pattern Clusters

DEFINITION EXTENDED COMBINED PATTERN SEQUENCES. *An Extended Combined Pattern Sequence (ECPC), or called Incremental Combined Pattern Sequence (ICPS), is a special combined pattern cluster with additional items appending to the adjacent local patterns incrementally.*

$$\mathcal{S}: \begin{cases} X_{\mathrm{p}} \to T_1 \\ X_{\mathrm{p}} \wedge X_{\mathrm{e},1} \to T_2 \\ X_{\mathrm{p}} \wedge X_{\mathrm{e},1} \wedge X_{\mathrm{e},2} \to T_3 \\ \cdots \\ X_{\mathrm{p}} \wedge X_{\mathrm{e},1} \wedge X_{\mathrm{e},2} \wedge \cdots \wedge X_{\mathrm{e,k-1}} \to T_k \end{cases},$$

Group 1 behavior

Group K behavior

where $\forall i$, $1 \leq i \leq k-1$, $X_{i+1} \cap X_i = X_i$ *and* $X_{i+1} \setminus X_i = X_{e,i} \neq \emptyset$, *i.e.,* X_{i+1} *is an* increment *of* X_i. *The above cluster of rules actually makes a sequence of rules, which can show the impact of the increment of patterns on the outcomes.*

Multi-group Pattern Relation

- Type A: Demographics differentiated combined pattern
  - Customers with the same actions but different demographics
    → different classes/business impact

$$\text{Type A:} \begin{cases} A_1 + D_1 & \rightarrow & \text{quick payer} \\ A_1 + D_2 & \rightarrow & \text{moderate payer} \\ A_1 + D_3 & \rightarrow & \text{slow payer} \end{cases}$$

Multi-group Pattern Relation

- Type B: Action differentiated combined pattern
  - Customers with the same demographics but taking different actions → different classes/business impact

$$\text{Type B:} \begin{cases} A_1 + D_1 & \rightarrow & \text{quick payer} \\ A_2 + D_1 & \rightarrow & \text{moderate payer} \\ A_3 + D_1 & \rightarrow & \text{slow payer} \end{cases}$$

Multiple Group Pattern Relations

An Example of Combined Pattern Clusters

Clusters	Rules	X_p	X_e		T	Cnt	$Conf$	I_r	I_c	$Lift$	$Cont_p$	$Cont_e$	$Lift$ of	$Lift$ of
		demographics	arrangements	repayments			(%)						$X_p \to T$	$X_e \to T$
$\mathcal{P}_1$	P_5	marital:sin	irregular	cash or post	A	400	83.0	1.12	0.67	1.80	1.01	2.00	0.90	1.79
	P_6	&gender:F	withhold	cash or post	A	520	78.4	1.00		1.70	0.89	1.89	0.90	1.90
	P_7	&benefit:N	withhold & irregular	cash or post & withhold	B	119	80.4	1.21		2.28	1.33	2.06	1.10	1.71
	P_8		withhold	cash or post & withhold	B	643	61.2	1.07		1.73	1.19	1.57	1.10	1.46
	P_9		withhold & vol. deduct	withhold & direct debit	B	237	60.6	0.97		1.72	1.07	1.55	1.10	1.60
	P_{10}		cash	agent	C	33	60.0	1.12		3.23	1.18	3.07	1.05	2.74
$\mathcal{P}_2$	P_{11}	age:65+	withhold	cash or post	A	1980	93.3	0.86	0.59	2.02	1.06	1.63	1.24	1.90
	P_{12}		irregular	cash or post	A	462	88.7	0.87		1.92	1.08	1.55	1.24	1.79
	P_{13}		withhold & irregular	cash or post	A	132	85.7	0.96		1.86	1.18	1.50	1.24	1.57
	P_{14}		withhold & irregular	withhold	C	50	63.3	2.91		3.40	2.47	4.01	0.85	1.38

Multi-Group Combined Patterns

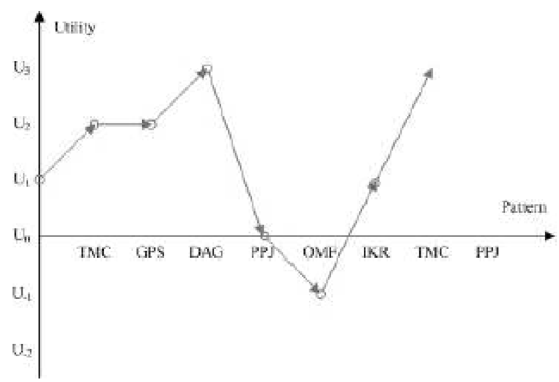


Figure 2: Pattern Evolution Chart

$$
\begin{cases}
TMC \rightarrow U_1 \\
TMC, GPS \rightarrow U_2 \\
TMC, GPS, DAG \rightarrow U_2 \\
TMC, GPS, DAG, PPJ \rightarrow U_3 \\
TMC, GPS, DAG, PPJ, OMF \rightarrow U_0 \\
TMC, GPS, DAG, PPJ, OMF, IKR \rightarrow U_{-1} \\
TMC, GPS, DAG, PPJ, OMF, IKR, TMC \rightarrow U_1 \\
TMC, GPS, DAG, PPJ, OMF, IKR, TMC, PPJ \rightarrow U_3
\end{cases}, \tag{6}
$$

Multi-Group Combined Patterns

$$
\begin{cases}
PLN \to T \\
PLN, DOC \to T \\
PLN, DOC, DOC \to T \\
PLN, DOC, DOC, DOC \to T \\
PLN, DOC, DOC, DOC, REA \to T \\
PLN, DOC, DOC, DOC, REA, IES \to T
\end{cases}
$$

Divergence vs. convergence of group behaviors

Statistical/Probabilistic Coupled Behavior Analysis

Coupled Hidden Markov Model-based Abnormal Coupled Behavior Analysis

Cao, L., Ou Y, Yu PS, Wei G. Detecting Abnormal Coupled Sequences and Sequence Changes in Group-based Manipulative Trading Behaviors, *KDD2010*.

Cao, Longbing, Yuming Ou, and S. Yu Philip. "Coupled Behavior Analysis with Applications." *IEEE Transactions on Knowledge & Data Engineering* 8 (2011): 1378-1392.

Pool manipulation

TABLE 1
An example of buy and sell orders

Investor	Time	Direction	Price	Volume
(1)	09:59:52	Sell	12.0	155
(2)	10:00:35	Buy	11.8	2000
(3)	10:00:56	Buy	11.8	150
(2)	10:01:23	Sell	11.9	200
(1)	10:01:38	Buy	11.8	200
(4)	10:01:47	Buy	11.9	200
(5)	10:02:02	Buy	11.9	250
(2)	10:02:04	Sell	11.9	500

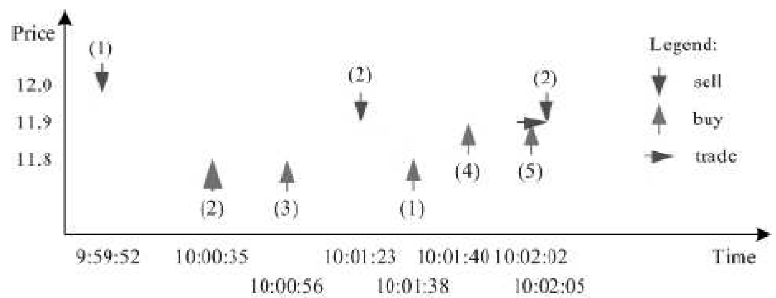
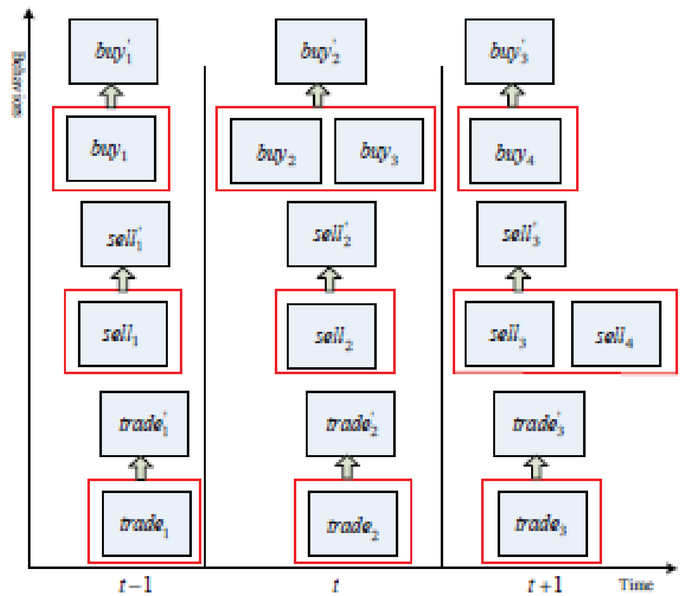


Fig. 1. Coupled Trading Behaviors



(a) An Example of Coupled Trading Behaviors in Stock Markets

Construct behavior sequences

$$\left\{\frac{Actor_i - Operation_i}{Attributes_i} \xrightarrow{\eta} \frac{Actor_j - Operation_j}{Attributes_j}\right\}_{i,j=1;winsize}^{I,J} \quad (12)$$

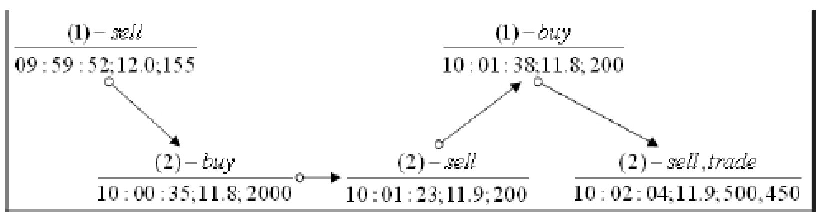


Fig. 2. Behavior sequences - Data Structure 1

$$Category : \{\frac{Actor_i - Operation_i}{Attributes_i} \xrightarrow{\eta} \frac{Actor_j - Operation_j}{Attributes_j}\}_{i,j=1;winsize}^{I,J} \quad (14)$$

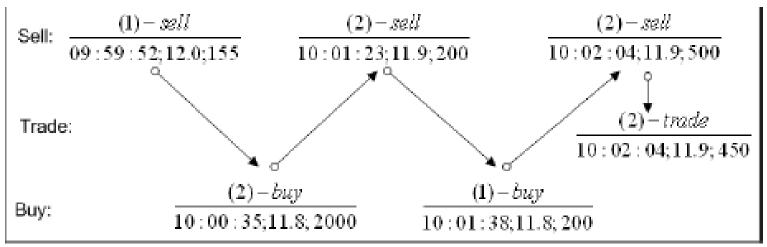


Fig. 3. Behavior sequences - Data Structure 2

CHMM Based Coupled Sequence Modeling

- Coupled behavior sequences
  - Multiple sequences

$$\Phi_1 = \{\phi_{11}, \ldots, \phi_{1T}\}$$
$$\Phi_2 = \{\phi_{21}, \ldots, \phi_{2F}\}$$
$$\Phi_C = \{\phi_{C1}, \ldots, \phi_{CG}\}$$

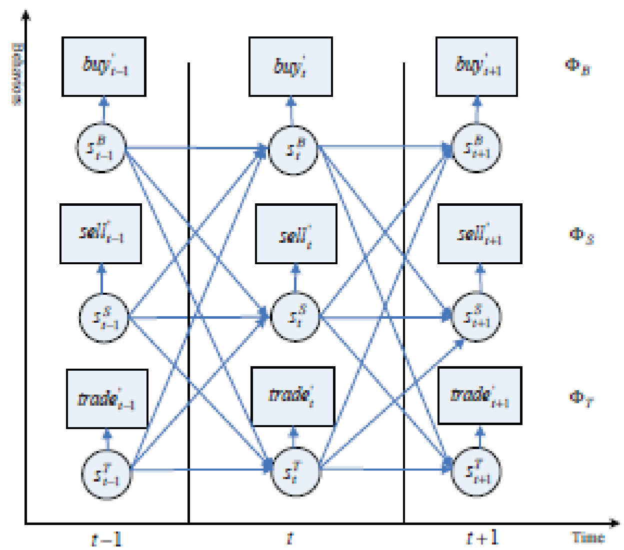
  - Coupling relationship

$$R_{ij}(\Phi_i, \Phi_j)$$

  - Behavior properties

$$R_{ij} \subset R, \qquad R_{ij}(\Phi_i, \Phi_j) = \varnothing$$

$$\phi_{ik}(p_{ik,1}, \ldots, p_{ik,L})$$



(b) The Structure of the CHMM

CBA - CHMM

$$CBA\ problem \rightarrow CHMM\ model \tag{15}$$

$$\Phi(\mathbb{B}_c)|category \rightarrow X \tag{16}$$

$$M(\Phi(\mathbb{B}_c))|\phi_{ik}([p_{ij}]_1, \ldots, [p_{ij}]_K) \rightarrow Y \tag{17}$$

$$f(\theta(\cdot), \eta(\cdot)) \rightarrow Z \tag{18}$$

$$Initial\ distribution\ of\ \Phi(\mathbb{B}_c)|category \rightarrow \pi \tag{19}$$

Framework: abnormal CBA

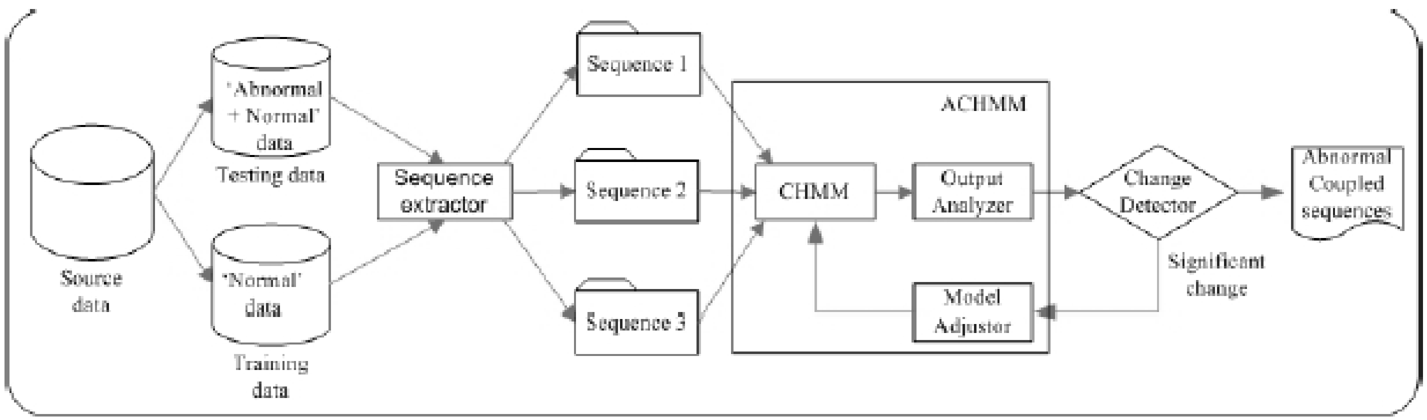


Fig. 5. Framework of abnormal coupled behavior detection

Hidden States

$$S^{buy} = \{Positive\ Buy, Neutral\ Buy, Negative\ Buy\}$$

$$S^{sell} = \{Positive\ Sell, Neutral\ Sell, Negative\ Sell\}$$

$$S^{trade} = \{Market\ Up, Market\ Down\}$$

Observation Sequences

Activity (A)

$$A = \{a_1, a_2, \ldots, \}$$

$$a_i = (a(t_i), p(t_i), v(t_i))$$

$$a(t_i) = \{buy \mid sell \mid trade\}$$

$$p(t_i) = \{buy\ price | sell\ price | trade\ price\}$$

$$v(t_i) = \{buy\ volume | sell\ volume | trade\ volume\}$$

Interval Activity (IA)

$$\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$$

$$A_i(a) = A_j(a)$$

$$\bar{p} = \frac{\sum_{i=1}^{n} p_i}{f} \quad f = |\mathcal{A}| = n \quad \bar{v} = \frac{\sum_{i=1}^{n} v_i}{f}$$

$$IA(\mathcal{A}, \bar{p}, \bar{v}, f) \xrightarrow{\text{quantization}} IA^{'}(p^{'}, v^{'}, f^{'})$$

Adaptive CHMM for Detecting Sequence Changes

Figure 3: Update Point of ACHMM

$$x_{ij}^{update} = (1-w)x_{ij}^{old} + w * x_{ij}^{new} \tag{15}$$

$$y_{ij}^{update} = (1-w)y_{ij}^{old} + w * y_{ij}^{new} \tag{16}$$

$$z_{ij'}^{update} = (1-w)z_{ij'}^{old} + w * z_{ij'}^{new} \tag{17}$$

$$\pi_{i}^{update} = (1-w)\pi_{i}^{old} + w * \pi_{i}^{new} \tag{18}$$

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

The Algorithm

Algorithm 1 Constructing observation sequences

Step 1: Segment the whole trading day into L intervals by a time window with the length $winsize$.

Step 2: Calculate IA for buy-order, sell-order and trade activities respectively in each window. They are denoted as IA_l^{buy}, IA_l^{sell} and IA_l^{trade}, respectively.

Step 3: Obtain $IA_l'^{buy}$, $IA_l'^{sell}$ and $IA_l'^{trade}$ by quantizing IA_l^{buy}, IA_l^{sell} and IA_l^{trade}.

Step 4: Obtain the trading activity sequnce IA^{buy} for buy-order by putting all $IA_l'^{buy}$ in a trading day together. Obtain IA^{sell} and IA^{trade} in the same way. We obtain

$$IA^{type} = IA_1'^{type}, IA_2'^{type}, \cdots, IA_L'^{type} \qquad (19)$$

where $type \in \{buy, sell, trade\}$. IA^{buy}, IA^{sell} and IA^{trade} are the observation sequences of CHMM in the day.

Step 5: Repeat Step 1-4 for each trading day

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Algorithm 2 Detecting abnormal trading sequences

Step 1: Construct trading sequences including training sequences $Seq_1, Seq_2, \cdots, Seq_K$ and test sequences $Seq_1^{'}, Seq_2^{'}, \cdots, Seq_{K'}^{'}$.

Step 2: Train the ACHMM model on the training sequences;

Step 3: Compute the mean (μ) and standard deviation (σ) of probability of training sequences according to the following formulas:

$$\mu = \frac{\sum_{i=1}^{K} Pr(Seq_i|ACHMM)}{K} \tag{20}$$

$$\sigma = \sqrt{\frac{1}{K}\sum_{i=1}^{K} Pr(Seq_i|ACHMM)) - \mu} \tag{21}$$

where K is the total number of training sequences, mean μ represents the centroid of model ACHMM, and the standard deviation σ represents the radius of model ACHMM.

Step 4: For each test sequence $Seq_i^{'}$, calculate its distance D_i to the centroid of model by

$$D_i = \frac{\mu - Pr(Seq_i^{'}|\mathcal{M})}{\sigma} \tag{22}$$

Consequently, $Seq_i^{'}$ is an exceptional pattern, if it satisfies:

$$D_i > \psi_0 \tag{23}$$

where ψ_0 is a given threshold.

- Benchmark Models
  - HMM-B
  - HMM-S
  - HMM-T
  - IHMM
  - CHMM
  - ACHMM

Evaluation

- Technical performance

$$
\begin{aligned}
Accuracy &= \frac{TP+TN}{TP+FN+FP+TN} && (43)\\
Precision &= \frac{TP}{TP+FP} && (44)\\
Recall &= \frac{TP}{TP+FN} && (45)\\
Specificity &= \frac{TN}{FP+TN} && (46)
\end{aligned}
$$

- Business performance

$$Return = ln\frac{p_t}{p_{t-1}} \qquad (48)$$

$$Abnormal\ Return = Return - (\gamma + \xi Return^{market}) \qquad (49)$$

EHAVIOR INFORMATICS
Discovering Behavior Intelligence

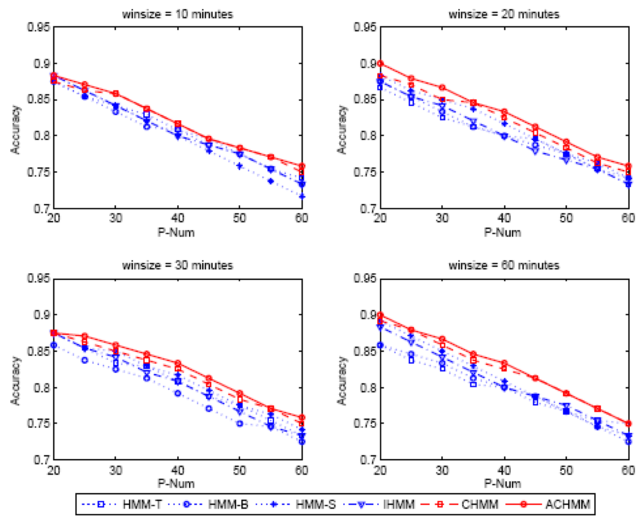


Figure 4: Accuracy of Six Models

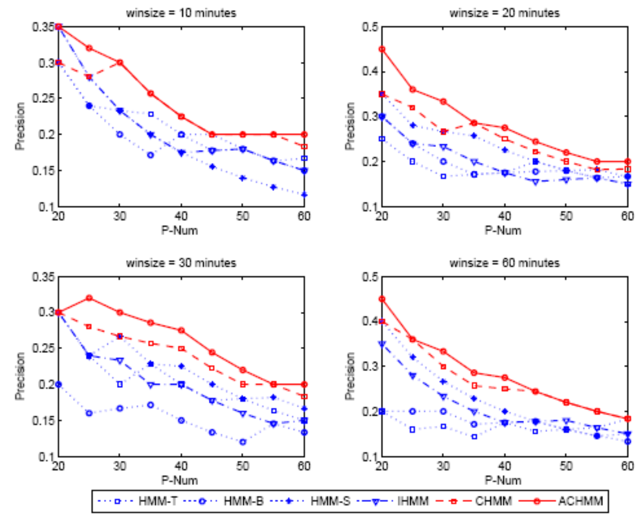


Figure 5: Precision of Six Models

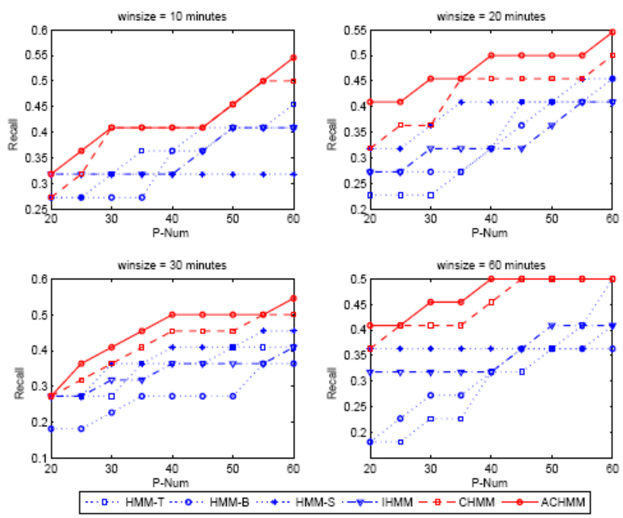


Figure 6: Recall of Six Models

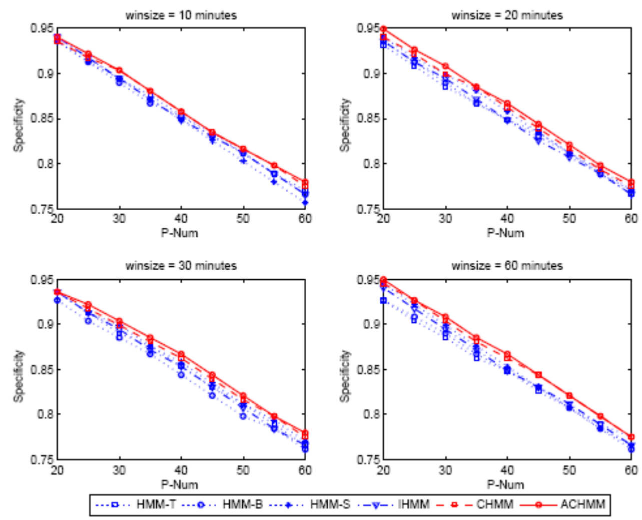


Figure 7: Specificity of Six Models

- Business Performance

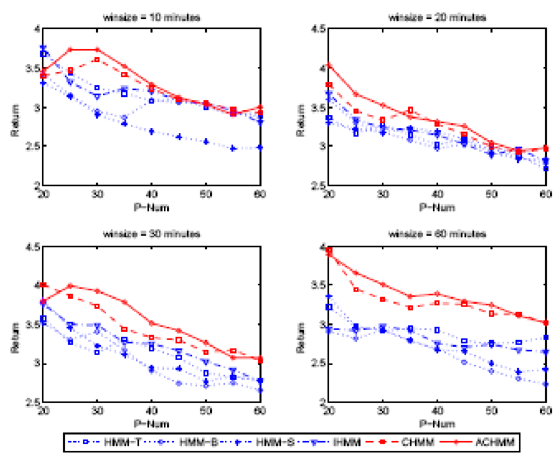


Fig. 9. Return of Six Models

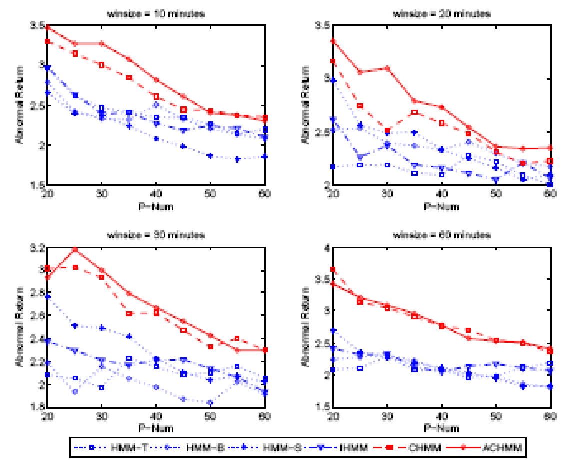


Fig. 10. Abnormal Return of Six Models

- Computational cost

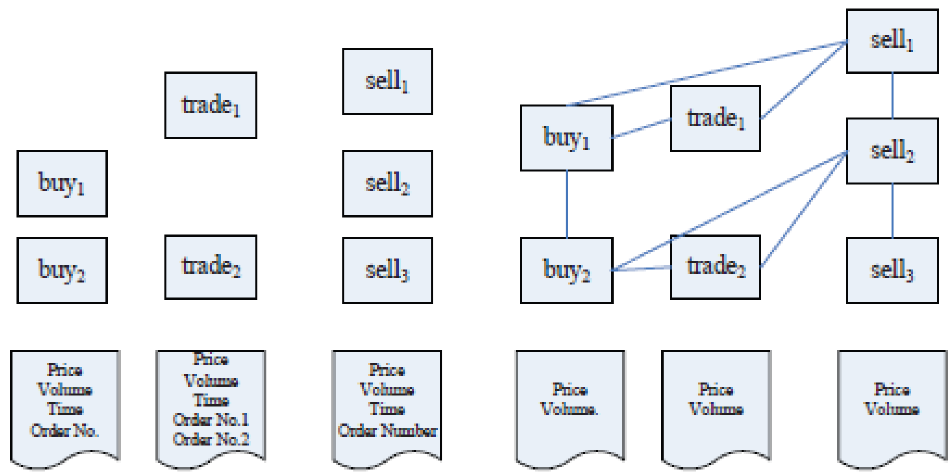
TABLE 5
Computational performance

		IHMM	CHMM	ACHMM
winsize =10 (m)	Training time (s)	0.574	11.978	11.988
	Test time (s)	0.056	1.296	3.576
winsize =20 (m)	Training time (s)	0.256	4.929	4.933
	Test time (s)	0.047	0.655	3.486
winsize =30 (m)	Training time (s)	0.206	4.121	4.119
	Test time (s)	0.042	0.447	2.429
winsize =60 (m)	Training time (s)	0.109	2.003	2.004
	Test time (s)	0.036	0.221	1.206

Conditional Probability Distribution-based Coupled Behavior Analysis

Yin Song, Longbing Cao, et al. Coupled Behavior Analysis for Capturing Coupling Relationships in Group-based Market Manipulation, *KDD* 2012, 976-984.

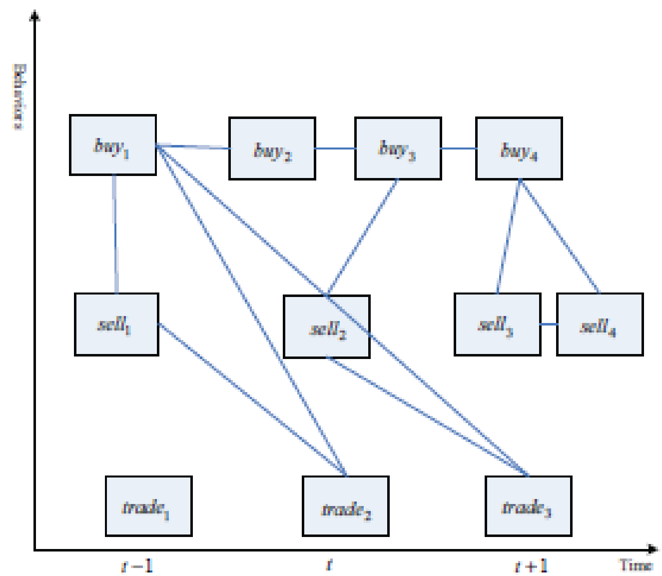
Yin Song and Longbing Cao. Graph-based Coupled Behavior Analysis: A Case Study on Detecting Collaborative Manipulations in Stock Markets, *IJCNN* 2012, 1-8.



(a) The Coupled Behaviors with Reference and Analysis Properties.

(b) Link Generation Using Reference Properties.

Graph-based Coupled Behavior Presentation



(c) The Structure of Graph-based Coupled Behavior Model

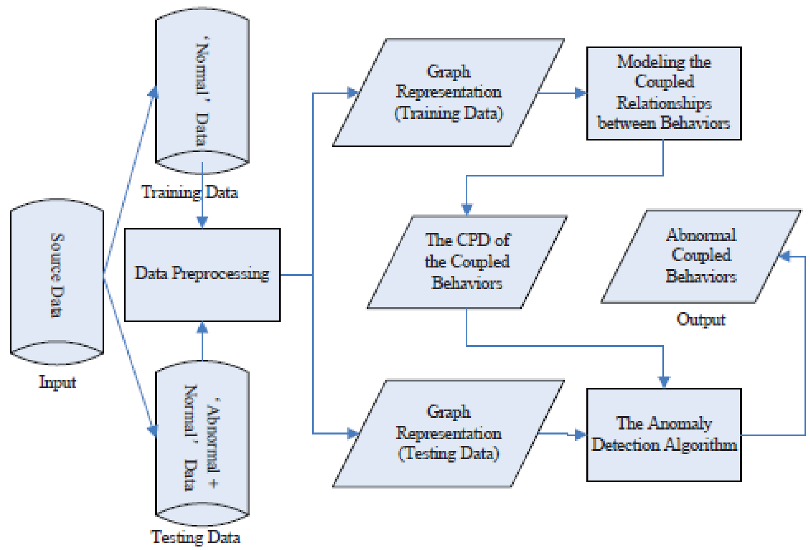
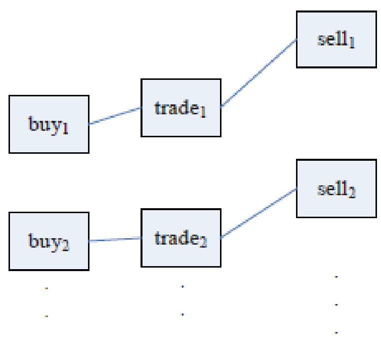


Figure 2: The Work Flow of the Proposed Framework.

Propositional Coupled Behavior

- CPD



(a) An Example of the Subgraphs for Each Target Behavior

	$X^{(t)}$	RF_1	RF_2	$\cdots$	RF_n
$trade_1$	x_1	rf_{11}	rf_{21}	$\cdots$	rf_{n1}
$trade_2$	x_2	rf_{12}	rf_{22}	$\cdots$	rf_{n2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

(b) An Example of the Relational Features for Each Target Behavior

$$p(X^{(t)}|RF_1, RF_2, \cdots, RF_n)$$

- Estimate *p(RF|X)*

$$p(RF_1|X^{(t)}) \quad p(RF_2|X^{(t)}) \quad \cdots, \; p(RF_n|X^{(t)})$$

- Estimate

CPD $p(X^{(t)}|RF_1, \cdots, RF_n)$

$$\alpha p(X^{(t)}) p(RF_1|X^{(t)}) p(RF_2|X^{(t)}) \cdots p(RF_n|X^{(t)})$$

- CBA problem → CPD problem

$$CBA\ problem \rightarrow SRL\ Modeling \qquad (5)$$

$$f(\theta(\cdot), \eta(\cdot)) \rightarrow the\ CPD\ p(X^{(t)}|RF_1, \cdots, RF_n) \qquad (6)$$

Relational Bayesian Classifiers (RBCs)

The CPD $p(X^{(t)}|RF_1, \cdots, RF_n)$ can be estimated as

$$\alpha p(X^{(t)})p(RF_1|X^{(t)})p(RF_2|X^{(t)}) \cdots p(RF_n|X^{(t)}) \quad (8)$$

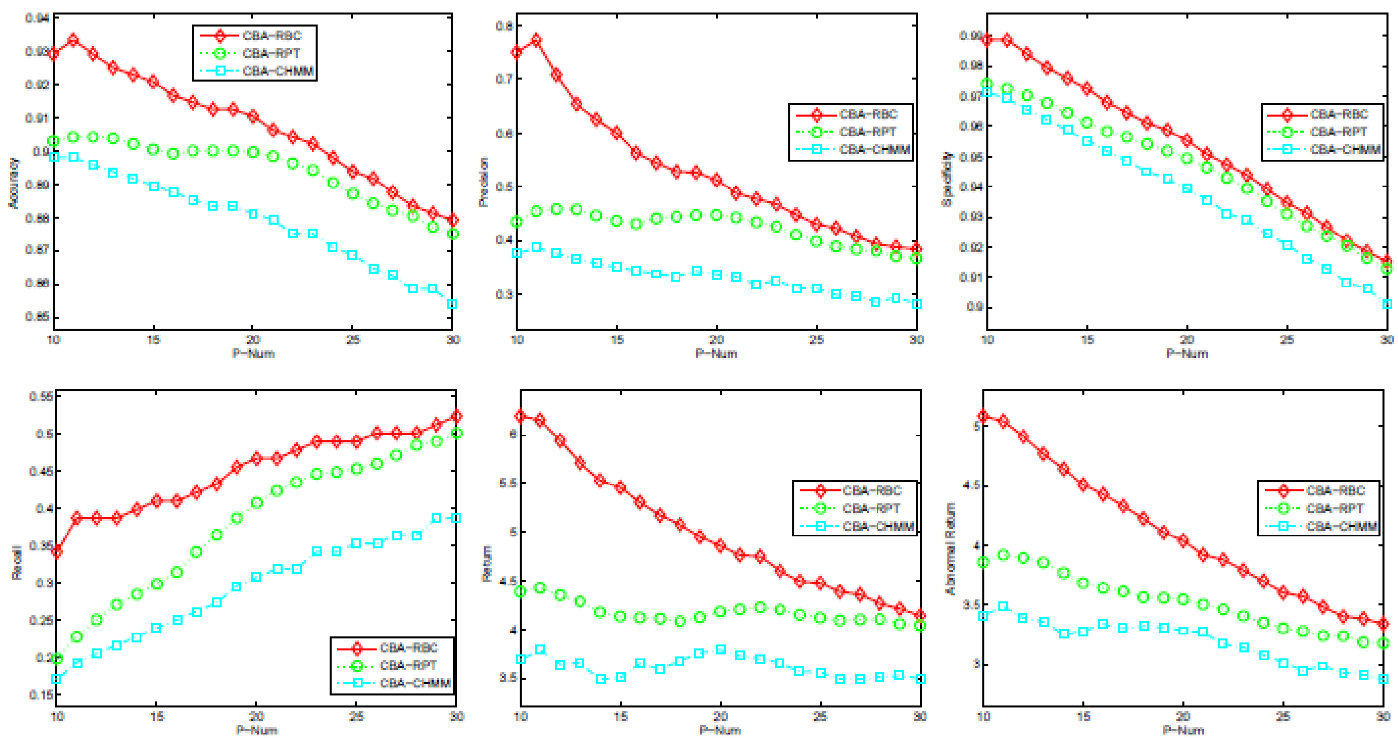
where α is the normalized constant.

- Conditional likelihood:

$$CL(\mathbf{b^k}) = \prod_{\mathbf{b_i^{(t)}} \in \mathbf{b^k}} p(X^{(t)} = x_{b_i^{(t)}} | rf_{1i}, rf_{2i}, \cdots, rf_{ni}; M)$$

Relational Probability Trees (RPTs)

The RPT algorithm uses aggregation functions (e.g, mode, count, proportion and degree) to transform the relational features of subgraphs to propositional features and use these features to construct probability trees.


Accuracy
Precision
Specificity
Recall
Return
Abnormal Return
P−Num
CBA−RBC
CBA−RPT
CBA−CHMM

Probabilistic Modeling of Rating Behaviors for Scalable Recommendation

Trong Dinh Thac Do and Longbing Cao. Metadata-dependent Infinite Poisson Factorization for Efficiently Modelling Sparse and Large Matrices in Recommendation, IJCAI2018.

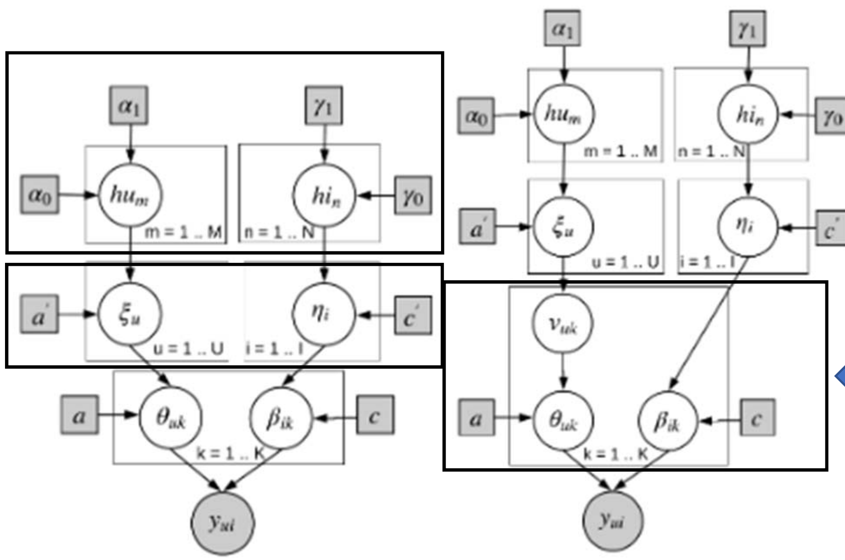
Trong Dinh Thac Do, Longbing Cao. Coupled Poisson Factorization Integrated with User/Item Metadata for Modeling Popular and Sparse Ratings in Scalable Recommendation. AAAI2018.

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Metadata-integrated Poisson Factorization (MPF) Metadata-integrated Infinite Poisson factorization (MIPF)

Enrich prior using user and item metadata

Enrich prior by modeling user behavior and item attractiveness



(a) MPF (b) MIPF

Figure 1: The graphical models of MPF and MIPF.

Using Bayesian Nonparametric techniques to automatically determines the number of latent components

Mpf - finite Metadata-integrated PF

(1) For the m^{th} user attribute in the metadata, sample the weight:

$$hu_m \sim Gamma(\alpha_0, \alpha_1) \tag{1}$$

(2) For the n^{th} item attribute, sample the weight:

$$hi_n \sim Gamma(\gamma_0, \gamma_1) \tag{2}$$

(3) For each user u, sample latent behavior:

$$\xi_u \sim Gamma(a', \prod_{m=1}^{M} hu_m^{f_{u_{u,m}}}) \tag{3}$$

(4) For each item i, sample latent attractiveness:

$$\eta_i \sim Gamma(c', \prod_{n=1}^{N} hi_n^{f_{i_{i,n}}}) \tag{4}$$

(5) For each component k in the PF factorization:
(a) Sample user's latent preference:

$$\theta_{uk} \sim Gamma(a, \xi_u) \tag{5}$$

(b) Sample item's latent feature:

$$\beta_{ik} \sim Gamma(c, \eta_i) \tag{6}$$

(6) Sample rating:

$$y_{ui} \sim Poisson\left(\sum_k \theta_{uk}\beta_{ik}\right) \tag{7}$$

MIPF – Metadata-integrated Infinite PF

(1) For the m^{th} user attribute, sample the weight:

$$hu_m \sim Gamma(\alpha_0, \alpha_1) \tag{8}$$

(2) For the n^{th} item attribute, sample the weight:

$$hi_n \sim Gamma(\gamma_0, \gamma_1) \tag{9}$$

(3) For each user $u(= 1, \ldots, M)$:
(a) Draw the user's latent behavior:

$$\xi_u \sim Gamma(a', \prod_{m=1}^{M} hu_m^{f_{u_{u,m}}}) \tag{10}$$

(b) For $k(= 1..\infty)$, draw stick-breaking proportion:

$$v_{uk} \sim Beta(1, a') \tag{11}$$

(c) For $k(= 1..\infty)$, set the user's latent preference:

$$\theta_{uk} = \xi_u . v_{uk} \prod_{l=1}^{k-1} (1 - v_{ul}) \tag{12}$$

(4) For each item $i(= 1...N)$:
(a) Draw the item's latent attractiveness:

$$\eta_i \sim Gamma(c', \prod_{n=1}^{N} hi_n^{f_{i_{i,n}}}) \tag{13}$$

(b) For $k = (1...\infty)$, set the item's latent feature:

$$\beta_{ik} \sim Gamma(c, \eta_i) \tag{14}$$

(5) For $u(= 1...M)$ and $i(= 1...N)$, draw

$$y_{ui} \sim Poisson\left(\sum_{k=1}^{\infty} \theta_{uk}\beta_{ik}\right) \tag{15}$$

Variational inference of MPF/MIPF

Algorithm 1 Variational Inference for MPF

Initialize the variational parameters $\{\zeta, \rho, \nu, \mu, \kappa, \tau, \phi\}$.
Set the number of components K.
Sample *shape* of user's latent behavior, and *shape* of item's latent attractiveness, as in Eqs. (22) and (24).
Sample *shape* of the weight of user's attribute (in metadata), and *shape* of the weight of item's attribute (in metadata), as in Eqs. (18) and (20).
repeat
    for each rating of user u to item i that $y_{ui} \neq 0$ **do**
        Update the multinominal as in Eq. (26).
    end for
    for each user **do**
        Update the latent preference as in Eqs. (27) and (28)
        Update *rate* of latent behavior as in Eq. (23).
        for each user attribute in metadata **do**
            Update *rate* of the weight as in Eq. (19)
        end for
    end for
    for each item **do**
        Update the latent feature as in Eqs. (29) and (30).
        Update *rate* of latent attractiveness as in Eq. (25).
        for each item attribute **do**
            Update *rate* of the weight as in Eq. (21).
        end for
    end for
until convergence

Properties

- MPF/MIPF improve precision when working with large and sparse data by integrating user/item metadata.

- MIPF efficiently estimates the number of latent components.

- The variational inference for MPF and MIPF applies to massive data.

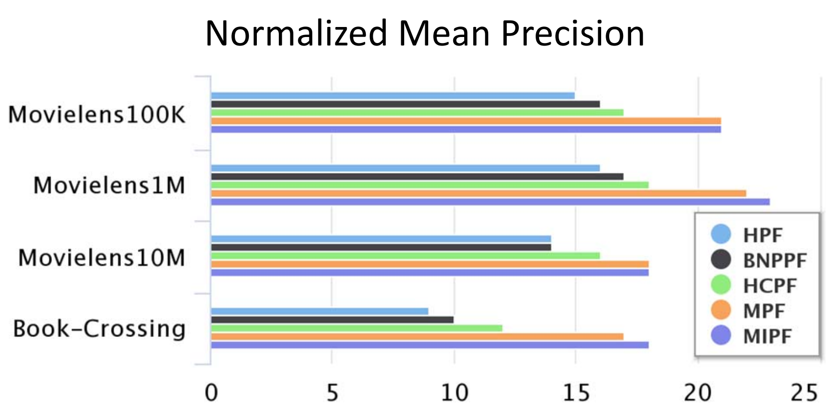
Results

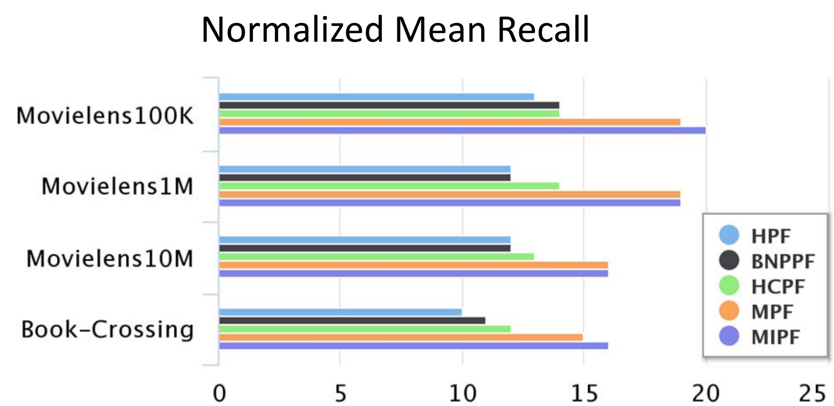
- Datasets:
  - (1) Movielens100K, Movielens1M and Movielens10M [Harper and Konstan, 2016].
  - (2) Book-Crossing [Ziegler et al., 2005].

- Baseline methods:
  - HPF [Gopalan et al., 2015] as it outperforms many baselines in MF including NMP, LDA and PMF.
  - Bayesian Nonparametric PF (BNPPF) [Gopalan et al., 2014a].
  - The latest PF: Hierarchical Compound PF (HCPF) [Basbug and Engelhardt, 2016].

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

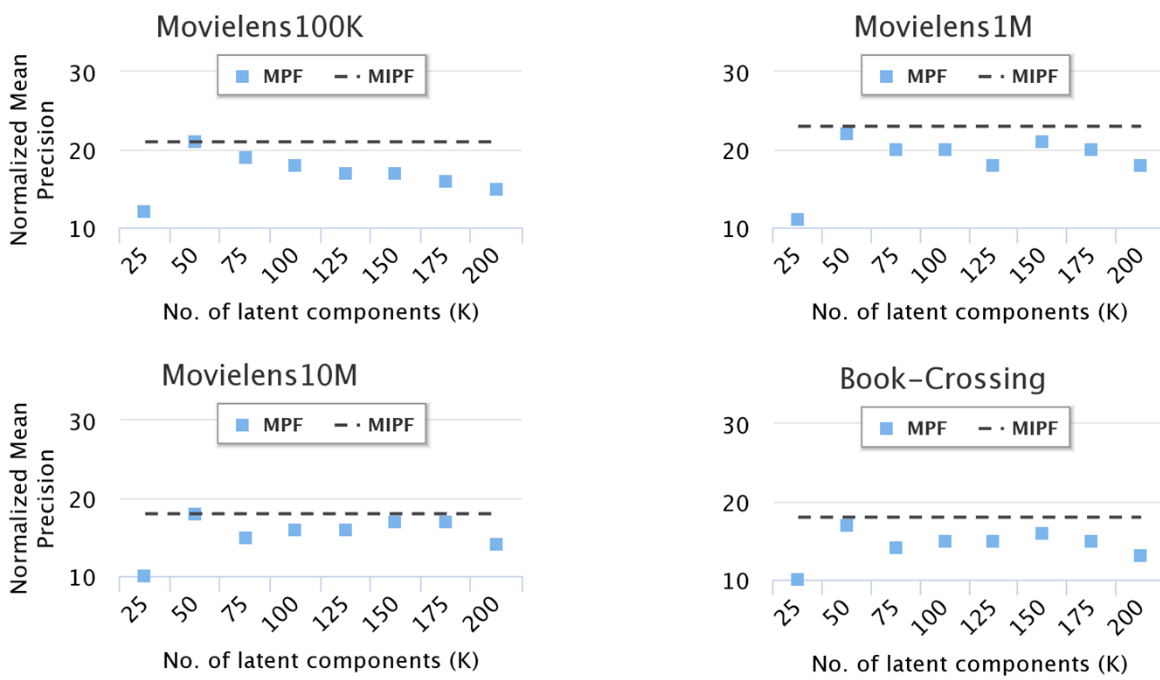
Results - How do MPF/MIPF significantly outperform other PF models?





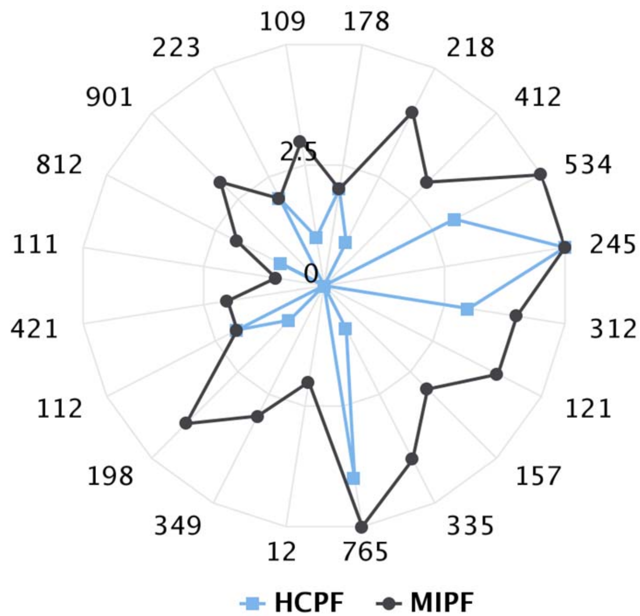
Top-20 Recommendation Compared with baselines

Results - How does MIPF effectively estimate the number of unbounded latent components?



Performance of top-30 recommendations made by finite model MPF and infinite model MIPF.

Results - How do MPF/MIPF deal with sparse items/users?



Example of MIPF in handling sparse items in comparison with HCPF.

Understanding Behavior Driving Forces: Intent, Choice, Attraction

Modeling User Choices

Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., and Wang, J. Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. ACM Trans. Inf. Syst., 2017.

Hu, L., Cao, W., Cao, J., Xu, G., Cao, L., & Gu, Z. (2014). Bayesian Heteroskedastic Choice Modeling on Non-identically Distributed Linkages. In R. Kuamr (Ed.), *Proceedings of the 2014 IEEE International Conference on Data Mining* (pp. 851-856)

Multi-objective Recommender Systems

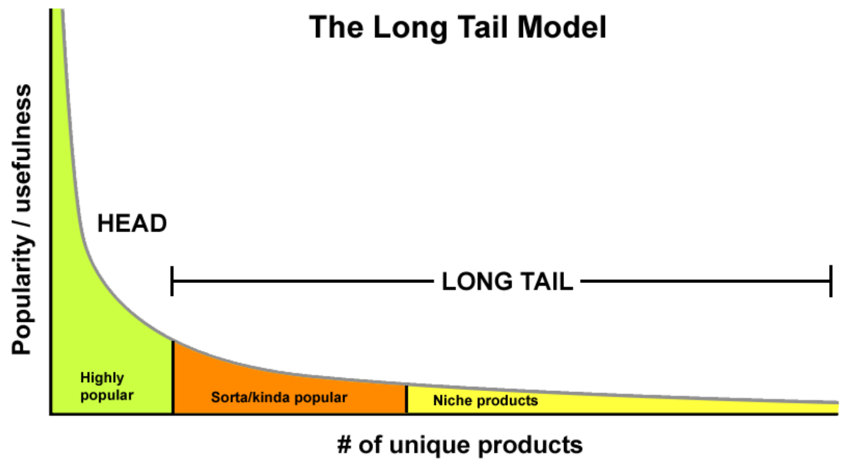
- Traditional RSs are built on single objective

- However, users' choices are determined by multiple aspects
  - Diversity

- To learn users' profile more comprehensively, we need to build new RSs to optimize multiple objectives for each aspect

Problems for Long-tail Users/Items

- Popularity Bias
  - Short-head users and items account for the majority of data, and models tend to fit these users and items. It overlooks the preference of users and items in the tail.
  - Specialty modeling is desirable
- Shilling Attack
  - Short-head items are well known by users. However, long-tail items have few data and they are more vulnerable to shilling attack.
  - Credibility modeling is desirable



RMRM : Joint Optimizing Credibility and Specialty

- Recurrent Mutual Regularization Model (RMRM) consists of two main components
  - C-HMF models user choices by emphasizing credibility
  - S-HMF models user choices by emphasizing specialty

- Each component leads to a different objective for optimization, so RMRM is a multi-objective recommenders systems

Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., and Wang, J. Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. ACM Trans. Inf. Syst., 2017

Classic Probabilistic MF & Heteroscedastic MF

- $P(\boldsymbol{U}_i) = N(\boldsymbol{U}_i | \boldsymbol{0}, \sigma_U^2 \boldsymbol{I})$
- $P(\boldsymbol{V}_j) = N(\boldsymbol{V}_j | \boldsymbol{0}, \sigma_V^2 \boldsymbol{I})$
- $P(Y_{ij} | \boldsymbol{U}_i, \boldsymbol{V}_j) = N(Y_{ij} | \boldsymbol{U}_i^{\mathrm{T}} \boldsymbol{V}_j, \sigma^2)$

- $P(\boldsymbol{U}_i) = N(\boldsymbol{U}_i | \boldsymbol{\mu}_U, \sigma_U^2 \boldsymbol{I})$
- $P(\boldsymbol{V}_j) = N(\boldsymbol{V}_j | \boldsymbol{\mu}_V, \sigma_V^2 \boldsymbol{I})$
- $P(Y_{ij} | \boldsymbol{U}_i, \boldsymbol{V}_j) = N(Y_{ij} | \boldsymbol{U}_i^{\mathrm{T}} \boldsymbol{V}_j, \sigma_{ij}^2)$

$$P(\boldsymbol{U}, \boldsymbol{V} | \boldsymbol{Y}) \propto P(\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V}) = \prod_{ij \in \boldsymbol{O}} P(Y_{ij} | \boldsymbol{U}_i, \boldsymbol{V}_j) \prod_i P(\boldsymbol{U}_i) \prod_j P(\boldsymbol{V}_j)$$

- *Loss function:*
  - $-\log P(Y_{ij}, \boldsymbol{U}_i, \boldsymbol{V}_j) = \underset{\boldsymbol{U},\boldsymbol{V}}{argmin} \left[\boxed{\underbrace{\sum_{ij} (Y_{ij} - \boldsymbol{U}_i^{\mathrm{T}} \boldsymbol{V}_j)^2}_{weighted\ loss}} + \underbrace{\lambda_U \sum_i \|\boldsymbol{U}_i\|^2 + \lambda_V \sum_j \|\boldsymbol{V}_j\|^2}_{regularization} \right]$

- *Loss function:*
  - $-\log P(Y_{ij}, \boldsymbol{U}_i, \boldsymbol{V}_j) =$
  - $\underset{\boldsymbol{U},\boldsymbol{V}}{argmin} \left[\underbrace{\sum_{ij} w_{ij} (Y_{ij} - \boldsymbol{U}_i^{\mathrm{T}} \boldsymbol{V}_j)^2}_{weighted\ loss} + \underbrace{\lambda_U \sum_i \|\boldsymbol{U}_i - \boldsymbol{\mu}_i\|^2 + \lambda_V \sum_j \|\boldsymbol{V}_j - \boldsymbol{\mu}_j\|^2}_{regularization} \right]$

Popularity Bias

Shilling Attack

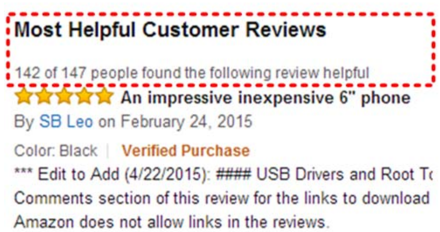
- model variance, i.e. weight on the loss : $w_{ij} = f(\sigma_{ij}^{-2})$

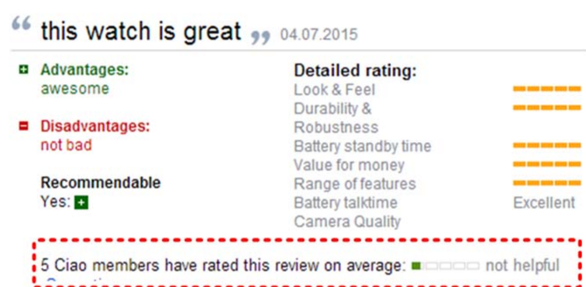
Credibility Enhancement

Shilling Attack

- C-HMF (Credibility-specific Heteroscedastic MF)
  - $\sigma_{ij}^2 = f^C(Y_{ij}) \propto \varphi_i^{-1}$ scores the *credibility* of each review
- Bayesian Reputation Modeling
  - *Reputation Score: Given the helpfulness scores h_i of a user i, the reputation score on this user is defined by:*

$$\varphi_i = \mathcal{R}(\boldsymbol{e}_i|\boldsymbol{h}_i) \overset{\text{def}}{=} \frac{r+\alpha}{r+s+\alpha+\beta}$$





Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., and Wang, J. Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. ACM Trans. Inf. Syst., 2017

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Specialty Enhancement

Popularity Bias

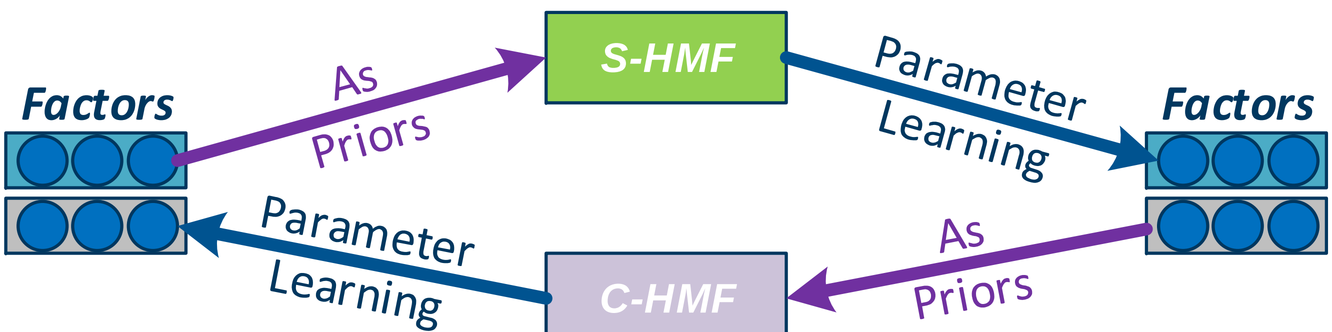
- S-HMF (Specialty-specific Heteroscedastic MF)
  - $\sigma_{ij}^2 = f^S(Y_{ij}) \propto {\psi_j}^{-1}$ scores the *specialty* of user choice, which tightly fits the choices over long-tail items

- Given all observed choices, the ***specialty score*** of a choice on an item j is measured by the *self-information*:
  - $\psi_j = -\,log\, \bar{p}(j|\alpha)$

Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., and Wang, J. Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. ACM Trans. Inf. Syst., 2017

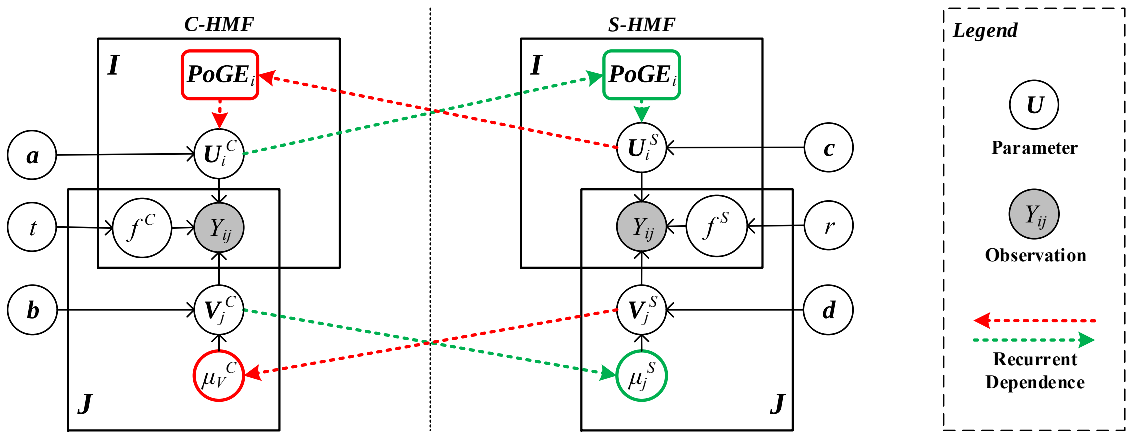
Recurrent Mutual Regularization

- A **recurrent mutual regularization process** couples S-HMF and C-HMF using the user and items factors learned from each other as the **empirical priors**



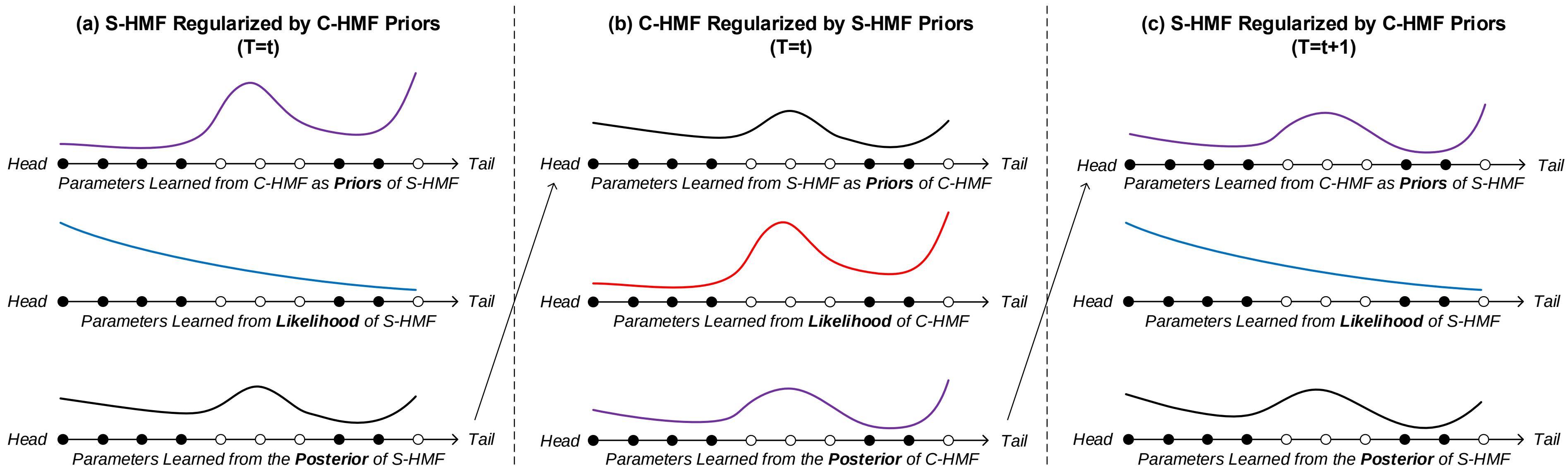
Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., and Wang, J. Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. ACM Trans. Inf. Syst., 2017

Graphical model of RMRM framework



Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., and Wang, J. Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. ACM Trans. Inf. Syst., 2017

Demonstration of the recurrent mutual Regularization process



$$\underset{\boldsymbol{U}^C,\boldsymbol{V}^C}{argmin}\left[\underbrace{\sum_{ij} w_{ij}\left(Y_{ij}-\boldsymbol{U}_i^{C\,\mathrm{T}}\boldsymbol{V}_j^C\right)^2}_{weighted\ loss}+\underbrace{\lambda_U\sum_i\left\|\boldsymbol{U}_i^C-\boldsymbol{U}_i^S\right\|^2+\lambda_V\sum_j\left\|\boldsymbol{V}_j^C-\boldsymbol{V}_j^S\right\|^2}_{regularization}\right]$$

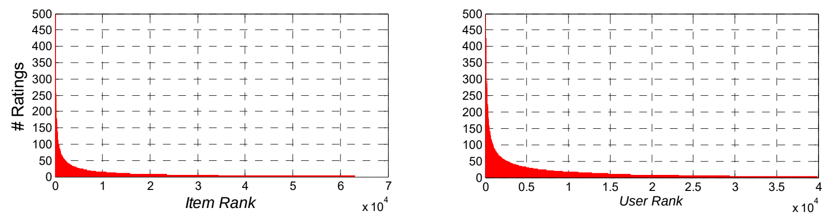
$$\underset{\boldsymbol{U}^S,\boldsymbol{V}^S}{argmin}\left[\underbrace{\sum_{ij} w_{ij}\left(Y_{ij}-\boldsymbol{U}_i^{S\,\mathrm{T}}\boldsymbol{V}_j^S\right)^2}_{weighted\ loss}+\underbrace{\lambda_U\sum_i\left\|\boldsymbol{U}_i^S-\boldsymbol{U}_i^C\right\|^2+\lambda_V\sum_j\left\|\boldsymbol{V}_j^S-\boldsymbol{V}_j^C\right\|^2}_{regularization}\right]$$

Mutual Regularization

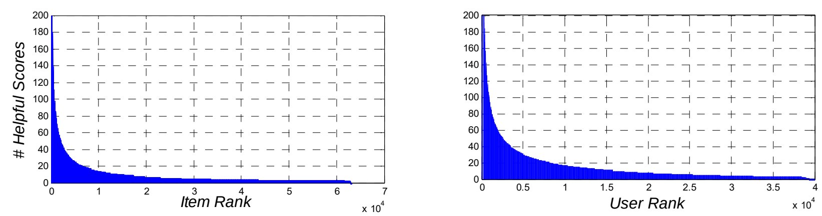
Hu, L., Cao, L., Cao, J., Gu, Z., Xu, G., and Wang, J. Improving the Quality of Recommendations for Users and Items in the Tail of Distribution. ACM Trans. Inf. Syst., 2017

Dataset: the Epinions

# users: 39,902	# items: 63,027
# trust links: 43,8965	# trusters / users: 11
max # of trusters: 1,713	# users with zero truster: 14,202
# ratings: 734,441	density: 0.029%
# ratings / users: 18	# ratings / items: 11
max # ratings of user: 1,809	max # ratings of item: 2,112



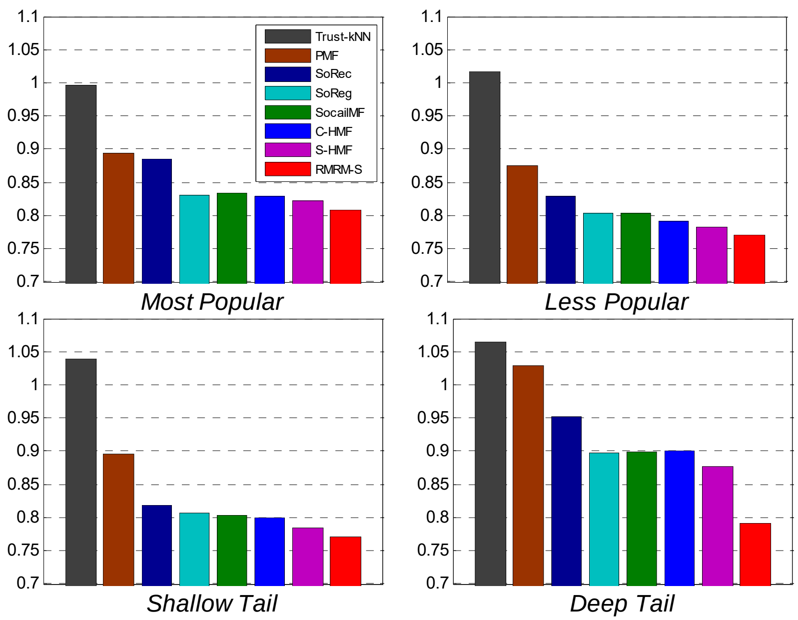
Long-tail distributions for the number of ratings of items and users (truncated from 0 to 500)



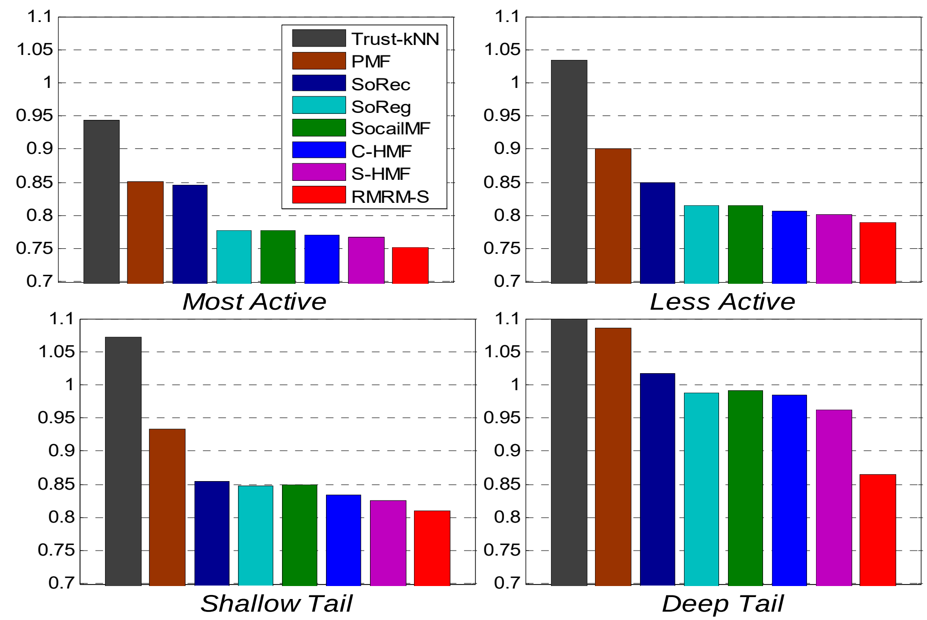
The distributions for the number of helpful scores w.r.t. items and users (truncated from 0 to 200)

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Rating Prediction on Long-tail Distributed Items and Users



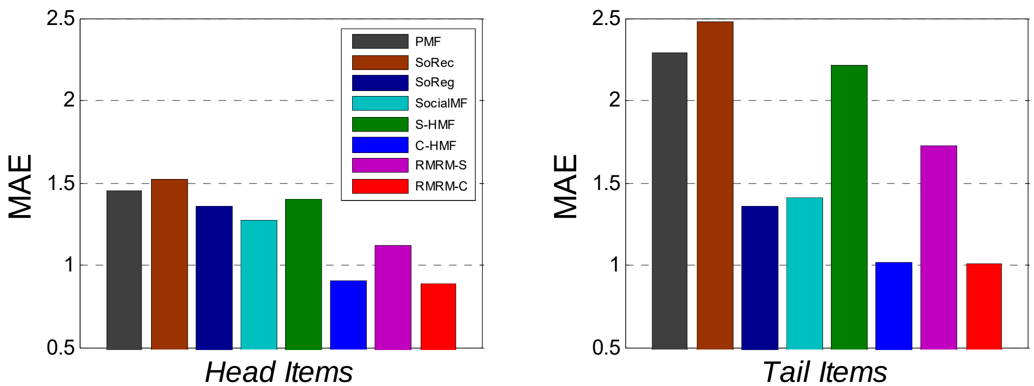
MAEs of rating prediction for the long-tail item distribution



MAEs of rating prediction for the long-tail user distribution

Shilling Attack Simulation

- To simulate such an environment
  - We created 1,000 virtual spam users to conduct the attack
  - We selected 100 items from the head (0%~20%) and the tail (20%~100%) as the attack targets.
- Nuke attack in the case of the average attack model



Outline

- 8 Statistical Modeling of Coupled Behaviors
- 9 Probabilistic Modeling of Sparse Rating Behaviors
- 10 Understanding Behavior Drivers: Choice and Attraction
- 11 Behavior Analysis with Recurrent Networks
- 12 Behavior Analysis in Visual Data
- 13 Behavior Learning from Demonstrations
- 14 Challenges and Prospects

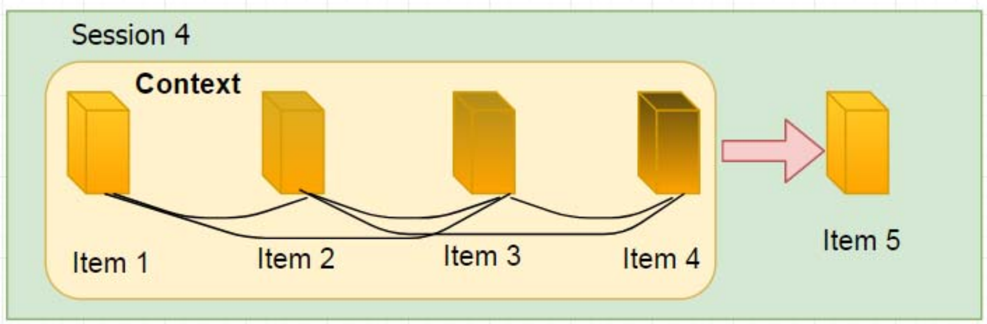
Representation of Next-item/choice in Recommendation within Session and Context

Shoujin Wang, Liang Hu, Longbing Cao, Xiaoshui Huang, Defu Lian and Wei Liu. Attention-based Transactional Context Embedding for Next-Item Recommendation. AAAI2018.

S. Wang, L. Hu, L. Cao, X. Huang. Perceiving the Next Choice with Comprehensive Transaction Embeddings for Online Recommendation, *ECML/PKDD2017*

Session context

- Session context consists of observed sequence that leads to the consequent actions.
  - e.g., clicked pages in browsing history, or chosen items in a transaction.



BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

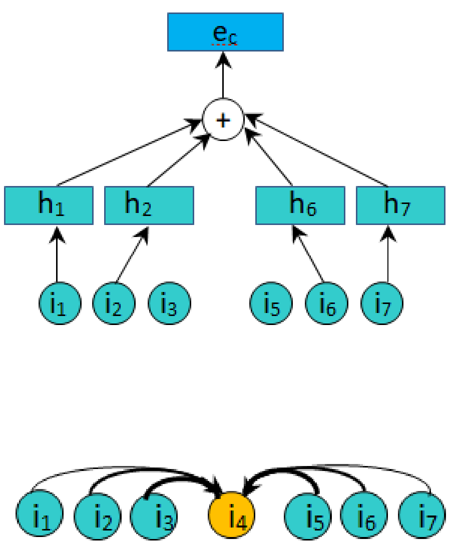
Session-based Recommender Systems Overview

- Session-based Recommender Systems
  - First-order dependency modeling
    - Markov chain models
    - Matrix factorization models
  - Higher-order dependency modeling
    - RNN-based model for session Modeling
    - Encoder-Decoder for Session Modeling
  - Loosely ordered dependency modeling
    - ATEM model
    - NTEM model

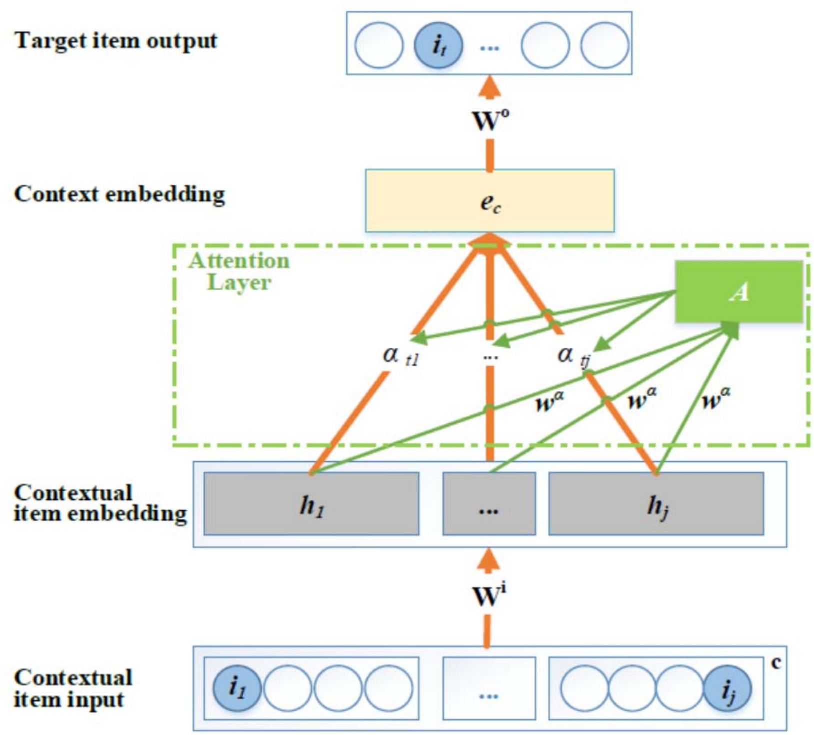
ATEM:

Weight transaction embedding with attention mechanism

Context items **contribute differently** to the next choice



ATEM



Wang, S., Hu, L., & Cao, L. *Attention-based Transactional Context Embeddings for Next-Item Recommendation*. AAAI2018

Experiments

- ATEM achieves best performance compared to baselines.
- Attention mechanism contributes greatly by comparing ATEM and TEM, a simplified model without attention mechanism.

Table 2: Accuracy comparisons on IJCAI-15

Model	REC@10	REC@50	MRR
PBRS	0.0780	0.0998	0.0245
FPMC	0.0211	0.0602	0.0232
PRME	0.0555	0.0612	0.0405
GRU4Rec	0.2283	0.3021	0.1586
ATEM	**0.3542**	**0.5134**	**0.2041**
TEM	0.3177	0.3796	0.1918

Table 3: Accuracy comparisons on Tafang

Model	REC@10	REC@50	MRR
PBRS	0.0307	0.0307	0.0133
FPMC	0.0191	0.0263	0.0190
PRME	0.0212	0.0305	0.0102
GRU4Rec	0.0628	0.0907	0.0271
ATEM	**0.1089**	**0.2016**	**0.0347**
TEM	0.0789	0.1716	0.0231

Experiments

- Test the robustness to the item order within a session
- ATEM is almost not affected when randomly disordering items.

Table 2: Accuracy comparisons on IJCAI-15

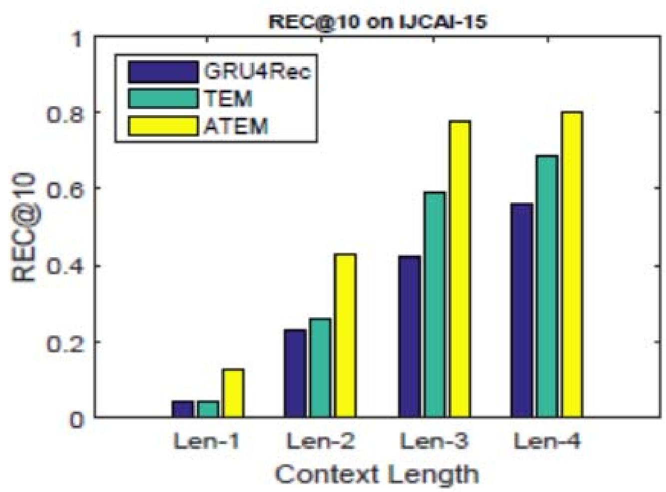
Model	**REC@10**	**REC@50**	**MRR**
PBRS	0.0780	0.0998	0.0245
FPMC	0.0211	0.0602	0.0232
PRME	0.0555	0.0612	0.0405
GRU4Rec	0.2283	0.3021	0.1586
ATEM	**0.3542**	**0.5134**	**0.2041**
TEM	0.3177	0.3796	0.1918

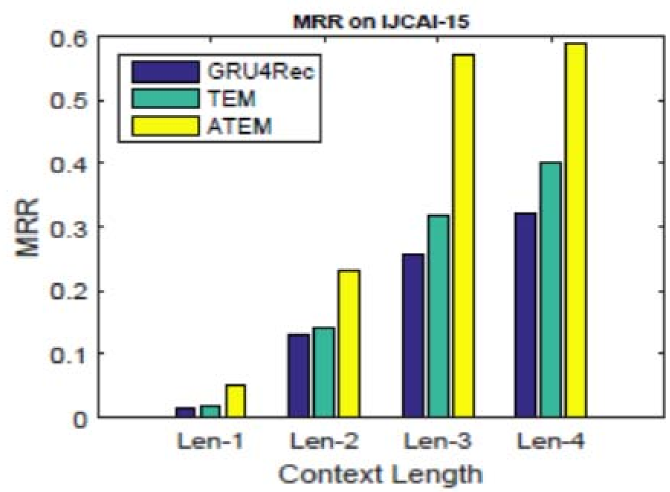
Table 4: Accuracy on disordered IJCAI-15

Model	**REC@10**	**REC@50**	**MRR**
PBRS	0.0500	0.0559	0.0185
FPMC	0.0151	0.0412	0.0183
PRME	0.0346	0.0389	0.0351
GRU4Rec	0.1636	0.2121	0.1022
ATEM	**0.3423**	**0.4981**	**0.1960**
TEM	0.2660	0.3012	0.1431

Experiments

- Test the effect of context length on recommendation performance
  - ATEM outperforms other methods on longer context, which proves attention mechanism effectively choose the most related items in context.





Attraction Learning

Liang Hu, Songlei Jian, **Longbing Cao**, Qingkui Chen. Interpretable Recommendation via Attraction Modeling: Learning Multilevel Attractiveness over Multimodal Movie Contents, IJCAI2018

Why modeling attraction?

- First, the attraction is the ***highlights*** that largely lead to a person's behaviour, selection and decision.
- For example,
  - An action may be taken due to the attraction to something interesting.
  - We often cannot recite a whole poem but we can always recall some impressive sentences;
  - We may not remember a whole song but we can hum some touching lyrics.
  - These highlights make a person to be attracted by the poem or the song.

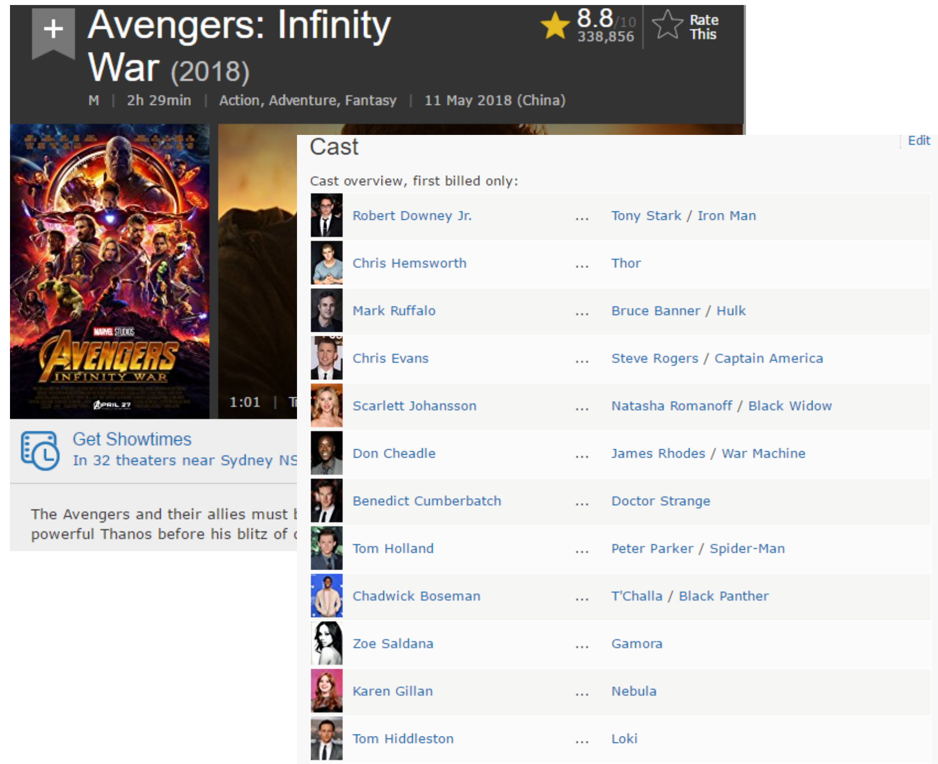
Why modeling attraction?

- Second, the attraction is a ***subjective feeling*** which is often different from person to person.
- For example,
  - Readers in Go community may be attracted by the target problem, i.e., Go playing, of this scientific paper while readers in AI community may be attracted by the technical methods.

Example: Attraction on Movies

- The internet movie has accounted for the major traffic in new media age.

- In particular, the ***story*** and the ***cast members***, e.g., actors, directors and writers, are two most important aspects of a movie to attract audience.
  - A person may be caught by some attractive words by the story of a movie. Only a few sentences of the core plot instead of all sentences actually attract a user.
  - Cast members of a movie are another very important factor to attract users.



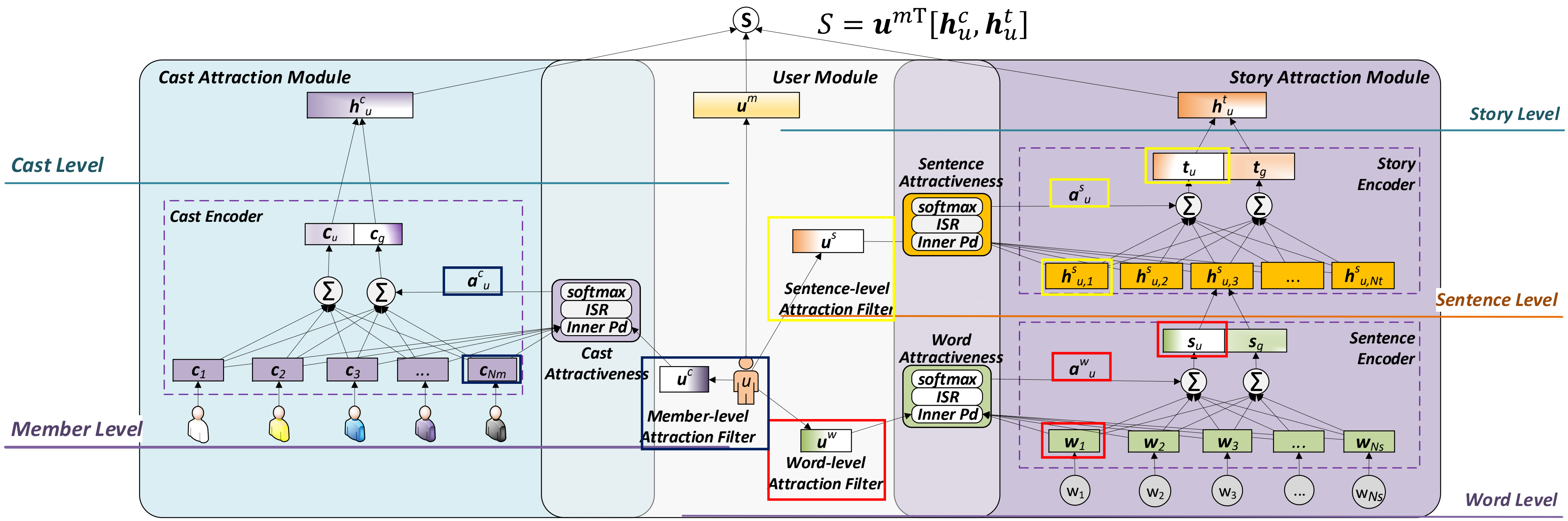
Hu, L., Jian, S., Cao, L., Chen. Q. Interpretable Recommendation via Attraction Modeling: Learning Multilevel Attractiveness over Multimodal Movie Contents. IJCAI2018

Multimodal and Multilevel Attraction Model

- One multilevel neural model on the movie story to capture
  - Word-level attraction: e.g. some character, some place
  - Sentence-level attraction: e.g. some core plot
  - Story-level attraction: e.g. like the movie to what extent

- The other multilevel neural model on the cast to capture
  - Member-level attraction: e.g. a fan of some actor
  - Cast-level attraction: e.g. attracted by the movie to what extent

Model Architecture



$$a_u^{c_i} = softmax\left(isr\left(\boldsymbol{u}^{c\mathrm{T}}\boldsymbol{c}_i\right)\right) \qquad \boldsymbol{c}_u = \sum a_u^{c_i}\boldsymbol{c}_i$$

$$a_u^{w_i} = softmax\left(isr\left(\boldsymbol{u}^{w\mathrm{T}}\boldsymbol{w}_i\right)\right) \qquad \boldsymbol{s}_u = \sum a_u^{w_i}\boldsymbol{w}_i$$

$$a_u^{s_i} = softmax\left(isr\left(\boldsymbol{u}^{s\mathrm{T}}\boldsymbol{h}_i^s\right)\right) \qquad \boldsymbol{t}_u = \sum a_u^{s_i}\boldsymbol{h}_i^s$$

Experiments

- The experiments are conducted on the real-world movie watch dataset MovieLens 1M. The model is evaluated from three aspects:
  - Recommendation accuracy
  - New movie recommendation
  - Interpretation of attraction on movies

Datasets

- We collect user watch records from the MovieLens 1M dataset.
  - https://grouplens.org/datasets/movielens/1m/
- Story and cast data are provided the mapping from MovieLens ID to DBPedia URI
  - https://github.com/sisinflab/LODrecsys-datasets/tree/master/Movielens1M

Augment information from DBPedia

- SPARQL Interface

```
PREFIX
movie:http://dbpedia.org/resource/Screwed_(2000_film)
select ?abstract ?director ?writer ?starring
 { movie: dbo:abstract ?abstract.
  optional { movie: dbo:director ?director }
  optional { movie: dbo:writer ?writer }
  optional { movie: dbo:starring ?starring }
 FILTER (langMatches(lang(?abstract),"en")) }
```

DBpedia Browse using Formats Faceted Browser Sparql Endpoi

About: Screwed (2000 film)

An Entity of Type : movie, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

Screwed is a 2000 American comedy film, written and directed by Scott Alexander and Larry Karaszewski. It stars Norm Macdonald, Dave Chappelle, Danny DeVito, Elaine Stritch, Daniel Benzali, Sarah Silverman, and Sherman Hemsley. The film was released by Universal Studios.

Property	Value
dbo:Work/runtime	81.0
dbo:abstract	Screwed is a 2000 American comedy film, written and directed by Scott Alexander and Larry Karaszewski. It stars Norm Macdonald, Dave Chappelle, Danny DeVito, Elaine Stritch, Daniel Benzali, Sarah Silverman, and Sherman Hemsley. The film was released by Universal Studios. (en)
dbo:director	dbr:Scott_Alexander_and_Larry_Karaszewski
dbo:distributor	dbr:Universal_Studios
dbo:imdbId	0156323
dbo:musicComposer	dbr:Michel_Colombier
dbo:producer	dbr:Robert_Simonds
dbo:releaseDate	2000-05-12 (xsd:date)
dbo:runtime	4860.000000 (xsd:double)
dbo:starring	dbr:Sarah_Silverman dbr:Danny_DeVito dbr:Norm_Macdonald dbr:Elaine_Stritch dbr:Sherman_Hemsley dbr:Daniel_Benzali

Statistics of the Enriched Dataset

# movies:	3,900	# users:	6,040
# watch record:	1,000,209	# cast:	9,398
movie story vocabulary	22,582	# sentences per story	10.2
# cast members per movie	6.44	# plays per cast	2.10

Table 1: Statistics of content-enriched MovieLens dataset

Training and Testing Sets

- **Released movie recommendation**: we randomly held out 20% user watch records as the testing set, and the remainder were served as the training set.
- **New movie recommendation:** we randomly selected 10% movies and held out all their watch records from the dataset, and the remainder of 90% movies and their watch records were used for training.
- For each hold-out test sample in above two testing sets, we randomly draw ten noisy samples to test whether the testing methods can rank the true sample at a top position out of noisy samples.

Comparison Methods

- ***CENTROID***: We create user profiles using the centroid of all word embedding vectors from the users' movie stories. Then, we rank recommendations by the similarity between the user profile and the controid of word embedding vectors of movie story.
- ***CTR***: Collaborative topic regression performs user regression over the latent topic distribution of movie stories learned from LDA.
- ***CWER***: Similar to CTR, we create the collaborative word embedding user regression (CWER) to perform regression over the centroid word embedding vector of each movie story initialized by GloVe embeddings.
- ***MLAM***: This is the full multilevel attraction model proposed in this paper.
- ***MLAM-S***: This is the single-modal version MLAM that only has the story attraction module.
- ***MLAM-C***: This is the single-modal version MLAM that only has the cast attraction module.

Ranking Performance

- Recommendation accuracy on released movies and new movies

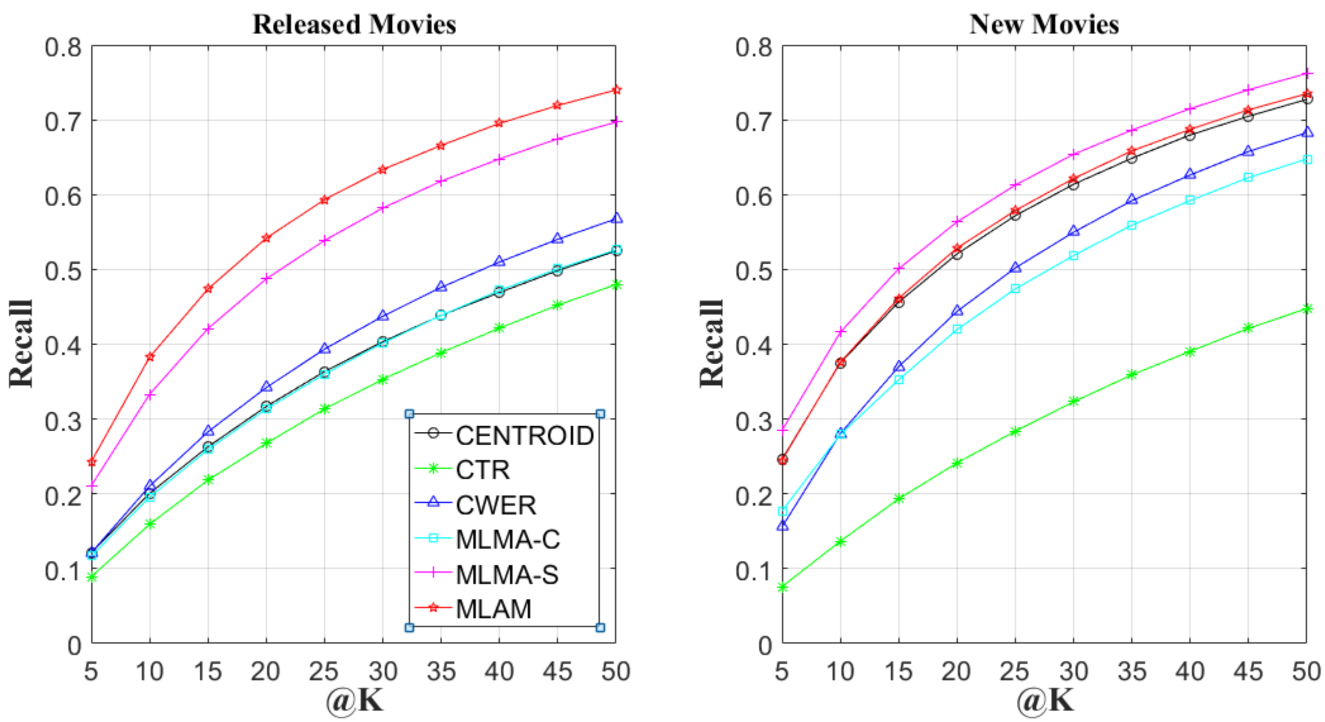
Method	MAP@5	MAP@20	MRR@5	MRR@20
CENTROID	0.1738	0.1481	0.0763	0.0958
CTR	0.1226	0.1069	0.0514	0.0692
CWER	0.1666	0.1580	0.0798	0.1089
MLAM-C	**0.4243**	**0.3963**	**0.2118**	**0.2398**
MLAM-S	0.3816	0.3451	0.1822	0.2093
MLAM	**0.4252**	**0.3997**	**0.2187**	**0.2464**

Table 2: Ranking performance on released movies (80% training)

Method	MAP@5	MAP@20	MRR@5	MRR@20
CENTROID	0.2381	0.2409	0.1623	0.1900
CTR	0.1056	0.1374	0.0798	0.1089
CWER	0.1971	0.2346	0.1461	0.1801
MLAM-C	0.1817	0.1664	0.1132	0.1370
MLAM-S	**0.3001**	**0.3059**	**0.2091**	**0.2371**
MLAM	**0.2573**	**0.2671**	**0.1794**	**0.2090**

Table 3: Ranking performance on new movies (90% training)

Recall on Release Movies and New Movies



Visualization and Interpretation

User 156	*Sentence level attractiveness*	Election is a 1999 American comedy-drama film directed and written by Alexander Payne and adapted by him and Jim Taylor from Tom Perrotta's 1998 novel of the same title. The plot revolves around a high school election and satirizes both suburban high school life and politics. The film stars Matthew Broderick as Jim McAllister, a popular high school social studies teacher in suburban Omaha, Nebraska, and Reese Witherspoon as Tracy Flick, around the time of the school's student body election. When Tracy qualifies to run for class president, McAllister believes she does not deserve the title and tries his best to stop her from winning. Election opened to acclaim from critics, who praised its writing and direction. The film received an Academy Award nomination for Best Adapted Screenplay, a Golden Globe nomination for Witherspoon in the Best Actress category, and the Independent Spirit Award for Best Film in 1999.
	Word level attractiveness	Election is a 1999 American comedy-drama film directed and written by Alexander Payne and adapted by him and Jim Taylor from Tom Perrotta's 1998 novel of the same title.
	Cast member attractiveness	Alexander Payne, Reese Witherspoon, Matthew Broderick, Jim Taylor
User 2163	*Sentence level attractiveness*	Election is a 1999 American comedy-drama film directed and written by Alexander Payne and adapted by him and Jim Taylor from Tom Perrotta's 1998 novel of the same title. The plot revolves around a high school election and satirizes both suburban high school life and politics. The film stars Matthew Broderick as Jim McAllister, a popular high school social studies teacher in suburban Omaha, Nebraska, and Reese Witherspoon as Tracy Flick, around the time of the school's student body election. When Tracy qualifies to run for class president, McAllister believes she does not deserve the title and tries his best to stop her from winning. Election opened to acclaim from critics, who praised its writing and direction. The film received an Academy Award nomination for Best Adapted Screenplay, a Golden Globe nomination for Witherspoon in the Best Actress category, and the Independent Spirit Award for Best Film in 1999.
	Word level attractiveness	The film received an Academy Award nomination for Best Adapted Screenplay, a Golden Globe nomination for Witherspoon in the Best Actress category, and the Independent Spirit Award for Best Film in 1999
	Cast member attractiveness	Alexander Payne, Reese Witherspoon, Matthew Broderick, Jim Taylor

Statistical attractiveness on movie ***Election (1999)*** w.r.t. sentences, words in the most attractive sentences and cast members. The larger size and deeper color of font denote the larger attractiveness weight is assigned.

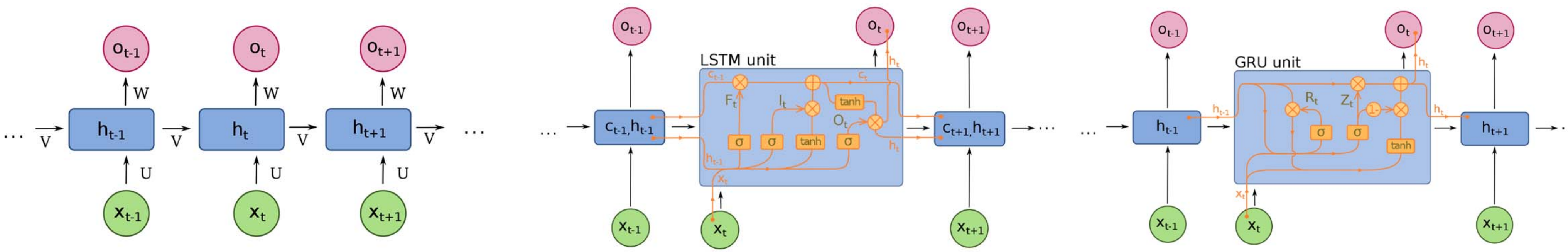
Open issues and directions

- More advanced approaches involving Psychology, Neuroscience, Brain science, are demanded to precisely model attraction.
- Attraction modeling on more data types as well as text, e.g., behaviors, images, videos, audios.
- Attraction is quite subjective, which changes with context
  - Incorporating contextual information for modeling context-aware attraction is more preferable

Behavior Analysis with Recurrent Networks

Recurrent neural networks

Basic RNNs **LSTM** **GRU**



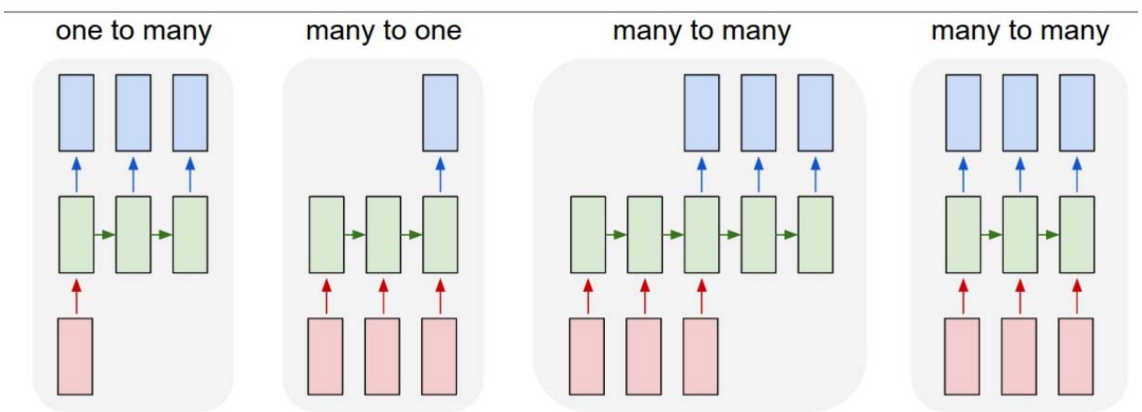
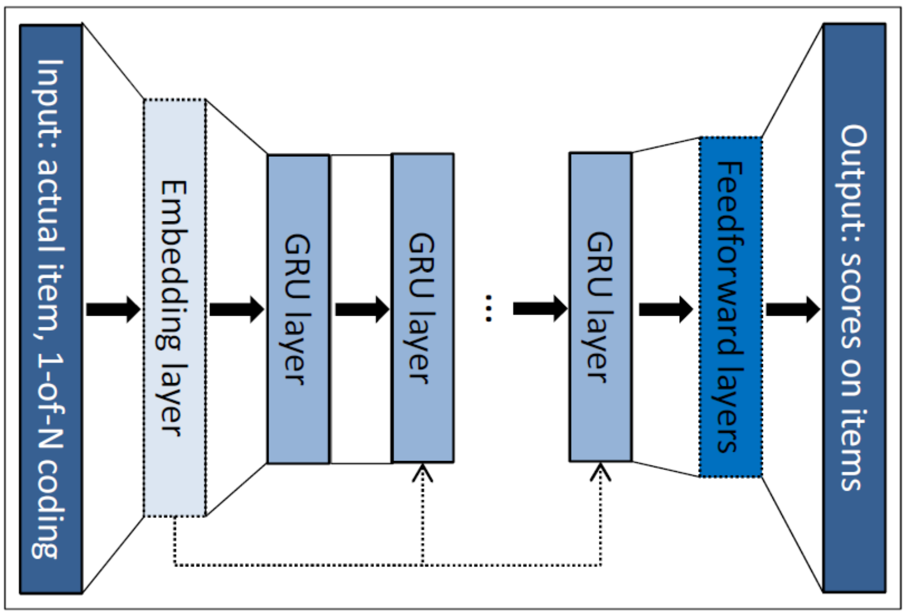


Image sources:
Wikipedia; http://karpathy.github.io

User click sequence – Session-based recommendation

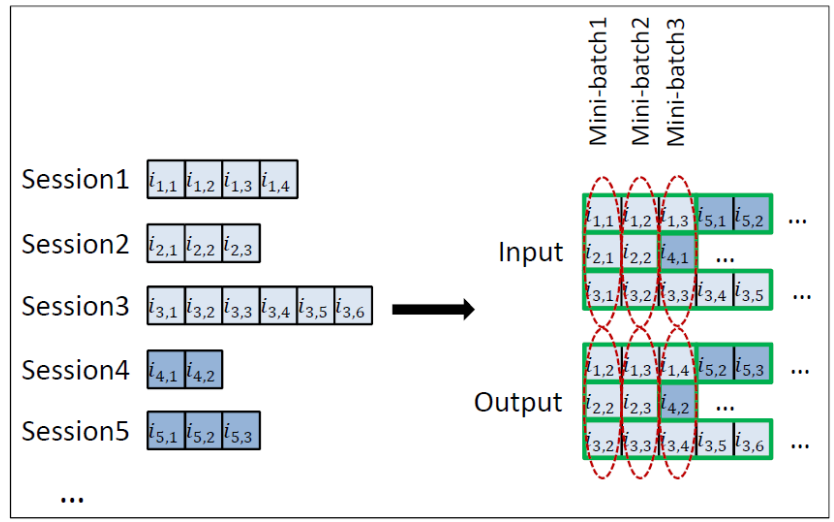
- The overall framework



Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. Session-based recommendations with recurrent neural networks. In *ICLR.* 2016.

- Adapting RNNs to RS

Session-parallel mini-batch creation

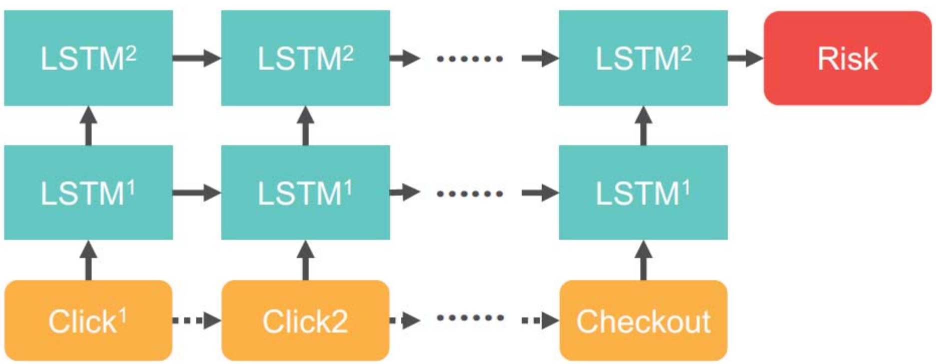


Pairwise ranking loss

$-\frac{1}{N_S} \cdot \sum_{j=1}^{N_S} \log\left(\sigma\left(\hat{r}_{s,i} - \hat{r}_{s,j}\right)\right)$ **or** $\frac{1}{N_S} \cdot \sum_{j=1}^{N_S} I\{\hat{r}_{s,j} > \hat{r}_{s,i}\}$

User click sequence - Fraud detection

- The overall framework



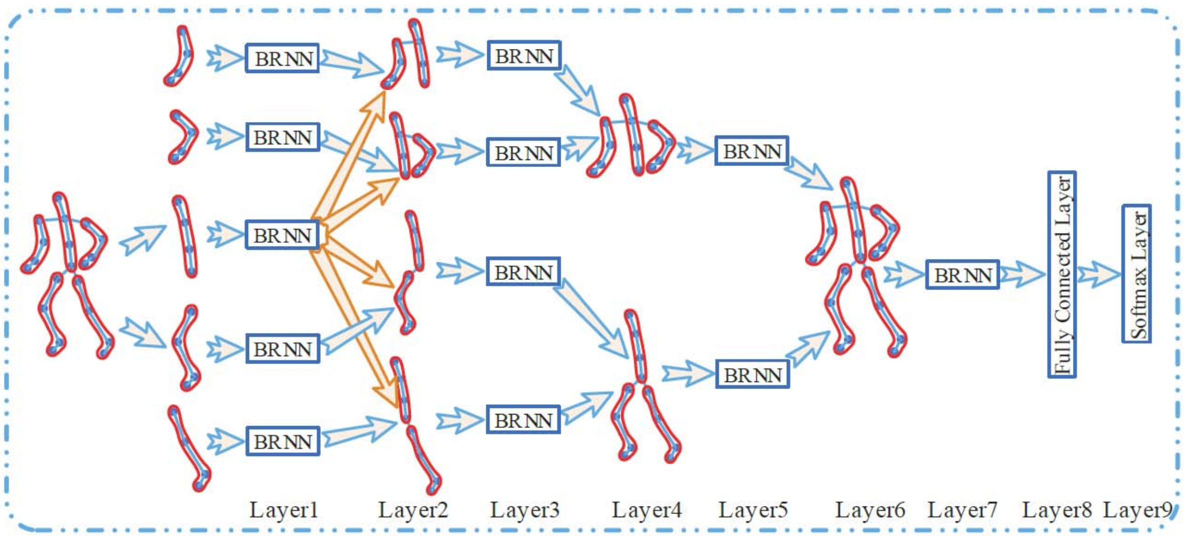
- Click embedding
  - One-hot encoding
  - Item2vec: Item – word, session – sentence
- Class-imbalance issue
  - Undersampling + cost-sensitive learning
- Concept drift
  - Incremental model updating

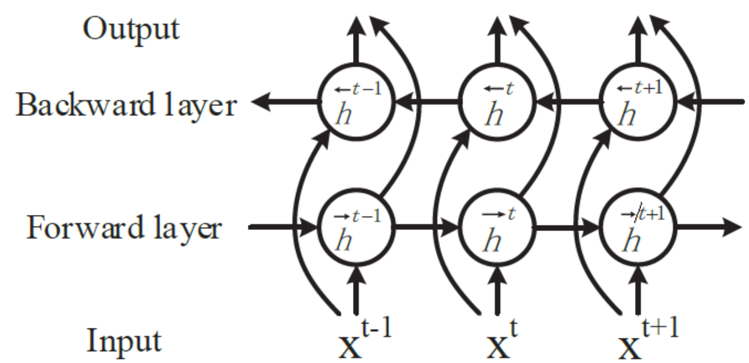
Wang, S., Liu, C., Gao, X., Qu, H., & Xu, W. (2017). Session-Based Fraud Detection in Online E-Commerce Transactions Using Recurrent Neural Networks. In *ECMLPKDD* (pp. 241-252).

Trajectory data - Human skeleton recognition

- Parts-based Hierarchical RNNs

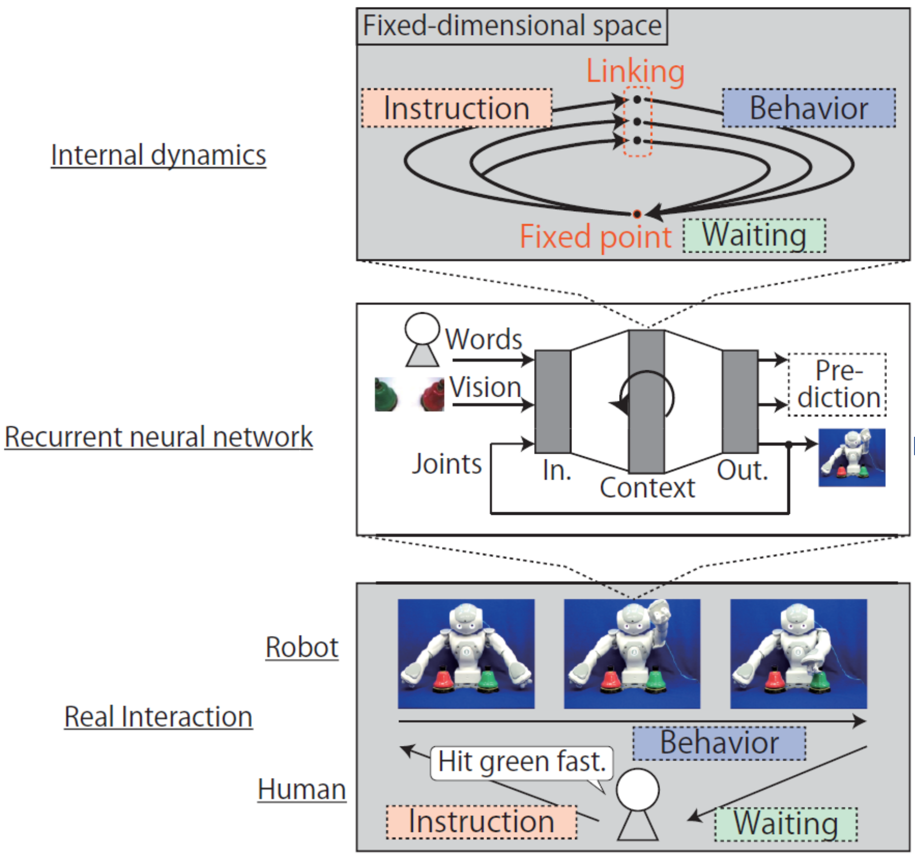
Bidirectional RNN

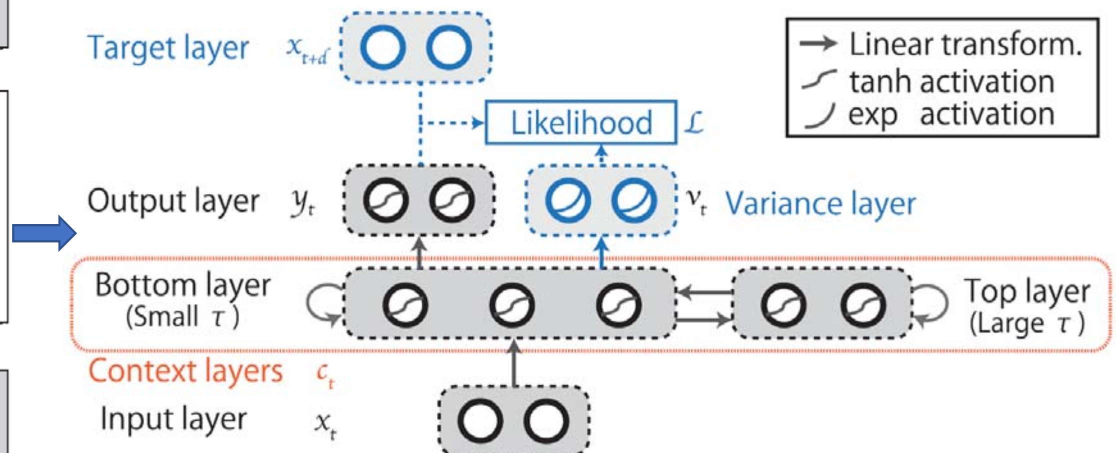




Du, Y., Wang, W., & Wang, L. (2015). Hierarchical recurrent neural network for skeleton based action recognition. In *CVPR* (pp. 1110-1118).

Human language-behavior sequence

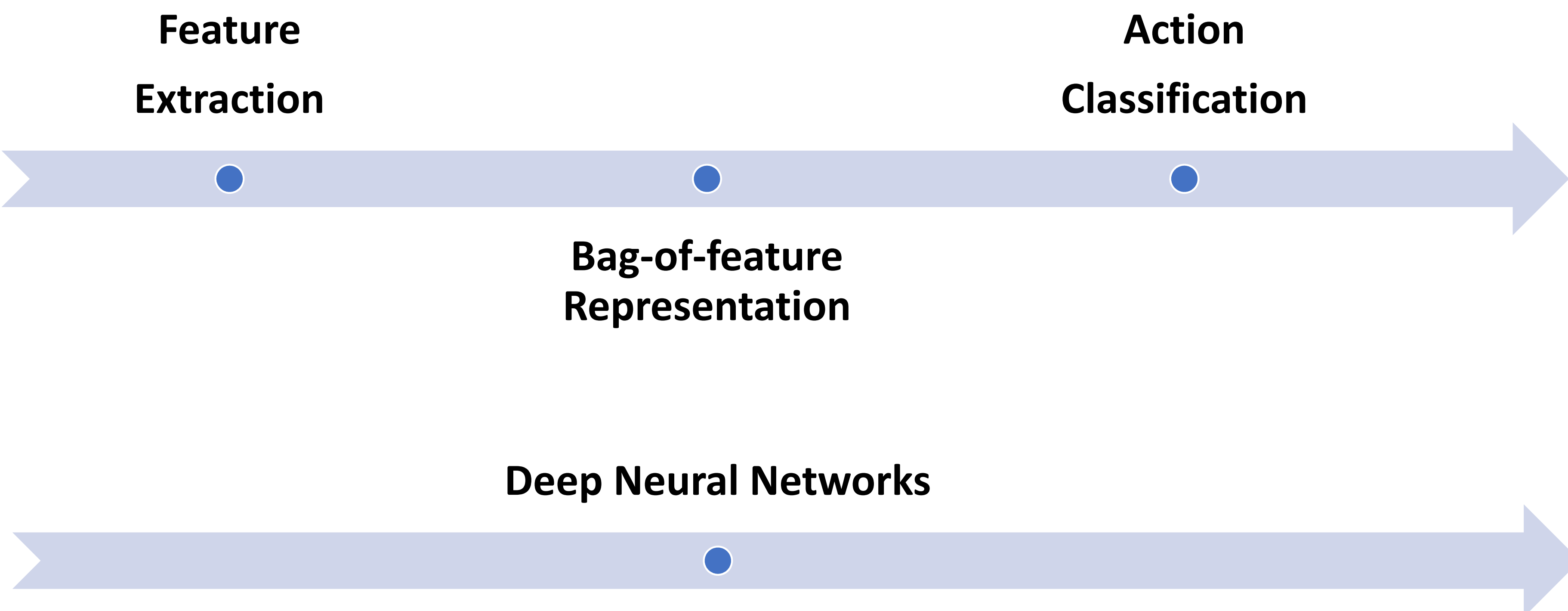




Yamada, T., Murata, S., Arie, H., & Ogata, T. (2016). Dynamical integration of language and behavior in a recurrent neural network for human–robot interaction. *Frontiers in neurorobotics*, *10*, 5.

Behavior Analysis in Visual Data

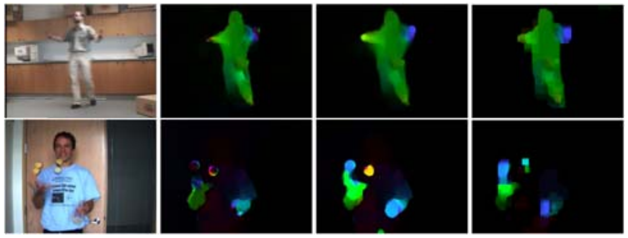
Action recognition



Some key basic concepts

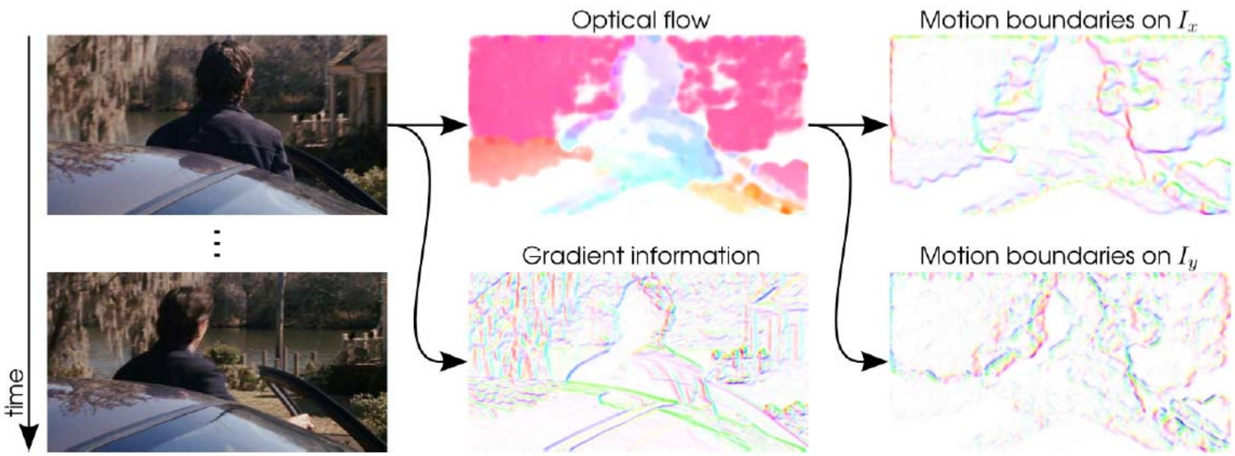
- Motion -> action -> activity
- Optical flow
  - 2D vectors of motion displacement

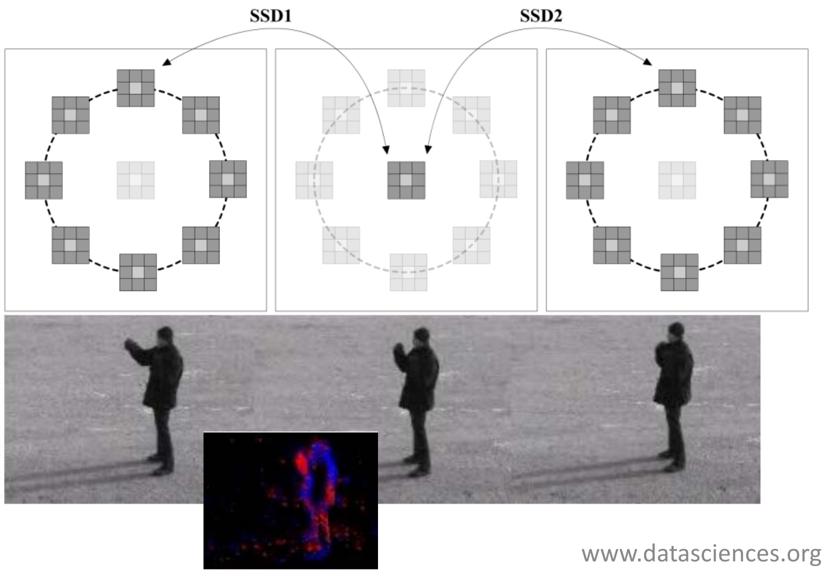
$$I(x, y, t) = I(x + dx, y + dy, t + dt)$$



Some key concepts

- Popular feature descriptors
  - Histograms of optic flow (HoF) – absolute motion
  - Histograms of oriented gradient (HoG) – appearance
  - Motion boundary histograms (MBH) – relative motion
  - Local binary/trinary patterns - motion





Space-time Feature Extraction

- Detect space-time interest points
- Extract histogram-based features of space-time volumes in the neighborhood of detected points → 100,000 features
  - histograms of oriented gradient (**HoG**) and optic flow (**HoF**) are respectively used
- Bag-of-features representations
  - Create a visual feature vocabulary
    - K-means clustering (k=4,000) with the above features
    - The cluster centers are used as a visual word
    - Each feature is assigned to the closest visual word
- Classification based on the BoF representations

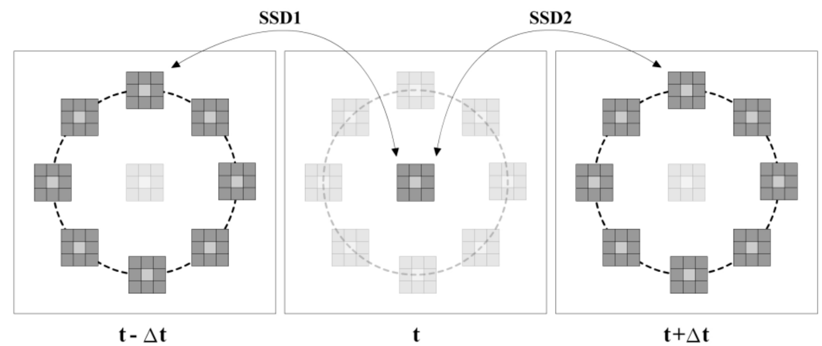
Task	HoG BoF	HoF BoF
KTH multi-class	81.6%	89.7%
Action AnswerPhone	13.4%	24.6%
Action GetOutCar	21.9%	14.9%
Action HandShake	18.6%	12.1%
Action HugPerson	29.1%	17.4%
Action Kiss	52.0%	36.5%
Action SitDown	29.1%	20.7%
Action SitUp	6.5%	5.7%
Action StandUp	45.4%	40.0%

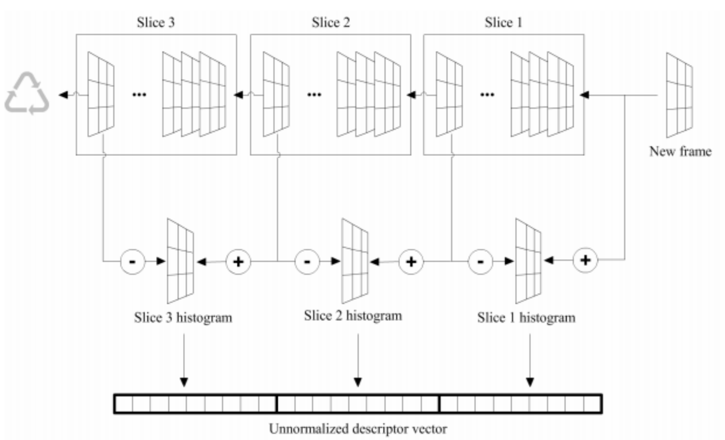
Video text scripts are also exploited

Laptev, I., Marszalek, M., Schmid, C., & Rozenfeld, B. (2008). Learning realistic human actions from movies. In *CVPR* (pp. 1-8).

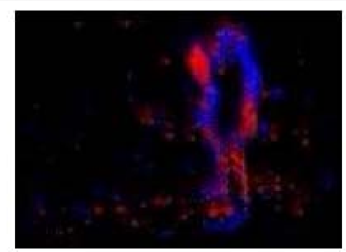
Local Pattern-based Features

- Trinary encoding of every pixel per frame
    - local self-similarities between frames
- Histograms of k frames per time slice
- Classification based on the histograms





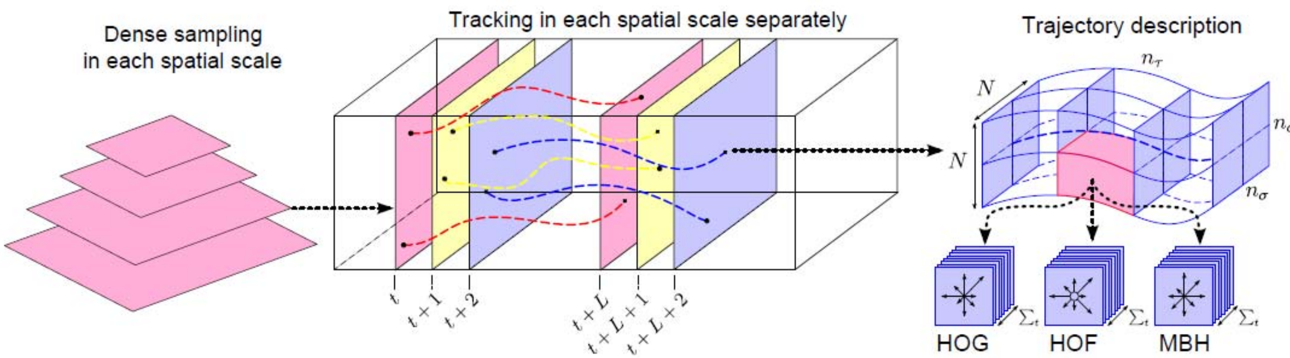
Class	LTP (single)	Laptev (combined) [11]	Laptev (single) [11]
Answer phone	**35.1%**	32.1%	26.7%
Get out car	32.0%	41.5%	22.5%
Hand shake	**33.8%**	32.3%	23.7%
Hug person	28.3%	40.6%	34.9%
Kiss	**57.6%**	53.3%	52.0%
Sit down	36.2%	38.6%	37.8%
Sit up	13.1%	18.2%	15.2%
Stand up	**58.3%**	50.5%	45.4%

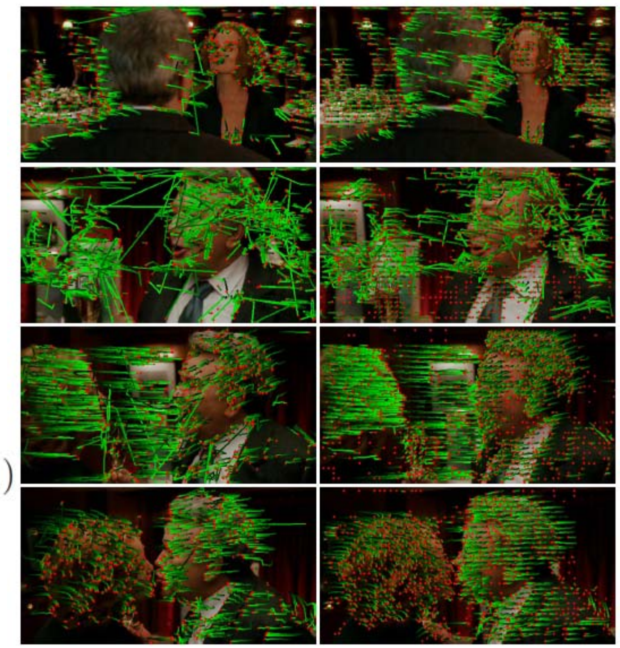


Yeffet, L., & Lior W. (2009). Local trinary patterns for human action recognition. In *CVPR* (pp. 492-497).

From Cuboids to Trajectories

Wang, H., Kläser, A., Schmid, C., & Liu, C. L. (2011). Action recognition by dense trajectories. In *CVPR* (pp. 3169-3176).





- Densely sampled points of subsequent frames are concatenated to form a trajectory : $(P_t, P_{t+1}, P_{t+2}, \dots)$
  - 5x5 dense sampling patches
  - Trajectory length of 15 (L=15) is used to avoid drifting
  - Shape of a trajectory encodes local motion patterns
    - Described by displacement vector: $S = (\Delta P_t, \dots, \Delta P_{t+L-1})$
      $\Delta P_t = (P_{t+1} - P_t) = (x_{t+1} - x_t, y_{t+1} - y_t)$

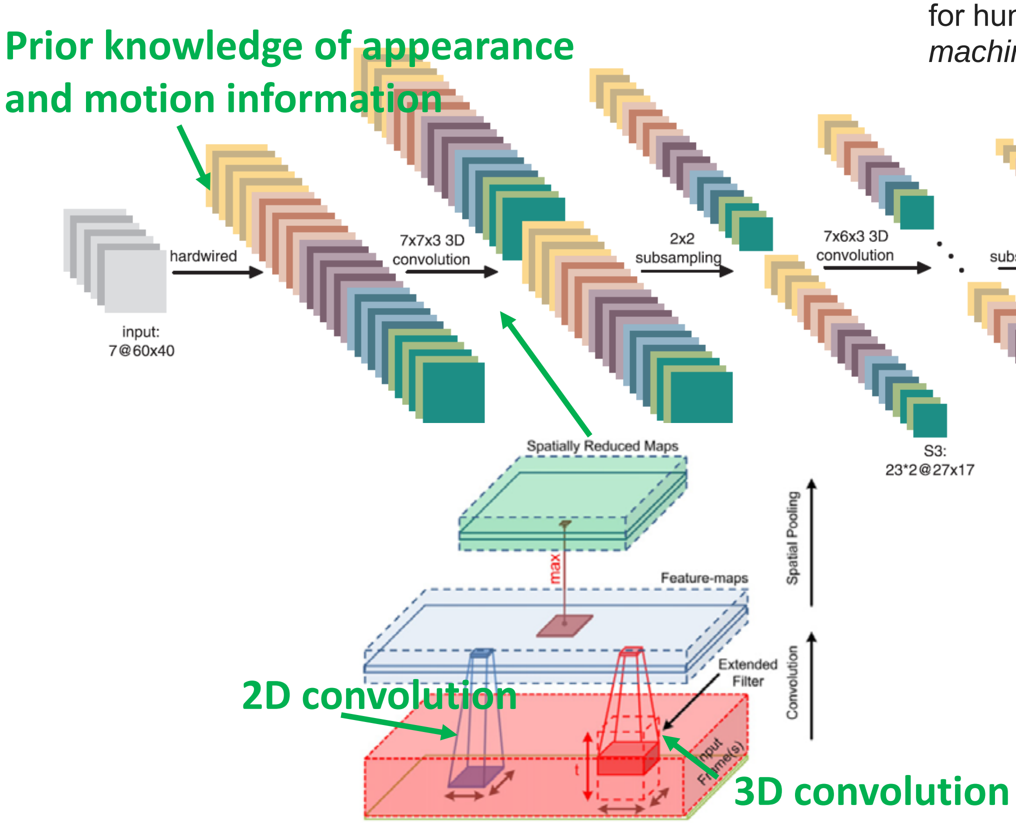
Action feature learning with deep architectures

- Spatiotemporal networks
- Multiple stream networks
- Deep generative networks
- Temporal coherency networks

Herath, S., Harandi, M., & Porikli, F. (2017). Going deeper into action recognition: A survey. *Image and vision computing*, *60*, 4-21.

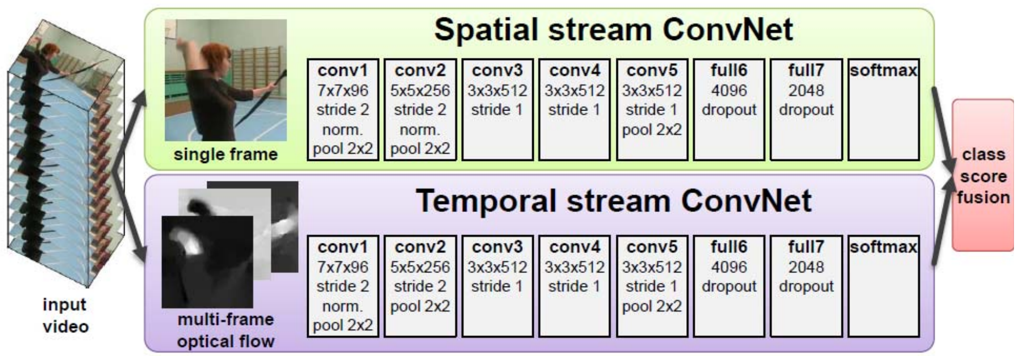
Spatiotemporal networks - 3D convolutional neural networks

Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, *35*(1), 221-231.



Multiple stream networks

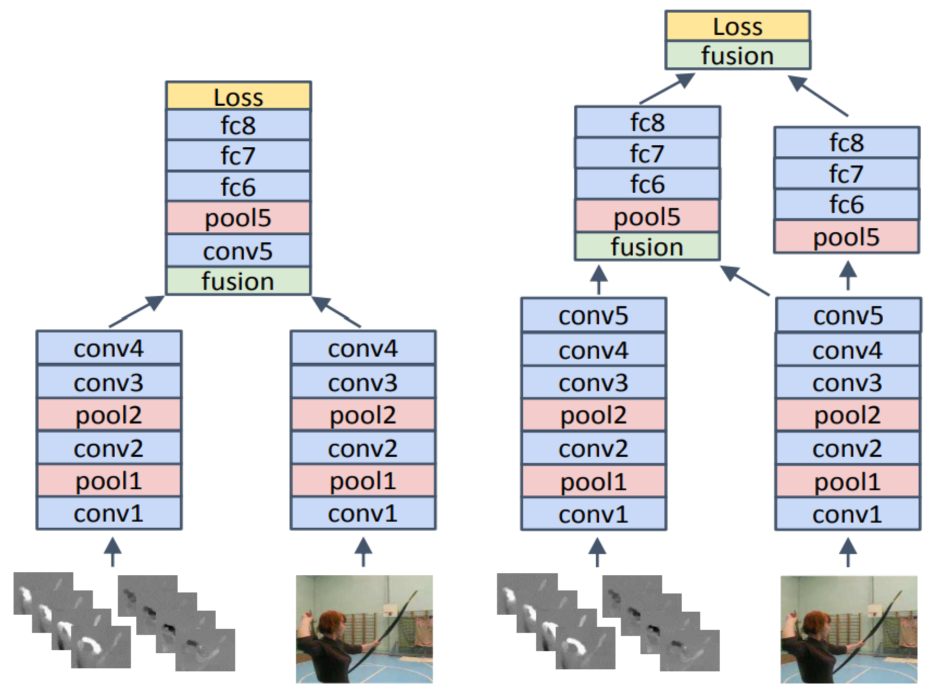
- Two-stream convolutional networks



Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In *NIPS* (pp. 568-576). 2014

Feichtenhofer, C., Pinz, A., & Zisserman, A. (2016). Convolutional two-stream network fusion for video action recognition. In *CVPR* (pp. 1933-1941).

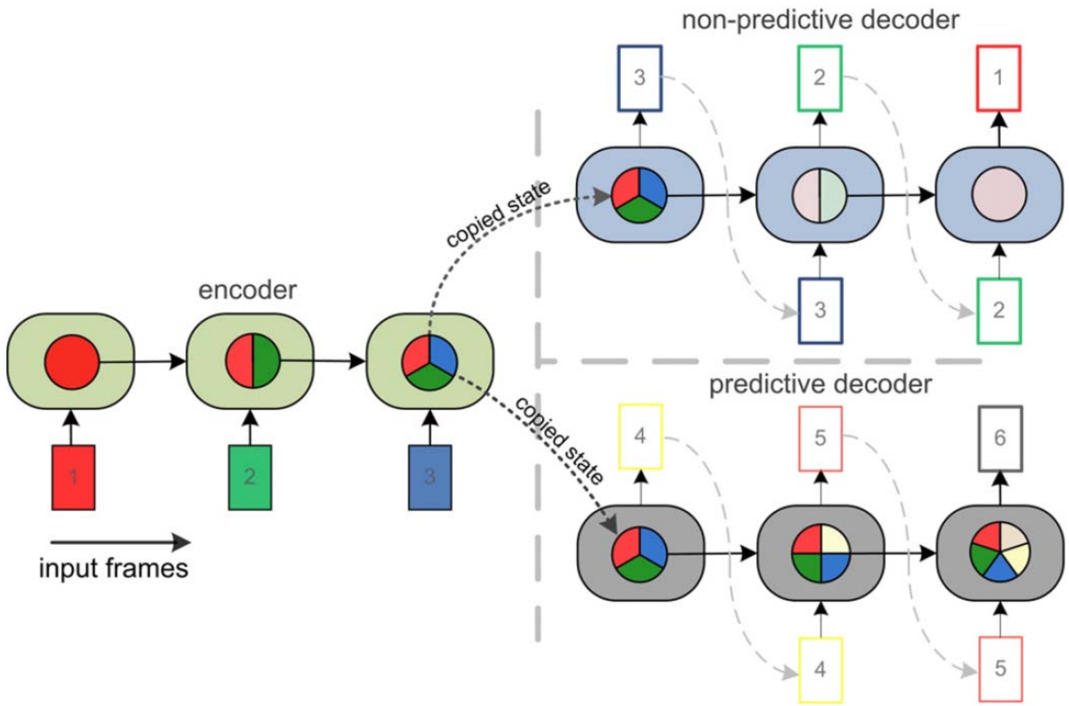
- Where to fuse the networks?



BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Deep generative networks - LSTM autoencoder

- Unsupervised learning of video representations

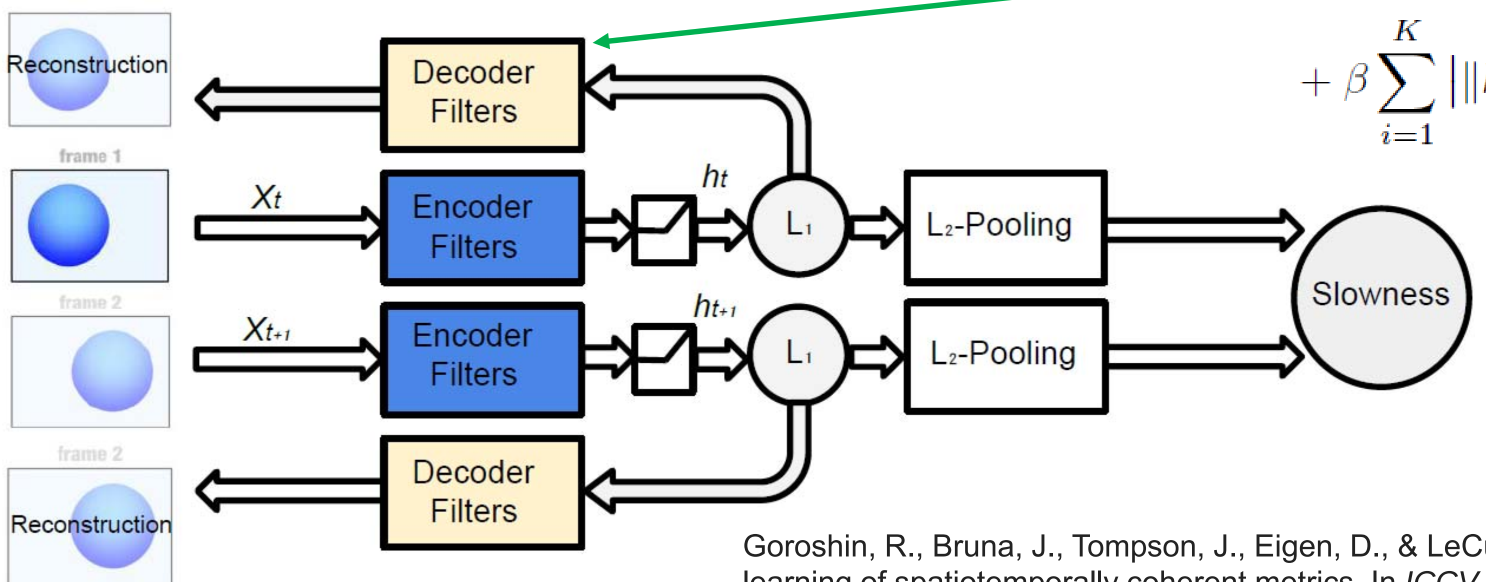


Srivastava, N., Mansimov, E., & Salakhudinov, R. (2015). Unsupervised learning of video representations using lstms. In *ICML* (pp. 843-852).

Temporal coherency networks

- Slowness as metric learning

$$L(x_t, x_{t'}, W) = \sum_{\tau=\{t,t'\}} \left(\|W_d h_\tau - x_\tau\|^2 + \alpha|h_\tau|\right) + \beta \sum_{i=1}^{K} \left| \|h_t\|^{P_i} - \|h_{t'}\|^{P_i} \right|$$



Goroshin, R., Bruna, J., Tompson, J., Eigen, D., & LeCun, Y. (2015). Unsupervised learning of spatiotemporally coherent metrics. In *ICCV* (pp. 4086-4093).

Behavior Learning from Demonstrations

Imitation learning

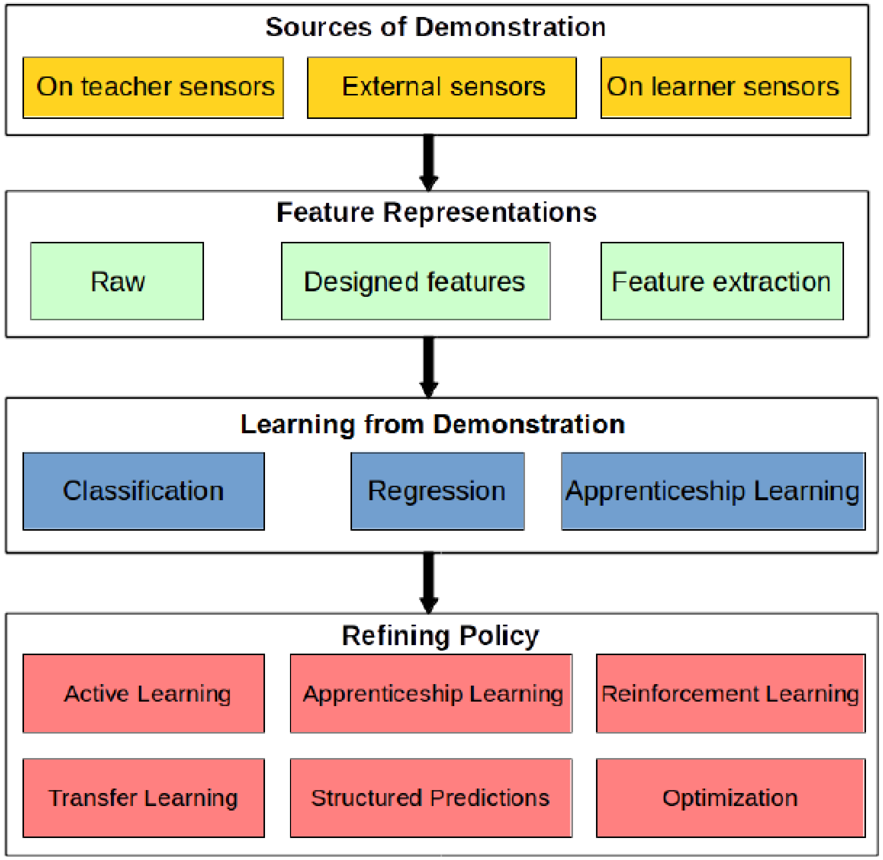
- Learning from demonstrations
  - An agent (a learning machine) is trained to perform a task from (expert) demonstrations by learning a mapping between observations and actions

Demonstration set: $D = \{x_i, y_i\}_{i=1}^{n}$

Policy learning: $u(t) = \pi(x(t), t, \alpha)$

- Learning from experiences

$$E = \{s_i, a_i, r_i, s_i'\}_{i=1}^{n}$$



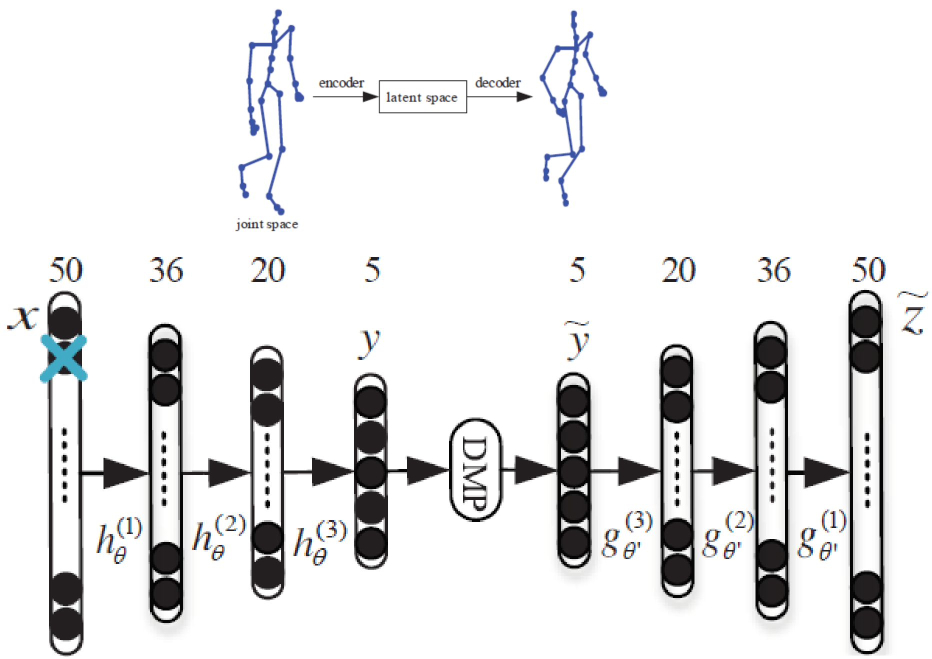
Hussein, A., Gaber, M. M., Elyan, E., & Jayne, C. (2017). Imitation learning: A survey of learning methods. *ACM Computing Surveys (CSUR)*, *50*(2), 21.

Key challenges

- Action trajectories violate the **IID** assumption adopted in most machine learning practices
- Noisy and erroneous signals in demonstrations
- Unseen circumstances
- Correspondence/matching between the learner and the teacher
- Computing power and memory limitations in on-board computers
- Variations in the task and the surroundingenvironment
- …

Representation learning from demonstrations

- AE-DMP (Autoencoded Dynamic Movement Primitive)



DMP: $$\mathbf{w}^\star = \operatorname*{argmin}_{\mathbf{w}} \sum_{i=1}^{m} L\left[\mathbf{f}_t^{(i)}, \mathbf{f}(s)^{(i)}\right]$$

AE: $$\theta^\star, \theta'^\star = \operatorname*{argmin}_{\theta,\theta'} \frac{1}{n}\sum_{i=1}^{n} L\left[\mathbf{x}^{(i)}, g_{\theta'}\left(h_\theta(\tilde{\mathbf{x}}^{(i)})\right)\right]$$

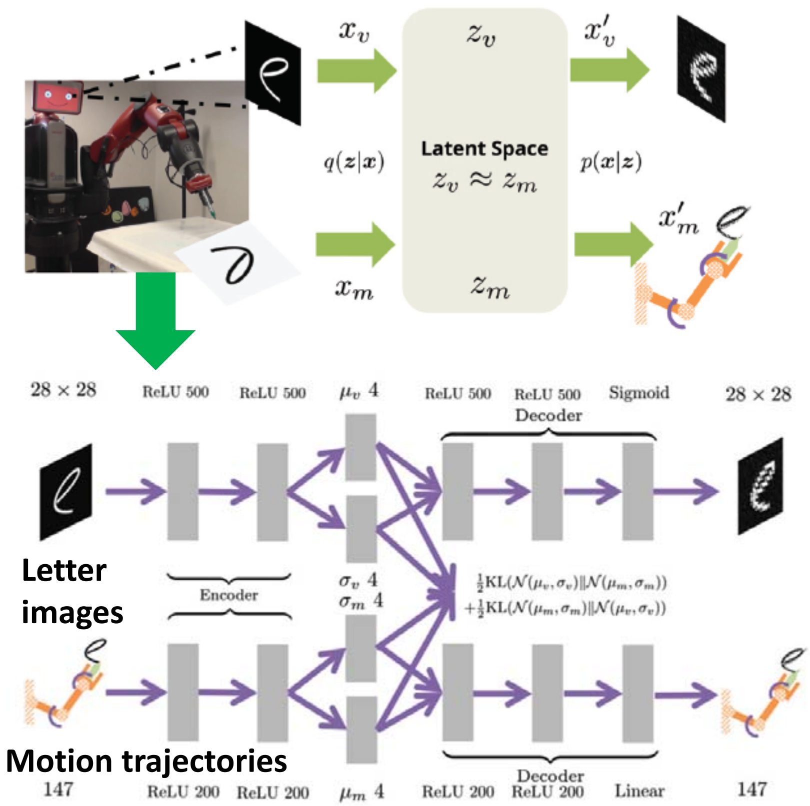
AE-DMP

$$\operatorname*{argmin}_{\mathbf{w},\theta,\theta'} \sum_{i=1}^{m}\left\{L(\mathbf{x}^{(i)}, \tilde{\mathbf{z}}^{(i)}) + \lambda L\left[\mathbf{f}_t^{(i)}, \mathbf{f}(s)^{(i)}\right]\right\}$$

Chen, N., Bayer, J., Urban, S., & Van Der Smagt, P. (2015). Efficient movement representation by embedding dynamic movement primitives in deep autoencoders. In *IEEE-RAS* (pp. 434-440).

Representation learning from multi-modal demonstrations

Yin, H., Melo, F. S., Billard, A., & Paiva, A. (2017). Associate Latent Encodings in Learning from Demonstrations. In *AAAI* (pp. 3848-3854).



- Robotic handwriting: robots take a **visual** sensor as inputs and output an **arm motion trajectory**, e.g., write a letter

Single-modal Encoding

$$\mathcal{L}_v(\boldsymbol{\theta}_v, \boldsymbol{\phi}_v, \boldsymbol{x}_v^i) = \mathrm{KL}[q_{\phi_v} \| p(\boldsymbol{z}_v|\boldsymbol{x}_v^i)] - \log p(\boldsymbol{x}_v^i)$$

Cross-modal Association

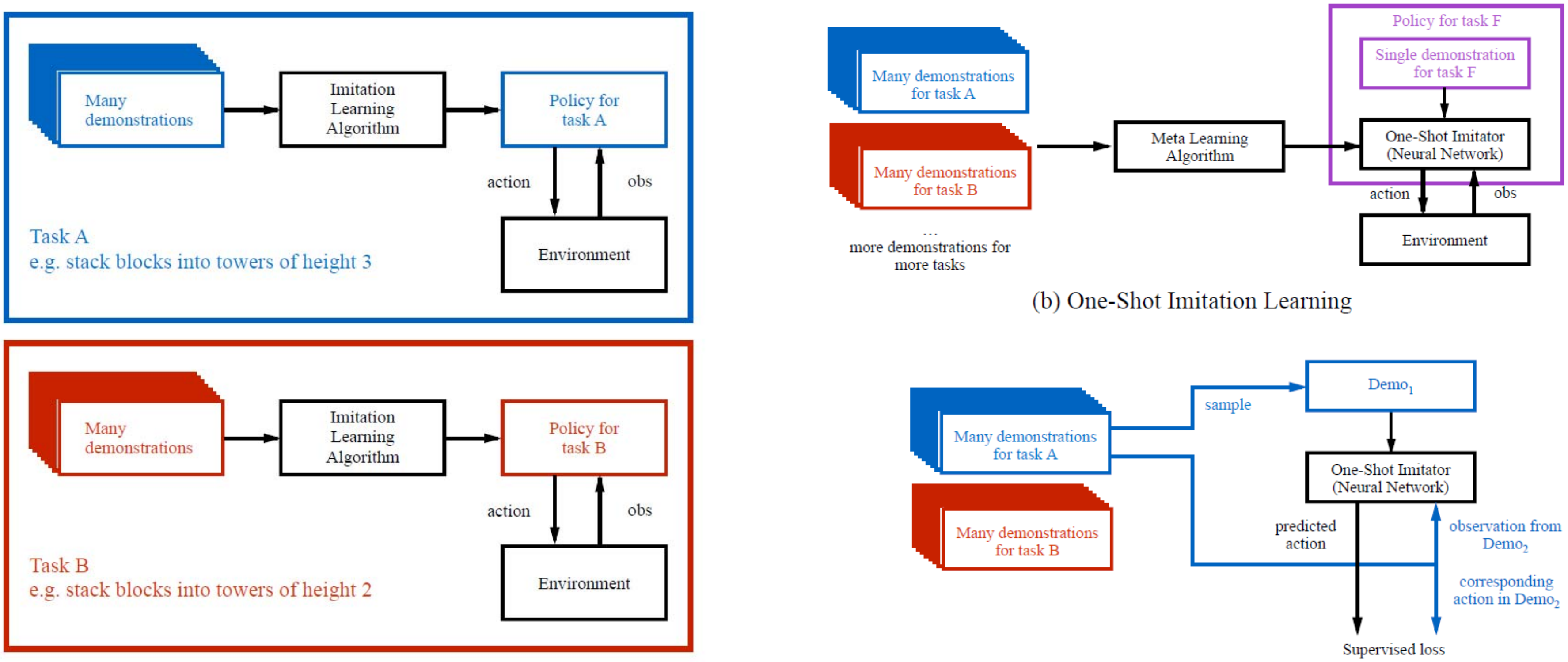
$$q_{\phi_v}(\boldsymbol{z}|\boldsymbol{x}_v^i) = q_{\phi_m}(\boldsymbol{z}|\boldsymbol{x}_m^i), \forall \boldsymbol{z}$$

$$\mathcal{L}_{assoc} = \mathrm{KL}(q_{\phi_v}(\boldsymbol{z}_v|\boldsymbol{x}_v^i) \| q_{\phi_m}(\boldsymbol{z}_m|\boldsymbol{x}_m^i)) + \mathrm{KL}(q_{\phi_m}(\boldsymbol{z}_m|\boldsymbol{x}_m^i) \| q_{\phi_v}(\boldsymbol{z}_v|\boldsymbol{x}_v^i))$$

Joint Objective

$$\mathcal{L}(\boldsymbol{\theta}_v, \boldsymbol{\theta}_m, \boldsymbol{\phi}_v, \boldsymbol{\phi}_m, \boldsymbol{x}_v^i, \boldsymbol{x}_m^i) = \mathcal{L}_v + \mathcal{L}_m + \lambda \mathcal{L}_{assoc}$$

One-shot imitation learning



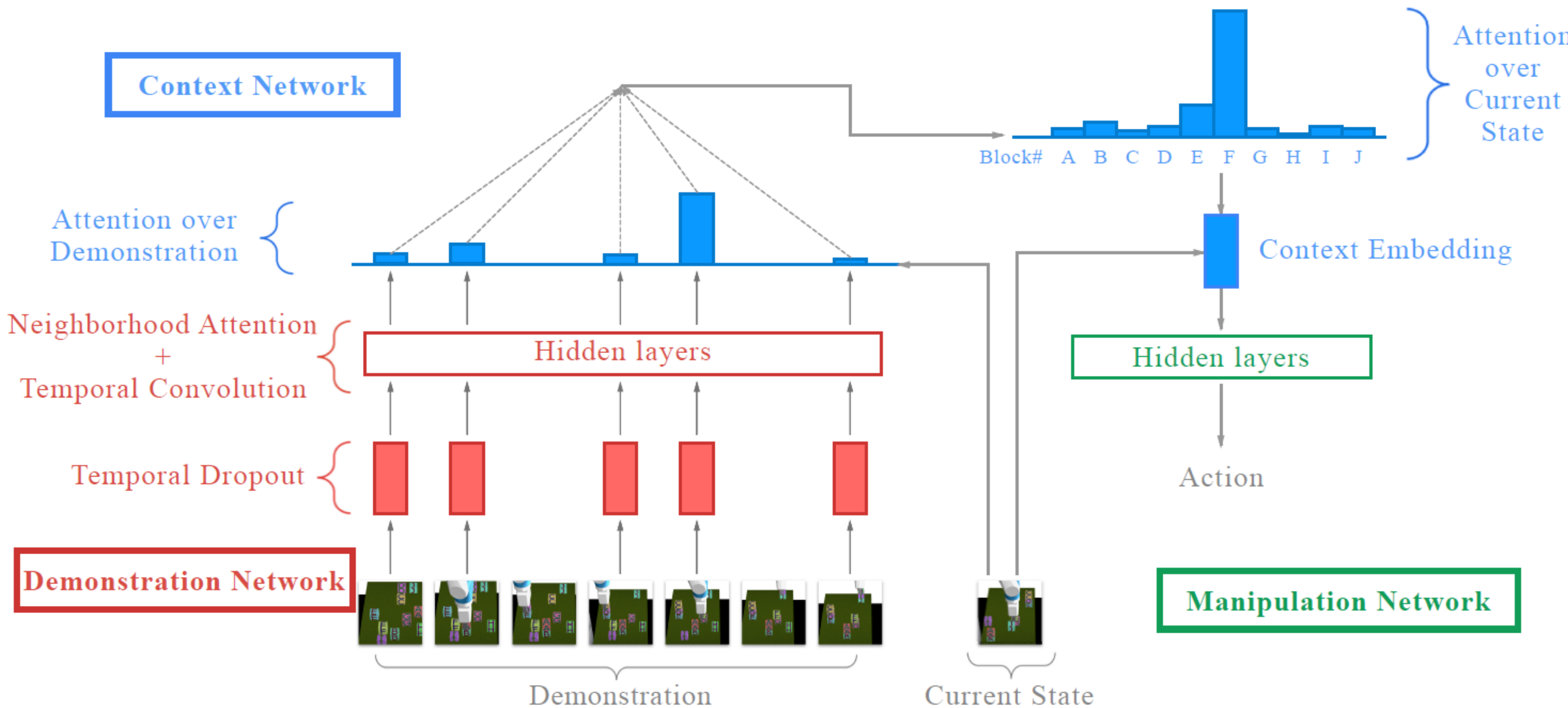
(a) Traditional Imitation Learning

(b) One-Shot Imitation Learning

(c) Training the One-Shot Imitator

Duan, Y., Andrychowicz, M., Stadie, B., Ho, O. J., Schneider, J.,... & Zaremba, W. (2017). One-shot imitation learning. In *NIPS* (pp. 1087-1098).

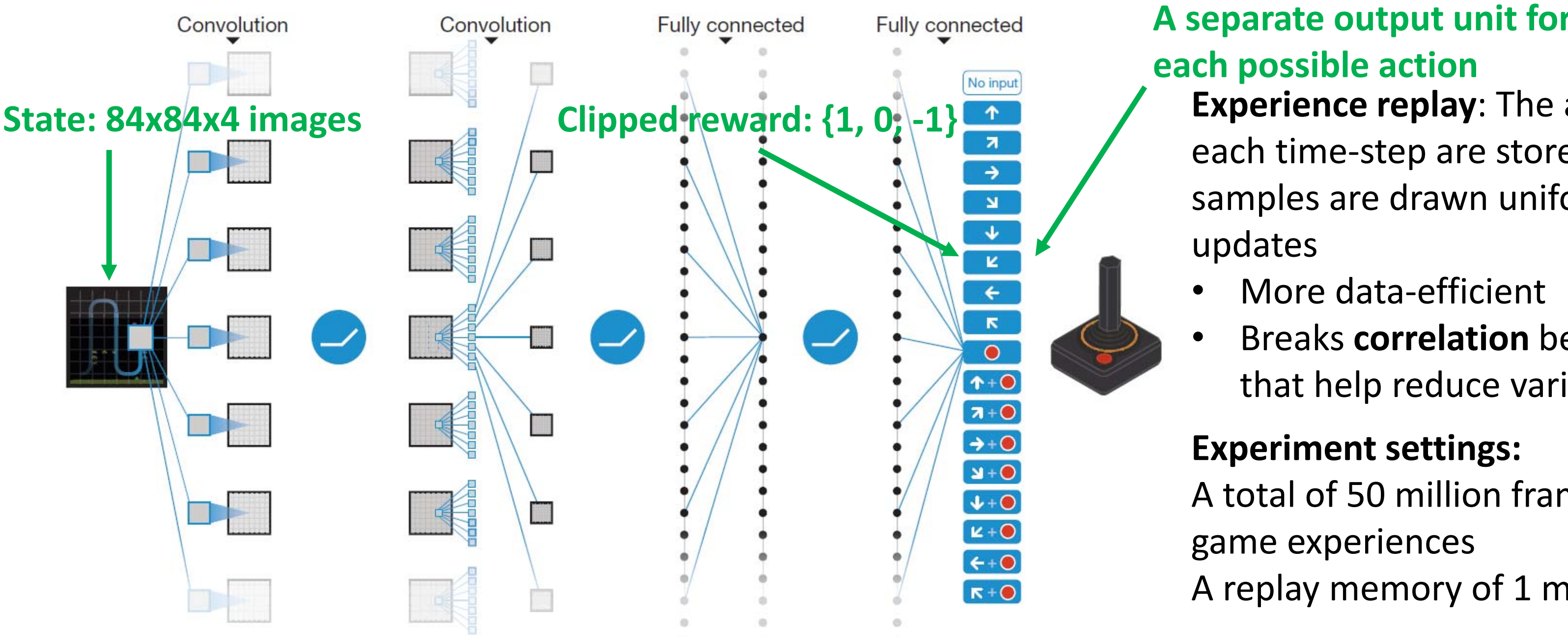
One-shot imitation learning



Duan, Y., Andrychowicz, M., Stadie, B., Ho, O. J., Schneider, J.,... & Zaremba, W. (2017). One-shot imitation learning. In *NIPS* (pp. 1087-1098).

End-to-end learning from experiences

- Deep Q-learning with experience replay



Experience replay: The agent's last N experiences at each time-step are stored, from which minibatch samples are drawn uniformly to perform Q-learning updates

- More data-efficient
- Breaks **correlation** between consecutive samples that help reduce variances of the updates

Experiment settings:

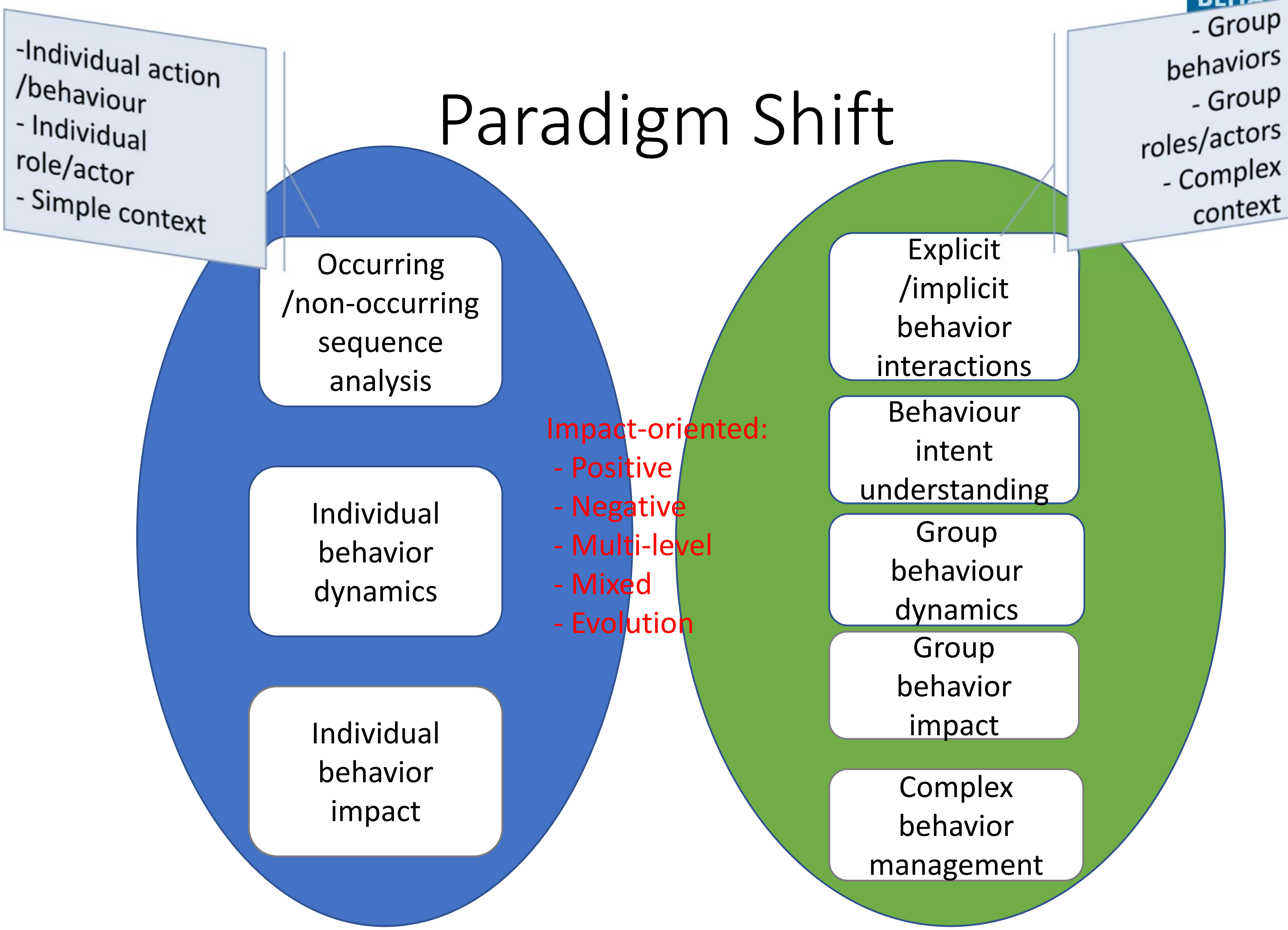
A total of 50 million frames, i.e., around 38 days of game experiences

A replay memory of 1 million most recent frames.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529.

Challenges and Prospects

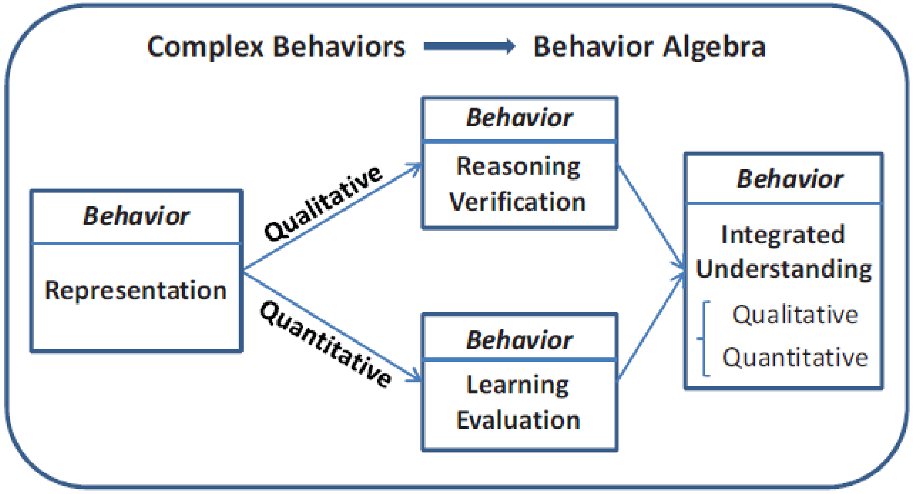
Paradigm Shift



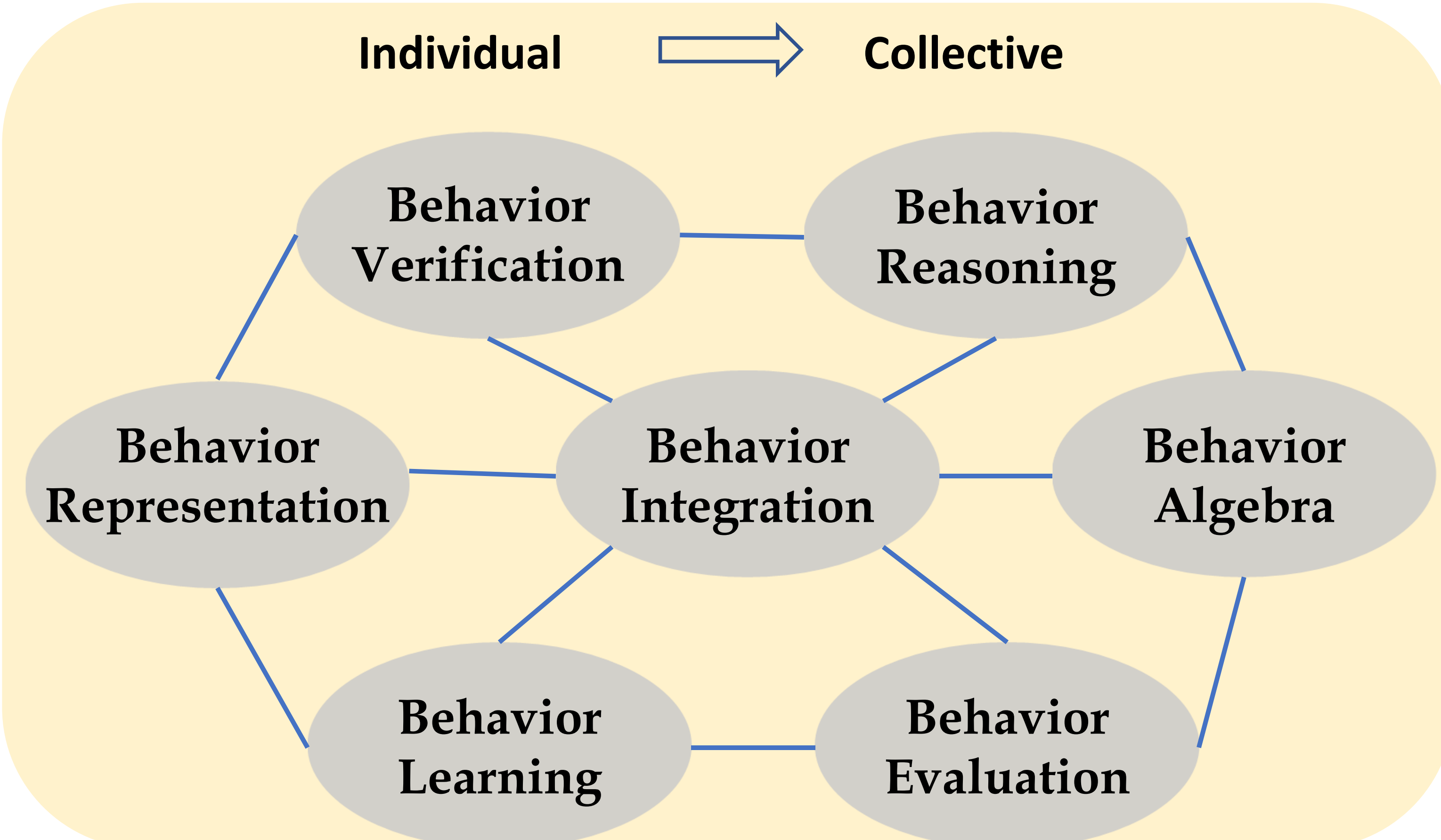
Modeling and Analysis of Complex Behaviors

- Qualitative behavior analytics
- Quantitative behavior analytics
- Integrated behavior analytics

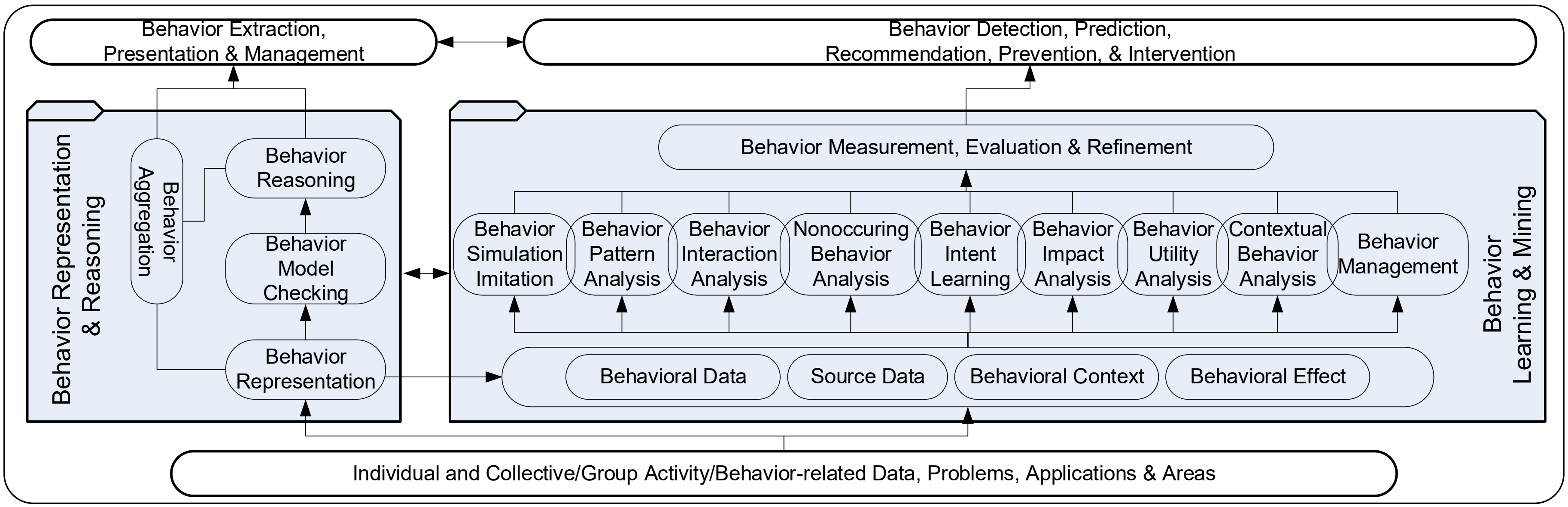
- Open issues:
  e.g., behavior reasoning, behavior learning, behavior evaluation, behavior integration at individual but more on group levels.



Modeling and Analysis of Complex Behaviors



Behavior Informatics – a big area



Individual Behaviour Learning

- Behaviour intent learning
- Non-occurring sequence/behaviour analysis
- Multiple behaviours/sequences coupled with complex interactions
- Behaviour impact learning
- Behaviour utility learning
- Early prediction of high impact/utility behaviours
- Next-best action prediction
- …

Group-oriented Behaviour Analytics

- Group behaviour intent learning
- Coupled group behaviour sequence modelling and analysis
- Explicit and implicit behaviour couplings in collective behaviours
- Heterogeneous behaviour analysis
- Social influence/impact modelling and analysis
- Individual and group behaviour evolution (e.g., divergence vs. convergence of group behaviors)
- Intervention of individual/group behaviours in groups
- …

Non-IID Behaviour Analytics

- Individual/group behaviours are non-IID, with varied and hierarchical couplings and heterogeneities between behaviours and between actors

Definition 4 *(Coupled Behaviors) Coupled behaviors $\mathbb{B}_c$ refer to behaviors $\mathbb{B}_{i_1j_1}$ and $\mathbb{B}_{i_2j_2}$ that are coupled in terms of relationships $f(\theta(\cdot), \eta(\cdot))$, where $(i_1 \neq i_2) \vee (j_1 \neq j_2) \wedge (1 \leq i_1, i_2 \leq I) \wedge (1 \leq j_1, j_2 \leq J_{max})$*

$$\mathbb{B}_c = (\mathbb{B}^{\theta}_{i_1j_1})^{\eta} * (\mathbb{B}^{\theta}_{i_2j_2})^{\eta} ::= \mathbb{B}_{ij}(\mathscr{E}, \mathscr{O}, \mathscr{C}, \mathscr{R})| \sum_{i_1,i_2=1}^{I} \sum_{j_1,j_2=1}^{J_{max}} f(\theta_{j_1j_2}(\cdot), \eta_{i_1i_2}(\cdot)) \odot (\mathbb{B}_{i_1j_1}\mathbb{B}_{i_2j_2}) \quad (5)$$

$$FM(\mathbb{B}) = \left(\begin{array}{cc|cc} \mathbb{B}_{11} & \mathbb{B}_{12} & \ldots & \mathbb{B}_{1J_{max}} \\ \mathbb{B}_{21} & \mathbb{B}_{22} & \ldots & \mathbb{B}_{2J_{max}} \\ \vdots & \vdots & \ddots & \vdots \\ \hline \mathbb{B}_{I1} & \mathbb{B}_{I2} & \ldots & \mathbb{B}_{IJ_{max}} \end{array}\right)$$

Cao, L. (2014). Non-IIDness Learning in Behavioral and Social Data. *The Computer Journal*, *57*(9), 1358-1370.
Cao, L., Ou, Y., Yu, P.S. Coupled Behavior Analysis with Applications, *IEEE Transactions on Knowledge and Data Engineering*, 24 (8): 1378-1392, 2012

Non-IID Learning, KDD2017 Tutorial: www.datasciences.org

References

It is very challenging to extract and list all references related to comprehensive behavior analysis, here we mainly list references related to ‘hard’ behavior analytics (i.e., action, activity and their sequences-centric). There are much more papers address the ‘soft’ behavior perspectives.

References

- Book, special issues and workshop proceedings
  - Longbing Cao. Behavior Informatics: A New Perspective. *IEEE Intelligent Systems (Trends and Controversies)*, 29(4): 62-80, 2014.
  - Longbing Cao, Yu, Philip S; Motoda, Hiroshi; Williams, Graham. Special issue on behavior computing (editorial), *Knowledge and Information Systems*, 37(2): 245-249, 2013
  - Longbing Cao; Hiroshi Motoda, Jaideep Srivastava, Ee-Peng Lim, Irwin King, Philip S. Yu, Wolfgang Nejdl, Guandong Xu, Gang Li, Ya Zhang (Eds.). Behavior and Social Computing, Proceedings of International Workshop on Behavior and Social Informatics and Computing, Lecture Notes in Computer Science, Vol. 8178, Springer, 2013
  - Longbing Cao, Philip S Yu (Eds). Behavior Computing: Modeling, Analysis, Mining and Decision, Springer, 2012.

References

- Concepts of behavior informatics and behavior computing

  - Longbing Cao, In-depth Behavior Understanding and Use: the Behavior Informatics Approach, *Information Science*, 180(17); 3067-3085, 2010
  - Longbing Cao. Behavior Informatics to Discover Behavior Insight for Active and Tailored Client Management. KDD2017, 15-16 (Industry invited talks), 2017.
  - Longbing Cao. Coupling Learning of Complex Interactions, Information Processing and Management, 51(2): 167-186 (2015).
  - Longbing Cao and Thorsten Joachims. Behavior Computing, *IEEE Intelligent Systems*, 29(4): 62-66, 2014.
  - Longbing Cao, Yanchang Zhao, Chengqi Zhang, Huaifeng Zhang. Activity Mining: from Activities to Actions, *International Journal of Information Technology & Decision Making*, 7(2): 259-273, 2008

References

- Behavior modeling and representation
  - Wang, C., Cao, L., & Chi, C. -H. (2015). Formalization and Verification of Group Behavior Interactions. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *45*(8), 1109-1124.
  - Wang, C., and Cao, L. (2012). Modeling and Analysis of Social Activity Process. In *Behavior Computing Modeling, Analysis, Mining and Decision* (pp. 21-35). Springer Science & Business Media. doi:10.1007/978-1-4471-2969-1_2

References

- Behavior sequence analysis

  - L. Wang, X. Bao, H. Chen, L. Cao. Effective Lossless Condensed Representation and Discovery of Spatial Co-location Patterns. *Information Sciences*, 2018
  - Lizhen Wang, Xuguang Bao, Longbing Cao. Interactive Probabilistic Post-mining of User-preferred Spatial Co-location Patterns, *ICDE2018*
  - X. Zhou, L. Chen, Y. Zhang, D. Qin, L. Cao, G. Huang, C. Wang: Enhancing online video recommendation using social user interactions. *VLDB J.* 26(5): 637-656 (2017)
  - Bin Shen, Longbing Cao, Min Yao, Yunjun Gao. Mining preferred navigation patterns by consolidating both selection and time preferences, *World Wide Web* 19(5): 979-1007 (2016).
  - Philippe Fournier-Viger, Cheng-Wei Wu, Vincent S. Tseng, Longbing Cao, Roger Nkambou. Mining Partially-Ordered Sequential Rules Common to Multiple Sequences, *IEEE Trans. Knowledge and Data Engineering*, 27(8): 2203-2216 (2015).
  - A. Fariha, C. Ahmed, C.K. Leung, M. Samiullah, S. Pervin, L. Cao: A new framework for mining frequent interaction patterns from meeting databases. Eng. Appl. of AI 45: 103-118 (2015)
  - Huaifeng Zhang, Yanchang Zhao, Longbing Cao, Chengqi Zhang and Hans Bohlscheid, Customer Activity Sequence Classification for Debt Prevention in Social Security, *Comput. Sci. & Technol.*, 24(6): 1000-1009, 2009

Note: many classic papers on sequence analysis typically focus on occurring sequential pattern mining.

References

- High-impact and combined behavior sequence analysis

  - Longbing Cao, Yanchang Zhao, Chengqi Zhang. Mining Impact-Targeted Activity Patterns in Imbalanced Data, *IEEE Trans. on Knowledge and Data Engineering*, 20(8): 1053-1066, 2008
  - Zhigang Zheng, Wei Wei, Chunming Liu, Wei Cao, Longbing Cao, Maninder Bhatia. An effective contrast sequential pattern mining approach to taxpayer behavior analysis, *World Wide Web* 19(4): 633-651 (2016).
  - Cao, L., Zhang, H., Zhao, Y., Luo, D., & Zhang, C. (2011). Combined Mining: Discovering Informative Knowledge in Complex Data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *41*(3), 699-712
  - Yanchang Zhao, Huaifeng Zhang, Longbing Cao, Jian Pei, Shanshan Wu, Chengqi Zhang and Hans Bohlscheid, Debt Detection in Social Security by Sequence Classification Using Both Positive and Negative Patterns, ECML-PKDD 2009, 648-663, 2009.
  - Yanchang Zhao, Huaifeng Zhang, Longbing Cao, Chengqi Zhang and Hans Bohlscheid. Mining Both Positive and Negative Impact-Oriented Sequential Rules From Transactional Data, PAKDD 2009, 656-663

References

- High utility behavior sequence analysis
  - Jingyu Shao, Junfu Yin, Wei Liu, Longbing Cao. Mining actionable combined patterns of high utility and frequency. DSAA 2015: 1-10
  - Junfu Yin, Zhigang Zheng, Longbing Cao, Yin Song, Wei Wei. Efficiently Mining Top-K High Utility Sequential Patterns, ICDM2013: 1259-1264.
  - Junfu Yin, Zhigang Zheng, Longbing Cao. USpan: An Efficient Algorithm for Mining High Utility Sequential Patterns, KDD 2012, 660-668

References

- Nonoccurring behavior analysis

  - Longbing Cao, Philip S. Yu, Vipin Kumar. Nonoccurring Behavior Analytics: A New Area. *IEEE Intelligent Systems* 30(6): 4-11 (2015)
  - Xiangjun Dong, Yongshun Gong and Longbing Cao. F-NSP+: A fast negative sequential patterns mining method with self-adaptive data storage, Pattern Recognition, 2018
  - Longbing Cao, Xiangjun Dong and Zhigang Zheng. e-NSP: Efficient Negative Sequential Pattern Mining. Artificial Intelligence, 235: 156-182, 2016.
  - Dong, X., Zheng, Z., Cao, L., Zhao, Y., Zhang, C., Li, J., Wei, W. & Ou, Y. (2011). e-NSP: efficient negative sequential pattern mining based on identified positive patterns without database rescanning, CIKM, 825-830
  - Zheng, Z., Zhao, Y., Zuo, Z., & Cao, L. (2010). An Efficient GA-Based Algorithm for Mining Negative Sequential Patterns. *Advances in Knowledge Discovery and Data Mining - Lecture Notes in Artificial Intelligence,* pp. 262-273
  - Zhigang Zheng, Yanchang Zhao, Ziye Zuo, Longbing Cao, Huaifeng Zhang, Yanchang Zhao, Chengqi Zhang. An Efficient GA-Based Algorithm for Mining Negative Sequential Patterns, PAKDD 2010, 262-273

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

References

- Non-IID, coupled, collective and group behavior analysis

  - Longbing Cao. Coupling Learning of Complex Interactions, Information Processing and Management, 51(2): 167-186 (2015)
  - Cao, L. (2014). Non-IIDness Learning in Behavioral and Social Data. *The Computer Journal*, *57*(9), 1358-1370.
  - Wei Cao, Longbing Cao. Financial Crisis Forecasting via Coupled Market State Analysis, *IEEE Intelligent Systems*, 30(2): 18-25 (2015).
  - Wang, C., Cao, L., Gaussier, E., Li, J., Ou, Y., & Luo, D. (2014). Coupled Behavior Representation, Modeling, Analysis, and Reasoning. *IEEE Intelligent Systems*, *29*(4), 66-69.
  - Wei Cao, Longbing Cao, Yin Song. Coupled Market Behavior Based Financial Crisis Detection, IJCNN2013
  - Cao, L., Ou, Y., Yu, P.S. Coupled Behavior Analysis with Applications, *IEEE Transactions on Knowledge and Data Engineering*, 24 (8): 1378-1392, 2012
  - Song, Y., Cao, L. et al. Coupled Behavior Analysis for Capturing Coupling Relationships in Group-based Market Manipulations. KDD 2012, 976-984, 2012.
  - Yin Song and Longbing Cao. Graph-based Coupled Behavior Analysis: A Case Study on Detecting Collaborative Manipulations in Stock Markets, IJCNN 2012, 1-8, 2012
  - Longbing Cao, Yuming Ou, Philip S YU, Gang Wei. Detecting Abnormal Coupled Sequences and Sequence Changes in Group-based Manipulative Trading Behaviors, KDD 2010, 85-94

References

- Statistical modeling of behaviors

  - Trong Dinh Thac Do and Longbing Cao. Metadata-dependent Infinite Poisson Factorization for Efficiently Modelling Sparse and Large Matrices in Recommendation, IJCAI2018
  - Trong Dinh Thac Do, Longbing Cao. Coupled Poisson Factorization Integrated with User/Item Metadata for Modeling Popular and Sparse Ratings in Scalable Recommendation. AAAI2018.
  - Song, Y., Cao, L. et al. Coupled Behavior Analysis for Capturing Coupling Relationships in Group-based Market Manipulations. KDD 2012, 976-984, 2012.
  - Yin Song and Longbing Cao. Graph-based Coupled Behavior Analysis: A Case Study on Detecting Collaborative Manipulations in Stock Markets, IJCNN 2012, 1-8, 2012

References

- Session-based next choice/next-best action recommendation

  - Shoujin Wang, Liang Hu, Longbing Cao, Xiaoshui Huang, Defu Lian and Wei Liu. Attention-based Transactional Context Embedding for Next-Item Recommendation. AAAI2018.
  - S. Wang, L. Hu, L. Cao, X. Huang. Perceiving the Next Choice with Comprehensive Transaction Embeddings for Online Recommendation, *ECML/PKDD2017*
  - L. Hu, L. Cao, S. Wang, G. Xu, J. Cao, Z. Gu: Diversifying Personalized Recommendation with User-session Context. *IJCAI 2017*: 1858-1864
  - Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. Session-based recommendations with recurrent neural networks. In *ICLR.* 2016.
  - Hu, L., Cao, W., Cao, J., Xu, G., Cao, L., & Gu, Z. (2014). Bayesian Heteroskedastic Choice Modeling on Non-identically Distributed Linkages. ICDM, 851-856, 2014.

References

- Action recognition and visual behavior analysis

  - Herath, S., Harandi, M., & Porikli, F. (2017). Going deeper into action recognition: A survey. Image and vision computing, 60, 4-21.
  - Feichtenhofer, C., Pinz, A., & Zisserman, A. (2016). Convolutional two-stream network fusion for video action recognition. In *CVPR* (pp. 1933-1941).
  - Du, Y., Wang, W., & Wang, L. (2015). Hierarchical recurrent neural network for skeleton based action recognition. In *CVPR* (pp. 1110-1118).
  - Srivastava, N., Mansimov, E., & Salakhudinov, R. (2015). Unsupervised learning of video representations using lstms. In *ICML* (pp. 843-852).
  - Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. In *NIPS* (pp. 568-576). 2014
  - Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, *35*(1), 221-231.
  - Wang, H., Kläser, A., Schmid, C., & Liu, C. L. (2011). Action recognition by dense trajectories. In CVPR (pp. 3169-3176)
  - Yeffet, L., & Lior W. (2009). Local trinary patterns for human action recognition. In CVPR (pp. 492-497)
  - Laptev, I., Marszalek, M., Schmid, C., & Rozenfeld, B. (2008). Learning realistic human actions from movies. In CVPR (pp. 1-8).

References

- Imitation learning, observation and experience-based representation
  - Hussein, A., Gaber, M. M., Elyan, E., & Jayne, C. (2017). Imitation learning: A survey of learning methods. ACM Computing Surveys (CSUR), 50(2), 21
  - Yin, H., Melo, F. S., Billard, A., & Paiva, A. (2017). Associate Latent Encodings in Learning from Demonstrations. In *AAAI* (pp. 3848-3854).
  - Duan, Y., Andrychowicz, M., Stadie, B., Ho, O. J., Schneider, J.,... & Zaremba, W. (2017). One-shot imitation learning. In *NIPS* (pp. 1087-1098).
  - Yamada, T., Murata, S., Arie, H., & Ogata, T. (2016). Dynamical integration of language and behavior in a recurrent neural network for human–robot interaction. *Frontiers in neurorobotics*, *10*, 5
  - Chen, N., Bayer, J., Urban, S., & Van Der Smagt, P. (2015). Efficient movement representation by embedding dynamic movement primitives in deep autoencoders. In *IEEE-RAS* (pp. 434-440).
  - Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529.

References

- Abnormal behavior detection, prediction and prevention
  - Wei Wei, Jinjiu Li, Longbing Cao, Yuming Ou, Jiahang Chen, Effective Detection of Sophisticated Online Banking Fraud in Extremely Imbalanced Data, *World Wide Web Journal,* 1-27, 2012
  - Song, Y., Cao, L., Yin, J., & Wang, C. (2013). Extracting discriminative features for identifying abnormal sequences in one-class mode, *IJCNN2013,* pp. 1-8.
  - Cao, L., Ou, Y., Yu, P., & Wei, G. (2010). Detecting Abnormal Coupled Sequences and Sequence Changes in Group-based Manipulative Trading Behaviors, pp. 85-93

BEHAVIOR INFORMATICS
...Discovering Behavior Intelligence

Acknowledgement

- We'd like to acknowledge all authors whose work has been mentioned in this slide for your contributions, in particular
  - Can Wang, Yin Song, Yanchang Zhao, Huaifeng Zhao, Yuming Ou, Zhigang Zheng, Junfu Yin, Xiangjun Dong, Guansong Pang, Shoujin Wang, Trong Dinh Thac Do, Liang Hu, etc.
  - Other authors whose work has been included in this tutorial

Thank you very much

- References:
  - IJCAI’2013 Tutorial: Behavior Informatics
  - KDD’2018 Tutorial: Behavior Analytics
  - AAAI’2019 Tutorial: Behavior Analytics
- www.behaviorinformatics.org
- www.datasciences.org
- Email: Longbing.Cao@uts.edu.au