Fuzzy Patterns

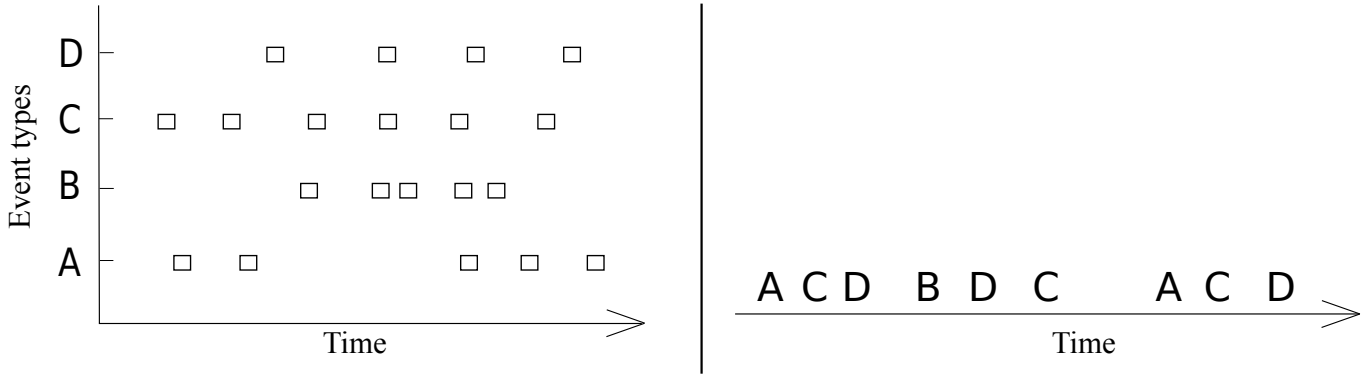
V. V. Vishnevskiy[1]

[1]Moscow State University

Plan

Basic notions

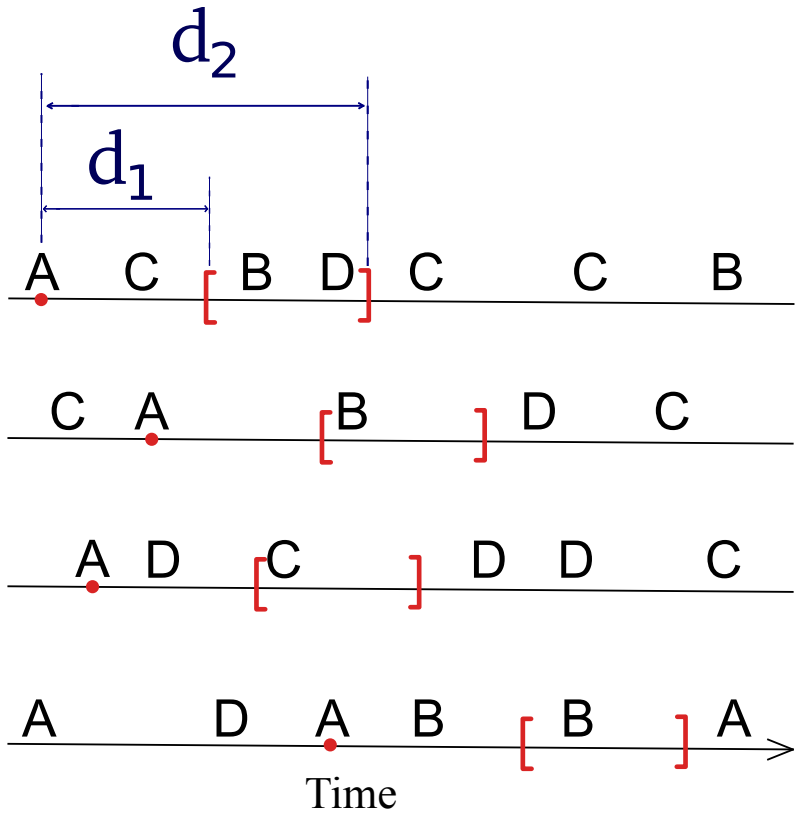
- Behavioral events(acts): $A, B, C, D \ldots$
- Each event occurs at certain time moments: $t_{A_1}, \ldots, t_{A_N}$.
- Pattern is a chain of events, that occur one after another quite often.



Plan

T-Patterns data types

- Events are joined with critical intervals. $A[d_1, d_2]B$.
- Critical interval relation means that event B occurs after event A in time span $[t_A + d_1, t_A + d_2]$, more often than usual.



T-Patterns detection procedure

Repeat while new patterns are detected:

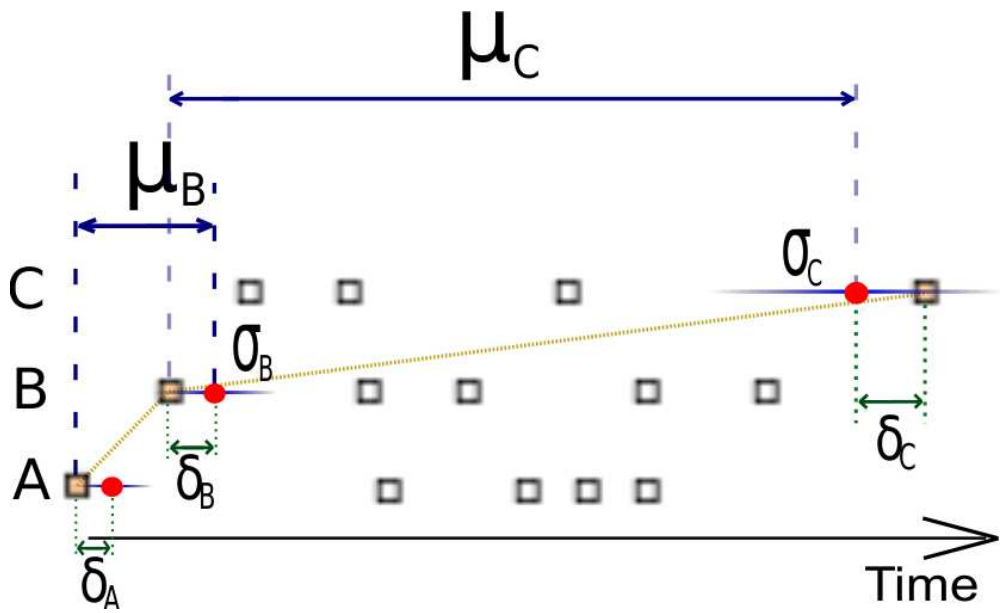
- For each two patterns try to join them with critical interval relation.
- Delete duplicate and incomplete patterns.

Proposed method

- Probabilistic pattern representation.
- Same iterative process.

Pattern representation

- Pattern consists of events.
- Each event in pattern characterized by mean shift and variance from previous event.
- $P = A[0, \sigma_A]B[\mu_B, \sigma_B]C[\mu_C, \sigma_C]$.



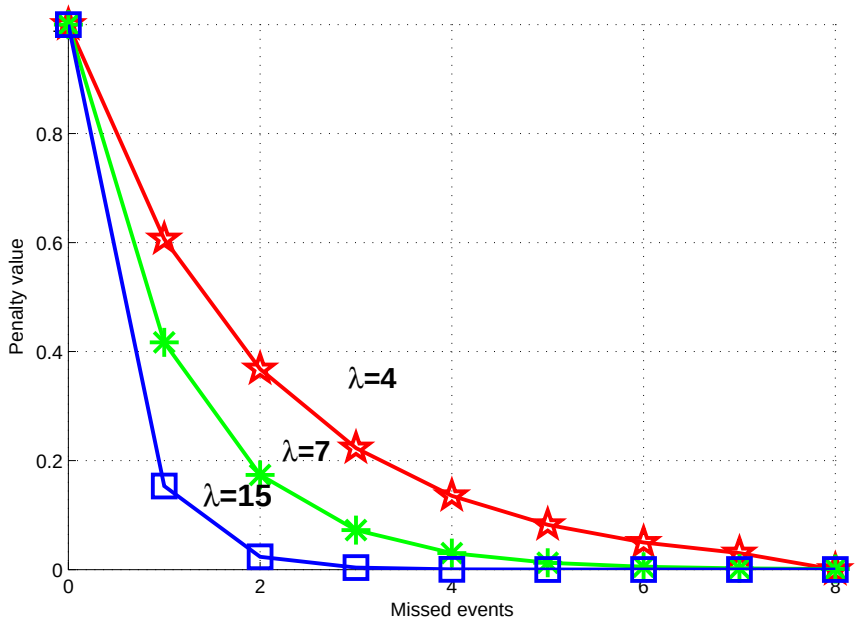
Plan

Loss function

- Penalty for missing x events in pattern of length N:

$$f_{LOSS}(x, N) = \begin{cases} \exp\left(-\frac{\lambda x}{N}\right), & x < N, \\ 0, & x = N. \end{cases}$$

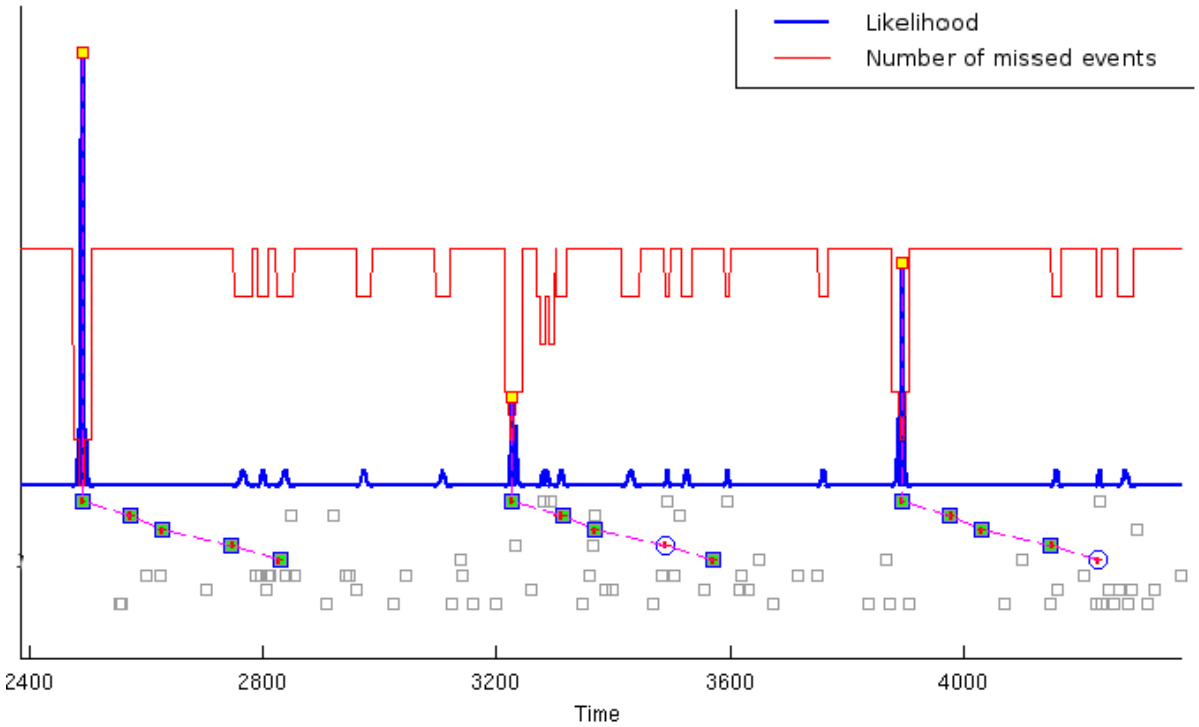
- λ defines level of pattern's fuzziness.



Likelihood

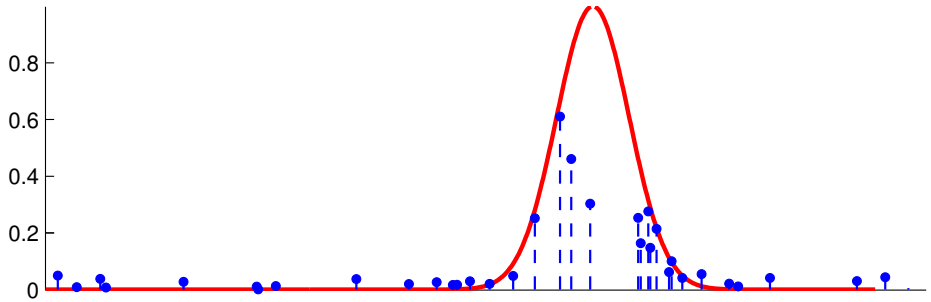
For every pattern P of length N, for every time moment $\varepsilon \in [0, N_t]$

$$L_P(\varepsilon) = f_{LOSS}(N_-, N) \prod_{i=1}^{N} \left(\frac{1}{\sqrt{2\pi}\sigma_i} \right) \prod_{i=1}^{N_+} \exp\left(-\frac{\delta_i^2}{2\sigma_i^2} \right)$$



Detecting co-occurrences

- Computing distribution of distances between two patterns.
- Searching μ and σ, that fits that distribution.
- Determine, if there is a relation between patterns.



Plan

Type of unnecessary patterns

- **Duplicates:** $(AB)(CD), ((AB)C)D$.
- **Incomplete copies:** BCD doesn't occur outside of $ABCD$.
- Similarity of patterns, using likelihood vector $\overrightarrow{L}$. Correlation coefficient:

$$cor(\overrightarrow{L_1}, \overrightarrow{L_2}) = \frac{\overrightarrow{L_1}\overrightarrow{L_2}^T}{\sqrt{\overrightarrow{L_1}\overrightarrow{L_1}^T}\sqrt{\overrightarrow{L_2}\overrightarrow{L_2}^T}}$$

Elimination of patterns

Consider patterns P_1 and P_2. If P_1 consists of all events, that are met in P_2 and $\exists m : cor(\overrightarrow{L_{P_1,1}}, \overrightarrow{L_{P_2,m}}) > \nu$, then P_1 is dropped.

Plan

Parameters of algorithm

Parameter	**Possible values**	**Default value**	**Has influence on**
ω	$[0, 1]$	0.995	Significance of pattern
μ	$[0, +\infty]$	3	Minimal pattern occurrences
μ	$[0, +\infty]$	6	Fuzziness of patterns
ν	$[0, 1]$	0.7	Similarity of patterns for elimination
M	$[0, N_t]$	None	Max time span between events in patterns

Experiments on real data

- Patterns found, using different methods:

T-Patterns	**T-Patterns found, using Fuzzy patterns**	**New fuzzy patterns**
21	18	3
87	84	21
96	80	13
58	57	21

- Histogram of pattern's lengths:

Pattern's length	2	3	4	5	6	7	8
T-Patterns	18	38	12	18	3	1	0
Fuzzy patterns	22	41	14	20	6	4	2

Conclusion

- Longer and more complex patterns are found.
- Statistical roots.
- Computational complex.