The Algorithm for Detection of Fuzzy Behavioral Patterns

Valera V. Vishnevskiy
valera.vishnevskiy@yandex.ru

Dmitry P. Vetrov
vetrovd@yandex.ru

Moscow State University,
Vorobyevy Gory, 1, 119992 Moscow, Russia

ABSTRACT
In this paper we present a new algorithm for the detection of fuzzy patterns in discrete time series. It generalizes the known approach by M. Magnusson to T-patterns detection. In contrast to the latter, our algorithm is able to find patterns where some elements can be absent in some occurrences of pattern. This makes possible to find soft stereotype in data which seems to be more natural in behavioral analysis.

Author Keywords
T-Patterns, behavior, fuzzy patterns, elementary behavioral acts.

INTRODUCTION
The problem of stereotypes detection in the behavior of humans and animals is extremely important in cognitive research since it allows to measure the complexity of behavior in quantitative terms, to monitor behavioral changes, etc. Here we focus on one possible approach to measuring behavior which is based on pattern detection. The behavior is represented as a sequence of events from a finite set of event types (e.g. the beginnings of behavioral acts) which occur at some moments of time. One or more events can occur at one moment of time. A pattern is a chain of events which occur one after another quite often. Such pattern allows to detect the repeated fragments of behavior.

A popular approach to pattern detection was proposed in [1] by Magnusson where the notion of T-pattern was established. The main drawback of this approach is the fact that patterns are assumed to be crisp, i.e. if at least one elements from the chain of events is missing the pattern is absent.

This complicates the process of searching the patterns in noisy chaotic data. It seems natural to assume that behavior is more complex than just a chain of events and depends on many factors that cannot be observed directly. In mathematics the traditional way to deal with unknown factors which cause influence on the studied process is to remove deterministic model with probabilistic one thus allowing the dependencies to be fuzzy and to contain the element of randomness. In the paper we establish probabilistic approach to pattern detection.

PROPOSED METHOD
General idea of our approach is based on the algorithm, proposed in [1]. It is iterative method, which consists of following repeated actions:

- Test every two patterns from the pattern set, whether there is significant co-occurrence among them (second pattern often occurs after the first one). If so, then these two patterns are joined together and added to the pattern set.
- Remove all duplicates and incomplete patterns from the pattern set.

That process goes on, until no more patterns are found. At first iteration the pattern set is a set of pseudo patterns, i.e. patterns of length 1.

Determining Data Types
Similarly to Magnusson's model, behavioral data is coded during an observation period$[1, N_t]$. At every time moment one or more events (behavioral acts) *can* take place. The set of time moments, when event E_i had appeared, we will denote as $ind(E_i)$. Formally speaking, we are searching for the temporal patterns in a discrete signal.

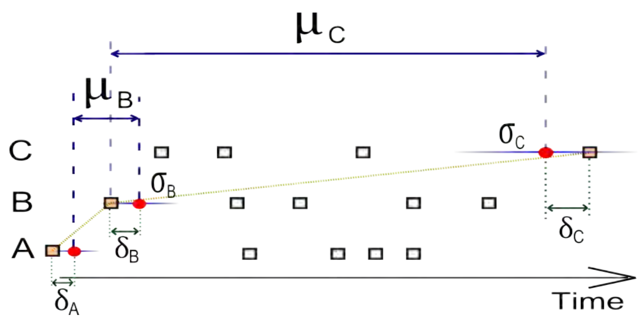

Figure 1. Example of pattern A[0;1]B[μ_B;σ_B]C[μ_C;σ_C].

To describe a pattern of length N, we use N pairs of parameters, which describe correspondent events: expected shift from previous event occurrence and its standard deviation (μ and σ). For the first event $\mu = 0$ and $\sigma = 1$. We will denote the pattern P, that consists of events $E_1, E_2, \dots, E_N$ in the following manner :

$$P = E_1[\mu_1, \sigma_1]E_2[\mu_2, \sigma_2] \dots E_N[\mu_N, \sigma_N].$$

For each pattern P of length N, for every time moment $\varepsilon \in [1, N_t]$, we compute the *likelihood function* L_P in the following way (see Figure 2):

$$L_P(\varepsilon) =$$

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

$$\delta_i = \min_{x \in ind(E_i)} \left|\varepsilon + \textstyle\sum_{j=1}^{i-1}[\mu_j + \delta_j] + \mu_i - x\right|,$$

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

$$N_- + N_+ = N,$$

where N_+ is the number of events, that occurred in the pattern at current time moment, N_- is the number of events, that are missing in pattern at current time moment, δ_i is distance between expected and observed position of i-th event occurrence(see Figure 1). Event is treated as missed, if $\exp\left(-\frac{\delta_i^2}{2\sigma_i^2}\right) < \exp\left(\frac{\lambda}{N}\right)$, i.e. when δ_i is too big. Here we assume that the position of the first event is fixed at ε.

We can control the level of pattern fuzziness, by changing λ: decreasing it, would allow more event gaps in pattern (see Figure 3).

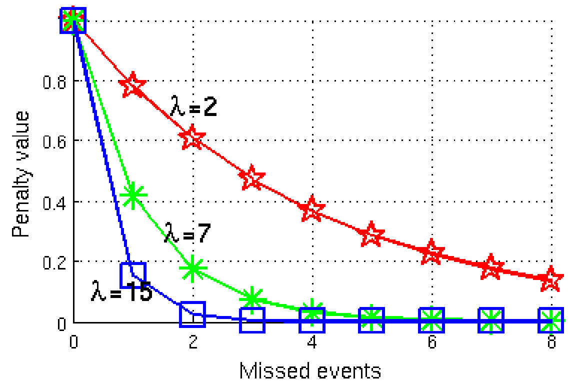

Figure 3. Loss function of pattern of length $N = 8$ and different λ.

The value of likelihood function at time moment ε can be interpreted as level of confidence that given pattern starts at that time moment. Finding significant maximums of the likelihood function, we can define the moments, when pattern begins.

Note, that we can compute the likelihood function fixing any event we want. For example, the likelihood of the pattern P, counted w.r.t. m-th event, we will denote as

$$L_{P,m}(\varepsilon) = L_P\left(\varepsilon + \textstyle\sum_{j=1}^{m} \mu_j\right).$$

Detecting Co-Occurrences

On that step we consider whether two patterns P_L and P_R should be merged to larger pattern. By computing the likelihood functions for P_L from the end and for P_R from the beginning, we find significant maximums of these likelihoods. Let $\{\alpha_i\}$, $\{\beta_j\}$ be the values of the likelihood

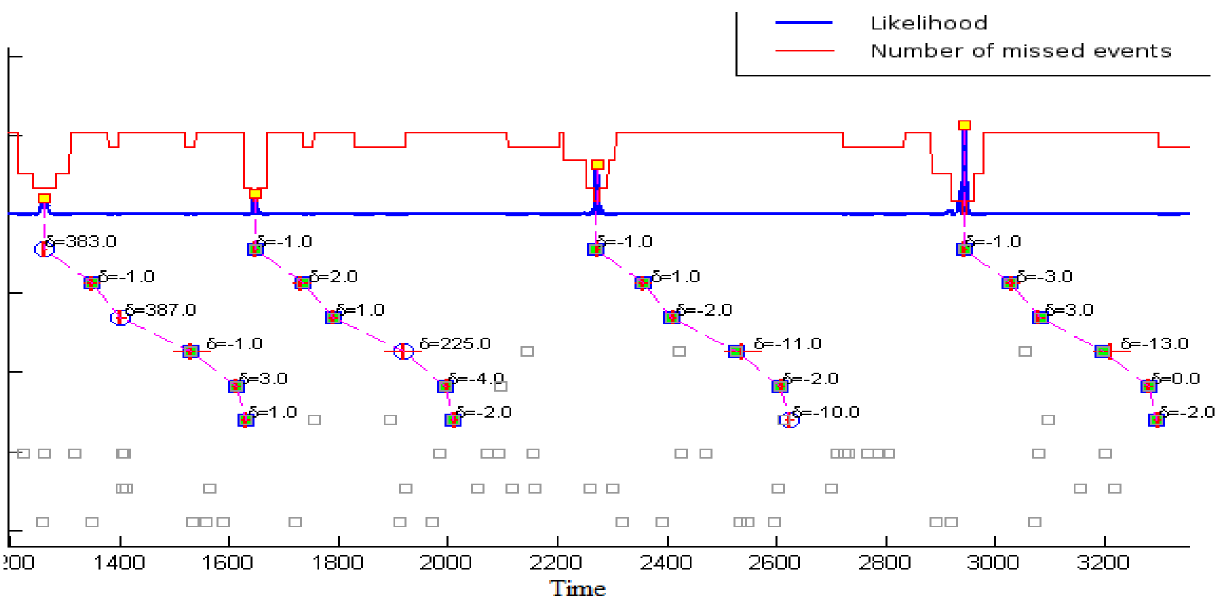

Figure 2. Example of likelihood for pattern
`A[0;1]B[84;3.5]C[54;2.0]D[120;3.2]E[82;2.0]F[16;2.3]`

maximums and $\{x_{L,i}\}$, $\{x_{R,j}\}$ indexes of these maximums for P_L and P_R respectively. We consider the distances between the occurrences of each pattern which are less than some predefined threshold M:

$$\rho = \{x_{R,j} - x_{L,i} | 0 \le x_{R,j} - x_{L,i} \le M\},$$

and used weights to increase the influence of those pattern occurrences that better correspond to the statistical model of pattern:

$$w_l = \log(1 + a_i \beta_j).$$

Next, we consider the following sum (see Figure 4):

$$k = \sum_{l=1}^{Q} w_l \, g_{\mu,\sigma}(\rho_l),$$

$$g_{\mu,\sigma}(\rho_l), = exp\left(-\frac{(\rho_l - \mu)^2}{2\sigma^2}\right),$$

$$S(\sigma) = \sigma\sqrt{2\pi},$$

$$Q = |\rho|.$$

Here $g_{\mu,\sigma}(\rho_l)$ is statistical model of co-occurrence. We are trying it on with different μ and σ, testing if there is a significant co-occurrence. Note, that because of computational complexity and the assumption that the distance between events in pattern should be small, only the co-occurrences that are shorter than M are considered.

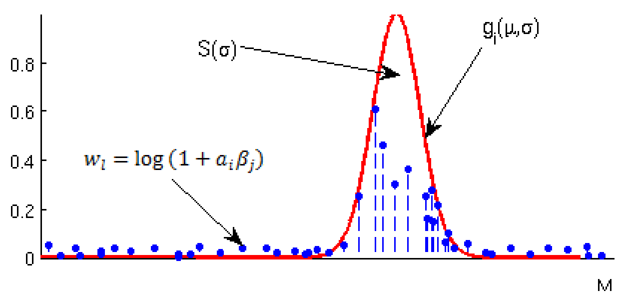

Figure 4. The distribution of distances between patterns and $g_i(\mu, \sigma)$ which maximizes the expression (1).

To test the significance of co-occurrences establish null-hypothesis that the two patterns are independent. Let

$$Y = \sum_{l=1}^{Q} X_l,$$

$$X_l = w_l g_{\mu,\sigma}(\rho_l) = w_l exp\left(-\frac{(\rho_l - \mu)^2}{2\sigma^2}\right).$$

Under null-hypothesis:

- w and ρ_l are independent random variables,
- ρ_l is uniformly distributed. $\rho_l \sim U[0, M]$.

Then it can be shown that

$$Y \sim \mathcal{N}\left(\frac{\sum_{i=1}^{Q} w_i}{M} S,\ \frac{1}{M^2}\left[MS\sqrt{2}\sum_{i=1}^{Q} w_i^2 - \frac{\left(\sum_{i=1}^{Q} w_i\right)^2}{Q} S^2\right]\right).$$

In order to perform test, we maximize the following value, using methods from [4]:

$$\frac{k - \mathbb{E}Y}{\sqrt{\mathbb{D}Y}} \to \max_{\mu,\sigma} \qquad (1)$$

If maximum value is greater, then the quantile of normal distribution with predefined significance level $\omega \in [0, 1]$, and the two patterns P_L and P_R co-occur sufficiently frequently, then we make decision that the statistically significant co-occurrence $P_L[\mu; \sigma]P_R$ takes place. And therefore, constructed pattern is added to pattern set. Parameter ω stands for significance of found patterns: the closer ω to 1, the more significant patterns are found.

While speaking “co-occur sufficiently frequently” we mean, that the sum of significant maximums of the likelihood function $L_{P_L[\mu;\sigma]P_R}(\varepsilon)$ is greater than η.

Removing Patterns

Similarly to Magnusson’s approach, our method can construct duplicate and incomplete patterns. That is why we need some mechanism to eliminate those patterns on each step.

Duplicated Patterns

The problem is that, one pattern can be constructed from different subpatterns. For example, pattern ABCD can be detected both by uniting (AB) and (CD), or (A) and (BCD). Generally they result to the same patterns, but due to complicated process of uniting, they could have slightly different likelihood functions.

Incomplete Patterns

While constructing patterns from subpatterns, it’s possible, that subpattern only appears as a part of constructed pattern. Therefore we don’t need to consider such subpatterns independently. For example, if in pattern ABCD, AB just doesn’t occur out of ABCD, then we don’t need pattern AB in pattern set, and likelihood functions of ABCD and AB should be very similar.

Considering described above examples, simple procedure of pattern elimination was proposed. First, let’s define the following values:

$$\overrightarrow{L_{P,\iota}} = \left(L_{P,i}(1), \ldots, L_{P,i}(N_t)\right) \text{ — vector } 1 \times N_t,$$

where $L_{P,i}(\varepsilon)$ — is the likelihood function of pattern P, at time moment ε, computed with respect to the i-th event. N_t is the length of the time period of the observation.

$$cor\left(\overrightarrow{L_1}, \overrightarrow{L_2}\right) = \frac{\overrightarrow{L_1}\,\overrightarrow{L_2}^T}{\sqrt{\overrightarrow{L_1}\,\overrightarrow{L_1}^T}\sqrt{\overrightarrow{L_2}\,\overrightarrow{L_2}^T}} \in [0,1],$$

is the correlation coefficient between likelihoods, therefore, the closer it is to 1, the more similar the likelihoods are.

Procedure of Elimination

We test every pair of non-pseudo patterns(P_L and P_R), from the pattern set, considering P_L as a duplicate or incomplete copy of P_R. Note that we don’t consider pseudo patterns(single event types), because they may be necessary for constructing new patterns.

The first test deals with duplicate patterns: if P_L and P_R consist of the same events, and $cor\left(\overrightarrow{L_{P_L,1}}, \overrightarrow{L_{P_R,m}}\right) > \nu$ (m is index of the first event of P_L in P_R), and $\left\|\overrightarrow{L_{P_L,1}}\right\| \leq \left\|\overrightarrow{L_{P_R,m}}\right\|$, then we remove P_L from the pattern set. The exclusion of incomplete copies is done in a similar manner.

In this section we introduced new parameter ν. In our experiments, $\nu = 0.7$ usually worked well.

Structural Parameters

The algorithm we derived in the paper has some parameters (see Table 1), that should be set manually. However, during experiments we discovered that the default values are often working well, or alternatively they can be set, according to prior information about considered behavioral time series, which would improve the performance leading to more interpretable patterns

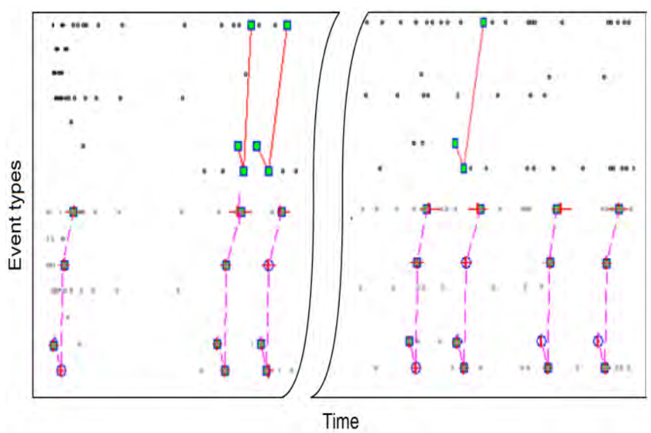

Figure 5. Comparison of the longest detected pattern (in the actual data), using T-Patterns (above) and fuzzy patterns (below). Gaps are illustrated as circles. Note that fuzzy patterns are longer and are observed more often than their crisp variants. That happens because of insufficient number of pattern occurrences without gaps.

Parameter	Possible values	Default value	Has influence on
ω	$[0, 1]$	0.95	Significance of patterns
ν	$[0, 1]$	0.7	How much similar patterns should be to be eliminated
M	$[0, N_t]$	None	Length of relations that connect patterns.
η	$[0, +\infty]$	3	Minimal pattern occurrences
λ	$[0, +\infty]$	6	Fuzziness of patterns

Table 1. Main algorithm parameters.

EXPERIMENTS AND COMPARISION WITH ANOTHER METHOD

To test implemented algorithm on real data, we used hamster behavioral data from open field test and recordings of grooming. Also we compared proposed algorithm with Magnusson's T-Pattern approach [1].

Each dataset was presented by set of pairs: event type (behavioral act) and time moment at which that event had started. Both grooming and open field data had, on average, 15-30 event types and each event type occurred 20-80 times. Every following event occurrence defined the end of previous event. Figure 5 contains an example of discovered fuzzy pattern and the closest T-pattern generated by Magnusson's algorithm on open field test data.

In general, the set of patterns found by our method contained[1] almost all patterns that were discovered using T-Patterns technique. At the same time, it didn't contain too much noisy patterns, which meant, that Fuzzy Patterns extended T-Patterns framework in a reasonable way. The typical example of the difference between two methods is shown in Figure 6. Moreover, fuzzy patterns that corresponded to some T-Pattern had greater likelihoods and fuzzy analogues of longest T-Patterns were always detected by our method.

In some cases we observed the situation when there were many fuzzy patterns that were fuzzy variations of the same T-Pattern. This effect could be eliminated by fine parameter tuning. Also the longest fuzzy pattern is the extension of the discovered T-Pattern, which seems reasonable.

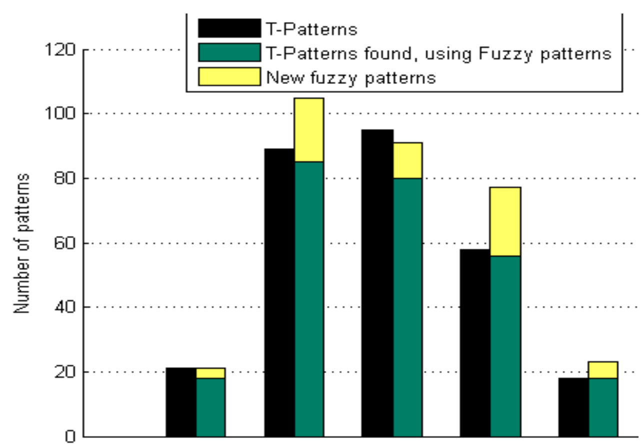

Figure 6. Number of patterns found using different methods on several real datasets. On the average, Fuzzy patterns method finds 93.6% of T-Patterns.

[1] In general we cannot say definitely that some fuzzy pattern corresponds to the specified T-pattern, because of different pattern representations.

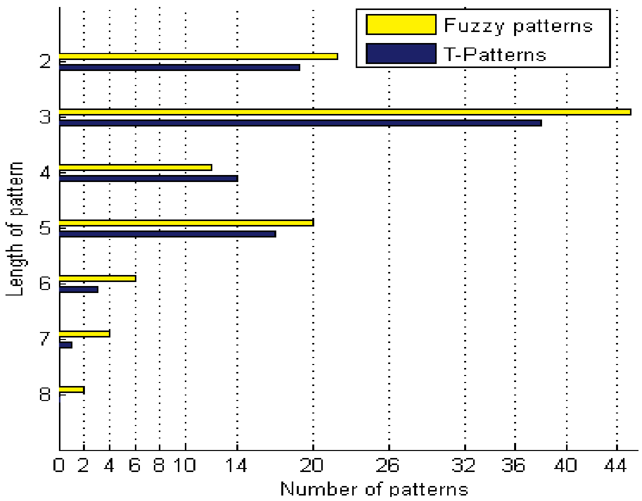

Figure 7. Histogram of pattern lengths.

CONCLUSION AND FUTURE WORK

Our proposed method for behavioral time patterns discovery, based on fuzzy pattern detection has shown promising results. It worked well on synthetic data (both when the distances between elements of pattern were generated from Gaussian and uniform distributions), and on actual data, detecting only those patterns, that were present in time series. The experiments show, that our algorithm is able to detect longer significant patterns in time series, than the algorithm based on T-Pattern detection. Figure 7 shows the distribution of the length of patterns found in grooming behavioral data.

Due to method's statistical roots, some patterns can be treated as noise. Also our method is computationally complex. The current version, implemented on MATLAB works approximately 100 times longer, than the algorithm, based on T-Pattern detection. One of the directions for future work is parallel implementation of the algorithm on multiprocessor computers or on Graphical Processing Units (GPUs).

We would like to thank Irina Zarayskaya for providing us behavioral data. The work was partly supported by Russian foundation for basic research (grant #08-01-00405), Russian President grant (MK3827.2010.9) and by Federal Target Program "Scientific and scientific-pedagogical personnel of innovative Russia in 2009-2013" (contact no. P1265).

REFERENCES

1. Magnusson, M.S. Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments, & Computers 23* (2000), 93-110.
2. Anolli, L., Duncan, S., Magnusson, M.S. Understanding Social Interaction: Discovering Hidden Structure with Model and Algorithms. In *The Hidden Structure of Interaction: From Neurons to Culture Patterns.* IOS Press (2005), 2-22.
3. Anolli, L. The Detection of the Hidden Design of Meaning. In *The Hidden Structure of Interaction: From Neurons to Culture Patterns*. IOS Press (2005), 23-50.
4. Boyd, S., Vandenberghe, L. *Convex Optimization*. Cambridge University Press (2004), 466-475.