

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

Паттерны и закономерности во временных рядах

В. В. Вишневский¹

¹МГУ, ВМиК, каф. ММП

План

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

① Введение
Где встречается задача

② Т-Паттерны
Качество поиска

③ Нечеткие паттерны
Правдоподобие паттерна
Статистический критерий
Удаление паттернов

④ Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной активности
Геном

План

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

Т-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

1 Введение

Где встречается задача

2 Т-Паттерны

Качество поиска

3 Нечеткие паттерны

Правдоподобие паттерна

Статистический критерий

Удаление паттернов

4 Другие алгоритмы

P-Patterns

Episodes

Мотивы в нейронной активности

Геном

Поиск закономерностей

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

Т-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

- Поведение – важна иерархия, точные временные интервалы,
- белковые структуры(биологические последовательности) – огромные объемы данных, биологическая информация о структуре,
- распространение сигналов по сетям(компьютерным и биологическим) – большие объемы данных, можно использовать кластеризацию,
- анализ потребительской корзины – не важна последовательность, *min_support*.

Данные

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача

T-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов

Другие алгоритмы

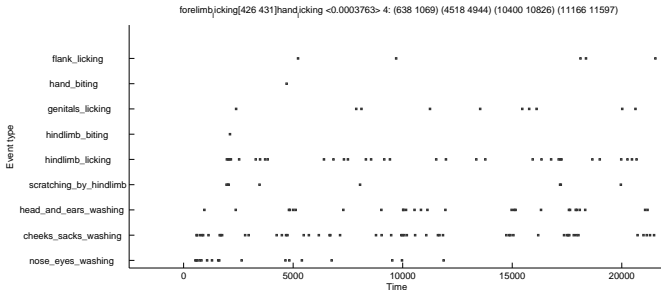
P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

- Элементарные события (event types): A, B, C, D ,
- у каждого события есть времена появления:
 t_{A_1}, \dots, t_{A_N}



Критический интервал

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

Т-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

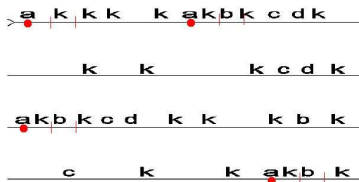
Episodes

Мотивы в нейронной
активности

Геном

События связываются в паттерны критическими интервалами.

Критический интервал – это связь между двумя паттернами, означающая, что второй паттерн встречается в некотором промежутке после первого чаще, чем ожидается.



$$\rho = P(\geq N_{AB}) = 1 - \sum_{i=0}^{N_{AB}-1} C_{N_A}^i (1 - P(\neg B)^d)^i P(\neg B)^{N_A-i}.$$

Процедура поиска

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

T-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

Пока добавляются новые паттерны:

- Для всевозможных пар паттернов, пытаемся найти связывающий их критический интервал,
- удаление неполных копий и паттернов-дубликатов.

План

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска

Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

① Введение
Где встречается задача

② Т-Паттерны
Качество поиска

③ Нечеткие паттерны
Правдоподобие паттерна
Статистический критерий
Удаление паттернов

④ Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной активности
Геном

Недостатки

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

T-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

- Реализован базовый алгоритм,
- возможно распараллеливание,
- алгоритм очень чувствителен к пропускам в паттернах.

Предпосылки

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

Т-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

- Еще раз: Т-Паттерны очень чувствительны к пропускам в данных,
- новый тип паттернов,
- схожий с Т-Паттернами метод поиска,
- правдоподобие паттерна в каждой точке.

Представление паттерна

Паттерны и
закономерности во
временных рядах

Вишневский

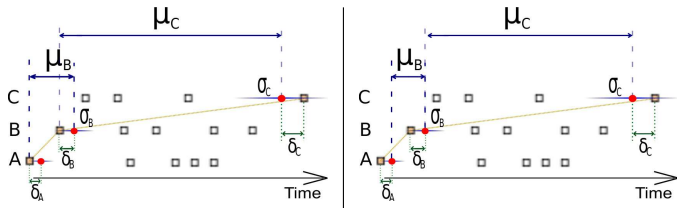
Введение
Где встречается задача
Т-Паттерны
Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов

Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

- Паттерн состоит из элементарных событий,
- каждое событие паттерна характеризуется смещением и разбросом от предыдущего события(гармошка),
- либо от предыдущего мат. ожидания(занавеска),
- $P = A[\mu_A, \sigma_A]B[\mu_B, \sigma_B]C[\mu_C, \sigma_C]$



План

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

① Введение
Где встречается задача

② Т-Паттерны
Качество поиска

③ Нечеткие паттерны
Правдоподобие паттерна
Статистический критерий
Удаление паттернов

④ Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной активности
Геном

Функция потерь

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

T-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes

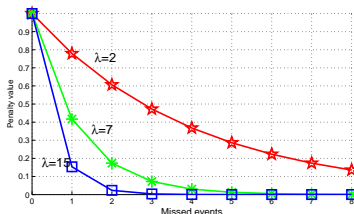
Мотивы в нейронной
активности

Геном

- $L_P(\epsilon) = \prod_{i=1}^{N_-} \left(\frac{1}{\sqrt{2\pi}\sigma_i} \right) f_{LOSS}(N_-, N) \prod_{i=1}^{N_+} \exp\left(-\frac{\delta_i^2}{2\sigma_i^2}\right),$
- Штраф за пропуск x событий из паттерна длины N :

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

- λ определяет уровень нечеткости паттернов.



Правдоподобие

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

T-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

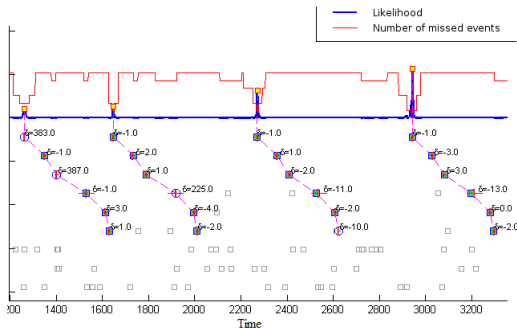
P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

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



Правдоподобие

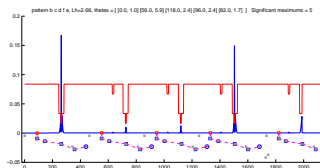
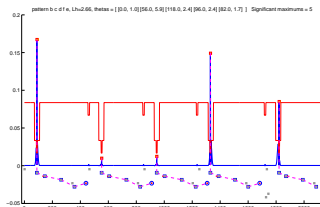
Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

Правдоподобие можно считать с конца, или начиная с i -го события.

$$L_{P,m} = L_P(\epsilon + \sum_{j=1}^m \mu_j)$$



План

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

- 1 Введение
Где встречается задача
- 2 Т-Паттерны
Качество поиска
- 3 Нечеткие паттерны
Правдоподобие паттерна
Статистический критерий
Удаление паттернов
- 4 Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной активности
Геном

Межточечное распределение

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

Т-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

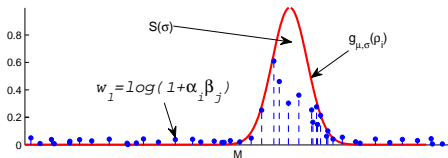
P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

- Рассматриваем распределение расстояний между концом левого и началом правого паттерна,
- отсечение окном ширины M ,
- вводим $g_{\mu,\sigma}(\rho_l) = \exp\left(-\frac{(\rho_l - \mu)^2}{2\sigma^2}\right)$ – это статистическая модель связи между событиями,
- подсчитываем $k = \sum_{l=1}^Q w_l g_{\mu,\sigma}(\rho_l)$.



Гипотеза о случайности распределения

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

- Предположив, что в данных нету закономерностей, следовательно:

- с.в. w и ρ_l – независимы,
- $\rho_l \in U[0, M]$,
- $Y = \sum_{l=1}^Q w_l g_{\mu, \sigma}(\rho_l)$,

- Используя Ц.П.Т, можно показать, что:

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

- ищем μ и σ :

$$\frac{k - \mathbb{E}Y}{\sqrt{\mathbb{D}Y}} \rightarrow \min_{\mu, \sigma},$$

- сравниваем с квантилью нормальмально распределения; ω .

План

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

- 1 Введение
Где встречается задача
- 2 Т-Паттерны
Качество поиска
- 3 Нечеткие паттерны
Правдоподобие паттерна
Статистический критерий
Удаление паттернов
- 4 Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной активности
Геном

Виды «лишних» паттернов

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

- **Дубли:** (AB)(CD), (ABC)D,
- **Неполные копии:** например, (BCD) не встречается вне (ABCD).
- похожесть паттернов по вектору правдоподобия \vec{L} , коэффициент корреляции:

$$\text{cor}(\vec{L}_1, \vec{L}_2) = \frac{\vec{L}_1 \vec{L}_2^T}{\sqrt{\vec{L}_1 \vec{L}_1^T} \sqrt{\vec{L}_2 \vec{L}_2^T}}$$

Процедура удаления

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

Т-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

Есть паттерн P_1 и P_2 . Если все события, которые
входят в P_1 так же входят в P_2 и
 $\exists m : cor(\overrightarrow{L_{P_1,m}}, \overrightarrow{L_{P_2,1}}) > \nu$, то P_1 удаляется.

План

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска

Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes
Мотивы в нейронной
активности
Геном

① Введение
Где встречается задача

② Т-Паттерны
Качество поиска

③ Нечеткие паттерны
Правдоподобие паттерна
Статистический критерий
Удаление паттернов

④ Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной активности
Геном

P-patterns

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

Т-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

Sheng Ma and Joseph L. Hellerstein. «Mining Partially Periodic Event Patterns With Unknown Periods». IBM T.J. Watson Research Center Hawthorne, NY 10532. [50902.pdf]

- Решается задача поиска паттернов в коммуникационных сетях,
- заданым окном ширины w ведется поиск ассоциативных правил,
- поиск периодов с помощью FFT и критерия χ^2 ,
- параметры:

δ	time tolerance,	predefined
w	time window	predefined
$minsup$	minimum support for p-pattern	predefined
p	period length	to be found
\mathbf{A}_1	subset of event types.	to be found

Особенности

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

Т-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

- Период всегда фиксирован,
- возможны две стратегии поиска: сначала периоды, потом ассоциации, либо сначала ассоциации, а потом периоды. Первый метод быстрее, второй более устойчив к шуму.

План

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

- 1 Введение
Где встречается задача
- 2 Т-Паттерны
Качество поиска
- 3 Нечеткие паттерны
Правдоподобие паттерна
Статистический критерий
Удаление паттернов
- 4 Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной активности
Геном

Episodes

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

Т-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

HEIKKI MANNILA HANNU TOIVONEN A. INKERI
VERKAMO. «Discovery of Frequent Episodes in Event
Sequences». Department of Computer Science, P.O. Box
26, FIN-00014 University of Helsinki, Finland

- Определяются 3 связи между событиями:



α



β



γ

Episodes α , β , and γ .

- Для каждого эпизода считается частота:

$$fr(\alpha, s, win) = \frac{|\{\mathbf{w} \in \mathcal{W}(s, win) \mid \alpha \text{ occurs in } \mathbf{w}\}|}{|\mathcal{W}(s, win)|}.$$

Паттерны и
закономерности во
временных рядах

Вишневский

Введение

Где встречается задача

Т-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

- Выбираются простейшие события,
- для каждой пары событий проверяют их связанность. (Confidence)

Особенности

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

- Нету временных интервалов. Проход окном;
- Определяется только структура,
последовательность связи в паттернах.

План

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

- ① Введение
Где встречается задача
- ② Т-Паттерны
Качество поиска
- ③ Нечеткие паттерны
Правдоподобие паттерна
Статистический критерий
Удаление паттернов
- ④ Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной активности
Геном

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

«Identifying repeating motifs in the activation of synchronized bursts in cultured neuronal networks».
Nadav Raichman, Eshel Ben-Jacob. Journal of Neuroscience Methods 170 (2008)96-110

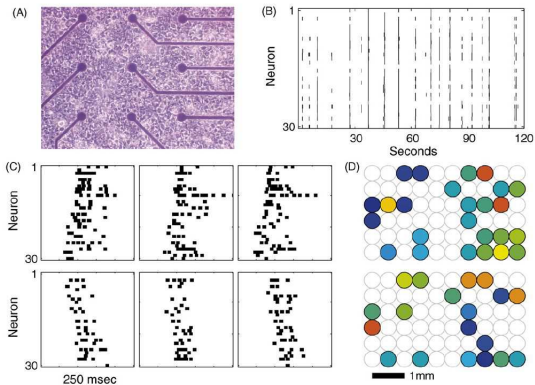
- В чашку Петри ставят датчики, каждый из которых может регистрировать активность нескольких нейронов;
- спайки объединяются по времени в пачки;
- каждая пачка представляет распространение активности по нейронам;
- метрика на множестве пачек;
- пачки кластеризуются. Каждый кластер определяет паттерн.

Визуализация

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

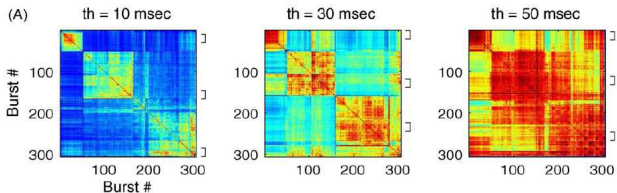


Дендограмма

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном



План

Паттерны и
закономерности во
временных рядах

Вишневский

Введение
Где встречается задача
Т-Паттерны
Качество поиска
Нечеткие паттерны
Правдоподобие
паттерна
Статистический
критерий
Удаление паттернов
Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной
активности
Геном

- 1 Введение
Где встречается задача
- 2 Т-Паттерны
Качество поиска
- 3 Нечеткие паттерны
Правдоподобие паттерна
Статистический критерий
Удаление паттернов
- 4 Другие алгоритмы
P-Patterns
Episodes
Мотивы в нейронной активности
Геном

Паттерны и закономерности во временных рядах

Вишневский

Введение

Где встречается задача

Т-Паттерны

Качество поиска

Нечеткие паттерны

Правдоподобие
паттерна

Статистический
критерий

Удаление паттернов

Другие алгоритмы

P-Patterns

Episodes

Мотивы в нейронной
активности

Геном

/*Добавить. FFT очень большие данные, не подходит
для поведения.*/*