

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ
КАФЕДРА МАТЕМАТИЧЕСКИХ МЕТОДОВ ПРОГНОЗИРОВАНИЯ

Отчет по дипломной работе

Параллельная реализация методов поиска скрытых
поведенческих закономерностей.

Вишневский Валерий
группа 517
научный руководитель: Ветров Дмитрий Петрович

Москва, Декабрь 2010

1 Введение

В предыдущих работах был программно реализован метод поиска скрытых закономерностей в поведенческих данных, основанный на алгоритме, предложенном М.С. Магнусоном в [1]. Данный метод хорошо зарекомендовал себя, однако имел ряд очевидных недостатков. Попытка устранить главный недостаток — чувствительность к шуму в исходных, была произведена нами введением вероятностной модели на паттерн [2].

Стоит отметить, что метод Магнусона поиска четких закономерностей достаточно вычислительно сложен. Вероятностный подход, эксплуатируемый нами в методе поиска нечетких паттернов требует еще большего количества вычислений, что и мотивирует нас на разработку параллельной версии предложенного метода.

1.1 Поиск четких паттернов

Дана разметка поведения животного во времени:

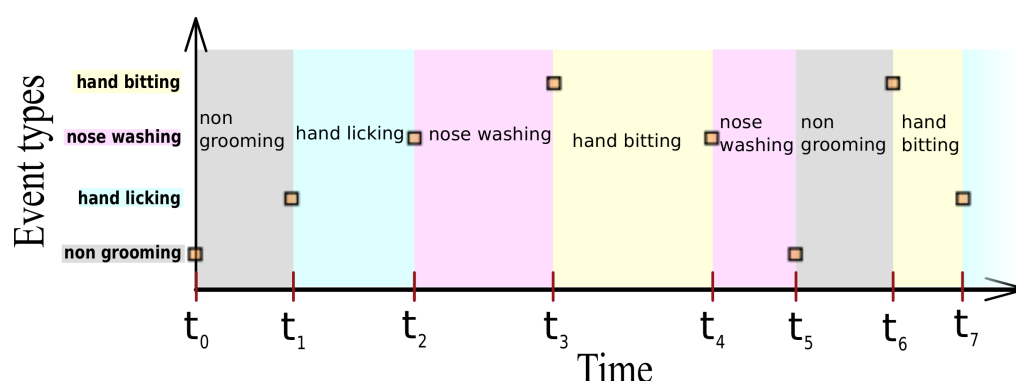


Рис. 1: Поведенческие данные.

Паттерн — это последовательность событий(поведенческих актов), повторяющихся один за другим достаточно часто. Для поиска паттернов используется следующая процедура:

Инициализируем множество паттернов псевдопаттернами(поведенческие акты). Потом итеративно повторяем:

- **Конструирование:** Для всех пар паттернов проверить, повторяется ли один за другим достаточно часто. Если да, то получаем новый паттерн.
- **Зачистка:** Удалить одинаковые паттерны, которые были сконструированы по-разному.

Ключевым моментом является этап конструирования паттернов. Здесь вводится понятие критического интервала $(A[d_1, d_2]B)$ — это связь между двумя паттернами, означающая, что второй паттерн(B) промежутке $[d_1, d_2]$ после появления первого паттерна(A) чаще, чем ожидается. Сразу отметим, что в данном определении дискретность вводится требованием события попасть в интервал $[d_1, d_2]$.

1.2 Поиск нечетких паттернов

В этом методе паттерн определяется как последовательность элементарных событий, где каждое событие паттерна характеризуется смещением и разбросом от предыдущего события(гармошка), либо от предыдущего мат. ожидания(занавеска),

$$P = A[\mu_A, \sigma_A]B[\mu_B, \sigma_B]C[\mu_C, \sigma_C]$$

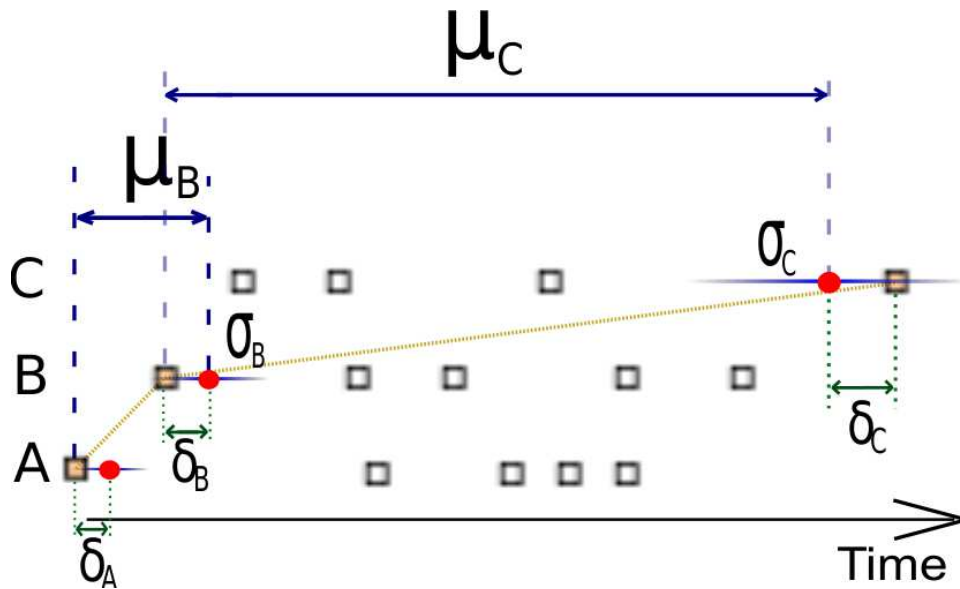


Рис. 2: Представление нечеткого паттерна.

Более того, для каждого паттерна длины N мы вводим штраф за пропуск x событий:

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

здесь λ определяет уровень нечеткости паттернов.

Сама процедура конструирования паттернов основанная на анализе распределения межточечных расстояний описана в [2].

2 Ожидаемые результаты

Будучи, по сути, переборными, оба метода поиска и четких и, тем более, нечетких паттернов вычислительно сложны. Однако в нашем случае основная часть вычислений(подсчет уровней значимости межсобытийных связей) может быть произведена параллельно, используя лишь часть исходных данных наблюдений. Именно такие задачи, как показывает практика(например, [3, 4]), особенно хорошо подходят для реализации на вычислительной архитектуре GPU(увелечение производительности на порядки: от 10 до 100 раз, в зависимости от задачи и реализации).

Такое увеличение производительности позволит запускать поиск закономерностей на более длинных массивах данных, учитывать большее число событий, вести более глубокий поиск взаимосвязей между событиями.

3 Прodelанная работа

Для реализации нашего метода на GPU была выбрана платформа CUDA(Fermi) от компании NVIDIA. Пока ведется работа по распараллеливанию метода поиска четких паттернов. На текущей версии реализован базовый метод определения критических интервалов в данных. Пока что наблюдается прирост производительности на один порядок(7-8 раз). Однако стоит заметить, что из-за ограничений платформы CUDA по работе с памятью, еще не реализована возможность обработки реальных длинных поведенческих закономерностей, на которых уже можно более объективно оценить прирост производительности параллельной версии метода.

Список литературы

- [1] M.S. Magnusson. Discovering hidden time patterns in behavior: T-patterns and their detection. — Behavior Research Methods, Instruments, Computers 2000, 32 (I), 93-110.
- [2] V.V. Vishnevskiy, D.P. Vetrov. The Algorithm for Detection of Fuzzy Behavioral Patterns —Proceedings of Measuring Behavior 2010,ISBN 978-90-74821-86-5
- [3] J.D. Hall, J.C. Hart. GPU Acceleration of Iterative Clustering. — University of Illinois at Urbana-Champaign, June 4, 2004.
- [4] J. Kruger, R. Westermann. Linear Algebra Operators for GPU Implementation of Numerical Algorithms. — Computer Graphics and Visualization Group, Technical University Munich