

УДК 519.85

**Использование теории графов для выявления регулярных последовательностей во
временном ряду дискретного пространства**

***Смирнов В.И.**, бакалавр*

*Россия, 105005, г. Москва, МГТУ им. Н. Э. Баумана
кафедра «Программное обеспечение ЭВМ и информационные технологии»*

*Научный руководитель: Тассов К.Л., старший преподаватель кафедры
«Программное обеспечение ЭВМ и информационные технологии»*

Россия, 105005, г. Москва, МГТУ им. Н. Э. Баумана

bauman@bmstu.ru

Введение

Рассмотрим некоторый объект, который совершает передвижения в n -мерном пространстве. Нет возможности осуществлять слежение за объектом непрерывно. Слежение производится через отдельные контрольные точки пространства, задающиеся постоянными координатами. Фиксируется время "прохождения" объектом контрольных точек. Поведение объекта между контрольными точками не определено. Необходимо выявить возможные регулярные, т.е. повторяющиеся через определенные промежутки времени, последовательности "прохождения" объектом контрольных точек.

В качестве объекта может выступать, например, человек, совершающий в выходной день прогулку и навещающих своих многочисленных друзей, чьи дома разбросаны по городу. Движение человека может рассматриваться как движение по двумерному пространству – карте города. В качестве контрольных точек, в таком случае, выступают дома друзей. В конце дня можно с определённой точностью восстановить маршрут человека, обзвонив его друзей и узнав время каждого из визитов. Обладая историей походов человека за несколько дней можно ответить на следующие вопросы: возможно ли, что человек движется каждый раз по одному и тому же маршруту или нет?

Математическое описание модели

Сформулируем описание задачи в математическом виде. Дан объект O , характеризующийся n параметрами. Объект осуществляет движение по n -мерному пространству, называемому пространством состояний:

$$S: \forall t \exists s_t \in S, \text{ где } s_t = (s_{t_1}^1, s_{t_2}^2, \dots, s_{t_n}^n)$$

– набор параметров, характеризующих объект O в момент времени t .

В рассмотренном примере карта выступает в качестве двумерного пространства состояний, а набор параметров включает значения долготы и широты.

$$\exists c_i \in C: C \subset S$$

– множество контрольных точек.

$$H_O^m = (c_{t_1}^1, c_{t_2}^2, \dots, c_{t_m}^m), \forall j \in [1, m]: c_{t_j}^j \in C$$

– известна история прохождения объектом O контрольных точек C , являющаяся цепочкой из m элементов – точек пространства состояний.

Несмотря на то, что само пространство состояний S является непрерывным, доступные данные ограничиваются дискретным подпространством, состоящим из контрольных пунктов.

Контрольные пункты могут быть расположены в пространстве как угодно, а время между двумя последовательными "прохождениями", в общем случае, не является константой.

Отклонения

При наличии чёткого выраженного регулярного маршрута объект осуществляет движение не "точно", т.е. возможны отклонения. Под отклонением понимается один из двух факторов или их совокупность:

1. Пространственный: в рассматриваемый момент времени объект зафиксирован в состоянии, которое отличается от предполагаемого из регулярного маршрута:

$$\exists i: c_{t_i}^{\text{ожидание}} \neq c_{t_i}^{\text{реальность}}$$

2. Временной: объект зафиксирован в состоянии, предполагаемом из регулярного маршрута, но факт фиксации произошёл в момент времени, отличный от предполагаемого:

$$\begin{aligned} \exists i: H_{\text{ожидание}}^m &= (\dots, c_{t_{i-1}}^{i-1}, c_a^i, c_{t_{i+1}}^{i+1}, \dots), \\ H_{\text{реальность}}^m &= (\dots, c_{t_{i-1}}^{i-1}, c_b^i, c_{t_{i+1}}^{i+1}, \dots), a \neq b \subset T. \end{aligned}$$

Методы прогнозирования временных рядов

Для анализа временного ряда могут применяться методы прогнозирования. Выявляя зависимости между элементами ряда, методы дают возможность определить, основываясь на k -предыдущих элементах, вероятное значение следующего элемента ряда.

Существуют общепризнанные и часто используемые способы прогнозирования временных рядов, такие как модели Бокса-Дженкинса (ARMA, ARIMA), метод анализа сингулярного спектра и т.д. Данные методы в оригинале не рассчитаны на работу с

дискретным n -мерным пространством, но существуют модификации, расширяющие их возможности в этом направлении.

Серьёзным недостатком является вид входных данных: методы предназначены для равноотстоящих временных рядов. Интерполяция данных, которая могла бы привести к необходимому виду, значительно снижает точность метода.

Теория графов

Пример с человеком и картой красиво ложится на теорию графов. Представить модель в терминах данного раздела возможно следующим образом.

Дан ориентированный граф $G = (V, E)$, где

V – множество вершин:

$$\forall v_i \in V: \exists c_j \in C \rightarrow v_i \leftrightarrow c_j$$

, т.е. вершины взаимнооднозначно сопоставлены с контрольными точками.

$E \subset (V \times V)$ – множество дуг – упорядоченных пар вершин.

$$\forall e_i = (v_j, v_{j+1}) \exists k: c_{t_k}^k \leftrightarrow v_j \wedge c_{t_{k+1}}^{k+1} \leftrightarrow v_{j+1},$$

т.е. две вершины соединены дугой, если в истории есть подряд идущие фиксации соответствующих состояний.

Дуги обладают двумя параметрами:

1. Вес – функция от количества проходов по дуге:

$$w = f_w(count).$$

2. Длина – функция от времени, которое прошло между прохождением соответствующими контрольными точками:

$$l = f_l(dt_1, dt_2, \dots, dt_n).$$

Обладая достаточной историей, можно задать граф с такими значениями параметров дуг, которые бы для заданной вершины давали вероятность того, в каком состоянии и через какое время объект будет зафиксирован в следующий раз, в случае, если объект движется по регулярному маршруту.

Условие определения следующего состояния путём рассмотрения исключительно текущего состояния и игнорирования предыдущих в явном виде приводит к Марковскому процессу, а учитывая дискретность пространства – к цепям Маркова.

Пути модификации метода

Яркий недостаток данного подхода состоит в отсутствии учёта направления движения объекта.

Рассмотрим случай из примера, когда прогуливающийся человек проходит за день всех друзей не по кругу, но несколько (как минимум трёх) друзей посещает сначала в одном направлении, а затем в обратном:

$$\exists i, j: i + 2 < j, c_{t_{i-1}}^{i-1} = c_{t_{j+1}}^{j+1}, c_{t_i}^i = c_{t_j}^j, c_{t_{i+1}}^{i+1} = c_{t_{j-1}}^{j-1} \subset H_O^m.$$

В этом случае при наличии выборки достаточной величины на центральном контрольном пункте будет получено два варианта развития событий с почти одинаковыми вероятностями, несмотря на то, что сам объект может никогда не останавливаться и менять направление на противоположное:

$$\neg \exists i: c_{t_i}^i = c_{t_{i+2}}^{i+2} \subset H_O^m.$$

Данный недостаток может быть устранён следующей модификацией: рассматривается не только текущее состояние, но и ряд предыдущих. Для описанного выше примера достаточно заглядывать ровно на один шаг назад. Однако, при использовании прошлого процесс перестаёт быть марковым, хотя математический аппарат цепей Маркова может оказаться полезным.

Рассмотрим, каким образом модификация может быть реализована для теории графов и для Марковских цепей.

Первый вариант – для графов: введение новых n дополнительных типов дуг:

$$E^n \subset (V \times V): \forall e_i^n = (v_j, v_{j+1}) \exists k: c_{t_k}^k \leftrightarrow v_j \wedge c_{t_{k+n}}^{k+n} \leftrightarrow v_{j+1},$$

соединяющих вершины, которые соответствуют контрольным точкам, которые были зафиксированы не подряд, но через ровно $n - 1$ состояние. В таком случае, вероятность обнаружения в следующем состоянии является функцией от вероятности обнаружения в следующем состоянии для текущего состояния, в после-следующем состоянии для предыдущего состояния и так далее:

$$p(c_i, c_j) = f(w_{e_i^1}, w_{e_{i-1}^2}, \dots, w_{e_{i-n}^n}).$$

Достоинством такого подхода является то, что он автоматически учитывает пропуски фиксаций, позволяя снизить их влияние на точность. Недостаток: необходимость обращения не только к рассматриваемой вершине, но и к истории, откуда берутся предыдущие вершины, а также к этим вершинам полученным из истории.

Второй вариант: сворачивания n -вершин истории в одну вершину v^n :

$$v_i^n = (c_{t_i}^i, c_{t_{i-1}}^{i-1}, \dots, c_{t_{i-n}}^{i-n}),$$

т.е. каждая вершина соответствует не одному состоянию, а n -состояний, зафиксированных подряд в какой-то момент истории. В таком случае, вместо каждой вершины из первого случая появляется m -вершин, где m – число различных путей которыми мы попали в данную вершину.

Данный способ менее нагляден при графическом представлении, т.к. теряется взаимно-однозначное отображение вершин и контрольных точек. Информация внутри вершин обладает избыточностью, но позволяет не заглядывать в историю и анализировать предыдущие вершины, обеспечивая доступ к математическому аппарату Марковских цепей.

Таким образом, рассмотренные пути модификации могут применяться для разных случаев: первый подходит для данных, в которых имеются пропуски и помехи, в то время как второй путь обладает более проработанным математическим аппаратом.

Заключение

В рамках данной статьи была рассмотрена задача выявления регулярных последовательностей во временном ряду дискретного пространства. Приведена формулировка задачи и наглядный пример. Были указаны существующие методы. Описана математическая модель, приведён метод выявления регулярных последовательностей во временном ряду с использованием теории графов и цепей Маркова. Рассмотрены модификации метода, устраняющие недостатки оригинала

Список литературы

1. Martin L. Markov Decision Processes: Discrete Stochastic Dynamic Programming. NY: Puterman John Wiley & Sons, 2014. 162 p.
2. Дурбин Р., Эдди Ш., Крог А., Митчисон Г. Анализ биологических последовательностей. Ижевск: Научно-издательский центр «Регулярная и хаотическая динамика», 2006. 480 с.
3. Винер Н. Нелинейные задачи в теории случайных процессов. М.: Рипол Классик, 2013. 164 с.
4. Смирнов В.И. Методы выделения повседневных маршрутов с использованием средств автоматического распознавания ГРЗ // Молодежный научно-технический вестник. Электрон. журн. 2014. № 12. Режим доступа: <http://sntbul.bmstu.ru/doc/746950.html> (дата обращения 31.03.2015).