

УПРАВЛЕНИЕ ДИНАМИКОЙ СЛОЖНОЙ СИСТЕМЫ

Кузнецов М. В., Герман О. В.

Кафедра информационных технологий автоматизированных систем, Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

E-mail: {kuznetsov, german}mishaky@mail.ru

Рассматривается общий подход к построению управления динамикой сложной системы. Особенность подхода состоит в том, что он строится практически исключительно на использовании вложенной структуры кластеров, так что управление реализуется за счет погружения от одного уровня кластеров к другому с принятием решений в пределах каждого уровня кластеров.

Введение

Предлагается концепция иерархии кластеров для реализации управления сложной системой, например, финансовой, производственной, транспортной и т.п. Для целей описываемого подхода существенное значение имеет таблица наблюдений за определенный период времени о функционировании системы. На основании этой таблицы строится система вложенных кластеров, с помощью которых и реализуемой техники выбора кластеров решается задача выбора управляющего воздействия для наблюдаемого состояния системы. При этом существенными механизмами являются выбор кластера на основании используемой метрики сходства и оценка близости на основании функции минимального штрафа байесовской классификации. Представленный материал составляет некую общую методологическую основу для реализации управления сложной системой на основании концепции, известной как управление на основе прецедентов.

I. Описание подхода

Отправной позицией нашего подхода является таблица наблюдений за динамикой сложной системы (например, финансовой, производственной, транспортной и т.п.) Система описывается вектором параметров $V(t)$, состояния векторов изменяются в результате реализуемых управлений $U(V, t)$, так что поведение системы описывается достаточно сложной траекторией. Мы принимаем (хотя это не является ограничением), что всегда принимается одно какое-нибудь управление. Например, при управлении рисками примерами управляющих воздействий могут быть продажа активов, инвестирование в производство и технологии, сокращение персонала, техническое перевооружение, сокращение объемов производства и др. [1]. Для учета динамики нам можно воспользоваться наблюдениями по разным фирмам за достаточно длительный период времени. Такие данные можно описывать многомерными массивами $V[t, n, Z_t]$, где $t = 0, 1, 2, \dots, T$ задает моменты наблюдения (дискретные величины),

Z_t определяет значения наблюдаемых критериев [2].

Имея таблицу наблюдений, мы можем выделить кластера – совокупности близких значений векторов $V(t)$ [3, 4]. Число кластеров задаем равным 4. Эти кластера будут иметь такой смысл: (1) нормальное функционирование (штатный) режим; (2) режим близкий к нормальному; (3) аварийный режим (катастрофа) и (4) опасный режим – дрейф в сторону катастрофы. Разумеется, таблица с исходными данными должна содержать достаточное число представителей всех этих режимов. В минимальном варианте число кластеров можно ограничить тремя, исключив кластер, соответствующий аварийному режиму, поскольку этот режим на практике возникает в редких случаях (например, предприятие становится банкротом в единичных случаях). Разбиение на заданное число кластеров – это хорошо разработанная задача анализа данных (data mining). Можно отметить например метод k -средних для ее решения. Решается следующая задача. Наблюдается текущее значение вектора параметров $V_i(t)$. Нужно определить, какое управление реализовать для этого значения $V_i(t)$. Эта задача решается в несколько этапов. На первом этапе определяется, к какому кластеру отнести $V_i(t)$. Разумеется, можно исходить из того, что каждый кластер описывается как множество реализаций многомерной случайной величины. Тогда можно говорить о задаче отыскания вероятности принадлежности к кластеру. Однако закон распределения многомерной случайной величины а priori не известен. Кроме того, параметры могут быть коррелированы, накопленных статистических данных недостаточно, а значения параметров могут быть зашумлены. Очевидно, следует найти другую рациональную оценку. Будем использовать функцию штрафа за неправильную классификацию [5] в виде

$$F_i = \sum C_{ik} \cdot P_k \cdot P(V|k).$$

Здесь F_i – это штраф за отнесение объекта V к кластеру i ; P_k – вероятность выбора кластера k ; $P(V|k)$ – условная вероятность появления объекта V в кластере k . На практике полагают, что $C_{ik} = 1$, если $i \neq k$, и $C_{ik} = 0$ в против-

ном случае. Вероятности P_k заменяют частотами появления кластеров. Поскольку мы определяем кластера, то их частоты пропорциональны числу членов в этих кластерах (т.е. мощности кластеров). Иначе обстоит дело с $P(V/k)$. Вместо $P(V/k)$ будем использовать величины, заменяющие их. А именно: полагаем, что вероятность принадлежности к кластеру обратно пропорциональна расстоянию до него. Расстояние от объекта до кластера можно рассчитать, например, на основе формулы Евклида или Махаланобиса [6], или какой-нибудь иной известной формулы. Таким образом мы в состоянии оценить штрафы за неправильную классификацию. Далее мы относим наблюдаемый вектор $V_i(t)$ в тот кластер, для которого вычисленная оценка штрафа минимальна. Таким образом первая задача решена. Вторая задача заключается в выборе управления в границах определенного кластера. Разумеется, в рамках одного и того же кластера могли быть реализованы различные управления и даже одно и то же по сути управление могло быть реализовано различным способом. Поэтому опять сталкиваемся с задачей принятия решения [7]. Сначала следует определить, какое управление реализовать. Теперь мы должны в рамках нашего кластера задать подкластеры векторов $V_i(t)$ по различным управлениям. Теперь это делается автоматически, поскольку каждый “подкластер” однозначно определяется тем управлением, которое в нем реализовано. Наконец, мы выходим на последнюю задачу: как реализовать выбранное управление. Здесь речь идет о фазификации нечеткой лингвистической переменной управления U_i [8]. Например, можно говорить о необходимости дополнительных инвестиций. Однако объем инвестиций не определен. Кроме того, различные резервы ограничивают размер инвестиций. Поэтому можно вести речь о незначительных инвестициях, средних инвестициях и крупных инвестициях. Нам нужно фазифицировать лингвистические значения, например, полагая, что незначительные инвестиции определены в диапазоне нечетких значений $[0.1 - 0.4]$, средние по размеру инвестиции определены в диапазоне $(0.4 - 0.66]$ и значительные – в диапазоне $(0.66 - 1.0)$. Исходя из этого, опять получаем разбиение на подкластеры и снова решаем задачу классификации. Очевидно, мы сохраняем ту же идею, что и выше – а именно: использование байесовской оценки штрафа за неправильную классификацию. Строго говоря, нужно уточнить значение нечеткой переменной управления после выбора подходящего кластера либо использовать сред-

нее интервальное значение. Например, мы выбрали подкластер $[0.1 - 0.4]$. Среднее интервальное значение в нем равно 0.25. Наконец, остается выполнить дефазификацию найденного нечеткого значения. Это в свою очередь зависит от определения функции принадлежности.

II. ВЫВОДЫ

Особенность описанного подхода состоит в возможности его полной автоматизации, за исключением сбора и подготовки таблиц с исходными данными. Составление таких таблиц следует выполнять за достаточно продолжительный период, чтобы обеспечить объективность формируемых решений. Разумеется, интересным является вопрос о работе с недостоверными данными наблюдений и даже необходимости восстановления отсутствующих данных. Этот вопрос является предметом обсуждения [9]. Для целей автоматизации описанного подхода весьма перспективным выглядит использование языка Python. Настоящая работа лишь указывает направление исследований, оставляя в качестве выбора как конкретные методологии, так и программные и математические средства. Поскольку исходные данные играют весьма существенную роль для качественного процесса управления системой, то вопросы подготовки данных (хотя и остались за пределами данной статьи) играют весьма существенную роль в реализации изложенного подхода.

1. Hopkin, P. Fundamentals of risk management: understanding, evaluating and implementing effective risk management / P. Hopkin// –Report of the Institute of Risk Management, England.2010. –352p.
2. Han, J. Data Mining. Concepts and techniques / J. Han, M. Camber, J. Pei// Elsevier, 2012. –740p.
3. Abonyi, J. Cluster analysis for Data Mining and System identification /J. Abonyi, B. Feil. //Berlin:BirkHauser, 2007. –304p.
4. Hoppner, F. Fuzzy cluster analysis /F. Hoppner, F. Klawonn, R. Kruse/John Willey Sons, 1999. –290p.
5. Theodoridis, S. Pattern recognition/S. Theodoridis, K. Koutroumbas//Elsevier, 2003. --700p.
6. Patrick, J. The Past, Present, and Future of Multidimensional Scaling/ J. Patrick, F. Groenen, I. Borg//USA:Econometric Institute Report EI, 2013. –25p.
7. Alter, S. Decision Support Systems: Current Practice and Continuing Challenges/Addison-Wesley Pub., 1980. –316p.
8. Mendel, J. Fuzzy logic systems for engineering: a tutorial/J. Mendel//Proceedings of the IEEE, vol.83(3), 1985. –p.p.345–377.
9. George, H. Robust Decision Trees: Removing Outliers from Databases/H. George// Proceedings of the AAAI Conference on artificial Intelligence. USA, 2018. –p.p. 175-181.