

Куч П.О. — рецензент Кисельов Г.Д.
 ННК “ІПСА” НТУУ “КПІ”, Київ, Україна

Пошук типових шаблонів в задачах мультимодальної взаємодії

В задачах мультимодальної взаємодії вирішуються проблеми прогнозування поведінки людини, для прийняття системою відповідних рішень. Деякі методи прогнозування базуються на пошуку відповідностей між реальними даними та типовими шаблонами. В даній роботі розглядається метод автоматичного визначення шаблонів, виходячи з тривалих спостережень за поведінкою.

Розглянемо деякий набір сервісів, кожен з яких має свою мітку. Через певні проміжки часу система отримує інформацію про кількість викликів кожного сервісу за цей проміжок часу. Виходячи з отриманих даних визначити характерні шаблони роботи з сервісами в залежності від часу доби.

Задача пошуку шаблонів зводиться до задачі кластеризації вхідного набору спостережуваних даних. Опишемо вимоги до алгоритму кластеризації:

- Невідома кількість кластерів.
- Важливим є визначення центру кластера.
- Обов'язковою є збіжність алгоритму.
- Максимально точна кластеризація.

Алгоритм, що максимально задовольняє даним вимогам – алгоритм формального елемента (ФорЕл). Розглянемо характерні особливості використання алгоритму ФорЕл для даної задачі.

Для алгоритму ФорЕл вхідними даними є множина точок в n -мірному просторі. В даній задачі елементом множини виступатиме вектор \bar{x} :

$$\bar{x} = (t, s_1, \dots, s_m)$$

де s_i , $i = \overline{1, m}$ – кількість викликів i -го сервісу, m – кількість сервісів, t – час доби в який знімалися виміри s_i .

В процесі роботи, алгоритм ФорЕл намагається покрити максимальну кількість точок мінімальною кількістю сфер заданого радіуса R . Це досягається зміщенням центрів сфер до центрів мас точок, що потрапляють в задану сферу. Початкові центри сфер вибираються з вхідної множини випадково. Точки належать одному кластеру, якщо вони потрапили в одну сферу.

Результатом роботи алгоритму ФорЕл є набір множин (кластерів), об'єднання яких дає вхідну множину. З вхідного набору кластерів вибираються множини, потужність яких суттєво більша потужності інших множин. Кожному з вибраних кластерів ставиться у відповідність його центр мас, що і буде типовим шаблоном в контексті даної задачі.

Використання даного алгоритму дозволить виділити типові шаблони роботи з сервісами в певний час доби. Крім того можлива динамічна корекція шаблонів: оскільки шаблони також є елементами n -мірного простору – їх можна розглядати як елементи вхідної множини разом із новими даними.

Література

1. Stephan S. Development of a novel context prediction algorithm and analysis of context prediction schemes. Kassel, Univ., Diss. 2008.
2. Шуметов В.Г., Шуметова Л.В. “Кластерный анализ: подход с применением ЭВМ”. ОрелГТУ, Орел, 2000. – 118 с.