

**А. В. Иванов** (Москва, ИПИ РАН). **Идентификация параметров в моделях процессов функционирования информационного портала.**

Информационный web-портал — это программная система, построенная на основе протоколов и технологий Интернет. Ее основная задача — организация доступа пользователей к информационным источникам и службам, объединенным на принципах федерации, и создание общей информационной среды, обеспечивающей эффективную работу пользователей с предоставленными ресурсами.

Как и у любой сложной системы, многие процессы функционирования информационного портала нуждаются в оптимизации [1]. Это, в свою очередь, требует использования математических моделей, адекватных динамике изучаемых процессов.

Основная функциональность по обработке пользовательских запросов и взаимодействию с информационными источниками и службами портала обеспечивается портальной подсистемой интеграции и поиска. Для оптимизации функционирования соответствующей программной службы ключевое значение имеют показатели эффективности источников и пользовательской активности.

Для моделирования процессов, определяющих указанные показатели, предлагается использовать стохастические модели с порогами, поскольку рассматриваемые показатели имеют такие характерные для нелинейных систем особенности, как зависимость возмущений от текущих значений показателя, цикличность процессов, скачкообразные изменения характеристик. Для учета таких явлений, с одной стороны, и для возможности использования в решениях полезных свойств линейных систем идентификации, с другой, предлагается подход на основе классификации состояний изучаемого показателя. Пространство значений показателя разбивается на области, и предполагается, что для значений показателя внутри области его динамика описывается простейшими линейными уравнениями, а при выходе показателя за границы области модель изменяется. Такой подход, по-видимому, впервые предложен в [2], где рассматривается модель авторегрессии с порогом (TAR, threshold autoregression).

Отличием предлагаемых моделей от классических моделей TAR, а также моделей более общего вида TARMA, является то, что значения моделируемого показателя оказываются доступными только косвенно, поэтому вместо моделей временных рядов используются модели динамических систем наблюдения.

Для формирования моделей предложены два подхода. В задаче моделирования эффективности источников используются два процесса с порогами — первый для формирования классифицирующей функции, определяющей текущий режим работы источника, и второй — для непосредственного моделирования наблюдений за источником.

В задаче моделирования активности пользователей используется один процесс с порогами, значения которого непосредственно моделируют соответствующий показатель. Обнаружено, что данные активности пользователей содержат явно выраженную периодическую составляющую, поэтому необходимо исключить ее влияние, чтобы при использовании модели первого порядка получить адекватные результаты моделирования. С этой целью реальные данные специально подготавливались: выделялся период колебаний, определялся усредненный уровень активности для этого периода (тренд), после чего он вычитался из наблюдаемых данных. Полученное отклонение от линии тренда затем использовалось в качестве исходных данных для идентификации параметров модели.

Предложены модели идентификации параметров в задачах оценки эффективности источника и активности пользователей на основе экспертной оценки.

Установлена возможность обоснованного выбора параметров рассматриваемых моделей за счет простейшего статистического анализа среды функционирования портала.

Полученные результаты могут быть использованы для решения задач оптимизации использования ресурсов подсистемы интеграции и поиска web-портала.

В докладе приводятся и сравниваются результаты многочисленных экспериментов, выполненных на основе реальных данных.

#### СПИСОК ЛИТЕРАТУРЫ

1. *Босов А. В.* Моделирование и оптимизация процессов функционирования Информационного web-портала. — Программирование, 2009, № 6, с. 53–66.
2. *Tong H., Lim K. S.* Threshold autoregression, limit cycles, and cyclical data (with discussion). — J. Royal Statistical Society, Series B, 1980, v. 42, p. 245–292.