

АНАЛИЗ ЗАДЕРЖЕК ПРИ ПЕРЕДАЧЕ СООБЩЕНИЙ В КОММУНИКАЦИОННОЙ СРЕДЕ ВЫЧИСЛИТЕЛЬНОГО КЛАСТЕРА

Козлов Владимир Дмитриевич

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: kozlov2.volodia2@gmail.com

Задержкой при пересылке сообщения между двумя узлами вычислительного кластера называется промежуток времени между отправкой сообщения на одном узле и его принятием на другом узле. Анализ задержек можно использовать для диагностики или оптимизации коммуникационной среды кластера [1]. Однако размеры современных распределённых вычислительных систем, состоящих из десятков тысяч узлов, делают сбор, хранение и обработку больших объёмов данных о задержках чрезвычайно затратными или просто невозможными. Для решения этой проблемы в работе [1] была предложена вероятностная модель задержек в виде смеси трёхпараметрических логнормальных (3LN) распределений, параметры которой оцениваются по конечной выборке.

Оценка параметров 3LN распределения является нетривиальной задачей. Показано [2], что функция правдоподобия 3LN не ограничена сверху, что делает невозможным применение метода максимума правдоподобия и вынуждает искать новые способы оценки параметров распределения по данным. То же справедливо и для смеси 3LN распределений.

В работе рассмотрены уже описанные в литературе методы оценки параметров 3LN распределения (метод локального максимума правдоподобия, метод моментов и др.), а также предложены новые (метод минимального расстояния). Приводится анализ и сравнение методов на синтетических и реальных данных. Также описываются и анализируются на реальных данных различные методы разделения смеси 3LN распределений — поиск локального максимума правдоподобия EM-алгоритмом, сеточные методы [3] и методы минимального расстояния. Доказывается идентифицируемость семейства 3LN распределений, гарантирующая корректность задачи разделения смесей в рассматриваемой модели.

Литература

1. Gorelov A., Maysuradze A., Salnikov A. Delay structure mining in computing cluster // CEUR Workshop Proceedings. Vol. 1482.

- Aachen: M. Jeusfeld c/o Redaktion Sun SITE, Informatik V, RWTH Aachen Germany Germany, 2015. P. 546–551
2. Hill B. The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic // Journal of the American Statistical Association. Vol. 58, № 301, 1963. P. 72–84
 3. Королев В. Ю., Назаров А. Л. Разделение смесей вероятностных распределений при помощи сеточных методов моментов и максимального правдоподобия // Автоматика и телемеханика. 2010. № 3, С. 98–116