

НОРМАЛИЗАЦИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Студент гр. 11307122 Добровольская Е. Д.

Кандидат физ.-мат. наук, доцент Гундина М. А.

Белорусский национальный технический университет, Минск, Республика Беларусь

Информационная избыточность данных ведет не только к увеличению требуемого объема памяти для хранения информации, содержащейся в базе данных. В литературе подробно рассмотрены вопросы избыточности данных; аномалий включения, модификации и удаления; нормализации данных; рассмотрены вопросы приведения базы данных к нормальной форме [1].

Известно, что избыточность – это термин из теории информации, означающий превышение количества информации, используемой для передачи или хранения сообщения, над его информационной энтропией. Под информационной энтропией будем понимать меру неопределенности (неупорядоченности) или сложности некоторой системы, например, какого-либо опыта, который может иметь разные исходы, а значит и количество информации.

Часто возникают сложности, связанные с обработкой информации, называемые аномалиями. Под аномалией в этом контексте будем понимать такую ситуацию в базе данных, которая приводит к противоречию в базе данных либо существенно усложняет обработку базы данных.

Причиной возникновения таких аномалий может быть излишнее дублирование данных, которое вызывается наличием функциональных зависимостей от не ключевых атрибутов.

Так, например, аномалия включения – это проблема, связанная с добавлением данных в базу данных. Аномалия модификации – это проблема, связанная с изменением данных. Аномалия удаления – это проблема, связанная с удалением данных из совокупности.

Нормализация в этом случае – это процесс разбиения или декомпозиции исходного отношения на несколько отношений с целью устранения нежелательных функциональных зависимостей, приводящих к возникновению избыточности хранения информации и аномалиям добавления, удаления, обновления.

Информационная избыточность является отрицательным фактором, влияющим на целостность базы данных, вынуждающим приводить базы данных к нормальной форме.

Цель нормализации: исключение избыточного дублирования данных, которое является причиной аномалий, возникших при добавлении, редактировании и удалении информации.

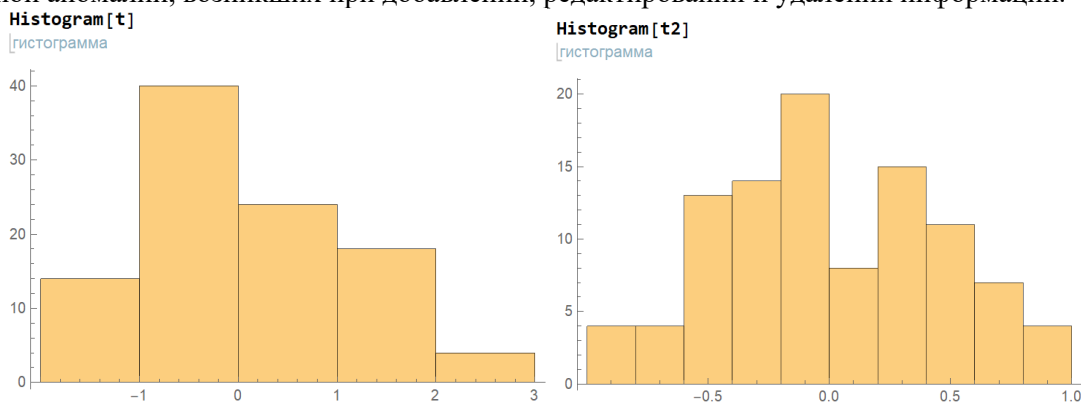


Рис. 1. Результат нормализации выборки

Для выборки также можно провести нормализацию, что позволит преобразовать данные к некоторым безразмерным величинам. Это позволит сравнивать различные данные после приведения их к единому виду. Для кластеризации данных используется следующий подход: вычитаем из текущего вектора данных среднее значение выборки и делим его на максимальное значение выборки. Полученный вектор является результатом смещения данных и последующего их масштабирования.

Литература

1. Чигарина, Е. И. Базы данных: учеб. пособие / Е. И. Чигарина. – Самара: Издво СГАУ, 2015. – 208 с.