

ПОСТРОЕНИЕ ВЗВЕШЕННОЙ ВЫБОРКИ W-ОБЪЕКТОВ ПРИ ОБРАБОТКЕ РЕЗУЛЬТАТОВ СОЦИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ¹⁾

Е. В. Волченко, канд.техн.наук,
Институт информатики и искусственного интеллекта Донецкого национального
технического университета
LM@mail.promtele.com

В данной работе рассматривается задача обработки результатов социологических исследований с целью формирования качественных выборок, отражающих свойства генеральных совокупностей. Обработка результатов социологических исследований проводится в данной работе путем обобщения и редукции данных [1]. Данная работа является продолжением исследований в области эффективной обработки статистических данных и посвящена исследованию возможности применения последовательного метода построения взвешенной выборки w-объектов [2] для обработки результатов социологических исследований.

Пусть в результате социологического исследования были получены данные о нескольких группах объектов, называемых в дальнейшем классами. Каждый объект имеет заданное множество числовых характеристик, называемых в дальнейшем признаками. Под классификацией объекта будем понимать его принадлежность одной из групп.

Основой предлагаемого метода являются выделение близких по значениям объектов выборки и их замена одним w-объектом, значения характеристик которого рассчитываются как средние значения соответствующих характеристик объединяемых объектов, а вес как количество объединяемых объектов. При этом объекты исходной выборки, использованные для построения некоторого w-объекта, из исходной выборки удаляются и соответственно не используются для построения других w-объектов. Процесс построения w-объектов заканчивается, когда в исходной выборке не останется ни одного объекта.

Особенностью данных, получаемых в результате социологических исследований, является разнотипность признаков объектов выборки. Такие показатели, как количество детей, количество единиц приобретаемого товара, наличие недвижимости, автомобиля и др., могут принимать только целочисленные значения. Значения признаков w-объектов, полученные по предложенному алгоритму, могут в силу построения принимать вещественные значения. Поэтому по окончании построения выборки w-объектов необходимо выполнять корректировку полученных результатов, округляя рассчитанные значения соответствующих признаков до целых.

Отметим, что принятое в данной работе требование наличия в исходной выборке выделенных классов объектов не является обязательным. Предварительную классификацию объектов исходной выборки можно выполнить путем автоматической кластеризации данных.

Для оценки качества предложенного подхода к построению взвешенной выборки был проведен ряд экспериментальных исследований. В качестве исходных данных использовались выборки объемом 1000 – 3000 объектов, описываемых 3 – 10 признаками. По исходным выборкам и взвешенным выборкам w-объектов рассчитывались статистические

¹⁾ Работа выполнена при содействии гранта Президента Украины для поддержки научных исследований молодых ученых №GP/F32/130 «Разработка теоретических основ и методов реализации открытых обучающихся систем автоматического распознавания: способы оптимизации обучающих выборок и методы построения взвешенных решающих правил классификации»

оценки и сравнивались со статистическими оценками генеральной совокупности. Так, гипотеза о равенстве математических ожиданий выборки w -объектов и генеральной совокупности подтвердилась на 6,3% больше, чем при сравнении математических ожиданий исходной выборки и генеральной совокупности. При этом сокращенная взвешенная выборка w -объектов составила 23% исходной выборки. Таким образом, результаты экспериментальных исследований позволяют говорить об эффективности применения последовательного метода построения взвешенной выборки w -объектов для обработки результатов социологических исследований.

1. Девятко И.Ф. Методы социологического исследования / И.Ф. Девятко – Екатеринбург: Издательство Уральского университета, 1998. – 208 с.

2. Волченко Е.В. Метод построения взвешенных обучающих выборок в открытых системах распознавания / Е.В. Волченко // Доклады 14-й Всероссийской конференции «Математические методы распознавания образов (ММРО-14)», Суздаль, 2009. – М.: Макс-Пресс, 2009. – С. 100 – 104.

