

Метод кластеризации текстовых данных на основе параллельной обработки

Чебаков С.В., Серебряная Л.В.*

Объединенный институт проблем информатики НАН Беларуси
Белорусский государственный университет информатики и
радиоэлектроники*

Работа посвящена организации параллельной обработки данных, которые сначала требуется разбить на кластеры, а затем объединить полученные решения в единое пространство. Поскольку в случае самообучения кластеры, составляющие обучающую выборку, заранее неизвестны и их требуется определить в ходе процедуры самообучения.

Определение 1. P_n есть частичное решение задачи кластеризации, если P_n представляет собой разбиение на кластеры некоторого отдельного подмножества начальных текстов.

Схему организации параллельных вычислений для выполнения процедуры кластеризации можно представить следующим образом. Имеется набор текстов произвольной тематики, поступающих с целью их кластеризации в динамическом режиме, где общее количество текстов неизвестно. Одновременно в обработке может находиться достаточно большое число текстов. Поэтому предлагается разбить все имеющиеся тексты на некоторое число отдельных подмножеств, содержащих равное число элементов. На каждом подмножестве в параллельном режиме выполняется построение собственных кластеров, т.е. формируются частичные решения. Эффективная реализация параллельных вычислений, включающих в себя алгоритмы формирования частичных решений, их объединения и получения конечного результата в виде разбиения всего множества поступающих текстов на отдельные кластеры, могла бы существенно уменьшить общее время решения задачи кластеризации.

На первом уровне решения поставленной задачи происходит разбиение на кластеры каждого отдельного подмножества начальных данных. Алгоритм кластеризации был представлен в работе [1]. Здесь он дополняется процедурой упорядочивания всех элементов по убыванию величины расстояния от них до центра кластера. Пусть A_1 и B_1 представляют собой множества кластеров, сформированных в параллельном режиме при кластеризации двух подмножеств начальных текстов P_1 и P_2 .

На втором уровне параллельных вычислений, оперируют непосредственно с построенными кластерами множеств A_1 и B_1 , для чего вводятся специальные операции.

1. Серебряная, Л.В., Чебаков, С.В. Методы автоматической классификации и кластеризации текстовой информации // Информатизация образования 2011. - № 2. – С. 52-61.

УДК 51(07.07)

Информационные компьютерные технологии как эффективное средство образовательного процесса студентов

Шляго Н.И., Климаневская Л.А.

Барановичский государственный университет

Сегодня информационные компьютерные технологии можно считать тем новым способом передачи знаний, который соответствует качественно-новому содержанию обучения и развития личности. Этот способ позволяет студенту с интересом учиться, находить источники информации воспитывает самостоятельность и ответственность при получении новых знаний, развивает дисциплину интеллектуальной деятельности.

Компьютерные технологии могут быть представлены в образовательном процессе следующими учебно-методическими ресурсами: электронные учебники, тренажерные системы, виртуальные лаборатории по дисциплине (виртуальный учебно-исследовательский комплекс), тестирующие комплексы.

Одной из новых форм внедрения компьютерных технологий является создание и использование электронных учебников, которые наглядно, красочно и с мобильным доступом информации предстанут перед студентами. Электронный учебник – это ресурс, содержащий структурированную интерактивную учебную информацию, систему упражнений для ее осмысления и закрепления, сценарии учебной работы.

Не менее эффективными в учебном процессе являются тестирующие комплексы, которые осуществляют контроль знаний и являются наиболее распространенными электронными ресурсами проверки знаний. Легкость и простота их использования позволяют быстро оценить способности либо уровень знаний тестируемого. Наилучший вариант использования тестирующих программ для студентов – это организация тестового контроля при непосредственном использовании электронного учебника. После изучения определённой темы, студент имеет возможность самостоятельно проверить свои знания с помощью тематических тестов, результаты которых попадают в базу данных, что даёт возможность преподавателю осуществлять контроль и проследить уровень усвоения материала. В конце каждого раздела предусмотрено итоговое тестирование, результаты которого отражают полную картину знаний каждого студента и могут быть использованы преподавателем при зачёте либо экзамене.