**УДК 338.24:330.332:658.152**

**DataMining для поиска ассоциативных правил и практического их применения**

Д.А. Скачко, e-mail: [dskachko@gmail.com](mailto:dskachko@gmail.com)  
Институт кибернетики имени В.М. Глушкова

Мы живем в эпоху информационных технологий, и именно информация является важнейшей единицей в нашей жизни. Но уже не достаточно просто владеть информацией – надо уметь ее понимать и обрабатывать. Для этих целей служит DataMining – процесс поиска знаний в большем объеме данных. Существует много подходов, методов и алгоритмов для решения разных задач в данной области. Среди решаемых проблем можно выделить задачи кластеризации, классификации, прогнозирования, поиска ассоциаций и многие другие. Но с развитием методов сбора данных возникают и проблемы анализа. Обычным пользовательским компьютерам уже не справиться с огромнейшим количеством данных за приемлемое время, даже суперкомпьютеры не дают такого большого увеличения производительности, как того требуют данные. Мы сталкиваемся перед очередной задачей – эффективно использовать имеющиеся вычислительные ресурсы для анализа и обработки данных.

С ростом популярности кластерных систем появилась нужда в масштабируемых алгоритмах, способных эффективно использовать все предоставленные ресурсы для увеличения не только производительности, но и для увеличения объемов обрабатываемых данных. Последнее время стали появляется так называемые ансамбли алгоритмов (ensemble), которые увеличивают результирующую производительность и точность вычислений. Один из самых популярных алгоритмов решения задачи поиска ассоциативных правил - алгоритм Apriori.

Поиск ассоциативных правил совсем не тривиальная задача, как может показаться на первый взгляд. Одна из проблем – алгоритмическая сложность при нахождении часто встречающих наборов элементов, т.к. с ростом числа элементов, экспоненциально растет число потенциальных наборов элементов. В данной работе было исследовано решения задачи поиска ассоциативных правил с помощью алгоритма Apriori.

Рассмотрено как базовые представители класса, так и алгоритм в параллельном исполнении. Также предоставлена характеристика класса алгоритма и описание его применения в практических задачах. Отметим что, несмотря на одинаково решаемые задачи отдельными классами алгоритмов, их эффективность зависит от области применения. Данный алгоритм может служить основой для дальнейшего практического применения.

СПИСОК ЛИТЕРАТУРЫ

1. Шеннон К. Работы по теории информации и кибернетике. -М.: Ин Основы параллельных вычислений для многопроцессорных систем
2. Бройнль Т. Паралельне програмування: Початковий курс: Навч. посібник / Втуп. Слово А. Ройтера; Пер. з нім. В.А. Святного. – К.: Вища шк., 1997. – 358с.остранная литература, 1963. –320с.