

Применение технологии data mining для анализа аудитории электронной рассылки

Мельников А.Ю., Кушнир Ю.В.

*Донбасская государственная машиностроительная академия,
г. Краматорск*

Одной из сфер применения технологии Data Mining (интеллектуального анализа данных) является web-mining – анализ действий посетителей интернет-сайтов с целью выявления их интересов и предпочтений. Если у сайта имеется собственная электронная рассылка, то помимо собственно анализа посетителей сайта отдельный интерес представляет анализ данных об аудитории (подписчиках) этой рассылки.

Сервис Subscribe.Ru по любой из своих рассылок может предоставить следующую статистическую информацию: динамику числа подписчиков за весь период существования рассылки; TOP-50 зон по количеству подписчиков; рейтинг зон 1-го уровня по количеству подписчиков; географию подписчиков; распределение возраста подписок.

Исходя из предоставленной информации, могут быть поставлены и решены задачи визуализации, прогнозирования и кластеризации. В качестве примера взята рассылка электронной версии газеты ДГМА «Академия» (<http://www.gazeta-academia.info>). Особый интерес представляют данные по распределению возраста подписок аудитории как по общему числу подписчиков, так и с учетом текущей даты. Здесь поставлена и решена задача кластеризации в среде аналитического пакета «Deductor» с использованием инструмента «Карты Кохонена». Карта Кохонена кластеризации по общему числу подписчиков с учетом текущей даты позволяет выбрать интересующее значение количества подписчиков, определить дату, когда было достигнуто данное количество подписчиков и просмотреть возраст всех подписок (рис. 1).

