

Методики анализа больших данных (BigData)

Гутич И.И., Жаворонков О.Н.

Белорусский национальный технический университет

Интерес к автоматизированным системам сбора, обработки, управления и анализа больших объемов данных проявляют все ведущие ИТ-компании в мире. Автоматизация обработки больших данных открывает перспективные возможности для освоения новых рынков для бизнеса и привлечения заказчиков. Ведущими игроками на мировом рынке в этой области являются: Amazon, Dell, eBay, EMC, Facebook, Fujitsu, Google, HitachiDataSystemsCorporation, HP, IBM, LinkedIn, Microsoft, NetApp, Oracle, SAP, SAS, SGI (SiliconGraphicsInc), Teradata, VMware, Yahoo.

Существует множество разнообразных методик анализа больших массивов данных:

- A/Btesting – методика, в которой контрольная выборка поочередно сравнивается с другими, выявляя оптимальную комбинацию показателей для достижения наилучшей ответной реакции потребителей;

- Associationrulelearning – набор методик для выявления взаимосвязей, ассоциативных правил, между переменными величинами в больших массивах данных;

- Classification – набор методик, предсказывающий поведение потребителей в определенном сегменте рынка;

- Crowdsourcing – методика сбора данных из большого количества источников;

- Datafusionanddataintegration – набор методик, позволяющий анализировать комментарии пользователей социальных сетей и сопоставлять с результатами продаж в режиме реального времени;

- Machinelearning – создание алгоритмов самообучения на основе анализа эмпирических данных;

- Optimization – набор численных методов для улучшения одного или нескольких показателей в сложных системах;

- Simulation – моделирование поведения сложных систем;

- Visualization – методы графического представления результатов анализа больших данных.

Умение оперировать большими объемами информации, анализировать взаимосвязи между ними и принимать взвешенные решения, придает потенциал компаниям из различных вертикалей для увеличения показателей доходности и прибыльности, повышения эффективности бизнеса.