

## Секция 4:

# ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ (BIG DATA)

## ПРОБЛЕМЫ ДОСТУПА К ДАННЫМ В ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ BIG DATA

В.Г. Чумак, В.М. Рамзаев, И.Н. Хаймович

Международный институт рынка

Сейчас в мире наблюдается экспоненциальный рост объема и разнообразия данных в экономических исследованиях. Полученные в результате исследований данные используются в совместных проектах разных стран с помощью специализированных новейших инструментов, в том числе и Big Data. В данной статье показано развитие применения методов и средств анализа данных и управления данными в финансовых коллективных областях, беспрецедентно быстрое расширение спектра задач, требующих решения на основе полученных данных, накопление опыта решения подобных задач и обеспечения возможности его междисциплинарного использования. Приведены возможности использования экономических коллекций данных в России.

Рассмотрим проблемы доступа к финансовым данным в России. Проблем здесь две: откуда их взять и для чего использовать. Интенсивное использование данных для решения финансовых задач сейчас только начинается, но в других странах мира, например в США, оно уже интенсивно идет. Рассмотрим проекты, которые осуществляются в других странах.

11 ноября 2014 года в США Национальный институт стандартов и технологий выпустил требования по сбору финансовых данных для интенсивного использования этих данных для анализа рисков и прогнозирования развития финансовых рынков [1-13]. В этих требованиях, сделанных по стандарту NIST для Big Data, содержится описание применения интенсивных данных в финансовой сфере, описание текущих решений, а также определены цели и проблемы использования Big Data для финансовых данных.

Согласно требованиям для финансовых данных, размещенных в Big Data, относят следующие данные:

- банковские, касающиеся коммерческих, розничных, кредитных карт, потребительского кредитования, корпоративных банковских услуг, банковских операций, финансирование торговли и глобальных платежей;

- данные в области ценных бумаг и инвестиций, такие как розничные брокерские услуги, услуги частным лицам, управление активами, институциональные брокерские, инвестиционно-банковские услуги, траст-банкинг, управление активами, ответственное хранение и клиринговых услуги;

- данные в области страхования, в том числе жизни, личного и группового имущества от несчастного случая.

Также в этих требованиях приводится один из подходов к разработке работоспособной стратегии в индустрии финансовых услуг. При работе с большими коллекциями данных в Big Data организация должна выполнить следующие этапы определения исходных условий для использования функции работы с большими данными в Интернет в пределах облачных систем государственных и частных финансовых учреждений, предлагающих финансовые услуги в пределах границ Соединенных Штатов; Великобритании, ЕС и Китая.

В соответствии с этой стратегией, каждое финансовое данное должно включать описание: людей (ресурсов), процессов (время/стоимость/рентабельность инвестиций), технологий (различных операционных систем, платформы), нормативного регулирования (в зависимости от различных и нескольких контролирующих органов).

Далее эти данные должны быть выявлены, проанализированы, оценены, рассмотрены, проверены и пересмотрены для работы в процессах:

1. инициации проекта;
2. оценки рисков;
3. анализа воздействия на бизнес;
4. разработка и тестирования стратегий обеспечения непрерывности бизнеса;
5. реагирования на чрезвычайные ситуации и операции;
6. разработки и реализации планов обеспечения непрерывности финансовых проектов;
7. проведения обучающих тренингов.

В настоящее время большие данные с методологией hadoop в облачных технологиях в интернете используются в составе гибридных систем как инструмент для проведения анализа риска и мошенничества, помимо помощи в организации процесса ("знай своего клиента"). Это те три области, где интенсивные данные из Big Data зарекомендовали себя хорошо.

В то же время традиционные клиент/сервер/хранилище данных/RDBM - системы используются для обработки, переработки, хранения и архивных подразделений финансовых данных. Недавно финансовое сообщество одобрило инициативу для представления финансовой отчетности по xbrl (расширяемый язык разметки, связанный с бизнесом) по состоянию на 13 мая 2013.

В настоящее время проблемные области, связанные с использованием хранилищ интенсивных данных в интернете под управлением Big Data, включают агрегирование и хранение данных (конфиденциальных и иных) от нескольких источников, которые могут создать административные и управленческие проблемы, касаются следующих вопросов:

- контроль доступа;
- управление/администрирование;
- право собственности на данные.

Однако, исходя из текущего анализа, эти проблемы широко известны и решаются в данный момент, например, по НИОКР (исследования и разработки).

В России практики обмена данными (сведения Сбербанка) для прогнозирования изменений на финансовых рынках и для анализа рисков в банковской сфере нет, поэтому можно использовать и американский стандарт, проведя необходимые модификации для финансовых рынков России.

На основе этих требований в США и ЕС осуществляются проекты в области использования интенсивных финансовых данных следующими организациями:

1. The IFRS, Securities and Markets Working Group, [www.xbrl-eu.org](http://www.xbrl-eu.org) - комплекс организаций, занимается вопросами безопасности и использования финансовых данных в рыночной области.

2. PCAOB <http://www.pcaob.org> - это некоммерческая корпорация, созданная Конгрессом для надзора и аудита открытых акционерных обществ в целях защиты интересов инвесторов и дальнейшей заинтересованности общественности в подготовке информативных, точных и независимых аудиторских отчетов.

3. <http://www.ey.com/GL/en/Industries/Financial-Services/Insurance> - организация, отслеживает основные нормативные изменения в финансовых документах, анализирует поведение развивающихся рынков, рассчитывает имеющуюся экономическую неопределенность, прогнозирует потребительские ожидания.

4. <http://www.treasury.gov/resource-center/fin-mkts/Pages/default.aspx> - сайт ресурсного центра (Казначейства) отвечает за широкий спектр мероприятий: консультирование

президента по экономическим и финансовым вопросам, поощрение устойчивого экономического роста, и поощрение более эффективного управления в финансовых институтах. Департамент казначейства осуществляет управление и обслуживание систем, критически важных для национальной финансовой инфраструктуры, такие как производство монет и валюты, выплата платежей для американской общественности, сбор доходов и привлечением заемных средств, необходимых для выполнения федерального правительства.

5. CFTC <http://www.cftc.org> -организация, занимающаяся вопросами страхования и кредитования.

6. SEC <http://www.sec.gov> -комиссия США по ценным бумагам и биржам для защиты инвесторов, поддержания справедливого, упорядоченного и эффективного рынка.

7. FDIC <http://www.fdic.gov>- федеральная депозитная страховая корпорация.

8. COSO <http://www.coso.org> - комитет спонсорских организаций комиссии Тредвея (coso) стремится обеспечить интеллектуальное лидерство через развитие инфраструктур и руководство на предприятии управления рисками, внутреннего контроля и сдерживания мошенничества.

9. isc2 International Information Systems Security Certification Consortium, Inc.: <http://www.isc2.org> – консорциум, проводящий широкие исследования в области управления предприятиями и их финансовыми потоками с учетом анализа финансовых рисков для предприятий.

10. ISACA Information Systems Audit and Control Association: <http://www.isca.org> - ассоциация, осуществляющая финансовую аналитику.

Этими организациями осуществляются проекты в США и в странах ЕС по стресс-тестированию финансовых рынков, т.е. они проводят оценку устойчивости финансовых институтов к неблагоприятным изменениям на рынке, а также оценку системного риска в системе ЕС по специальным унифицированным методикам, сценариям и вырабатывает ключевые предложения по упражнениям. Также выпускают агрегированные статистические данные о банковском секторе ЕС, кредитном риске, операционном риске, рыночном риске и деятельности надзорных органов и мероприятиях с 2007 по 2013 год в виде документов, размещенных на сайте <https://www.eba.europa.eu/supervisory-convergence/supervisory-disclosure/aggregate-statistical-data>.

Например, приводятся данные по кредитным рискам банков ЕС, которые осуществляются по стресс-тестам в режиме реального времени в виде итоговых таблиц (см. рис.1).

Table Data related to Credit Risk

| Data related to Credit Risk*                            |  | 2010                                | 2011    | 2012    | 2013    |         |
|---|--|-------------------------------------|---------|---------|---------|---------|
| Credit Institutions: Own Funds Requirements             | Own Funds Requirements Credit Risk % of Total Own Funds Requirements | 90,02 %                             | 89,07 % | 90,81 % | 90,45 % |         |
| Credit Institutions: Distribution by Approach           | % Number   | Standardised Approach (SA)          | 91,97 % | 92,23 % | 92,35 % | 91,42 % |
|   |  | Foundation IRB Approach (FIRB)      | 7,63 %  | 7,37 %  | 7,38 %  | 8,42 %  |
|   |  | Advanced IRB Approach (AIRB)        | 0,39 %  | 0,40 %  | 0,27 %  | 0,15 %  |
|   | Own Funds Requirements % of Own Funds Requirements Credit Risk**     | Standardised Approach (SA)          | 68,77 % | 69,40 % | 71,52 % | 69,92 % |
|   |  | Foundation IRB Approach (FIRB)      | 18,76 % | 18,27 % | 15,07 % | 16,29 % |
|   |  | Advanced IRB Approach (AIRB)        | 12,47 % | 12,34 % | 13,41 % | 13,79 % |
| Credit Institutions: Distribution by IRB Exposure Class | Exposures % of Risk Weighted Assets                                  | Central Governments & Central Banks | 2,85 %  | 1,30 %  | 2,09 %  | 1,13 %  |
|   |  | Institutions                        | 17,53 % | 21,22 % | 21,02 % | 16,62 % |
|   |  | Corporate                           | 37,64 % | 39,70 % | 37,37 % | 42,78 % |
|   |  | Retail                              | 13,90 % | 14,15 % | 14,55 % | 17,69 % |
|   |  | Equity                              | 6,43 %  | 7,43 %  | 7,53 %  | 8,17 %  |
|   |  | Securitisation positions            | 2,13 %  | 1,63 %  | 1,07 %  | 1,05 %  |
|   |  | Other Non Credit Obligation Assets  | 0,12 %  | 0,06 %  | 0,01 %  | 0,17 %  |

[data as per 31.12.]

\* Data provided by the OeNB.

\*\* If an institution uses more than one approach, it will be counted accordingly.

Рисунок 1 - Пример оценки кредитных рисков в ЕС на 2013 год, собранные по коллекции интенсивных данных

Также этими организациями ведется реестр кредитных организаций ЕС. Данных по российскому финансовому рынку и кредитным организациям в этих проектах нет, но эти данные могут быть полезны в инвестиционных проектах совместного кредитования, например, в ракетно-космической отрасли при проектировании и производстве малых космических аппаратов (Россия-Германия).

Одним из крупнейших финансовых проектов в США с интенсивным использованием данных в Big Data является проект прогнозирования системного риска кредитования в банковской сфере [14-16]. Исходными данными для данного проекта в США являются документы о совместных кредитах, размещенные на сайте (<http://www.sec.gov/>).

В России и регионах прогнозирование системного риска кредитования (MIDAS по-русски) в сегодняшних сложных экономических условиях является чрезвычайно актуальным. Возможно сформулировать цели такого проекта:

- 1) определение наиболее значимых финансовых институтов, в случае потери ликвидности которых экономическая система переходит в неуправляемое состояние;
- 2) определение компаний, наиболее зависимых от кредитных ресурсов (в том числе внешних);
- 3) определение последствий изменений параметров и индикаторов финансовой системы (ключевая ставка, изменение курса валюты, изменение стоимости облигаций и т.д.);
- 4) проведение стресс-тестов для определения устойчивости всей системы в целом.

Исходными данными для проектов могут быть: требования регулятора SEC (аналог бывшей ФСФР, сейчас ЦБ РФ), которые обязывают эмитентов ценных бумаг раскрывать свою отчетность. В данной отчетности применяется стандарт GAAP, достаточно похожий на МСФО, что позволяет сказать об определенной структуре данных, которые легко встраиваются в систему интенсивной обработки данных. В российской практике требование ЦБ РФ по раскрытию отчетности исполняется эмитентами ценных бумаг, кредитными учреждениями и др.

Использование интенсивных финансовых данных может, например, проходить по следующему алгоритму: 1) формирование графа сокредитования; 2) анализ критических хабов; 3) визуализация отношений корпорации с другими компаниями и ключевые персоны.

На основании исходных данных формируется *граф сокредитования*. Вершинами графа являются организации. Две вершины соединяются ребром, если соответствующие организации предоставляют некоторый совместный кредит (возможно, вместе с другими организациями). Ребру приписывается вес в зависимости от того, сколько совместных кредитов (на какую сумму) предоставлено парой организаций. Для каждой вершины вычисляется нормализованная центральность собственного вектора (центральность является мерой важности вершины в графе). При таком измерении центральности вклад в важность вершины дают важные вершины, с которыми она соединена. Вершины с наивысшими показателями центральности с высокой вероятностью являются критическими хабами для сети сокредитования. Критические хабы затем подлежат более детальному анализу:

- отношения с другими компаниями (дочерние компании, конкуренты, заемщики и т.д.);
- руководство (занимаемые позиции и даты для каждой из персон, руководство другими компаниями);
- агрегированные финансовые данные (займы, инвестиции и т.д.).

В итоге удастся в Америке создать графы отношений конкретных банков с другими компаниями и банками, а также выделить ключевые рискованные проекты в банковской сфере. Примеры отношений с другими компаниями и ключевые персоны для Citigroup приведены на рисунках 2 и 3.

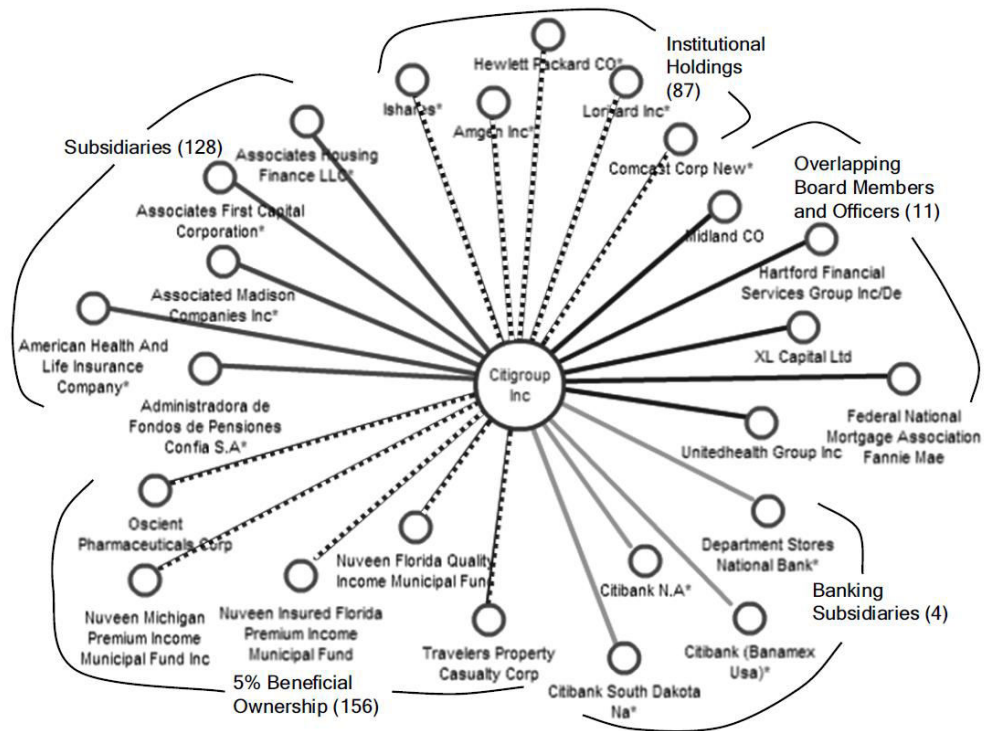


Рисунок 2 - Пример связи банка Citigroup с другими банками

| Position                             | 2005                    |    | 2006 |    | 2007 |    | 2008                |    | 2009         |          |
|--------------------------------------|-------------------------|----|------|----|------|----|---------------------|----|--------------|----------|
|                                      | H1                      | H2 | H1   | H2 | H1   | H2 | H1                  | H2 | H1           | H2       |
| <a href="#">Diana L Taylor</a>       |                         |    |      |    |      |    |                     |    |              | Boar.    |
| <a href="#">Richard D Parsons</a>    | Board Member            |    |      |    |      |    |                     |    |              |          |
| <a href="#">Alain J Belda</a>        | Board Member            |    |      |    |      |    |                     |    |              |          |
| <a href="#">Jerry A Grundhofer</a>   |                         |    |      |    |      |    |                     |    |              | Board Me |
| <a href="#">Franklin A Thomas</a>    | Board Member            |    |      |    |      |    |                     |    |              |          |
| <a href="#">Gerald R Ford</a>        | Board Member            |    |      |    |      |    |                     |    |              |          |
| <a href="#">Robert B Willumstad</a>  | Presid...               |    |      |    |      |    |                     |    |              |          |
| <a href="#">Michael E Oneill</a>     |                         |    |      |    |      |    |                     |    |              | Board Me |
| <a href="#">C Michael Armstrong</a>  | Board Member            |    |      |    |      |    |                     |    |              |          |
| <a href="#">Vikram S Pandit</a>      |                         |    |      |    |      |    | CEO, Citigroup Inc. |    |              |          |
| <a href="#">Roberto Hernandez</a>    | Board Member            |    |      |    |      |    |                     |    |              |          |
| <a href="#">Timothy C Collins</a>    |                         |    |      |    |      |    |                     |    |              | Boar.    |
| <a href="#">John M Deutch</a>        | Board Member            |    |      |    |      |    |                     |    |              |          |
| <a href="#">Lawrence R Ricciardi</a> |                         |    |      |    |      |    |                     |    | Board Member |          |
| <a href="#">William S Thompson</a>   |                         |    |      |    |      |    |                     |    |              | Board Me |
| <a href="#">Sallie Krawcheck</a>     | Chief Financial Officer |    |      |    |      |    |                     |    |              |          |
| <a href="#">William R Rhodes</a>     | Senior Vice Chairman    |    |      |    |      |    |                     |    |              |          |

Рисунок 3 - Определение ключевых персон для банка Citigroup в рамках проекта МИДАС

В итоге можно сделать вывод, что интенсивное использование финансовых данных с применением Big Data в России позволит решать задачи совместного кредитования в банковской сфере и своевременно прогнозировать риски, а также вырабатывать согласованные управленческие решения до наступления кризисных ситуаций. Правильно и своевременно принимаемые решения позволят избежать синергетических эффектов на финансовом рынке и сделают рынки более предсказуемыми и стабильными.

#### Литература

1. DRAFT NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements// National Institute of Standards and Technology, 2014, 300 p.
2. “Big Data is a Big Deal”, The White House, Office of Science and Technology Policy. <http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal> (accessed February 21, 2014)
3. Shantenu Jha, Judy Qiu, Andre Luckow, Pradeep Mantha, and Geoffrey C. Fox, A Tale of Two Data-Intensive Approaches: Applications, Architectures and Infrastructure, in 3rd International IEEE Congress on Big Data Application and Experience Track. June 27- July 2, 2014. Anchorage, Alaska. <http://arxiv.org/abs/1403.1528>.
4. Judy Qiu, Shantenu Jha, Andre Luckow, and Geoffrey C. Fox, Towards HPC-ABDS: An Initial High-Performance Big Data Stack. August 8, 2014. <http://grids.ucs.indiana.edu/ptliupages/publications/nist-hpc-abds.pdf>.
5. Geoffrey Fox, Judy Qiu, and Shantenu Jha, High Performance High Functionality Big Data Software Stack, in Big Data and Extreme-scale Computing (BDEC). 2014. Fukuoka, Japan.
6. <http://www.exascale.org/bdec/sites/www.exascale.org/bdec/files/whitepapers/fox.pdf>.
7. Geoffrey C. Fox, Shantenu Jha, Judy Qiu, and Andre Luckow, Towards an Understanding of Facets and Exemplars of Big Data Applications. July 20, 2014. <http://grids.ucs.indiana.edu/ptliupages/publications/OgrePaperv9.pdf>.
8. Geoffrey Fox and Wo Chang, Big Data Use Cases and Requirements. August 10, 2014. <http://grids.ucs.indiana.edu/ptliupages/publications/NISTUseCase.pdf>.
9. Geoffrey Fox. INFO 590 Indiana University Online Class: Big Data Open Source Software and Projects. 2014 [accessed 2014 December 11]; Available from: <http://bigdataopensourceprojects.soic.indiana.edu/>.
10. <http://www.whitehouse.gov/mgi>
11. <http://www.whitehouse.gov/open>
12. NIST Integrated Knowledge EditorialNet (NIKE). <http://xpdb.nist.gov/nike/term.pl>
13. <https://rd-alliance.org/group/metadata-standards-directory-working-group.html>
14. S. Balakrishnan, V. Chu, M. A. Hernandez, H. Ho, R. Krishnamurthy, S. Liu, J. Pieper, J. S. Pierce, L. Popa, C. Robson, L. Shi, I. R. Stanoi, E. L. Ting, S. Vaithyanathan, and H. Yang. Midas: Integrating Public Financial Data. In SIGMOD, pages 1187-1190, 2010.
15. A. Arasu, C. Re, and D. Suciu. Large-Scale Deduplication with Using Dedupalog. In ICDE, pages 952-963, 2009.
16. L. Chiticariu, Y. Li, S. Raghavan, and F. Reiss. Enterprise Information Extraction: Recent Developments and Open Challenges. In SIGMOD, pages 1257-1258, 2010.