

МОДЕЛИ КЛАСТЕРИЗАЦИИ БОЛЬШИХ ДАННЫХ

*Курдюков Виталий Владимирович, докторант PhD, кафедра
«Информатика и ИБ»*

*Евразийский национальный университет имени Л.Н. Гумилева, г. Астана.
(Научный руководитель - Казиев Г.З., докт. техн. наук, профессор)*

Переезжая по мосту, соединяющему одну часть города с другой, редко обращаем внимание на то, из чего сделаны эти сооружения, но экономим массу времени и сил, благодаря архитекторам и строителям мостов. Мир технологий никогда не имел границ – его движущая сила – человеческий ум и стремление к созиданию и объединению. С наступлением эры данных, перед всем человечеством открываются невероятные возможности для работы, инновационных решений и строительства.

Мосты являются сложными структурами, и должны быть тщательно разработаны [1]. Они могут соединять берега, перекидываясь через реки, непроходимые ущелья, обеспечивают более прямой маршрут, при этом существенно уменьшают время в пути. Для строительства мостов используют развитие технологий и совершенствование материалов. Мост является древнейшим изобретением человечества. Big Data – главный термин, фигурирующий на всех профессиональных конференциях, анализу данных, прогностической аналитике, интеллектуальному анализу данных. Термин используется в сферах, где актуальна работа с качественно большими объемами данных, где постоянно происходит увеличение скорости потока данных в организационный процесс: экономике, производстве, строительстве, банковской деятельности, маркетинге, телекоммуникациях.

Вместе со стремительным накоплением информации в строительстве, быстрыми темпами развиваются и технологии анализа данных. Большие данные - это многообразие получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, получаемые путём распределения по многочисленным узлам вычислительной сети.

В качестве определяющих характеристик для больших данных отмечают три «V»:

1. Объём (англ. *volume*, величина физического объёма).
2. Скорость (англ. *velocity* как скорость прироста, так и необходимость высокоскоростной обработки и получения результатов).

3. Многообразие (англ. *variety*, в смысле возможности одновременной обработки различных типов структурированных и полу структурированных данных).

На данный момент к трём «V» добавляют ещё две «V».

1. Достоверность (*veracity*, понимая под этим целостность данных и уровень доверия к результатам).

2. Ценность (*value*, информация полезная и ценная, собираемые данные должны оправдывать затраты на их обработку и анализ).

Big Data – это конкретный набор подходов и технологий, призванных решить данные задачи. В основе одного из таких подходов лежит система распределенных вычислений, где обработка больших объемов данных требует для себя не одну высокопроизводительную машину, а целую группу таких машин, объединенных в кластер.

Мосты состоят из пролётных строений и опор. Пролётные строения служат для восприятия нагрузок и передачи их опорам; на них может располагаться проезжая часть, пешеходный переход, трубопровод. Опоры переносят нагрузки с пролётных строений на основания моста. Пролётные строения состоят из несущих конструкций балок и ферм, плиты проезжей части. Статическая схема пролётных строений может быть арочной, балочной, рамной, вантовой или комбинированной; она определяет тип моста по конструкции. Обычно пролётные строения прямолинейны, однако в случае необходимости при постройке эстакад и дорожных развязок им придают сложную форму: либо кольцевую, либо спиралеобразную. Пролётные строения поддерживаются опорами, каждая из которых состоит из фундамента и опорной части. Формы опор могут быть весьма разнообразными. Схема моста - формула, в которой последовательно представлены размеры расчётных пролётов - расстояния между центрами опорных частей пролётных строений. Если несколько последовательных опорных частей имеют одинаковый размер, указывается их количество, помноженное на размер каждого. Например, схема моста $10+4 \times 15+8$ метра значит, что у первого пролётного строения моста расчётный пролёт — 10 метров, четыре следующих — по 15 метров, каждый и пятый — 8 метра.

Главнейшими материалами для строительства мостов являются камень, цемент, сталь, алюминий, дерево, кирпич, композиты. Бетон является популярным для всех видов строительства моста из-за своей доступности и прочности. Бетон не требует большого ухода, хотя он имеет тенденцию разрушаться от соленой воды и эрозии. Внешне бетон непривлекателен, поэтому архитекторы редко используют его в проектах, считая его скучным, серым. При использовании на длинных пролетах, бетон усиливается стальными решетками. Сталь является одним из самых популярных материалов для мостов и может

быть использована для большого диапазона расстояний, которые не возможны с другими материалами. Сталь в сто раз долговечнее, чем бетон и весит меньше. Стальные мосты подвержены ржавчине и коррозии, однако, и, как правило, они требуют специального техобслуживания. Многие мосты окрашены в различные цвета – это помогает улучшить их внешний вид. Алюминий слишком мягкий материал по сравнению со сталью, но он часто используется там, где высокая прочность не прерогатива. Алюминий устойчив к коррозии, и более эстетически привлекательный, чем сталь, поэтому он часто превращается в покрытие и защиту других материалов бетона или стали. Натуральный камень является одним из древнейших материалов для строительства моста. Этот материал способен противостоять эрозии ветра и воды. Такие материалы, как гранит и известняк очень привлекательны и внешне, плюс ко всему они будут служить в течение многих столетий практически без технического обслуживания. В то же время, камень имеет высокие затраты на установку. Он часто используется для создания опор моста и нижних колонн, при том, что верхняя часть моста может быть построена из более доступных и легких материалов. Дерево не так надежно, как другие строительные материалы для мостов, и должно быть использовано только на относительно простых структурах. Это один из более доступных строительных материалов. С деревом легко работать с использованием основных инструментов и оборудования. Деревянные мосты, в первую очередь, выигрывают внешне, благодаря природной красоте. Но дерево может разбухать и гнить под воздействием влаги, но жизнь деревянного моста будет длиться дольше, при условии обеспечения защиты от дождя при химической обработке. Композитные изделия из армированных волокном полимеров (FRP) являются одним из новейших материалов, которые используются для строительства мостов. Они весят на восемьдесят процентов меньше, чем сталь, но такие же прочные и долговечные. В то время как даже самый сильный стальной или железобетонный мост потребует значительного технического обслуживания на протяжении многих лет, FRP практически не требует заботы. Такие мосты не подвержены коррозии в морской воде, что делает их превосходным выбором для строительства в воде.

Большие данные в строительстве работают с объемами информации, которые нельзя обработать средствами традиционных СУБД. Как правило, это объемы, превышающие один петабайт. Основной идеей работы с таким объемом информации является параллельная обработка (один из видов обработки информации, когда несколько операций могут выполняться одновременно). Variety - многообразие. Технологии Big Data работают с различными данными. Это могут быть как структурированные отчеты фирм по необходимым материалам строительства мостов, так и аудио-видео файлы, изображения,

информация из социальных сетей (отзывы, комментарии). Velocity - скорость. Когда говорят о скорости, подразумевают не только скорость и быстроту строительства объекта, но и скорость реакции на ее изменение. Технологии Больших Данных быстро реагируют на изменения уже имеющихся данных и корректируют результат с поправкой на них. В идеале - в режиме реального времени. Veracity - достоверность. При строительстве мостов, в таком объеме данных, с которыми работают Big Data, особую ценность представляет именно отделение достоверной информации от недостоверной, так как от этой информации зависит правильность принятых решений. Value - ценность. Собираемые данные должны оправдывать затраты на их обработку и анализ при строительстве даже самого маленького моста. Информация должна быть ценной и полезной. Сегодня к большим данным обратился малый и средний бизнес. С дальнейшим проникновением информационных технологий в строительство, производство, бизнес-среду и повседневную жизнь каждого из нас подлежащие обработке информационные потоки продолжают непрерывно расти.

Наиболее характерной стороной современности является непрерывное развитие сети постоянных путей сообщения. Транспортные связи обеспечивают нужды промышленности, сельского хозяйства и торговли, способствуют международному разделению труда и сближению между странами и народами. Железные и автомобильные дороги, нефтепроводы и газопроводы прокладываются в различных географических и климатических зонах, преодолевают реки, горные хребты, тайгу и болота. Новые дороги и трубопроводы во многих случаях проходят по застроенной местности, сооружаются в городах и пересекают существующие железные и автомобильные дороги. Прокладка транспортных трасс через любые препятствия обычно сопряжена с необходимостью выполнения сложных строительных работ. Поэтому именно мосты, представляют собой инженерные сооружения, позволяющие проложить транспортную магистраль над встретившимися препятствиями.

Известны разнообразные типы мостов, различающиеся по назначению, конструкции и архитектуре. Компонировка и конструкция моста в значительной степени определяется характером препятствия, преодолеваемого транспортной магистралью в месте постройки сооружения. Чаще всего мосты строят через водные преграды: реки, озера, морские проливы. Объем работ и стоимость строительства определяются размерами моста, топографией и геологией местности, режимом водотока под сооружением. При прокладке дорог наибольшее число мостов обычно приходится строить через многочисленные малые реки, которые пересекают трассу дороги. Когда расчетные расходы воды, которые должны быть пропущены под сооружениями, невелики, то при

ограниченной высоте насыпи мосты заменяют водопропускными трубами, укладываемыми в тело насыпи. Земляное полотно дороги при этом полностью не прерывают, что позволяет получить экономию в стоимости и сократить сроки строительства. Размеры труб стандартизированы, а их конструкция допускает широкое использование строительных механизмов при монтаже.

В настоящее время широко применяют железобетонные и металлические трубы. Сложность и ответственность расчетов при проектировании любого сооружения существенно зависит от его масштабов. При увеличении линейных размеров быстро возрастает собственный вес и масса отдельных элементов. Постоянная нагрузка и динамические силы, развивающиеся при вынужденных колебаниях, начинают играть преобладающую роль при проверках прочности и устойчивости частей сооружения. Особого внимания заслуживают вопросы проектирования мостов больших пролетов. Само понятие «мост большого пролета» требует некоторого уточнения и должно связываться с системой и типом сооружения. К малым и средним обычно относят мосты, имеющие пролеты от ста до ста пятидесяти метров. Эта условная граница имеет в виду главным образом балочные и арочные системы. Для современных больших висячих мостов характерны пролеты шестьсот – семьсот метров и более.

При проектировании мостов больших пролетов возрастает роль расчетов на устойчивость и динамических расчетов, связанных с воздействием ветровых и сейсмических сил, а также расчетов по уточненным расчетным схемам. Особое значение приобретает также вопрос об использовании типов мостовых конструкций и материалов с надлежащими прочностными и весовыми характеристиками, выбранными с учетом климатических условий в месте расположения сооружений.

Кластеризация – разбиение множества объектов на группы (кластеры), основываясь на свойствах этих объектов [2]. Кластер мостов представляет собой группу объектов, имеющих общие признаки. Целью алгоритмов кластеризации является создание классов, которые максимально связаны внутри себя, но различны друг от друга. Поэтому характеристиками кластера мостов можно назвать два признака: внутренняя однородность – документы внутри одного кластера должны быть максимально схожи между собой; внешняя изолированность – документы из одного кластера должны быть как можно меньше схожи с документами из другого кластера [3].

Для этапа кластеризации характерно отсутствие каких-либо различий, как между переменными, так и между записями. Ищутся группы наиболее близких, похожих записей. Методы автоматического разбиения на кластеры редко используются сами по себе, просто для получения групп схожих объектов. Анализ только начинается с разбиения на кластеры, чем оно вызвано. Громадное

достоинство кластерного анализа в том, что он позволяет производить разбиение объектов не по одному параметру, а по целому набору признаков. Так же, кластерный анализ, в отличие от большинства математическо-статистических методов, не накладывает никаких ограничений на вид рассматриваемых объектов и позволяет рассматривать множество исходных данных практически произвольной природы. Это имеет большое значение для строительства мостов и прогнозирования конъюнктуры, при наличии разнородных показателей, затрудняющих применение традиционных эконометрических подходов. Кластерный анализ позволяет рассматривать достаточно большой объём информации и резко сокращать, сжимать большие массивы информации, делать их компактными и наглядными. Задача кластеризации состоит в разделении исследуемого множества объектов на группы «похожих» объектов называемыми кластерами. Слово кластер английского происхождения (cluster), переводится как сгусток, пучок, группа. Решение задачи разбиения множества элементов на кластеры называют кластерным анализом. В задаче кластеризации отнесение каждого из объектов данных осуществляется к одному (или нескольким) из заранее неопределённых классов. Разбиение объектов данных по кластерам осуществляется при одновременном их формировании. Определение кластеров и разбиение по ним объектов данных выражается в итоговой модели данных, которая является решением кластеризации и строительством необходимого моста. В данную серию включают средства массово-параллельной обработки неопределённо структурированных данных, прежде всего, решениями категории NoSQL, алгоритмами Map-Reduce, программными каркасами и библиотеками проекта Hadoop.

По пропускаемой нагрузке мосты делятся на железнодорожные, автомобильные, метромосты, пешеходные, велосипедные, комбинированные. Мосты различают по материалу, из которого выполнены пролетные строения - деревянные, железобетонные, металлические. А так же по основной статической схеме: рамные, балочные, арочные, комбинированные, висячие. Мосты подразделяются по уровню расположения проезда относительно конструкций пролетного строения - езда понизу, поверху, посередине.

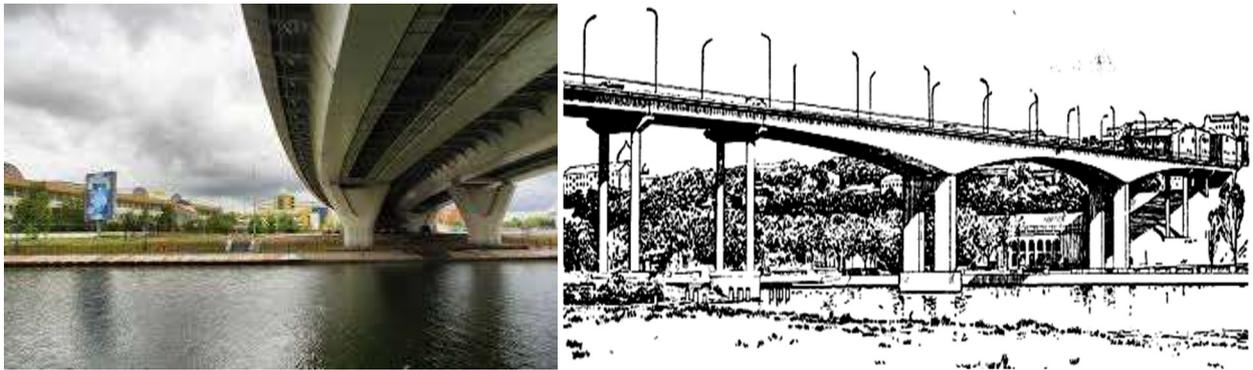


Рисунок 1 – Городской мост

В районе Евразийского университета открыта развязка, связывающая улицы А. Иманова, Л. Гумилева и К. Мунайпасова. Теперь мосты на проспектах Манаса и Абылай-хана значительно разгрузятся. Великолепным открытием стал мост М-2, достойное продолжение М-1 и других начинаний, связанных со строительством различных переправ. Сейчас автомобилисты получили возможность поворачивать налево прямо из микрорайона Молодежного, не выезжая на проспект Республики. Проект расширения улицы Гастелло удалось реализовать, и современной скоростной магистралью соединился аэропорт и район старого торгового рынка. Это, очень здорово для жителей и гостей столицы. Главное, что мост М-2 уже работает. Закончили его раньше срока. Генеральным подрядчиком выступило АО "Трансстроймост". Ширина шести полосного моста составляет почти 39 метров, длина — 362 метра. Объем инвестиций в строительство моста составил четыре миллиарда тенге. Особенное отличие моста — его покрытие. Это тонкослойное покрытие из смеси метилметакрилата и эпоксидной смолы. Эта технология позволяет делать покрытие толщиной всего 15 миллиметров. В то время как толщина асфальта в 6—7 раз больше. Соответственно и вес тоже. Новое покрытие не только легче (это позволяет строить более длинные и прочные мосты), но и надежнее. Строители, сдавшие мост в эксплуатацию дают гарантию на семьдесят пять лет. Кроме того, покрытие имеет характерную шероховатость, которая позволяет обеспечить дополнительное сцепление автомобилей с трассой. Время не стоит на месте, оно стремительно движется вперед. И за Акордой уже строится следующий мост. На основании генерального плана развития города Астаны до 2030 года предусмотрено размещение 30 мостов к уже существующим семнадцати, - сообщили в управлении пассажирского транспорта. Именно сколько необходимо построить автомобильных и пешеходных мостов в Астане.

Создание образа моста относится к сфере деятельности архитектора. В образе сооружения автор выражает свою творческую позицию, свое понимание стоящей перед ним проблемы. Поиск художественного образа

представляет собой главную задачу архитектурного проектирования. Примером выразительного образа мостового сооружения являются балочные мосты, образную характеристику которых можно определить как движение формы над препятствием. Необходимо отметить дизайнерский метод. Дизайн, или художественное конструирование, - это эстетическое и функциональное оформление предметной среды [4]. Применительно к мостам художественное конструирование - это проектирование и размещение элементов конструкции в пространстве при комплексном решении инженерных и эстетических задач. Выделение дизайна из общей системы методов архитектурного проектирования носит весьма условный характер, однако сделать это позволяют следующие его особенности. Первая состоит в безусловном соблюдении принципа наибольшей конструктивной и функциональной рациональности. Любые отступления от этого принципа, с целью достижения художественного эффекта, уже не являются художественным конструированием. Вторая и главная особенность, которая определяет художественное конструирование, состоит в том, что архитектор работает с архитектурными формами, в то время как дизайнер может иметь дело с конструкцией. Архитектору необходимо решить задачу соотношения моста и окружающей среды. Под окружающей средой понимается как естественная природа, так и самая разнообразная городская застройка. Поиск наиболее удачного соотношения моста со сложившимися городскими районами не подчиняется точным законам. Вопрос о том, как они должны соотноситься между собой, в разное время решался по-разному в зависимости от преобладающих эстетических взглядов.



Рисунок 2

Кластеризация Big Data — результат технической революции. Необходимость в аналитической работе с большими данными заметно изменило лицо ИТ-индустрии и стимулирует появление новых программных и аппаратных платформ [5]. Для анализа больших объемов данных применяются самые

высокотехнологические методы: искусственные нейронные сети — модели, построенные по принципу организации и функционирования биологических нейронных сетей; методы предиктивной аналитики, статистики и Natural Language Processing (направления искусственного интеллекта и математической лингвистики).

За последние годы в целом по Казахстану внимание уделяется формированию транспортных путей сообщения. Число мостов, их размеры считаются показателем сложности любой дороги. Реку не обойдешь и не объедешь, её нужно пересекать. Поэтому мосты необходимо строить очень прочно, надёжно, качественно, чтобы исправно несли свою службу в любой день и час. Спроектировать мост подвластно только профессиональным мостостроителям, только они могут искать и находить революционные технические решения.

Литература:

1. Аджия Сасси Перино Джорджо Фараджана Мосты / перевод с англ.
2. И. Бочкова.- Балашиха: Изд. ООО Астрель. 2004. – 176 с.
3. Котов А., Красильников Н. Кластеризация данных. 2006.
4. Казиев Г.З., Курдюков В.В. Международная научно-практическая конференция «Интеграция отечественной науки в мировую: структурные преобразования и перспективные направления». Санкт-Петербург 2016. - с.113
5. Пунин А.Л. Эстетические проблемы мостостроения /А.Л. Пунин // Вест.мостостроения. – 1998. - № 3-4. – С. 5-6.
6. Jain A., Murty M., Flynn P. Data Clustering: A.Review. // ACM Computing Surveys. 2009. Vol. 31, no. 3.