

НАПРАВЛЕНИЕ «БОЛЬШИЕ ДАННЫЕ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, ФИНАНСОВЫЕ ТЕХНОЛОГИИ И КИБЕРБЕЗОПАСНОСТЬ»

Задание №1. «Прогнозирование оттока пользователей»

Разработчик задания: ГАОУ Центр поддержки одаренных детей «Стратегия»

Компании, работающие с клиентами, заинтересованы в увеличении клиентской базы. Большое количество постоянных клиентов свидетельствует о качественно выстроенных бизнес-процессах в отношении клиентов. Постоянная клиентская база и увеличение ее объема свидетельствует о развитии компании, ее стабильном экономическом положении. В то же время, при наличии более лучших предложений, клиентам свойственно переходить в другие компании, прекращать рано или поздно пользоваться услугами той компании, в которой они изначально были. При оттоке клиентов возможно привлекать новых клиентов, но это, как правило, более дорогостоящий процесс, чем удержание имеющихся. Удерживать клиентов необходимо заблаговременно до решения прекращения договора с компанией - в противном случае, предпринятые меры не помогут их сохранить. Также нужно уметь определять именно тех пользователей, которые могут расторгнуть отношения с компанией в ближайшее время. *Адресное удержание и прогноз оттока с солидным горизонтом* позволит решить задачу оптимально с точки зрения стабильности бизнеса.

Анализ поведения клиентов проводится с целью оценки ключевых характеристик аудитории (понять пользователей). Для этого описывается целевая аудитория, способы привлечения пользователей, делается прогнозирование оттока клиентов и предлагаются способы удержания.

Как узнать, есть ли отток? Используются базовые метрики. Доля возвращаемости RR Return Rate – доля аудитории, которая уже пользовалась сервисом ранее в определенном временном периоде, и решила вернуться. Доля оттока Churn rate - доля пользователей, которые перестали пользоваться сервисом (количество ушедших пользователей / количество всех пользователей).

Какую модель использовать? Предлагается использовать модель бинарной классификации и небольшой горизонт прогнозирования (2 недели).

Какие взять данные для выполнения задания? Для выполнения задания вам будут предоставлены обезличенные данные пользователей услуг образовательного Центра.

Проектное задание

Необходимо проанализировать наличие оттока в течение года, формализовать модель прогнозирования оттока по имеющимся данным, провести прогноз оттока пользователей по обучающей выборке за выбранный период, провести анализ качества модели.

Статьи, ссылки

<https://habr.com/company/plarium/blog/345310/>

https://courses.openedu.ru/courses/coursev1:mipt+DATA_AN+fall_2018/courseware

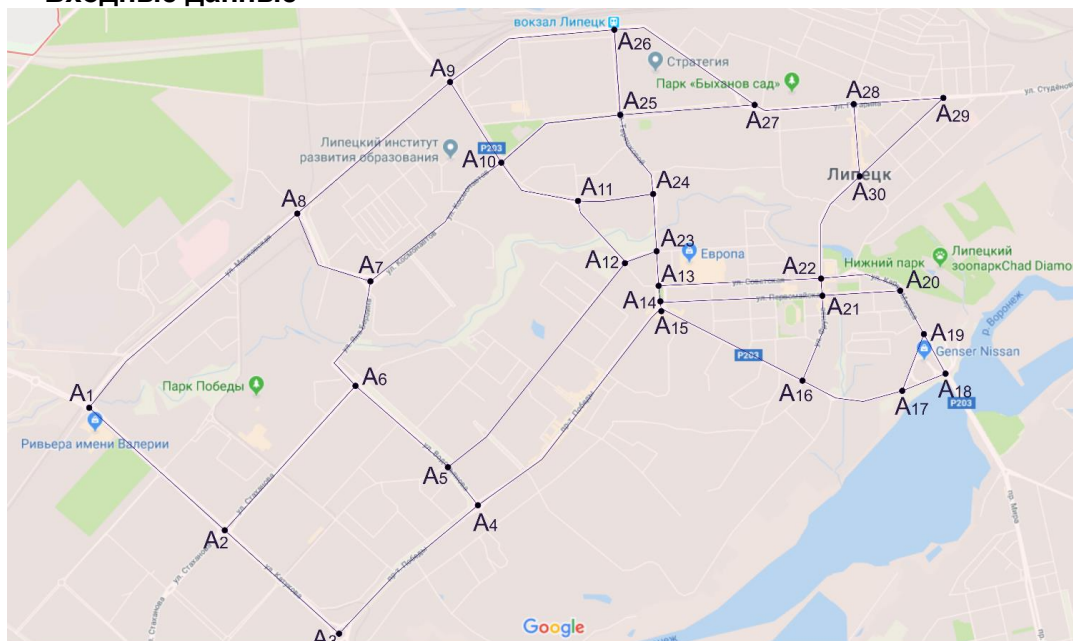
НАПРАВЛЕНИЕ «БОЛЬШИЕ ДАННЫЕ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, ФИНАНСОВЫЕ ТЕХНОЛОГИИ И КИБЕРБЕЗОПАСНОСТЬ»

Задание № 2. «Моделирование и оптимизация транспортной системы Липецка»

Разработчик задания: **ФГБОУ ВО «Елецкий государственный университет им. И.А. Бунина»**

Большинство задач, относящихся к анализу данных, в конечном счете сводится к решению задач оптимизации. К ним относится и задача оптимизации транспортной системы. Рост числа транспортных средств на дорогах ведет к необходимости развития транспортной сети, выбора режимов движения на светофорах, выбора параметров модифицируемых дорог. Математическое моделирование, в том числе имитационное, позволяет заранее спрогнозировать и описать состояние транспортной сети и дать ответ на вопрос, в каком именно направлении необходимо ее развитие. Для этого необходима разработка адекватных математических моделей процессов взаимодействия участников дорожного движения с такими элементами транспортной инфраструктуры, как дорожная сеть, системы организации дорожного движения, системы управления движением. На примере конкретной транспортной сети (или ее части) планируется с помощью аппарата имитационного моделирования решить задачу поиска равновесного распределения потоков по путям и рассмотреть возможные сценарии, связанные, например, с эффектами, возникающими при перекрытии дорог, росте нагрузки на конкретные участки путей и т.д.

Входные данные



Перед вами – упрощенная схема транспортной сети г. Липецка.

Транспортная система задается в виде ориентированного графа. Информацию о наличии дорог с односторонним движением можно взять на [Яндекс карте Липецка](#). Каждое ребро имеет параметр «время прохождения», определяемый, к примеру, формулой p/q , где p – коэффициент пропускной способности участка сети; q – плотность потока.

Проектное задание

Необходимо разработать алгоритмы решения задач оптимизации и их компьютерную реализацию.

Расчет равновесного распределения потоков по путям и изучение эффективности различных сценариев возможного изменения имеющейся сети дорог.

Задача 1. Создание имитационной модели транспортной системы г. Липецка (в виде ориентированного графа).

Задача 2. Реализация алгоритма поиска кратчайшего маршрута между любыми двумя вершинами (принадлежащими одному связному подграфу).

Задача 3. Создание имитационной модели, позволяющей определять плотность потока на каждом участке транспортной сети.

Задача 4. Включение в имитационную модель сценария перекрытия участка сети (исключения ребра) и пересчета плотности потока во всей системе.

* Задача 5. Включение в имитационную модель возможности изменять параметры транспортной системы (например, изменение характера движения на участках – одностороннего/двустороннего; режима работы светофоров; строительство новых участков пути).

Решение данного задания предполагает доклад и презентацию

Теоретическая база

Теория графов

Программирование (произвольный язык)

Теория игр

Теория массового обслуживания

НАПРАВЛЕНИЕ «БОЛЬШИЕ ДАННЫЕ, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, ФИНАНСОВЫЕ ТЕХНОЛОГИИ И КИБЕРБЕЗОПАСНОСТЬ»

Задание № 3. «Преобразование текста в структурированные признаки производства»

Разработчик задания: ПАО «Новолипецкий металлургический комбинат»

Термин «Большие данные» достаточно сложен и многогранен. Определение «больших данных» предполагает не только обработку значительных объемов информации. При работе с большими данными не стоит забывать о двух важных аспектах: многообразии данных, причем как хорошо, так и не полностью структурированных; и стремительный прирост поступления информации из разных источников.

Многие организации, такие как общественные, финансовые, производственные начали накопление информации еще до возможности оцифровки. Еще до появления первых компьютеров сбор данных выполнялся на бумажных носителях: картотеки, справочники, архивы были выдающимися для своего времени накопителями информации. Позже, появление персональных вычислительных машин внесло свой вклад: сбор данных стал возможным в большем объеме. Однако при увеличении массовости накопления данных зачастую не была решена одна проблема – проверка корректности данных и полный разбор вносимой информации. Данные, в том числе и о производственном процессе, могут быть внесены в произвольной форме в виде примечания или комментариев. Впоследствии, при возникновении необходимости, требуется проработка автоматизированных решений для проведения анализа текста и получения реальных, или по-другому, оцифрованных признаков производства.

Проектное задание

Участникам предлагается поучаствовать в процессе разбора текста и выделения признаков хода производства готовой стали.

Для примера предлагается набор комментариев, вносимых сталеварами во время выплавки стали (около 2000 плавов), для которых требуется выполнить лингвистический разбор с целью выделения признаков или выполняемых операций, которые можно выделить из комментариев. Следует отметить, что текст может содержать ошибки/опечатки.

Чтобы быть лучше знакомыми с предметной областью, проведем разбор одной из строк вручную:

0 .. трубок использовали для прожигания стального шва. Тликв.= 1527. Плавка с Аргона. Разлита полностью. Машина 11 погружной стакан КсС2ц5Z-60 партия: 121403-337. Машина 11 на 26 метрах замена воронки V. р. - 0,8 м/мин. (3-й сляб).

0 .. трубок использовали для прожигания стального шва – для прожигания шибера (устройство в днище для слива стали) стального шва использовано 0 трубок.

Тликв. – температура ликвидуса, температура полного плавления стали

Плавка с Аргона – плавка получена с установки продувки аргоном

Машина 11 на 26 метрах замена воронки V. р. - 0,8 м/мин. (3-й сляб) – разливка плавки производилась на установке номер 11. На двадцать шестом метре отлитого слитка проводилась замена промежуточной воронки, скорость разливки при этом снижена до 0,8 м/мин.

Еще несколько дополнительных примеров:

5 трубок использовали для прожигания стального ковша Тликв.=1519. Разлита полностью. Плавка с обработкой на УПК. На переходе с пл.74437 на пл.74440 перековшовка.

Плавка с обработкой на УПК – плавка обработана на установка «печь-ковш» с возможностью подогрева стали. Другое название – АПК (агрегат «печь-ковш»)

На переходе с пл.74437 на пл.74440 перековшовка – при переходе с плавки на плавку выполнена замена промежуточного ковша (п/к)

.0. трубок использовали для прожигания стального ковша Тликв.=1517 Разлита полностью. 5-е слябы +100 мм. так как на переходе на плавку 74274 прожигание, открытая струя.

5-е слябы +100 мм. – длина слябов (пятых, последних в плавке) увеличена на 100 мм.

Так как на переходе на плавку 74274 прожигание, открытая струя – при переходе на следующую плавку происходило прожигание шибера стального ковша, разливка началась без установки защитной трубы.