

КОНЦЕПЦИЯ КУРСОВОГО ПРОЕКТА

по дисциплине «Системы поддержки принятия решений»

Введение

Существует несколько определений интеллектуальных систем поддержки принятия решений (*ИСППР*), которые, в общем-то, определяются одним и тем же функционалом. В общем виде, *ИСППР* — это такая система, которая assisteрует ЛПР (Лицам, Принимающим Решения) в принятии этих решений, используя инструментарию дата майнинга, моделирования и визуализации, обладает дружелюбным интерфейсом (G)UI, устойчива по качеству, интерактивна и гибка по настройкам.

Зачем нужны СППР:

1. Сложность в принятии решений.
2. Необходимость в точной оценке различных альтернатив.
3. Необходимость предсказательного функционала.
4. Необходимость мультипоточкового входа (для принятия решения нужны выводы на основе данных, экспертные оценки, известные ограничения и т.п.).

Первые СППР (тогда еще без И) выросли из СПТ (Систем Процессинга Транзакций), в середине 60-х — начале 70-х. Тогда эти системы не обладали никакой интерактивностью, представляя собой, по сути, надстройки над РСУБД, с некоторым (совсем не большим) функционалом численного моделирования. Одной из первых систем можно назвать *DYNAMO*, разработанную в недрах MIT и представлявшую собой систему симуляции каких-либо процессов на основе исторических транзакций. После выхода на рынок мейнфреймов IBM 360 стали появляться и условно-коммерческие системы, применявшиеся в оборонке, спецслужбах и НИИ.

С начала 80-х уже можно говорить о формировании *подклассов СППР*, таких как MIS (Management Information System), EIS (Executive Information System), GDSS (Group Decision Support Systems), ODSS (Organization Decision Support Systems) и др. По сути, эти системы представляли собой фреймворки, способные работать с данными на различных уровнях иерархии (от индивидуального до общеорганизационного), а внутрь можно было внедрить какую угодно логику.

Примером может служить разработанная Texas Instruments для United Airlines система GADS (Gate Assignment Display System), которая поддерживала принятие решений в Field Operations — назначение гейтов, определение оптимального времени стоянки и т.п.

В конце 80-х появились *ПСППР* (Продвинутые — Advanced), которые позволяли осуществлять «what-if» анализ и использовали более продвинутый инструментарий для моделирования.

Наконец, с *середины 90-х* на свет стали появляться и *ИСППР*, в основе которых стали лежать инструменты статистики и машинного обучения, теории игр и прочего сложного моделирования.

Многообразие СППР

На данный момент существует несколько способов *классификации* СППР, опишем 3 популярных:

По области применения

- Бизнес и менеджмент (прайсинг, рабочая сила, продукты, стратегия и т.п.).
- Инжиниринг (дизайн продукта, контроль качества...).
- Финансы (кредитование и займы).
- Медицина (лекарства, виды лечения, диагностика).
- Окружающая среда.

По соотношению данные\модели (методика Стивена Альтера)

- FDS (File Drawer Systems — системы предоставления доступа к нужным данным).
- DAS (Data Analysis Systems — системы для быстрого манипулирования данными).
- AIS (Analysis Information Systems — системы доступа к данным по типу необходимого решения).
- AFM(s) (Accounting & Financial models (systems) — системы расчета финансовых последствий).
- RM(s) (Representation models (systems) — системы симуляции, AnyLogic как пример).
- OM(s) (Optimization models (systems) — системы, решающие задачи оптимизации).
- SM(s) (Suggestion models (systems) — системы построения логических выводов на основе правил).

По типу используемого инструментария

- Model Driven — в основе лежат классические модели (линейные модели, модели управления запасами, транспортные, финансовые и т.п.).
- Data Driven — на основе исторических данных.
- Communication Driven — системы на основе группового принятия решений экспертами (системы фасилитации обмена мнениями и подсчета средних экспертных значений)
- Document Driven — по сути проиндексированное (часто — многомерное) хранилище документов.
- Knowledge Driven — внезапно, на основе знаний. Причем знаний как экспертных, так и выводимых машинно.

Архитектура СППР

Несмотря на такое многообразие вариантов классификаций, требования и атрибуты СППР хорошо ложатся в 4 сегмента:

1. Качество.
2. Организация.
3. Ограничения.
4. Модель.

На схеме ниже (рисунок 1) покажем, какие именно требования и в какие сегменты ложатся.



Рисунок 1 – Требования к СППР

Отдельно отметим такие важные атрибуты, как масштабируемость (в подходе agile никуда без этого), способность обрабатывать плохие данные, юзабилити и user-friendly interface, нетребовательность к ресурсам.

Архитектура и дизайн ИСППР

Существует несколько подходов к тому, как архитектурно представить СППР. Пожалуй, лучшее описание разности подходов — «кто во что горазд». Несмотря на разнообразие подходов, осуществляются попытки создать некую унифицированную архитектуру, хотя бы на верхнем уровне.

Действительно, СППР вполне можно разделить на 4 больших слоя:

1. Интерфейс.
2. Моделирование.
3. Data Mining.
4. Data collection.

А уж в эти слои можно вложить различные инструменты. На схеме ниже представлено видение архитектуры, с описанием функционала и примерами инструментов (рисунок 2).

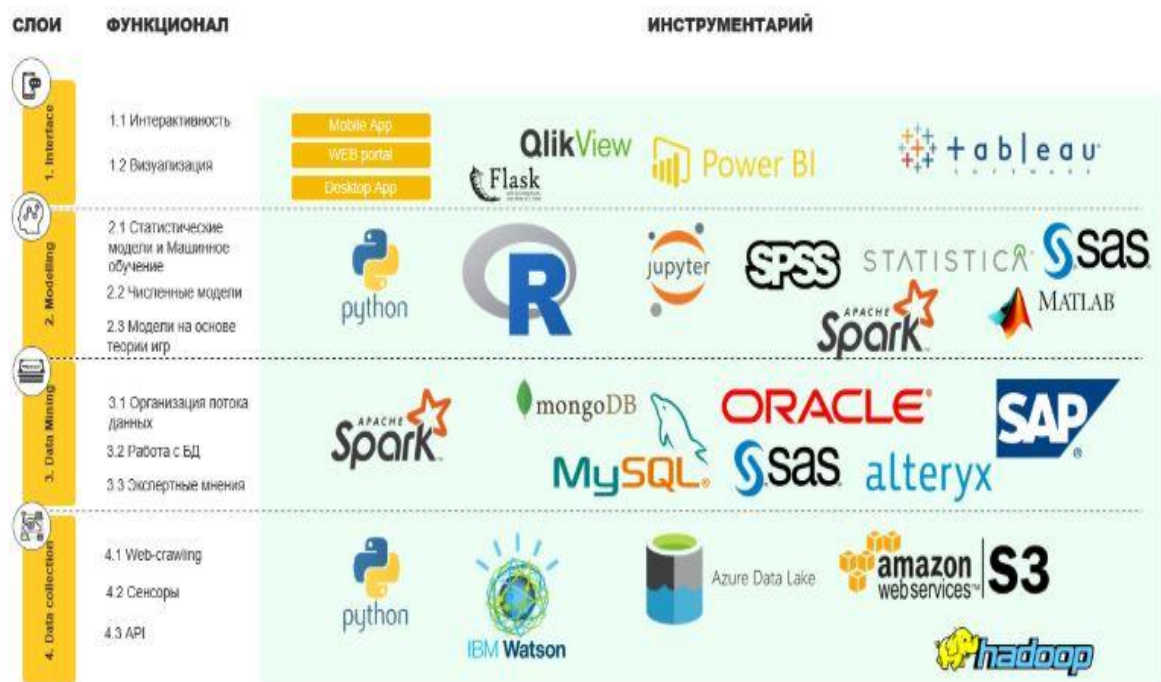


Рисунок 2 – Архитектура СППР

С архитектурой более или менее понятно, перейдем к дизайну и собственно построению СППР. При построении ИСППР необходимо придерживаться следующих шагов:

1. Анализ домена (собственно, где мы будем нашу ИСППР использовать).
2. Сбор данных.
3. Анализ данных.
4. Выбор моделей.
5. Экспертный анализ\интерпретация моделей.
6. Внедрение моделей.
7. Оценка ИСППР.
8. Внедрение ИСППР.
9. Сбор обратной связи (на любом этапе, на самом деле).

На схеме это выглядит так (рисунок 3).

Оценивать ИСППР можно двумя способами. Во-первых, по матрице атрибутов, которая представлена выше. Во-вторых, по критериальному чек-листу, который может быть любым и зависеть от вашей конкретной задачи. В качестве примера такого чек-листа можно привести следующее (рисунок 4).

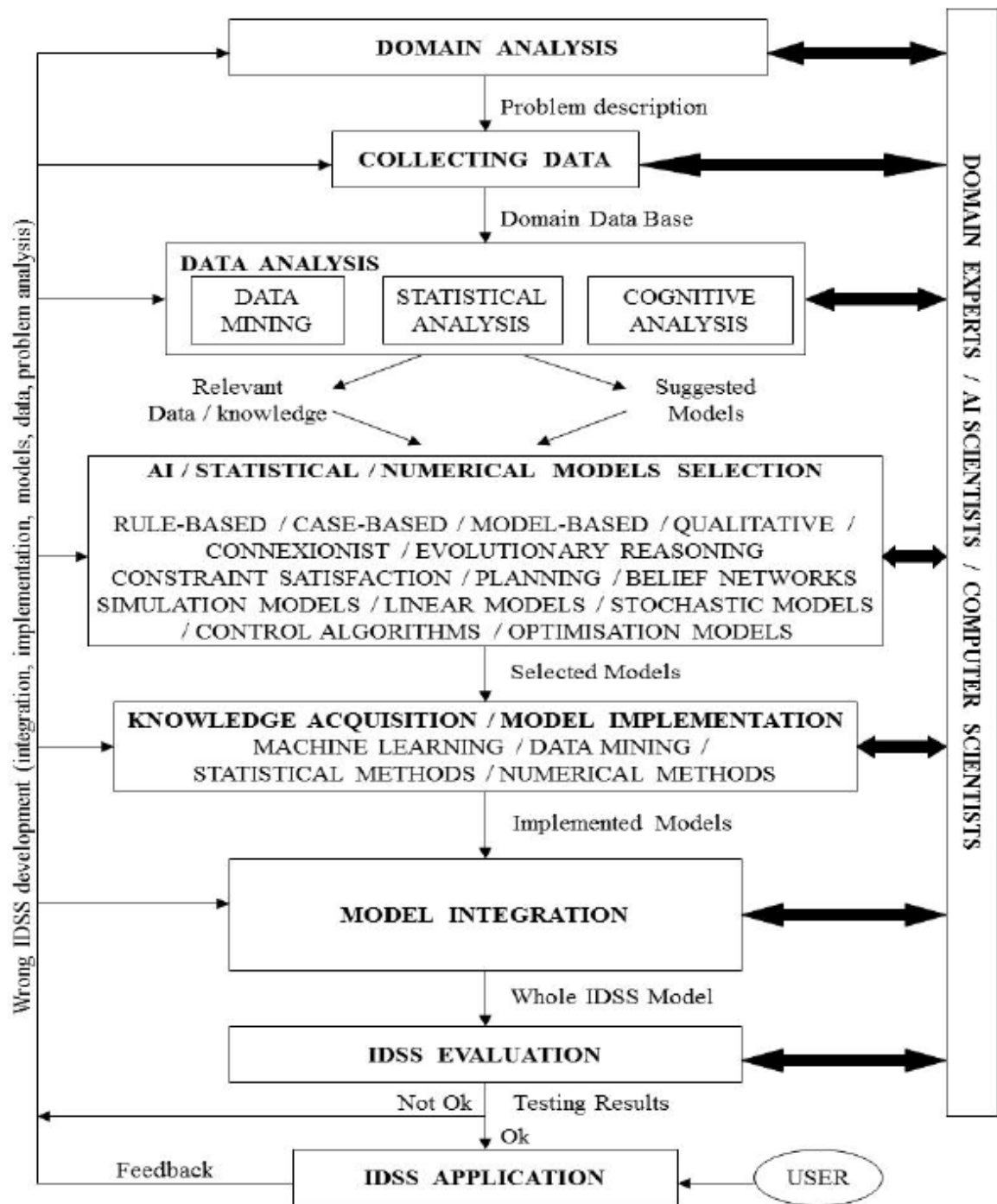


Рисунок 3 – Этапы разработки СИПР



Рисунок 4 – Чек-лист оценки СППР

Классические примеры СППР

Допустим, есть у нас задача: менеджеру по продажам стальной продукции надо еще на этапе получения заявки от клиента понимать, какого качества готовая продукция поступит на склад и применить некое управляющее воздействие, если качество будет ниже требуемого.

Алгоритм действий:

Шаг 0. Определяем целевую переменную (ну, например, содержание оксида титана в готовой продукции).

Шаг 1. Определяемся с данными (выгружаем из SAP, Access и т.п.)

Шаг 2. Собираем фичи\генерим новые.

Шаг 3. Определяем процесс data flow.

Шаг 4. Выбираем и обучаем модель, запускаем ее крутиться на сервере.

Шаг 5. Определяем *feature importance*.

Шаг 6. Определяемся со вводом новых данных. Пусть наш менеджер их вводит, например, руками.

Шаг 7. Пишем простой web-based интерфейс, куда менеджер вводит руками значения важных фич, это крутится на сервере с моделью, и в тот же интерфейс выводится прогнозируемое качество продукции.

Подобные «простые» алгоритмы также использует **IBM** в своей СППР **Tivoli**, которая позволяет определять состояние своих супер-компьютеров (Watson в первую очередь): на основе логов выводится информация по перформансу Watson, прогнозируется доступность ресурсов, баланс cost vs profit, необходимость обслуживания и т.п.

Компания **ABB** предлагает своим клиентам DSS800 для анализа работы электродвигателей той же ABB на бумагоделательной линии.

Финская **Vaisala**, производитель сенсоров для Минтранса Финляндии использует ИСППР для предсказания того, в какие периоды необходимо применять анти-обледенитель на дорогах во избежания ДТП.

Финская **Foredata** предлагает ИСППР для HR, которая помогает принимать решения по годности кандидата на позицию еще на этапе отбора резюме.

В аэропорту Дубай в грузовом терминале работает СППР, которая определяет подозрительность груза. Алгоритмы на основе сопроводительных документов и вводимых сотрудниками таможни данных выделяют подозрительные грузы: фичами при этом являются страна происхождения, информация на упаковке, конкретная информация в полях декларации и т.п.

Применение нейронных сетей

Кроме простого ML, в СППР отлично может применяться и Deep Learning.

Некоторые примеры можно найти в ВПК, например, в американской **TACDSS** (Tactical Air Combat Decision Support System). Там используются нейронные сети и эволюционные алгоритмы, помогающие в определении свой-чужой, в оценке вероятности попадания при залпе в данный конкретный момент и прочие задачи.

В немного более реальном мире можно рассмотреть такой пример: в сегменте **B2B** необходимо определить, выдавать ли кредит организации на основе пакета документов. Это в B2C вас оператор замучает вопросами по телефону, проставит значения фич у себя в системе и озвучит решение алгоритма, в B2B несколько посложнее.

ИСППР там может строиться так: потенциальный заемщик приносит заранее согласованный пакет документов в офис (ну или по email присылает сканы, с подписями и печатями, как положено), документы сканливаются в OCR, затем передаются в NLP-алгоритм, который дальше уже делит слова на фичи и сканливает их в NN. Клиента просят попить кофе (в лучшем случае), за это время как раз все и обсчитается и выведет на экран девочке-операционисту зеленый или красный смайлик. Ну или желтый, если вроде ок, но нужно больше справок богу справок.

Подобными алгоритмами пользуются также в МИД: анкета на визу + прочие справки анализируются прямо в посольстве \ консульстве, после чего сотруднику на экране высвечивается один из 3 смайликов: зеленый (визу выдать), желтый (есть вопросы), красный (соискатель в стоп-листе). Если вы когда-нибудь получали визу в США, то решение, которое озвучивает вам сотрудник консульства — это именно результат работы алгоритма в совокупности с правилами, а никак не его личное субъективное мнение о вас.

В тяжелых доменах известны также СППР на основе нейронок, определяющие места накопления буфера на производственных линиях, Общие Нечеткие Нейронные Сети на основе мин-макса (GFMMNN) для кластеризации потребителей воды и другие.

Вообще стоит отметить, что NN как нельзя лучше подходят для принятия решений в условиях неопределенности, т.е. условиях, в которых и живет реальный бизнес. Алгоритмы кластеризации также хорошо вписались.

Примерные темы курсовых проектов (работ)

1. Интеллектуальная СППР оптовой торговли (например: прогноз продаж, сегментация клиентов, ассортиментная политика, анализ потребительской корзины, программы лояльности, оценка эффективности маркетинговых действий, анализ аномалий, пресечение мошеннических действий персонала).
2. Интеллектуальная СППР для интернет-бизнеса (например: построение рекомендательных систем для персонализации пользователей веб-сайтов с целью повышения лояльности покупателей, и, как следствие, повышение продаж, выявление случаев мошенничества и т.д.).
3. Интеллектуальная СППР для телекоммуникационного бизнеса (к примеру, анализ доходности и риска клиентов).
4. Интеллектуальная СППР для промышленного производства (прогнозирование качества производимого изделия в зависимости от измеряемых параметров технологического процесса, планирование ремонтов).
5. Интеллектуальная СППР для медицины (диагностика заболеваний, оценка диагностических тестов, выявление побочных эффектов).
6. Интеллектуальная СППР для банковской деятельности (наиболее распространенная задача — кредитный скоринг, прогноз остатков на счетах и др.).
7. Интеллектуальная СППР в области энергетики (главная задача — прогноз потребления электроэнергии).
8. Интеллектуальная СППР страхового бизнеса.
9. Интеллектуальные СППР государственных учреждений.
10. Интеллектуальная СППР в маркетинговых, исследовательских агентствах.

Содержание курсового проекта (работы)

1. Введение.
2. Анализ предметной области. Постановка задачи.
3. Сбор, анализ и подготовка данных.
4. Разработка модели данных.
5. Разработка модели/прототипа СППР на основе выбранного подхода.
6. Анализ результатов.
7. Заключение.
8. Список использованных источников.