**Анализ данных и машинное обучение в авиации**

С развитием информационных технологий, объем обрабатываемых данных непреклонно растет. Объем информации стал настолько велик, что обработка всей информации стандартными программно-аппаратными средствами стал крайне неэффективен. Другими словами, большие данные (Big Data) – стал проблемой. По большей части, анализа и обработки данных. Проблема хранения и структурирования рассматриваться в данной статье не будет.

Под термином интеллектуального анализа данных подразумевается конкретный набор методик и технологий, повышающих эффективность обработки данных. На данный момент, исходя из предложений на рынке, основным средством является система распределенных вычислений, где обработка больших объемов информации требует для себя кластер высокопроизводительных вычислительных систем. Такие системы используют следующие методы анализа:

* Методы класса Data Mining – “обучение” системы некоторым ассоциативным правилам, принципам и категориям новых данных на основе уже проанализированных данных. Сюда же относятся кластерный и регрессионный анализ;
* Краудсорсинг – использование вычислительных мощностей некоторого круга привлеченных людей на основании публичной оферты, но не связанных трудовым договором;
* Машинное обучение – использование обучаемых моделей (с учителем и без), основанных на статистическом анализе и извлекающих знания из конкретного набора данных;
* Искусственные нейронные сети, генетические алгоритмы;
* Имитационное моделирование;
* И др.;

В данной статье, рассматриваются такие методы анализа больших данных как “Data mining” (рус. интеллектуальный анализ данных) и “Machine learning” (рус. машинное обучение), из-за их активного развития и эффективности.

1. **Машинное обучение**

Характерной чертой современных АСУ, является степень автоматизации процессов управления. Существует множество задач, в частности, в управлении воздушным движением, решение которых, в большинстве случаев, производиться классическими численными методами. Но большинство этих методов имеют важное ограничение, а именно, их использование возможно только при наличии готовой математической модели. Так, например, невозможно точно формализовать некоторые действия операторов систем УВД и других автоматизированных систем из-за когнитивных процессов, происходящих при оценке оператором текущей обстановки. Исходя из этого, актуальным вектором развития, видится, создание подсистем “помощников” или “советчиков” по принятию решений. Наиболее современные системы управления, уже сейчас, предлагают сервисы для автоматизированной генерации ситуационных сообщений и подсказок для оператора АСУ, та же функциональность присутствует и в системах диагностики для инженера технической эксплуатации.

Данные системы, для полноценного функционирования, предполагают наличие информационной базы. К примеру, при 6-ти часовом перелете на самолете Boeing 737, с помощью множества датчиков и сенсоров, генерируется, примерно, 240 TB данных. Если взять во внимание всю авиационную инфраструктуру, то поступающая от нее информация представляется в виде крупномасштабной и неструктурированной смеси различных данных, которые могут разделяться на следующие типы:

* Радиолокационные;
* Метеорологические;
* Географические;
* Технические;
* Текстовые;
* Другие.

Все эти данные используются для увеличения эффективности грузоперевозок, безопасности и потребительской лояльности. К примеру, использование машинного зрения для решения метеорологических задач.

Реализация таких систем, в большинстве случаев, основана на использовании методов машинного обучения. Выделяют два вида обучения:

* Индуктивное – выявление закономерностей в эмпирических данных и на основе предыдущих вариантах решения множества сходных задач;
* Дедуктивное – имитация поведения эксперта в некоторой области, на основе формализованных знаний, с помощью запрограммированных процедур принятия решений.

Второй вид относится к экспертным системам, которые, в рамках данной статьи не рассматривается.

Найти применение для такой функциональности можно, например, в системах “CPDLC”, которые предназначены для обмена сообщениями между пилотом ВС и диспетчером. Основополагающие требования для данных систем будут выглядеть, примерно, следующим образом:

* Интеллектуальный анализ данных о текущей воздушной обстановке;
* Формирование сообщений исходя из текущей ситуации;
* Представление сообщений в виде подсказки оператору;
* Корректировка результата оператором, в случае необходимости;
* Обеспечение приема и передачи сообщений провайдеру.

Основными источниками информации служат подсистемы первичной и вторичной радиолокационной обработки, а также метеорологические телеграммы и плановые данные.

1. **Интеллектуальный анализ данных**

Под определением Data Mining принято понимать процесс выявления потенциально полезных данных или закономерностей, скрытых в больших объемах исходных данных.

“Скрытые” данные характеризуются следующим образом:

* Ранее неизвестные;
* Практически полезные;
* Интерпретируемые;
* Нетривиальные.

Использование методов интеллектуального анализа данных оправдано для решения задач связанных с поиском паттернов, групировкой объектов по ассоциативным правилам, классификацией или кластеризации объектов и т.д. Таким образом можно выделить два основных типа задач:

* Описательный – дать наглядное описание закономерностей;
* Предсказательный – предсказать случаи данных для которых еще нет.

Первый тип задач характеризуется “обучением с учителем”, при котором обучение происходит за счет обучающей выборки – подготовленных данных, содержащих входные и выходные векторы. При использовании второго типа, применяется метод “обучение без учителя”, при котором выходного вектора не существует на момент обучения и подбирается, в процессе обучения, автоматически.

Эти методы интеллектуального анализа позволяют вырабатывать решения для задач в самых разных областях, которые можно отнести к описательным или прогнозируюмым категориям проблем. Цель прогнозирующих методов состоит в построении модели путем анализа базы данных системы для прогнозирования ценности характеристики через значения других характеристик.

На текущий момент, можно найти примеры использования таких алгоритмов для решения задач связанных с моделированием и улучшением аэродинамических характеристик ВС, предсказывание траекторий воздушных судов, обнаружение потенциально-конфликтных ситуаций. В частности, одним из этапов развития организации воздушного движения является концепция свободного полета, которая позволяет пилотам воздушного судна выбирать наиболее эффективные траектории полета, а также, планировать свой полет. Данная концепция, помимо преимуществ, связанных со снижением нагрузки на центры управления воздушным движением и сокращению задержек при выполнении рейса, создает множество конфликтных ситуаций между различными воздушными судами. Обнаружение и разрешение КС представляется серьезной проблемой для УВД. Для решения этой задачи предлагается использовать один из новых видов эволюционного алгоритма, известного как империалистический конкурентный алгоритм. Задачей которого будет поиск надежной и оптимальной раскраски графа, представляющего собой некоторую связанную область скопления летательных аппаратов. Алгоритм раскраски графа является частным случаем алгоритма разметки графа, при котором каждому элементу графа присваивается, в соответствии с определенными ограничениями, метка, традиционно называющаяся “цветом”. Особо отмечается, что империалистический алгоритм, подобно другим эволюционным алгоритмам, основан на соперничестве некоторых групп объектов, в данном случае используется метафора колоний и империалистического государства, которое должно привести колонии к схождению к глобальному экстремуму целевой функции. Моделирование работы алгоритма проводилось с использованием модели случайных полетов, где все воздушные суда были вынуждены летать на одной высоте и с постоянной скоростью. Для каждого ВС, в качестве маневров избегания потенциально конфликтных ситуаций, были приняты небольшие и мгновенные изменения курса. Моделирование показало, что рассматриваемый алгоритм, в большинстве случаев, обеспечивает оптимальные решения, а в качестве оценки эффективности выступала метрика минимального отклонения от курса рассчитываемая по следующей формуле:

Где N - количество ВС, а Ei - показывает характеристики каждого воздушного судна. В таблице 1 показаны результаты моделирования для различного количества воздушных судов.

|  |  |  |  |
| --- | --- | --- | --- |
| Количество ВС | Количество узлов | Количество оптимальных траекторий полета | Эффективность алгоритма (%) |
| 2 | 2 | 1 | 99.5 |
| 6 | 6 | 3 | 98 |
| 12 | 12 | 7 | 97 |
| 30 | 30 | 15 | 96 |

Таблица 1: Оценка эффективности империалистического алгоритма

Исходя из этих данных, можно сказать что этот алгоритм решения задачи кажется эффективным и надежным для разрешения конфликтных ситуаций.

1. **Выводы**

Использование подобных методов Mashine Learning и Data Mining, перспективы их использования видятся значительно шире, отраженного в данной статье. Например некторые алгоритмы позволяют предсказать, в зависимости от стимулирующего сигнала, например сообщения OLDI о передачи борта от зоны к зоне, что необходимо принять борт и выработать подсказку в виде подготовленного сообщения для диспетчера и отправлению по каналу CPDLС определенного сообщения ВС. Решение подобных задач с помощью алгоритмов машинного обучения позволят автоматизировать выработку решений и после одобрения оператора отправить их по линии связи, что, в значительной степени снизит нагрузку последнего и уменьшит влияние человеческого фактора. Также повысится эффективность и безопасность грузоперевозок за счёт оптимизации управления воздушным движением.