Mediterra software — Algorithms portfolio

Machine learning statistics

Technologies: C++, SAS

URL: https://www.salford-systems.com

- Developing modules for statistical calculations for Salford Systems company;
- Implementation of CHAID decision tree building algorithm;
- Implementing approach of multi-threaded calculations using large datasets and multiple datasources;
- Improving and optimizing developed algorithms for better performance and resources allocation;
- Implementation a set of General Linear Model (GLM) modules using different distributions (normal, poisson, gamma, inverse gaussian) and link functions;
- Porting/onverting SAS modules to C++ modules;

Correcting markup for translated text

Technologies: C#, .NET

Duration: 6 months

Implemented an algorithm for a translation application, allowing to automatically keep HTML and other markup during translation from one language to another. The translation itself is done by a human, and is often optimized by retrieving an appropriate translation from the database of previously existing translations. The work includes:

- Developing algorithm of auto-markup of translated text based on markup of original (not translated) text;
- Profiling and optimization of existing and new auto-markup algorithms;
- Developing logic of text parsing according to specific grammar/text structure;
- Improving parsing/auto-markup functionality using parameters, knowledge base and user-experience;

Automotive Routing Quality Assessment Software

Technologies: Python, MySQL, sqlalchemy, geoalchemy, nosetests, git

Duration: 1 year

The quality evaluation system has been developed, which is now being used to make decisions about making changes in the routing service. If the quality has dropped down, then the changes are not made, if the quality gets increased, then there is a green light for deployment of new data/software making effect on the QoS. The things involved in the development:

- Yandex's mapreduce cluster computations technology
- statistical modeling, including analytical and computational approaches, bootstrapping, resampling and Monte-Carlo methods
- sqlalchemy, geoalchemy ORMs

Fantasy Football Predictive Modelling

Technologies: Python, Matlab, R

Duration: 3 months

Business goal of that project was to build a statistical model that could predict Fantasy Football players results for nearest game based on history data of points players earned in previous games and current status of player (injuries, season, player efficiency and so on). The built statistical model takes in into consideration a lot of different parameters (input variables) such as:

• Players parameters (height, weight, age and so on) and finds possible correlation between these parameters and player efficiency;

• Player statistics for previous games as well as previous season; it allows to predict how good/bad player is currently in earning points as well as current trend and possible points amount player could earn next game;

• Efficiency of all players is calculated in order to find overvalued or under valued players and compare players with each other (having a kind of rating as an output);

Among technical issues we solved were finding models that most correctly describes and predicts players efficiency (as the result specific models were built for specific player types).the project. Implemented interactions within the Charts using the jQuery library so that the planned tasks could be accomplished in order, in a timely fashion, and with the desired effect.