

С.Сейфуллиннің 125 жылдығына арналған «Сейфуллин оқулары – 15: Жастар, ғылым, технологиялар: жаңа идеялар мен перспективалар» атты халықаралық ғылыми-теориялық конференциясының материалдары = Материалы Международной научно-теоретической конференции «Сейфуллинские чтения – 15: Молодежь, наука, технологии – новые идеи и перспективы», приуроченной к 125-летию С.Сейфуллина. - 2019. - Т.II, Ч 1 - С.163-165

ИСПОЛЬЗОВАНИЕ ИНСТРУМЕНТОВ DATA MINING В ПРОГНОЗИРОВАНИИ УРОЖАЙНОСТИ

Шекеров Н.Т.

«Data mining (рус. добыча данных, интеллектуальный анализ данных, глубокий анализ данных) — собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Термин введён Григорием Пятецким-Шапиро в 1989 году.» [1-2].

Data mining, это прежде всего [2:

- наличие достаточно крупной базы данных;
- предполагается, что в базе данных находятся некие «скрытые знания».
- Необходимо разработать методы обнаружения знаний, скрытых в больших объёмах исходных «сырых» данных. В текущих условиях глобальной конкуренции именно найденные закономерности (знания) могут быть источником дополнительного конкурентного преимущества.
- Что такое «скрытые знания»? Это должны быть обязательно знания:
- ранее неизвестные — то есть такие знания, которые должны быть новыми (а не подтверждающими какие-то ранее полученные сведения);
- нетривиальные — то есть такие, которые нельзя просто так увидеть (при непосредственном визуальном анализе данных или при вычислении простых статистических характеристик);
- практически полезные — то есть такие знания, которые представляют ценность для исследователя или потребителя;
- доступные для интерпретации — то есть такие знания, которые легко представить в наглядной для пользователя форме и легко объяснить в терминах предметной области

Другими словами наличие крупной базы дает нам возможность исходить из чего либо перед решением задачи. Это значить с нуля запустить нельзя, нужна обязательная точка старта. Например, рассмотрим простую ситуацию игры змейки в рисунках.

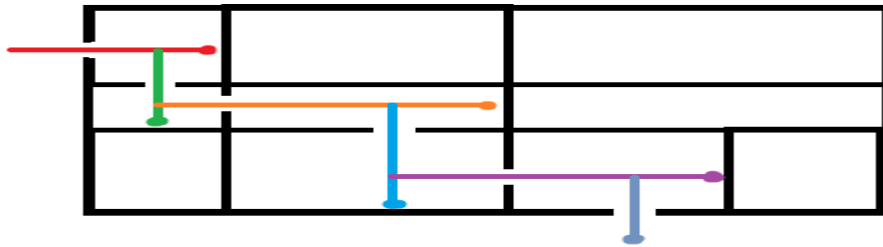


Рисунок 1. “Попытки” змейки выбраться из лабиринта.

Красным цветом показана первая попытка, в начале змейка просто идет вперед, сталкивается с препятствием и игра начинается заново. Зеленым цветом показана вторая попытка, в прежней попытке змейка в момент поворота не предприняла никаких решений, и в итоге проиграла, эти данные были сообщены и обработаны и зеленая змея в точке разворота, развернулась, в следующих попытках змея будет принимать более точные решения, так как знание прежнего передается следующей змее. Эта была ситуация принятия решений на основе случая (CBR- Case based reasoning). Именно вот поэтому наличие базы знаний обязательно для дальнейшего продвижения. В прогнозирование урожайности мы выбираем для прогноза это метод [1,4].

Итак, у нас есть задача «как увеличить урожайность?». У нас есть база, где указана разные урожайности разных времен, это будет служить нам как точка старта (напомню, что для начала нам нужна точка старта). Допустим, что есть 5 факторов влияние урожайности (их могут сколько угодно, чем больше, тем лучше и точнее результат) они:

1. Участие человеческого фактора.
2. Время посева.
3. Место посева.
4. Преимущественная погода в период роста посева.
5. Наличие пестицидов.

Построим модель задачи на программе для создания онтологий Protégé.

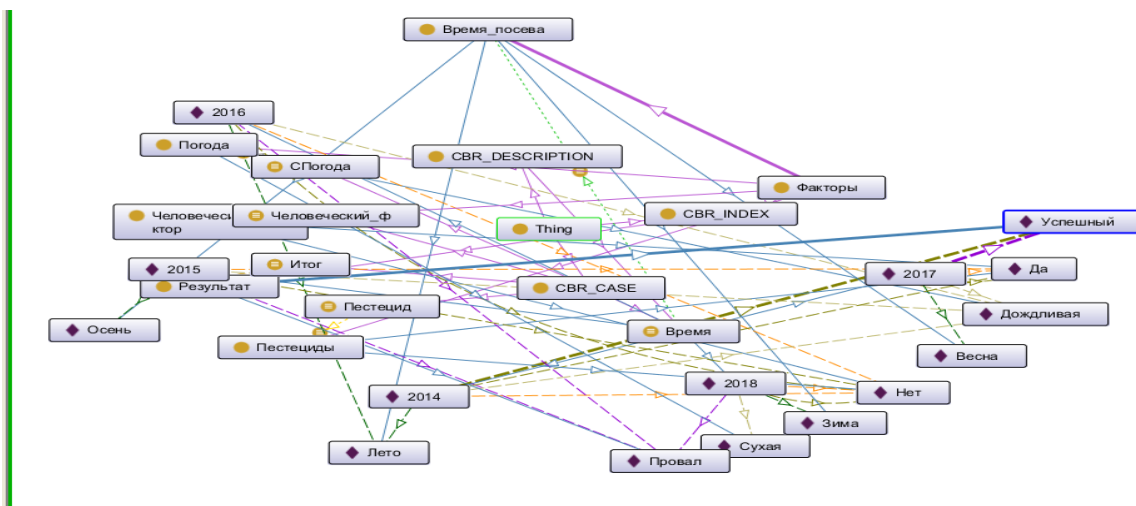


Рисунок 2. Отношения данных к результату

Теперь имея базу знаний можно приступить к прогнозу, для этого используем программу Jcolibri (Рисунок 3).

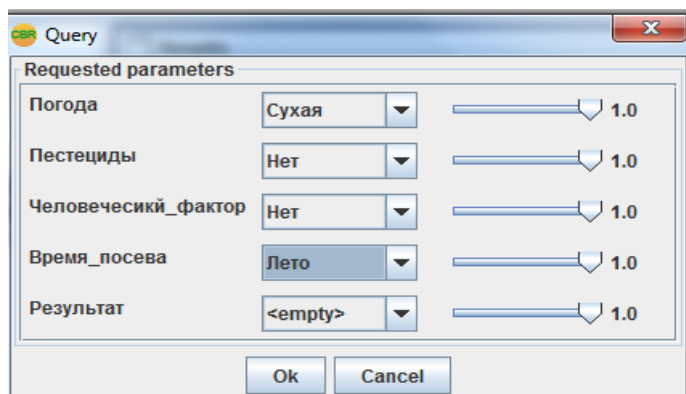


Рисунок 3. Программа для прогноза.

Список литературы

- 1.Plumbley M.D. and F.Fallside. The introduction case-based_reasoning//ACM Transactions on Information and System Security. Vol. 40, No. 1, March 2015 P.303-319
- 2.https://en.wikipedia.org/wiki/Data_mining
- 3.Предсказание на основе прецедентов// https://en.wikipedia.org/wiki/Case-based_reasoning
4. Protégé – ontology creator//<https://protege.stanford.edu/>

*Научный руководитель: Ст.
преподаватель Калдарова М.Ж*