STUDNET

ПРИМЕНЕНИЕ ГЕНЕТИЧЕСКОГО АЛГОРИТМА ДЛЯ РЕШЕНИЯ ЗАДАЧИ ОПТИМАЛЬНОГО РАСПРЕДЕЛЕНИЯ КАПИТАЛА НА ОСНОВЕ КРИТЕРИЯ КЕЛЛИ

USAGE OF A GENETIC ALGORITHM IN CASE OF OPTIMAL CAPITAL DISTRIBUTION BASED ON THE KELLY CRITERION

УДК 004.8

Ткачев Илья Петрович, магистрант, Санкт-Петербургский государственный экономический университет, г. Санкт-Петербург
Tkachev Ilya Petrovich, skaatt@mail.ru

Аннотация

Проблемы корректного прогнозирования ставок актуальны для спортивной отрасли. В данной статье обозначена формальная постановка задачи оптимального распределения капитала для множества одновременных спортивных событий на основе критерия Келли. Приведено решение задачи с помощью генетического алгоритма.

Annotation

Nowadays problems of correct sports bets forecasting are still relevant to the sports industry. This article describes a formal definition of the optimal capital distribution problem based on the Kelly criterion for a set of simultaneous sports events. The problem was solved using a genetic algorithm.

Ключевые слова: ставки, критерий Келли, оптимизация портфеля, генетический алгоритм, прогнозирование, вероятностные модели.
Key words: sports bets, Kelly criterion, capital distribution, genetic algorithm, forecasting, stochastic models.

1. Постановка задачи оптимального распределения капитала

На сегодняшний день оборот в отрасли спортивных ставок оценивается ежегодно от 700 млрд до 1 трлн долларов. Из них около 70 % данного бизнеса приходится на футбол. Актуальным в данной связи до сих пор остается вопрос максимизации прибыли для букмекера с точки зрения оптимального определения рисков. Аспект максимизации прибыли важен при этом и для игрока. Таким образом существует потребность в создании предсказательных моделей для удовлетворения спроса обоих сторон.

В данной статье предлагается рассмотреть возможность оптимального управление капиталом для одновременных нескольких событий, а именно долгосрочного возврата из ряда ставок на события. Предлагается

использовать целевую функцию на базе генетического алгоритма с использованием критерия Келли. Данная статья рассматривает проблему, с которой сталкиваются те, кто делает ставку на спортивные события с фиксированными коэффициентами. Игрок ожидает получить положительный выигрыш за определенный период времени (положительное математическое ожидание) в результате серии ставок. В качестве примера можно рассмотреть следующий жизненный сценарий: каждую субботу в примерно одно и то же время играют более 40 матчей. Есть вероятность, что многие из них дадут возможность сделать выгодные ставки для зрителей. Подход к оптимальному управлению капитала для событий, происходящих одновременно мог бы решить эту проблему. Решение этой проблемы также имеет широкое применение в инвестиционной стратегии и азартных играх.
Для вывода целевой функции в статье будут использоваться следующие математические обозначения, приведенные в Таблице 1.
Таблица 1. Математические обозначения, используемые в данной статье.

p_i	Вероятность события, полученная с помощью некоторой математической модели
r_i	Коэффициент букмекера на событие ($r_i > 1$)
B_t	Состояние капитала в какой-то момент времени t, где B_0 начальное значение
Q_t	$\log\left(\frac{B_t}{B_0}\right)$
x	Некоторая доля капитала игрока, которая может быть поставлена в какой-то момент времени t на конкретное событие
z_i	Полученная прибыль (убыток) с какой-либо ставки в пуле
G_i	Конкретный набор исходов для набора событий, соответствующие пулу ставок
S	Множество различных G_i
$R(S)$	Совокупный доход от некоторого набора S
$U(x_1, \dots, x_n)$	Целевая функция полезности, представляющая сумму логарифмов $R(S)$, умноженных на соответствующие вероятности

Обозначим целевую функцию, которая позволит создать вероятностные модели. Предполагается, что имеется несмещенная модель машинного обучения, которая дает оценки $\{p_i\}$ для событий. Вероятности модели можно сравнить с коэффициентами букмекеров $\{r_i\}$, логично, что $r_i > 1$, иначе ставка не имеет смысла. При $\{i: r_i * p_i > 1\}$ предполагается положительное

математическое ожидание игрока, однако в условиях риска (завышенной оценки модели, ее смещенность в отношении $\{p_i\}$ и/или заниженной оценки $\{r_i\}$ нет гарантии выполнения данного условия. Поэтому часто игроку необходимо понять, как разместить капитал для максимизации прибыли и минимизации риска. Данное условие можно отразить выражением:

$$Q_n = \log\left(\frac{B_n}{B_0}\right) = \log(B_n) - \log(B_0) \quad (1)$$

Для частного случая одного события функция полезности – это максимизация ожидаемого значения $E[Q_n]$, если обозначить $\delta Q_t = Q_t - Q_{t-1}$, обозначенное выражение будет выглядеть:

$$E[Q_n] = Max(\sum_{i=1}^{n} E[\delta Q_t]), \quad (2)$$

$$E[\delta Q_t] = p\log(1 - x + rx) + (1 - p)\log(1 - x) \quad (3)$$

Оптимальное значение x для формулы (3):

$$x = \frac{pr - 1}{r - 1} \quad (4)$$

Однако данное утверждение (4) справедливо для ситуации наличия одного единственного события в какой-то времени t.

Теперь обобщим поставленную задачу на множество событий с исходами G_i для каждого с вероятностью p_i и случайной величиной z_i, представляющий заработок, полученный с каждой ставки $z_i = r_i - 1$ для выигрышной ставки и $z_i = -1$ для проигранной. Проблема заключается в размещении долей капитала для максимизации ожидаемого роста капитала или, другими словами, максимизации логарифма полезности (log-utility function) для всех возможных совместных событий $S = \{G_1, \dots, G_n\}$.

Введем $R(S)$:

$$R(S) = 1 + \sum_{i=1}^{n} z_i(S)x_i, \quad (5)$$

где S конкретный набор результатов n событий. Тогда $R(S)$ общий доход от множества S. Таким образом общая полезность:

$$U(x_1, \dots, x_n) = Max(1 + \sum_{S} \Pr(S)\ \log\{R(S)\}) \quad (6)$$

при ограничении

$$\sum_{i \epsilon S} x_i = 1 \quad (7)$$

Ограничение (7) не допускает того, что размер пропорций капитала не превысил единицы, то есть того, что есть в распоряжении.

2. Оптимизационный генетический алгоритм решения

Данная задача принадлежит NP-классу (non-deterministic polynomial), тогда ее решение может быть найдено локально. Точные алгоритмы в данном случае уже могут не справиться с решением поставленной задачи, поэтому для корректной реализации применен эвристический метод решения – генетический алгоритм. Этот алгоритм состоит из следующих этапов:

1. Генерация. Создание новой популяции в размере 100 особей.
2. Размножение (кроссовер). На данном этапе вся популяция сортируется по возрастанию целевой функции и формируется группа из 1/3 лучших индивидов. Далее 30% от общего числа популяции скрещивается с другим решением из ранее сформированной выборки лучших индивидов. Скрещивание происходит путем добавления 40% входных данных от индивида из лучшей популяции. Если в результате скрещивания решение новой особи не допустимо (получено +inf, -inf), то процедура повторяется снова, однако уже с новым случайно выбранным индивидом из числа эталонных.
3. Мутации. Из популяции после проведенного кроссовера выбирается 20% случайных образцов, у которых «мутируют» 50% данных, причем амплитуда изменения генов распределена равномерна с параметрами $Uniform{\sim}(-1.1, +1.1)$, где 1 означает отсутствие изменений. Если в результате мутации решение новой особи не допустимо, то процедура повторяется пока не будет получено иное при других псевдослучайных параметрах амплитуды изменения геномов.
4. Отбор. «Выжившие» формируются с помощью bootstrap-sampling (выборка с возвратом) взвешенная на вероятность величины функции индивида относительно суммы всех индивидов в популяции.

3. Практическая симуляция на реальных данных

Применим сформированную функцию по обозначенному алгоритму на реальных данных. Представленный набор данных собран за период с 30.11.2019 по 5.12.2019 и состоит из букмекерских коэффициентов на спортивные события (*book_prop*), вероятности для событий, полученные логистической регрессией (p):

Таблица 1. Входные исходные данные для симуляции работы алгоритма

home_y	*away_y*	*p*	*book_prop.*
Cagliari	Sampdoria	0.41	2.24
Napoli	Bologna	0.84	1.47

Parma Calcio 1913	AC Milan	0.62	3.62
Lazio	Udinese	0.43	1.32
Juventus	Sassuolo	0.41	1.24
Inter	SPAL 2013	0.91	1.26
Verona	Roma	0.4	1.88
Espanyol	Osasuna	1	2.23
Mainz 05	Eintracht Frankfurt	0.37	2.89
Hoffenheim	Fortuna Duesseldorf	0.39	1.77
Borussia M.Gladbach	Freiburg	0.54	1.6
Liverpool	Everton	0.36	1.39
Chelsea	Aston Villa	0.93	1.31
Southampton	Norwich	0.98	1.77
Manchester United	Tottenham	0.35	2.64

В результате 100 эволюций алгоритма, были получены результаты, отраженные на Рисунке 1.

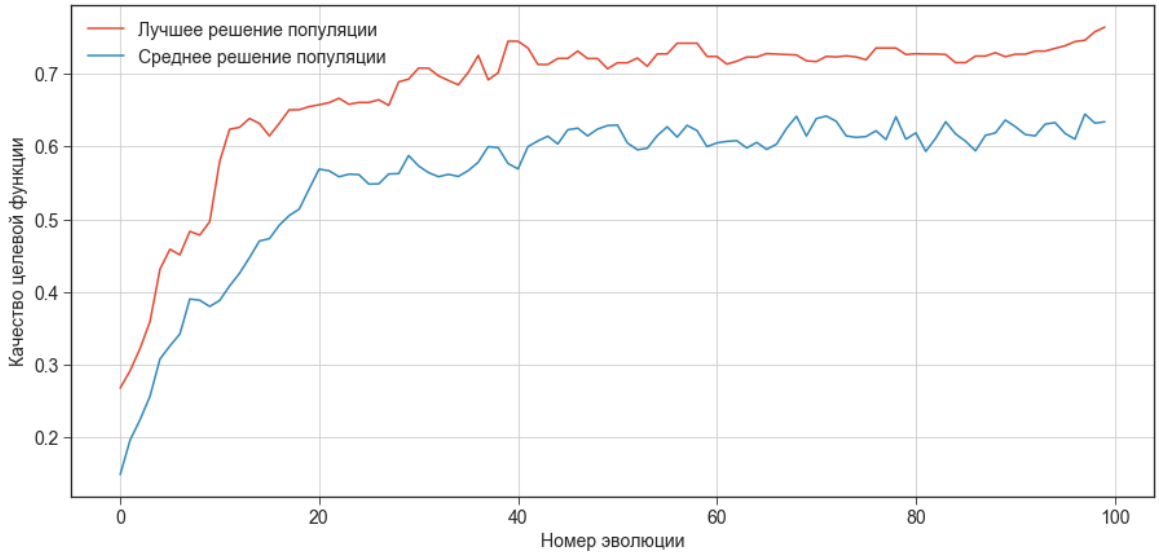


Рисунок 1. Динамика алгоритма

Исходя из полученного результата, можно заключить, что алгоритм сошелся уже примерно на 40-й эволюции, а лучшее решение было получено на 100-й эволюции, которое равно 0.75. Полученные параметры лучшего решения x отражены в Таблице 2.

Таблица 2. Результаты лучшего решения работы алгоритма

p	*book_prop.*	*x*
0.41	2.24	0
0.84	1.47	0
0.62	3.62	0.16
0.43	1.32	0.01
0.41	1.24	0
0.91	1.26	0
0.4	1.88	0

1	2.23	0.76
0.37	2.89	0
0.39	1.77	0
0.54	1.6	0.01
0.36	1.39	0
0.93	1.31	0
0.98	1.77	0.05
0.35	2.64	0

Анализируя получившиеся результаты можно сделать вывод, что алгоритм нашел такие ситуации, когда модель дала уверенное предсказание, а букмекер соответственно предложил высокий коэффициент на этот матч (низкая вероятность события). Таким образом произошло расхождение в оценке вероятности с точки зрения модели и букмекера — они получились ассиметричными (обратнопропорциональными, но не детерминированными). В качестве возможного дальнейшего направления данного исследования — улучшение полученного алгоритма за счет более точного подбора гиперпараметров для кроссовера и мутации. Результатом подобных модернизаций станет ускорение сходимости алгоритма.

Литература

1. Уитроу С. (2007) алгоритмы оптимального распределения ставок на многие одновременные события. Прикладной. Государственник. (2007) 56, Part 5, PP.607-623. Имперский колледж Лондона, Великобритания
2. Breiman, L. (1961) оптимальные системы азартных игр для благоприятных игр. В Proc. 4-й симпозиум Беркли. Математическая статистика и вероятность, вып. 1, стр. 65-78. Berkeley: University of California Press.
3. Browne, S. andWhitt, W. (1996) выбор портфеля и байесовский критерий Келли. Совр. Заявл. Вероятно., 28, 1145 – 1176.
4. Cover, T. M. and Thomas, J. A. (1991) элементы теории информации. Нью-Йорк: Уайли.
5. Crowder, M., Dixon, M., Ledford, A. and Robinson, M. (2002) динамическое моделирование и прогнозирование матчей английской футбольной лиги для ставок. Статистик, 51, 157-168.
6. Dixon, M. J. and Coles, S. G. (1997) моделирование футбольных результатов ассоциации и неэффективности на рынке ставок на футбол. Прикладной. Государственник., 46, 265–280.
7. Эдельман Д. (2000) о финансовой ценности информации. Энн. ОПС Рез., 100, 123-132.

8. Goddard, J. (2005) регрессионные модели для прогнозирования целей и результатов в профессиональном футболе. Int. Джей Прогноз., 21, 331–340.
9. Haigh, J. (2005) критерий Келли и Сравнение ставок в ставках на спред. Статистик, 49, 531-539.
10. Hakansson, N. H. (1970) оптимальные инвестиционные и потребительские стратегии в условиях риска для класса функций полезности. Эконометрика, 38, 587-607.
11. Kelly, Jr, J. L. (1956) a new interpretation of information rate. Белл Сист. Тех. J., 35, 917-926.
12. Kullback, S. and Leibler, R. A. (1951) Об информации и достаточности. Энн. Математика. Государственник., 22, 79–86.
13. MacLean, L. C., Ziemba, T. and Blazenko, G. (1992) рост и безопасность в динамическом инвестиционном анализе. Mangmnt Sci., 38, 1562–1585.
14. Паундстоун, У. (2005) Формула Фортуны: Нерассказанная история научной системы ставок, которая победила казино и Уолл-Стрит. Нью-Йорк: Хилл и Ван.
15. Sharpe, W. F. (1994) The Sharpe ratio. Дж. Пртфол. Мангмнт, 21, 49-58.
16. Thorp, E. O. (1984) The Mathematics Of Gambling. Секокус: Лайл Стюарт.
17. Thorp, E. O. (1997) критерий Келли в блэкджеке, спортивных ставках и фондовом рынке. Эдвард О. Торп И Партнеры, Ньюпорт-Бич. (Доступно от http://www.bjmath.com/bjmath/thorp/paper.htm-да.)

Literature

1. Whitrow C. (2007) Algorithms for optimal allocation of bets on many simultaneous events. Appl. Statist. (2007) 56, Part 5, pp. 607–623. Imperial College London, UK
2. Breiman, L. (1961) Optimal gambling systems for favourable games. In Proc. 4th Berkeley Symp. Mathematical Statistics and Probability, vol. 1, pp. 65–78. Berkeley: University of California Press.
3. Browne, S. andWhitt,W. (1996) Portfolio choice and the Bayesian Kelly criterion. Adv. Appl. Probab., 28, 1145 – 1176.
4. Cover, T. M. and Thomas, J. A. (1991) Elements of Information Theory. New York: Wiley.
5. Crowder, M., Dixon, M., Ledford, A. and Robinson, M. (2002) Dynamic modelling and prediction of English Football League matches for betting. Statistician, 51, 157–168.
6. Dixon, M. J. and Coles, S.G. (1997) Modelling association football scores and inefficiencies in the football betting market. Appl. Statist., 46, 265–280.

7. Edelman, D. (2000) On the financial value of information. Ann. Ops Res., 100, 123–132.
8. Goddard, J. (2005) Regression models for forecasting goals and results in professional football. Int. J. Forecast., 21, 331–340.
9. Haigh, J. (2005) The Kelly criterion and bet comparisons in spread betting. Statistician, 49, 531–539.
10. Hakansson,N. H. (1970) Optimal investment and consumption strategies under risk for a class of utility functions. Econometrica, 38, 587–607.
11. Kelly, Jr, J. L. (1956) A new interpretation of information rate. Bell Syst. Tech. J., 35, 917–926.
12. Kullback, S. and Leibler, R. A. (1951) On information and sufficiency. Ann. Math. Statist., 22, 79–86.
13. MacLean, L. C., Ziemba, T. and Blazenko, G. (1992) Growth versus security in dynamic investment analysis. Mangmnt Sci., 38, 1562–1585.
14. Poundstone,W. (2005) Fortune's Formula: the Untold Story of the Scientific Betting System that Beat the Casinos and Wall Street. New York: Hill and Wang.
15. Sharpe, W. F. (1994) The Sharpe ratio. J. Prtfol. Mangmnt, 21, 49–58.
16. Thorp, E. O. (1984) The Mathematics of Gambling. Secaucus: Lyle Stuart.
17. Thorp, E. O. (1997) The Kelly criterion in blackjack, sports betting and the stock market. Edward O. Thorp & Associates, Newport Beach. (Available from http://www.bjmath.com/bjmath/thorp/paper.htm.)