Секция «Искусственный интеллект и машинное обучение в цифровой экономике»

Эмпирический Байесовский подход для решения задачи информационного воздействия

Научный руководитель – Гринева Наталья Владимировна

Шачнева (Карпова) Анна Владимировна

Аспирант

Финансовый университет, Факультет прикладной математики и информационных технологий, Кафедра системного анализа и моделирования экономических процессов, Москва, Россия

E-mail: annavkarpova@mail.ru

Байесовский подход лежит в основе многих современных методов классификации, ранжирования и оценки. Нередко его приложение можно встретить в банковской и страховой сферах в основе скоринговых моделей. В данном случае предлагается использовать возможности Байесовского подхода к решению задачи предсказания влияния информационного воздействия на рыночные процессы. Для этого на основе статистических данных о внешних источниках информации (появлявшихся ранее в СМИ информационных поводах) и их влиянии на волатильность ценных бумаг необходимо построить классификатор[2].

Имеется простая выборка $X^l{=}(x_i;y_i)_{i=1}^l$ из распределения $p(x,y) {=}P_y p_y(x)$. Требуется построить эмпирические оценки априорных вероятностей P_y и функций правдоподобия $p_y(x)$ для каждого из классов $y{\in}Y$, затем построить алгоритм $\alpha(x)$[1].

Таким образом, для исследования предлагается взять выборку новостных поводов за 2017 год. Это легко сделать, пользуясь известными системами мониторинга средств массовой информации (например, Медиалогия или Bloomberg)

Каждый новостной повод можно охарактеризовать по ряду факторов:

1) Хорошие или плохие новости
2) Тип события: корпоративная новость[3], изменение политической или экономической конъюнктуры.
3) Степень распространения новости

Введем вероятностное пространство (Ω,F,P). Обозначим в этом пространстве $\in\Omega$ – новостной повод. На основании приведенных характеристик новостные поводы делятся на группы, в которых они схожи по определённым признакам. Согласно этим характеристикам введём разбиение пространства Ω на множества $B = \{B_i\}_{i=1}^n$:

Введём случайные величины: Y_j $-$ $-$количество новостных поводов в j$-$й группе;

$$X_{i,j} = \begin{cases} X_{1,j}{-}\text{количество новостных поводов повлиявших на волатильность из j}{-}\text{й группы} \\ X_{2,j}{-}\text{количество новостных поводов не повлиявших на волатильность из j}{-}\text{й группы} \end{cases}$$

Исходя из данных, мы можем построить совместное эмпирическое распределение вероятностей $P(X_{i,j}, Y_j)$. Здесь и далее, записывая $P(A)$, будем иметь ввиду эмпирическую вероятность события A определяемую как отношение количества исходов удовлетворяющих событию A к общему количеству исходов.

Далее мы применим эмпирический Байесовский подход, который основывается на формуле Байеса:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \tag{1}$$

Данная формула позволяет переоценить вероятность события A, учитывая тот факт, что произошло событие B.

Из определения условной вероятности можем записать: $P(X_{i,j}, Y_j) = P(X_{i,j}|Y_j)P(Y_j)$. Осуществив ряд преобразований , получим:

$$P(X_{i,j}|Y_j) = \frac{P(X_{i,j}, Y_j)}{\sum_{i=1}^{2} P(X_{i,j}, Y_j)} \tag{2}$$

Выраженная величина является рейтингом новостного повода из j-го множества. Для каждого новостного повода Байесовским методом находится рейтинг – вероятность того, что повод окажет влияние при условии, что он принадлежит данной группе. Исходя из её значений, осуществляется деление новостных поводов на категории, например: “способны повлиять на волотильность ценных бумаг” и “не способны повлиять на волотильность ценных бумаг”.

Источники и литература

1) Воронцов, К.В. Математические методы обучения по прецедентам (теория обучения машин) [Электронный ресурс] / К. В. Воронцов //Курс лекций – 2016.

2) Сидоров С.П., Дате П., Балаш В.А. Использование новостной аналитики в GARCH моделях. Прикладная эконометрика 29(1), 2013г.

3) Солодухина А.В., Репин Д.В. Влияние корпоративных новостей на рыночную стоимость компаний. Корпоративные Финансы №1(**??**) 2009