

ВИКОРИСТАННЯ МЕТОДУ ХОЛЬТА ДЛЯ АНАЛІЗУ ЧАСОВИХ РЯДІВ

Факультет комп'ютерних систем
Національного авіаційного університету

Розглянуто питання представлення часових дискретних даних та аналізу їх за допомогою методу Хольта. Визначено результати використання методу, виведені коефіцієнти шляхом тестових наборів даних і дано оцінку ефективності прогнозування. Створені розрахунковий лист в середовищі Microsoft Excel для дослідження даного методу та підбору коефіцієнтів та дослідження ефективності методу

Вступ

Аналітичні технології – це методи, які на основі деяких моделей, алгоритмів, математичних теорем дозволяють за відомими даними оцінити значення невідомих характеристик та параметрів.

Найпростішим прикладом є теорема Піфагора, яка дозволяє за довжинами сторін прямокутника визначити довжину діагоналі. Іншим прикладом аналітичної технології є способи, за допомогою яких оброблює інформацію людський мозок. Навіть мозок дитини здатний вирішувати такі задачі, які не під силу сучасним комп'ютерам, як, наприклад, розпізнавання обличчя в наговпі чи контроль десятків м'язів при грі у футбол. Унікальність мозку складається в тому, що здатний вчитися розв'язанню нових задач (гри в шахмати, водінню автомобіля), але погано пристосований для розв'язання великих об'ємів інформації (обрахування кореня з числа 463761). На практиці часто зустрічаються більш складні задачі ніж вирахування квадратного кореня. Таким чином людині для розв'язання цих задач потрібні додаткові методи та інструменти.

Аналітичні технології в першу чергу потрібні людям, які приймають важливі рішення – керівники, аналітики, експерти, консультанти. Прибуток компанії в більшості випадків залежить від якості цих рішень – точності прогнозів, оптимально обраних стратегій.

Наївні моделі прогнозування

При створенні «наївних» моделей припускаємо, що деякий останній період часового ряду найкращим чином описує майбутнє цього прогнозного ряду. Тому в таких моделях прогноз як правило є простою функцією від прогнозованої змінної в минулому.

Найпростіша модель має вигляд:

$$Y(t+1)=Y(t),$$

що відповідає твердженню «завтра буде як сьогодні».

Без сумніву від такої примітивної моделі не варто чекати великої точності. Вона не враховує механізми, які визначають прогнозні дані, і не захищені від випадкових флуктуацій, вона не враховує сезонні коливання і тренди. Можливо будувати «наївні» моделі і по-іншому

$$Y(t+1)=Y(t)+[Y(t)-Y(t-1)],$$

$$Y(t+1)=Y(t)*[Y(t)/Y(t-1)].$$

Такими способами ми намагаємося привчити модель до можливих трендів.

Середні та ковзні середні

Самою простою моделлю, яка ґрунтується на простому усередненні є:

$$Y(t+1)=(1/(t))*[Y(t)+Y(t-1)+...+Y(1)], \quad (1)$$

І на відміну від самої простої «наївної» моделі, як відповідає принцип «завтра буде як сьогодні», цій моделі відповідає принцип «завтра буде так, як в середньому за останній час». Така модель більш стійка до флуктуацій, тому що у ній згладжуються випадкові зміни відно-

сно середнього. Але все таки цей метод ідеологічно настільки ж примітивний як і «наївні» моделі і йому притаманні майже ті самі недоліки.

Середні та ковзні середні

В приведеній вище формулі ми пропустили, що ряд опосередковується за достатньо великим інтервалом часу. Але як правило свіжі значення часового ряду краще описують прогноз ніж більш старі значення. Тоді у якості прогнозу можна використовувати «ковзне середнє»

$$Y(t+1) = (1/(T+1)) * [Y(t) + Y(t-1) + ... + Y(t-T)],$$

Зміст полягає в тому, що використовується середнє значення за останні T періодів і тільки на цих значеннях будується прогноз.

Також можна використати метод експоненціальних середніх, коли більш свіжі значення мають більшу вагу ніж старі.

$$Y(t+1) = \alpha * Y(t) + (1 - \alpha) * Y(t), \quad (2)$$

де $Y(t+1)$ – прогноз на майбутній період часу; $Y(t)$ – поточне значення в момент t ; α – постійна згладжування ($0 \leq \alpha \leq 1$)

В цьому методі є внутрішній параметр α , який визначає залежність прогнозу від більш старих даних, причому вплив даних на прогноз експоненціально зменшується з «віком» даних. Залежність впливу даних на прогноз при різних коефіцієнтах α приведена на рис.1.

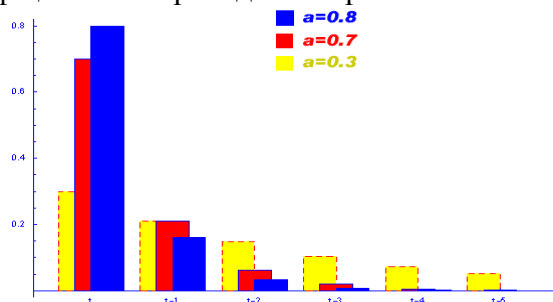


Рис. 1. Вплив змінних в залежності від часу

Видно що при $\alpha \rightarrow 1$, експоненціальна модель прямує до простої «наївної» моделі. При $\alpha \rightarrow 0$, прогнозна величина стає рівною попередньому прогнозу.

Якщо виконується прогнозування с використанням моделі експоненціального згладжування, то на деякому тестовому

наборі будуються прогнози при $\alpha = [0.01, 0.02, \dots, 0.98, 0.99]$ и визначається при якому α точність вище. Це значення використовується при прогнозуванні в майбутньому.

Хоч описані моделі («наївні» алгоритми, методи що основані на середніх, ковзних середніх і експоненціального згладжування) використовуються при прогнозування не дуже складних ситуацій, не рекомендовано їх використовувати окремо із причин примітивності і неадекватності.

Між іншим ці алгоритми можливо використовувати як додатковий інструмент для передобробки даних, а саме для виділення трендових складових.

Використання методу Хольта

Хольт запропонував удосконалений метод експоненціального згладжування [3]. В запропонованому алгоритмі значення рівня і тренду згладжуються за допомогою експоненціального згладжування, причому параметри згладжування у них різні.

$$\begin{cases} \Omega_t = \alpha \cdot Y_t + (1 - \alpha)(\Omega_{t-1} - T_{t-1}) \\ T_t = \beta \cdot (Q_t - Q_{t-1}) + (1 - \beta)T_{t-1} \\ Y_{t+p} = \Omega_t + pT_t \end{cases}$$

де Ω_t – прогнозне значення для поточного моменту часу, T_t – визначення тенденції значень; Y_t – значення часового ряду; Y_{t+p} – прогнозне значення на t періодів у майбутнє.

Тут перше рівняння описує згладжений ряд загального рівня.

Друге рівняння потрібне для оцінки тренду.

Третє рівняння визначає прогноз на p періодів часу вперед.

Постійні згладжування в методі Хольта ідеологічно грають ту саму роль, що і постійна в просто експоненціальному згладжуванні. Підбираються вони шляхом перебору по цим параметрам з якимось кроком. Далі буде використано адаптивний підбір параметрів. Після підбору па-

ри, яка дає найбільшу точність моделі на тестовому наборі, її використовують для реального прогнозування. Спробуємо дослідити і визначити найбільш ефективні параметри для використання даного методу.

Дослідження ефективності методу

У якості тестового набору візьмемо результати динаміки валютного курсу по одній з валютних на протязі 2007 року. Для узагальнення даних часових рядів використовується інструмент «японська свічка». Кожна свічка відповідає якомусь інтервалу часу (30 хвилин, година, чотири години, день). Свічка складається з 5 елементів: період свічки (t), відкриття ($open$), закриття ($close$), максимальне значення ($high$), мінімальне значення (low).

Прогноз вважаємо справдженим, якщо точка прогнозу попала в свічку. Для використання прогнозу нам потрібно, щоб для кожного моменту часу була одна точка. З чотирьох можливих оберемо одну.

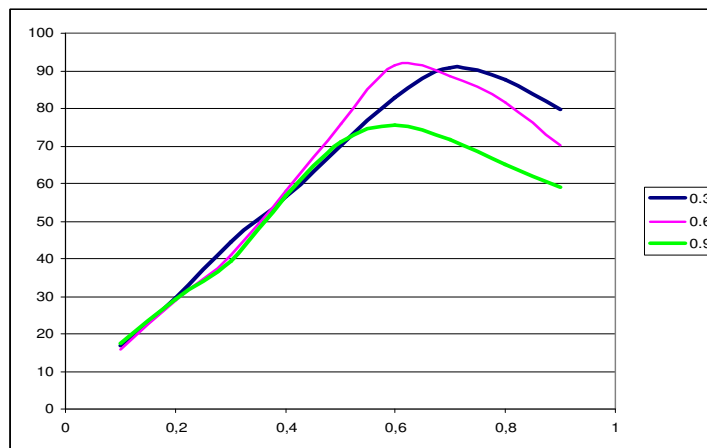


Рис. 2. Графік залежності ефективності від коефіцієнта $\alpha \in [0.1..0.9]$.

Якщо відобразити дані по таблиці на графіку, то можемо зробити деякі висновки, а саме:

Крива ефективності має свою форму і невиразну залежність α від β . Тобто досліджувати ефективність кожного коефіцієнта можна окремо.

Найбільш ефективним проміжком α є $[0.6;0.7]$.

Для β ширина ефективності більша і лежить в межах $[0.3;0.6]$.

З точки зору інформативності параметр $close$ вважається найбільш важливим. Далі ми дослідимо ефективність цього вибору..

Визначення найбільш ефективної пари α і β .

Спробуємо визначити коефіцієнти α і β . Спочатку ми розглянемо грубу зміну параметру для визначення самоподібності. Для цього для трьох значень β проведемо зміни α в проміжку $[0.1,0.2...0.9]$.

Таблиця 1. Ефективність в % в залежності від різних значень α і β

$\alpha \backslash \beta$	0.3	0.6	0.9
0.1	16,73	15,95	17,51
0.2	29,57	28,79	29,18
0.3	44,36	40,86	39,31
0.4	56,42	57,98	56,81
0.5	70,04	75,88	71,21
0.6	82,88	91,44	75,49
0.7	90,66	88,72	71,61
0.8	87,55	81,71	64,98
0.9	79,77	70,04	59,14

Дослідимо ефективність ще раз для звужених проміжків. Візьмемо зміну α $[0.61, 0.62..0.69]$, а для β знову три значення $[0.3,0.45,0.6]$.

Таблиця 2. Ефективність в % в залежності від різних значень α і β

$\beta \backslash \alpha$	0.3	0.45	0.6
0.61	84,05	89,11	91,05
0.62	86,77	91,44	91,05
0.63	86,77	93,11	91,44

0.64	87,55	93,77	90,27
0.65	87,16	93,39	90,66
0.66	87,94	93,77	91,44
0.67	89,49	94,55	90,66
0.68	90,27	95,72	89,88
0.69	79,77	94,16	89,88

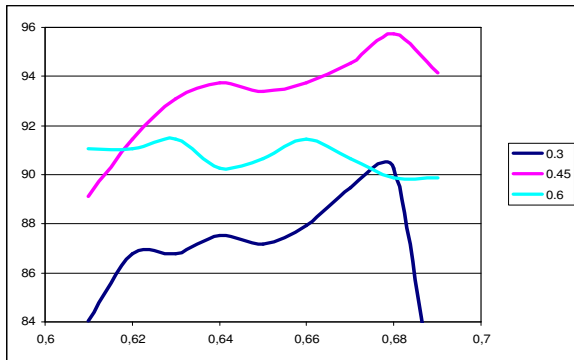


Рис. 3. Графік залежності ефективності від коефіцієнта $\alpha \in [0.61..069]$

З даних цього експерименту видно пара $[0.68;0.45]$ має найбільший коефіцієнт ефективності. Хоча при значенні $\beta=0,6$ наша модель має рівномірну ефективність, що також не менш важливо. Стабільність дозволить зменшити вплив індивідуальності ряду. Для дослідження цього питання підставимо значення за 2006 рік.

Таблиця 3. Ефективність в % в залежності від різних значень α і β

$\beta \backslash \alpha$	0.3	0.45	0.6
0.61	84,82	90,27	91,44
0.63	87,55	92,22	93,01
0.65	87,94	93,37	93,39
0.67	88,72	95,33	93,01
0.69	88,33	94,55	92,61

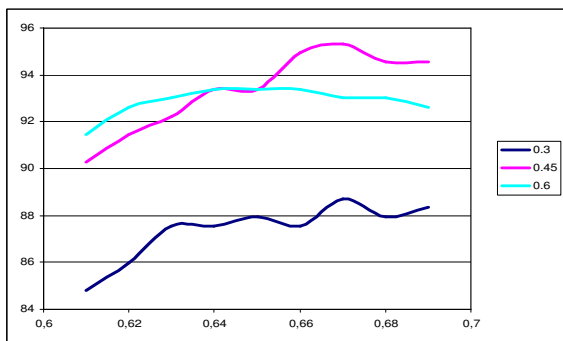


Рис. 4. Графік залежності ефективності від коефіцієнта $\alpha \in [0.61..069]$ за 2006р

Для цієї предметної області пара $[0.67;0.45]$ має найбільший коефіцієнт ефективності. Хоча також видно стабільність моделі для значення $\beta=0,6$.

Визначення найбільш вдалих вхідних значень

Для визначення ефективної пари α і β ми брали у якості вхідних даних рівень закриття (close). Спробуємо інші вхідні дані: open, high, low, $(open+close)/2$, $(high+low)/2$.

	2006	2007	2008
Open	45,14	52,14	49,22
High	52,14	54,86	48,45
Low	49,42	43,97	45,74
Close	94,55	95,72	94,19
$(open+close)/2$	65,76	67,32	70,16
$(high+low)/2$	62,66	63,42	64,34

Незаперечливий факт ефективності точки закриття (close) у якості вхідних даних.

Висновки

Метод Хольта показує гарні результати при прогнозуванні часових економічних рядів.

У якості коефіцієнти α і β методу вибрана пара $[0.67;0.45]$. Вважаю, що для всіх аналогічних за походженням даних оптимальна пара буде така або близька до неї. Це значно спрощує пошук коефіцієнтів.

Вхідними даними вважаю закриття кожною свічки. Саме ця точка має найбільшу ефективність при прогнозування рівня наступної.

Список літератури

1. Джон Э. Ханк, Дин У. Уччерн, Артур Дж. Райтс. Бизнес-прогнозирование. – Вільямс – 2006. – 663 с.
2. Эндрю Сигел. Практическая бизнес-статистика. – Вільямс. – 2008. – 556 с.
3. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. – М.: Финансы и статистика. – 2003 – с 415.