Analiza szeregów czasowych i prognozowanie

Dr inż. Krzysztof Krawiec

Zakład Inteligentnych Systemów Wspomagania Decyzji Instytut Informatyki, Politechnika Poznańska

http://www-idss.cs.put.poznan.pl

1 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Cel

Zbudowanie modelu pewnego zjawiska/procesu w oparciu o obserwowane zmiany w czasie pewnych mierzalnych wielkości opisujących ten proces.

Cele:

- weryfikacja intuicyjnych przypuszczeń,
- wyjaśnianie natury zjawiska,
- predykcja (ekstrapolacja) przebiegu lub jego składowych => prognozowanie.

2 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Szereg czasowy

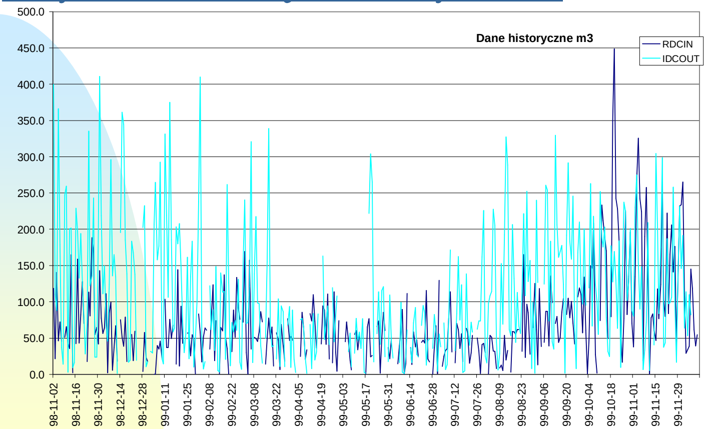
W statystyce: szczególny przypadek *szeregu dynamicznego* (nazwanego tak przez analogię do szeregu rozdzielczego).

Rozróżnia się szeregi czasowe:

- *momentów* (np. liczba ludności pewnego miasta w latach od 1990 do 1999 liczona wg stanów na 31 grudnia),
- *okresów* (np. roczna produkcja spinaczy w Polsce w latach od 1990 do 1999).

3 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Przykład 1: Obciążenie aktywności



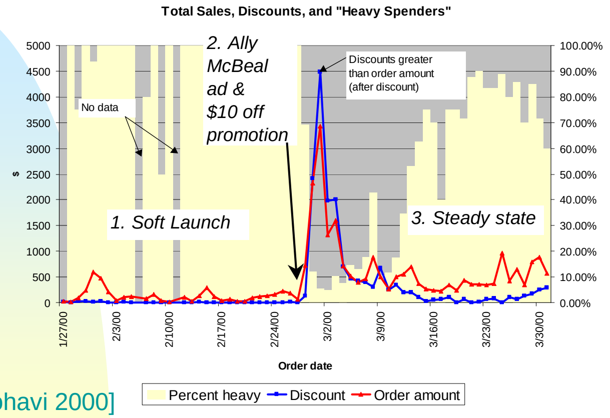
4 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Przykład 2: Odwiedziny witryny WWW

- Gazelle.com - a legwear and legcare web retailer
- Soft-launch: Jan 30, 2000
- Hard-launch: Feb 29, 2000 with an Ally McBeal TV ad on 28th and strong $10 off promotion
- „clickstream analysis”

5 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

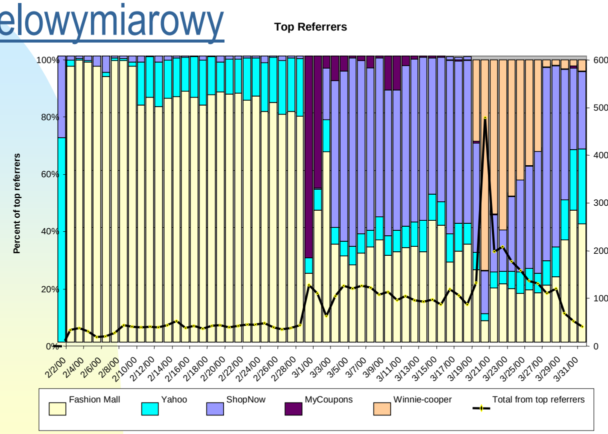
Przykład 2:



[Kohavi 2000]

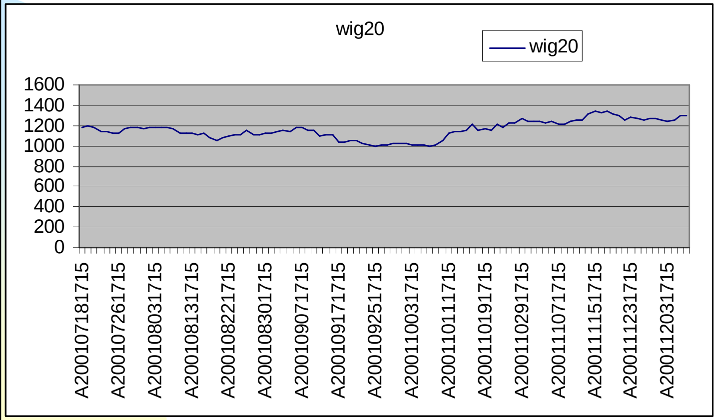
6 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Przykład – szereg czasowy okresów wielowymiarowy



7 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Przykład 3: Szereg czasowy momentów



8 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Spostrzeżenia

- taki a nie inny kształt danych historycznych nie jest (nie był) jedynym możliwym scenariuszem (wariantem),
- na nasz przebieg wpływają także czynniki zewnętrzne względem modelu, których w nim nie uwzględniamy

ZATEM:

- konieczność probabilistycznego ujęcia modelowania szeregów czasowych => statystyka.

Proces stochastyczny a szereg czasowy

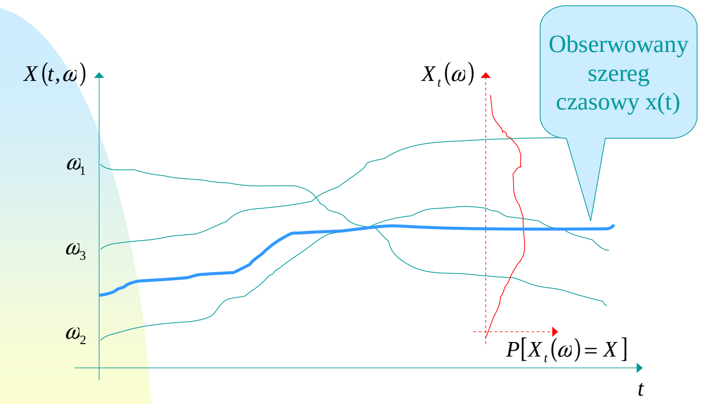
Proces stochastyczny $X(t,\omega)$ to zbiór losowych funkcji czasu (t) przyporządkowanych *zdarzeniom* ω.

Gdy ustalamy:

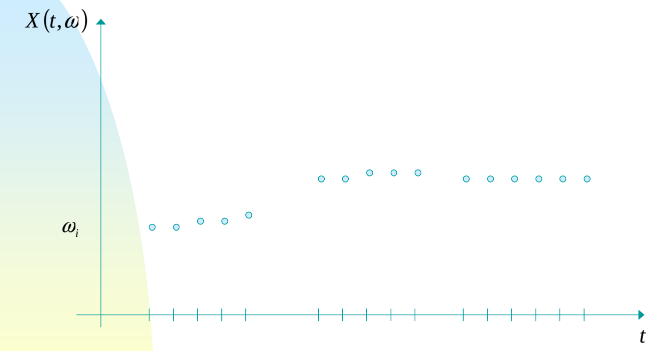
- t - dostajemy *zmienną losową* $X_t(\omega)$
- ω - dostajemy funkcję czasu x(t). Funkcje x(t) to *realizacje* procesu stochastycznego $X(t,\omega)$, *szeregi czasowe (SC)*.

W szczególności, gdy czas t przyjmuje wartości dyskretne (np. naturalne), funkcje x(t) nazywamy *dyskretnymi szeregami czasowymi* (DSC).

Proces stochastyczny a szereg czasowy



Przykład - dyskretny szereg czasowy



Typowy przebieg procesu analizy SC

- Zbieranie i przygotowywanie danych
- Analiza
- Modelowanie = odkrywanie struktury przebiegu czasowego (np. przez dekompozycję)
- Prognozowanie, w tym m.in.:
  - weryfikowanie prognoz,
  - szacowanie jakości prognoz,

=> przebieg wykładu

13 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Przebieg wykładu

- Zbieranie i przygotowywanie danych
- Analiza szeregów czasowych
- Dekompozycja szeregu czasowego
- Prognozowanie
- Prognozowanie w przedsiębiorstwie
- Uwagi końcowe
- Studium przypadku

14 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Zbieranie i przygotowywanie danych

15 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Założenia – metoda pozyskiwania

- Jednorodny charakter danych:
  - szeregi okresów: okresy powinny być równej długości,
  - szeregi momentów: momenty powinny być równo odległe w czasie.

  Bardzo istotne! Większość metod zakłada „jednorodność” danych.

- Dane historyczne powinny dotyczyć okresu, w którym parametry modelu opisującego proces były (choćby w przybliżeniu) stałe.

16 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Założenia – ilość danych

- Dane historyczne powinny jak najlepiej reprezentować poszukiwane regularności. W ogólności obowiązuje zasada „im więcej, tym lepiej”.
- W szczególności, jeżeli obserwujemy (lub domyślamy się istnienia) cyklicznej zmienności w analizowanym procesie, zbierane dane powinny obejmować minimum jeden pełen okres cyklu.

17 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Dane rzeczywiste = Problemy

- nieregularny charakter kalendarza,
- wartości brakujące, spowodowane np.
  - dniem wolnym od pracy,
  - awarią systemu zbierania danych,
- zmiana charakteru/parametrów modelowanego procesu
  Przykład: Dane historyczne dotyczące produkcji (*total*). Czy należy ignorować fakt, iż na pewnym etapie asortyment produktów uległ zmianie?
- wiarygodność danych

18 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Dane rzeczywiste = Problemy (c.d.)

- obserwacje odstające – przyczyny:
  - Błędy w danych statystycznych (np. przy wprowadzaniu),
  - Silny wpływ rzadko występującego czynnika zewnętrznego (np. jednorazowa realizacja bardzo dużego zamówienia, tzw. *elephant order*),
  - Silny wpływ rzadko występującego czynnika wewnętrznego (np. awaria urządzenia produkcyjnego),

19 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Środki zaradcze

- nieregularny charakter kalendarza:
  - standaryzacja długości miesiąca,
  - standaryzacja liczby dni roboczych w tygodniu.

Przykład:

Miesiąc	*Sprzedaż(szt)*	*Sprzedaż po standaryzacji*	*L. dni*
sty	632	612	31
lut	380	407	28
mar	439	425	31
kwi	434	434	30
maj	445	431	31
cze	542	542	30

20 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Problem

Manipulacje na poszczególnych wartościach dezaktualizują dane agregatowe (np. podsumowania).

Miesiąc	*Sprzedaż(szt)*	*Sprzedaż po standaryzacji*	*L. dni*
sty	632	616	31
lut	380	410	28
mar	439	428	31
kwi	434	437	30
maj	445	434	31
cze	542	546	30
Suma:	2872	2872	30.23

=> konieczność korekty.

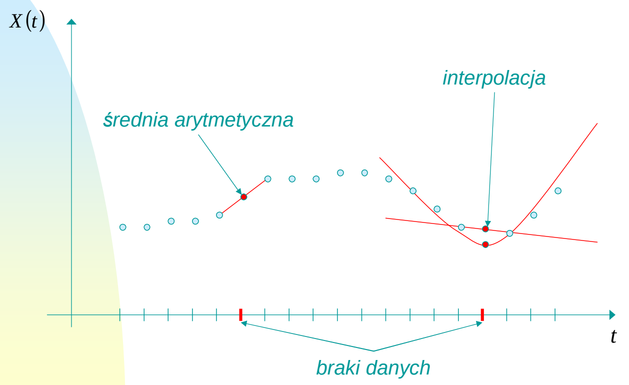
Korekta

Miesiąc	*Sprzedaż(szt)*	*Sprzedaż po standaryzacji*	*L. dni*		Wskaźnik		
sty	632	612	31	615.0108	0.967742	0.971847	614.2075
lut	380	407	28	409.4048	1.071429	1.075974	408.87
mar	439	425	31	427.1989	0.967742	0.971847	426.641
kwi	434	434	30	436.4111	1	1.004242	435.8411
maj	445	431	31	433.0376	0.967742	0.971847	432.472
cze	542	542	30	545.0111	1	1.004242	544.2993
Suma:	2872	2850	30.17	2866.074	0.995776		2862.331

Przetwarzanie obserwacji odstających i brakujących

- uwzględnić w modelu,
  - trudne, ale najbardziej pożądane.
- pominąć (eliminować),
  - wada: redukcja zbioru danych,
- zastąpić średnią (arytmetyczną) obserwacji sąsiednich,
  - wada: słabe „podparcie” statystyczne,
- interpolować funkcją złożoną,

Przetwarzanie obserwacji odstających



Interpolacja wielomianowa

Metoda Lagrange'a:
- oszacowanie $f(t_0)=\sum_{i=1}^{N} f(t_i)P_i^L(t)$
- współczynniki $P_i^L(t)=\frac{\prod_{j\neq i}(t-t_j)}{\prod_{j\neq i}(t_i-t_j)}$

Gdzie:
- N – liczba węzłów interpolacji
- t_0 – numer chwili, dla której dokonuje się szacunku nieznanej wartości szeregu czasowego,
- $f(t_0)$ – wartość tego szacunku
- P_i^L – współczynniki wielomianu Lagrange'a

Założenie: liczba brakujących punktów jest niewielka.

25 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Przebieg wykładu

- Zbieranie i przygotowywanie danych
- Analiza szeregów czasowych
- Dekompozycja szeregu czasowego
- Prognozowanie
- Prognozowanie w przedsiębiorstwie
- Uwagi końcowe
- Studium przypadku

26 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Analiza szeregów czasowych

27 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Cel

Opisanie ilościowe dynamiki szeregu czasowego (SC)

Podstawowe miary stosowane w analizie SC to przyrosty:
- absolutny
- stosunkowy

oraz ich średnie.

28 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Przyrost absolutny

Różnica w poziomie zjawiska zanotowana w dwóch różnych okresach/momentach,

$$y_t - y_0$$

gdzie y_t - poziom zjawiska w okresie/chwili t, y_0 - poziom odniesienia, w tzw. okresie podstawowym.

29 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Ciągi przyrostów absolutnych

Za pomocą przyrostów można utworzyć ciąg przyrostów absolutnych

- jednopodstawowy, lub

$$y_1 - y_0, y_2 - y_0, \ldots, y_n - y_0$$

- łańcuchowy

$$y_1 - y_0, y_2 - y_1, \ldots, y_n - y_{n-1}$$

30 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Ciągi przyrostów absolutnych - przykład

Samochody osobowe (stan na dzień 31 XII)

Lata	Liczba samochodów	Przyrosty absolutne jednopodstawowy	Przyrosty absolutne łańcuchowy
1990	5261	0	-
1991	6112	851	851
1992	6505	1244	393
1993	6771	1510	266
1994	7153	1892	382

Za: [Kassyk-Rokicka 1998]

31 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Przyrost stosunkowy

Stosunek przyrostu absolutnego do poziomu zjawiska w okresie wyjściowym,

$$\frac{y_t - y_0}{y_0}$$

gdzie y_t - poziom zjawiska w okresie/chwili t, y_0 - poziom odniesienia, w tzw. okresie podstawowym.

32 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Ciągi przyrostów stosunkowych

Za pomocą stosunków można utworzyć ciąg przyrostów względnych

- jednopodstawowy, lub
- łańcuchowy

$$\frac{y_1 - y_0}{y_0}, \frac{y_2 - y_0}{y_0}, \ldots, \frac{y_n - y_0}{y_0}$$

$$\frac{y_1 - y_0}{y_0}, \frac{y_2 - y_1}{y_1}, \ldots, \frac{y_n - y_{n-1}}{y_{n-1}}$$

33 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Ciągi przyrostów stosunkowych- przykład

Samochody osobowe (stan na dzień 31 XII)

Lata	Liczba samochodów	Przyrosty stosunkowe jednopodstawowy	łańcuchowy
1990	5261	0	-
1991	6112	0.162	0.162
1992	6505	0.236	0.064
1993	6771	0.287	0.041
1994	7153	0.360	0.056

Za: [Kassyk-Rokicka 1998]

34 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Przebieg wykładu

- Zbieranie i przygotowywanie danych
- Analiza szeregów czasowych
- Dekompozycja szeregu czasowego
- Prognozowanie
- Prognozowanie w przedsiębiorstwie
- Uwagi końcowe
- Studium przypadku

35 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Dekompozycja szeregu czasowego i konstrukcja modelu

36 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Idea

Poszukujemy
- adekwatnego, i
- możliwie oszczędnego

modelu wyjaśniającego obserwowane szeregi.

Na przykład: spośród kilku możliwych do zastosowania modeli wybieramy ten, który charakteryzuje się najmniejszą liczbą parametrów.

Podłoże filozoficzne: brzytwa Ockhama

(*Occam's razor; William of Ockham 1285-?*)

One should not increase, beyond what is necessary, the number of entities required to explain anything.

Nie należy mnożyć bytów ponad potrzebę.

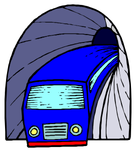
Overfitting

Problem: nawet jeżeli zbudujemy model bardzo dobrze (tj. z małym błędem) wyjaśniający dane historyczne, to nie mamy gwarancji, czy generowane przez niego prognozy będą równie dobre.

W praktyce: model zbyt dokładnie „dopasowany" (*overfitted*) do danych historycznych/uczących zazwyczaj sprawuje się na nowych danych (tu: w przyszłości) gorzej niż model prostszy, gorzej dopasowany do danych historycznych (oczywiście do pewnego stopnia).

Powód: indukcyjny charakter procesu konstruowania modelu zjawiska i prognozowania.

Overfitting – przykład

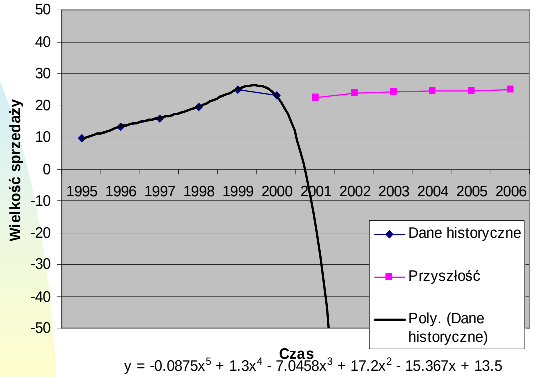


Once upon a time there was a little girl named Emma. She had never eaten banana in all her life nor had she ever taken a journey on a train. On one occasion circumstances made necessary for her to journey from New York to Pittsburgh alone. To relieve Emma's anxiety her mother gave her a large bag of bananas to eat on her railway journey west. At Emma's first bite of her banana, the train plunged into a tunnel. At the second bite, the train broke into the daylight again. Emma, being a bright little girl, takes a third bite. Lo! Into a tunnel. A fourth bite and into the daylight again. And so on all the way to Pittsburgh (and all the way to the bottom of the bag of bananas).

Is Emma justified in saying to the people who met her at the station, ``Every odd bite of a banana makes you blind; every even bite puts things right again?

[N. H. Hanson, za: H.Bensusan, *Odd bites into bananas don't make you blind -- learning about simplicity and attribute addition*]

Overfitting - przykład



Ogólne założenie

Obserwowany przebieg X(t) składa się z:

- części systematycznej (której model chcemy zbudować),
- szumu.

Uwaga: „składa się” niekoniecznie oznacza w tym kontekście sumę.

Szum utrudnia analizę procesu. Może wynikać z:

- błędów pomiarowych,
- interakcji procesu z otoczeniem,
- nieprzewidywalnych zdarzeń losowych.

41 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Podstawowa struktura SC

- *stały* (przeciętny) poziom zmiennej,
- *trend* (*tendencja rozwojowa*) - reprezentuje ogólny monotoniczny kierunek rozwoju zjawiska; może być
  - linowy, lub
  - nieliniowy, np. logarytmiczny, wykładniczy
- *składowa okresowa* (*wahania okresowe*) - składnik powtarzający się cyklicznie,
- *szum* (*zakłócenia, wahania przypadkowe*).

42 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Podstawowa struktura SC

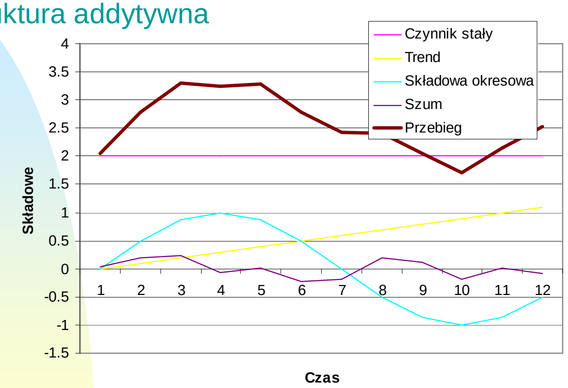
W szczególności, składowa okresowa może być wypadkową wahań:

- *cyklicznych* – o niestałym okresie i/lub amplitudzie (np. cykl koniunkturalny gospodarki, cykl rozwoju populacji nabywców danego produktu, etc.), i/lub
- *sezonowych* – o dość stałym okresie i amplitudzie (wynikają z zachowań ludzi wynikających z „kalendarza”, np. rytm pracy w skali tygodnia, dnia, pory roku, święta, etc.)

43 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Ilustracja struktury SC na przykładzie

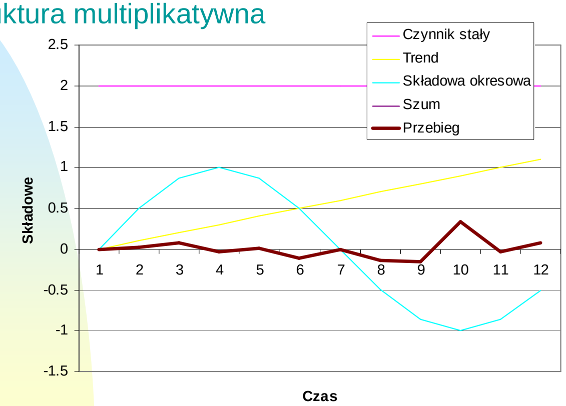
Struktura addytywna



44 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

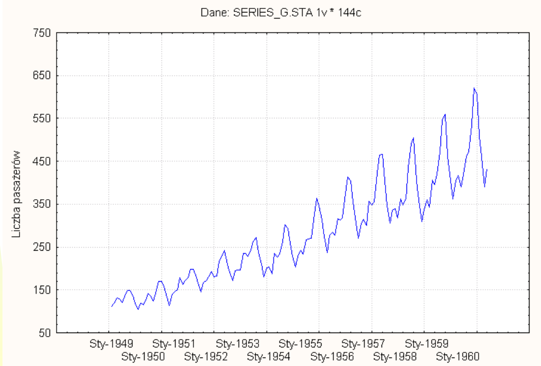
Ilustracja struktury SC na przykładzie

Struktura multiplikatywna



Przykład: seria G

Liczba pasażerów pewnych linii lotniczych w USA w poszczególnych miesiącach w latach 1949-1960.



[Box & Jenkins 1976]

Taksonomia metod analizy/modelowania SC

- analiza trendu,
- analiza sezonowości,
- dopasowanie funkcji (aproksymacja),
- analiza widmowa (Fouriera),
- wyrównywanie wykładnicze,
- dekompozycja sezonowa,
- podejścia niekonwencjonalne (sieci neuronowe, heurystyki tworzone *ad hoc*, specjalizowane, etc.)

(Ograniczamy się do metod ilościowych)

Wyodrębnianie i analiza trendu

Dwie grupy metod:

- *mechaniczne* (średnie ruchome)
- *analityczne*, oparte na metodzie najmniejszych kwadratów.

Trend – metody mechaniczne

Wygładzanie - polega na lokalnym (w czasie) uśrednieniu przebiegu.

Najbardziej popularna metoda: średnia ruchoma (*moving average*):

- Wartość przetworzona X'(t) jest pewną średnią *n* próbek „wokół" *t* (czyli dokładniej próbek X(t-n/2),...,X(t+n/2).

Najczęściej stosuje się:

- średnią arytmetyczną (wygładzanie arytmetyczne),
- medianę.

Średnie ruchome

Średnia ruchoma 3-okresowa:

$$\bar{y}_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}$$

Problem: jak liczyć średnią ruchomą z parzystej liczby okresów (momentów) ?

Średnia ruchoma – ogólna definicja

Średnia ruchoma *n*-okresowa zdefiniowana jest następująco:

- dla *n* nieparzystych

$$\bar{y}_t^n = \frac{1}{n}\sum_{\tau=t-\lfloor n/2\rfloor}^{t+\lfloor n/2\rfloor} y_\tau$$

- dla *n* parzystych (tzw. *średnie ruchome scentrowane*)

$$\bar{y}_t^n = \frac{1}{n}\left(\sum_{\tau=t-n/2+1}^{t+n/2-1} y_\tau + \frac{1}{2}y_{t-n/2} + \frac{1}{2}y_{t+n/2}\right)$$

Na przykład dla *n*=4:

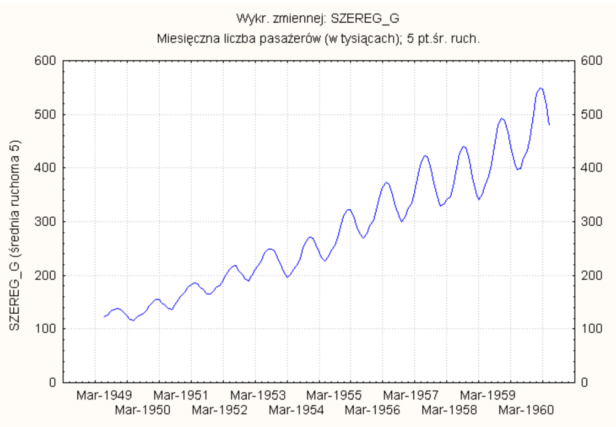
$$\bar{y}_3^4 = \frac{\frac{1}{2}y_1 + y_2 + y_3 + y_4 + \frac{1}{2}y_5}{4}$$

Średnia ruchoma – podsumowanie

Nie należy przesadzać – zbyt długa średnia ruchoma prowadzi do „zamazania" tendencji rozwojowej i skrócenia przebiegu czasowego (=> utrata danych).

Inna wada: brak możliwości przedstawienia głównej tendencji w formie matematycznej funkcji trendu.

Wygładzanie arytmetyczne - przykład



Trend - metoda analityczna

Gdy istnieje wyraźna, monotoniczna zależność -> dopasowanie funkcji (aproksymacja).

$$y_t = f(t) + z_t$$

gdzie: f(t) – funkcja trendu, z_t – składnik resztowy

Problem: nie znamy z góry postaci analitycznej trendu. Zatem niezbędne jest tu aprioryczne założenie.

Regresja liniowa

Funkcja f(t) jest liniowa:

$$y_t = at + b + z_t$$

Wartości parametrów a i b poszukujemy przez minimalizację składnika resztowego:

$$\min \sum_t z_t^2 = \sum_t (y_t - f(t))^2 = \sum_t (y_t - at - b)^2$$

Stosując matematyczne warunki minimalizacji powyższego wyrażenia dostajemy:

$$a = \frac{n\sum y_t t - \sum t \sum y_t}{n\sum t^2 - \left(\sum t\right)^2}, \quad b = \frac{\sum y_t - a\sum t}{n}$$

Wówczas oszacowanie:

$$y_t' = at + b$$

Regresja liniowa - przykład

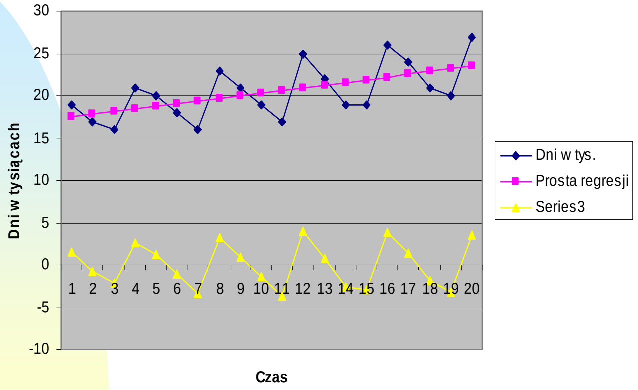
Dane:

Liczba dni nie przepracowanych z powodu choroby

Lata	Kwartały	Dni w tys.
1991	I	19
	II	17
	III	16
	IV	21
1992	I	20
	II	18
	III	16
	IV	23
1993	I	21
	II	19
	III	17
	IV	25
1994	I	22
	II	19
	III	19
	IV	26
1995	I	24
	II	21
	III	20
	IV	27

Za: [Kassyk-Rokicka 1998]

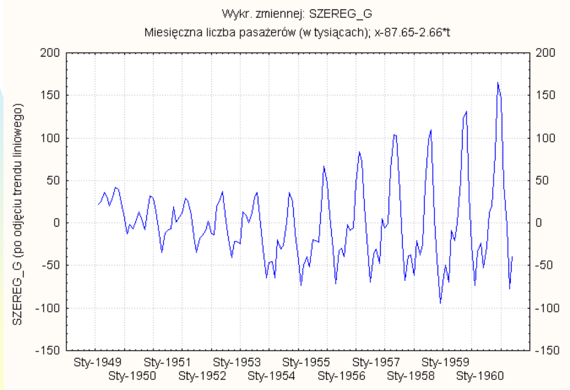
Regresja liniowa - przykład



57 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Trend liniowy - przykład

Szereg G po odjęciu trendu liniowego



58 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Ocena jakości dopasowania (regresji)

Problem: regresja liniowa da zawsze *jakiś* wynik (wartości *a* i *b*). Czy to oznacza, że zależność liniowa zawsze istnieje ?

⇒Konieczność oceny jakości dopasowania.

Stosując regresję wyjaśniamy tylko część zmienności; pewna część zmienności zjawiska pozostaje niewyjaśniona.

59 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Ocena jakości dopasowania

Współczynnik determinacji informuje, jaka część zaobserwowanej (w próbie) całkowitej zmienności y_t została wyjaśniona (przez *f(t)*) względem *t*.

$$r^2 = \frac{\sum_t (f(t)-\bar{y})^2}{\sum_t (y_t-\bar{y})^2}, \quad r^2 \in \langle 0,1 \rangle$$

Współczynnik indeterminacji informuje, jaka część zaobserwowanej (w próbie) całkowitej zmienności y_t **nie** została wyjaśniona względem *t*.

$$\varphi^2 = \frac{\sum_t (f(t)-y_t)^2}{\sum_t (y_t-\bar{y})^2} = \frac{\sum_t z_t^2}{\sum_t (y_t-\bar{y})^2}, \quad \varphi^2 \in \langle 0,1 \rangle$$

Można dowieść, że: $r^2 + \varphi^2 = 1$

60 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Ocena jakości dopasowania – przykład

Czas	Lata	Kwartały	Dni w tys.	Prosta regresji	Kwadrat odchylenia modelu od średniej			Kwadrad odchylenia danych historycznych od średniej
1	1991	I	19	17.51	8.91	Współczynniki równar		2.25
2		II	17	17.83	7.14	prostej regresji y=at+l		12.25
3		III	16	18.14	5.56	a	0.31	20.25
4		IV	21	18.46	4.17	b	17.20	0.25
5	1992	I	20	18.77	2.99			0.25
6		II	18	19.09	2.00			6.25
7		III	16	19.40	1.21	Średnia		20.25
8		IV	23	19.71	0.62	20.5		6.25
9	1993	I	21	20.03	0.22			0.25
10		II	19	20.34	0.02			2.25
11		III	17	20.66	0.02			12.25
12		IV	25	20.97	0.22			20.25
13	1994	I	22	21.29	0.62			2.25
14		II	19	21.60	1.21			2.25
15		III	19	21.91	2.00			2.25
16		IV	26	22.23	2.99			30.25
17	1995	I	24	22.54	4.17			12.25
18		II	21	22.86	5.56			0.25
19		III	20	23.17	7.14			0.25
20		IV	27	23.49	8.91			42.25
					65.69	Suma odch. Kw.		195
					Współczynnik determinacji			0.34
					Współczynnik indeterminacji			0.66

61 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Ocena jakości dopasowania

Wariancja składnika resztowego – miara niedopasowania funkcji regresji do regresji empirycznej (danych)

$$S^2(z)=\frac{\sum_{i=1}^{n} z_i^2}{n-\lambda-1}=\frac{\sum_{i=1}^{n}\left(y_i-a_y x_i-b_y\right)^2}{n-\lambda-1},\quad S(z)=\sqrt{S^2(z)}$$

Gdzie:

- n – liczba badanych okresów (momentów),
- λ – liczba niezależnych zmiennych w modelu regresji (dla regresji liniowej λ=1)

Przesłanka: miarą wielkości przeciętnego błędu losowego popełnianego przy estymacji nieznanego parametru z populacji generalnej za pomocą estymatora jest odchylenie standardowe estymatora.

62 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Ocena jakości dopasowania – przykład

Kwartały	Dni w tys.	Prosta regresji	Składnik resztowy	Kwadrat składnika resztowego			
I	19	17.51	1.49	2.21	Współczynniki równania		
II	17	17.83	-0.83	0.69	prostej regresji y=at+b		
III	16	18.14	-2.14	4.59	a	0.31	
IV	21	18.46	2.54	6.47	b	17.20	
I	20	18.77	1.23	1.51			
II	18	19.09	-1.09	1.18			
III	16	19.40	-3.40	11.56			
IV	23	19.71	3.29	10.80			
I	21	20.03	0.97	0.94			
II	19	20.34	-1.34	1.80			
III	17	20.66	-3.66	13.37			
IV	25	20.97	4.03	16.23			
I	22	21.29	0.71	0.51			
II	19	21.60	-2.60	6.76			
III	19	21.91	-2.91	8.49			
IV	26	22.23	3.77	14.22			
I	24	22.54	1.46	2.12			
II	21	22.86	-1.86	3.45			
III	20	23.17	-3.17	10.06			
IV	27	23.49	3.51	12.35			
		Suma kwadratów reszt		129.31			
		Wariancja składnika resztowego		7.18			
		chylenie standardowe składnika resztowego		2.68			

63 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Obliczanie błędów szacunku parametrów

Zagadnienie pokrewne w stosunku do oceny jakości dopasowania, dotyczy jednak jakości oszacowania parametrów a i b.

Przeciętny błąd losowy oszacowania współczynnika trendu liniowego (a) i wyrazu wolnego (b):

$$S(a)=\frac{S(z_t)}{\sqrt{\sum_{t=1}^{n} t^2-n\bar{t}^2}}$$

$$S(b)=S(z_t)\sqrt{\frac{\sum_{t=1}^{n} t^2}{n\left(\sum_{t=1}^{n} t^2-n\bar{t}^2\right)}}$$

64 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Inne postaci trendu

- logarytmiczny,
- wykładniczy,
- wielomianowy (prosty),
- ilorazowy,

65 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Analiza okresowości

Cel: wyodrębnienie składowej okresowej (sezonowej, cyklicznej).

Pytanie: czy w ogóle istnieje? jak odróżnić ją od składnika losowego?

Ważne rozgraniczenie:

- okres zmian cyklicznych znany => wskaźniki wahań okresowych,
- okres zmian cyklicznych nieznany => (np.) analiza autokorelacji.

66 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Analiza okresowości (sezonowości)

Wskaźniki wahań okresowych (sezonowości) – uzyskujemy przez porównanie wyrazów szeregu empirycznego (pierwotnego) z szeregiem teoretycznym (reprezentującym trend).

Przypomnienie: okresowość może mieć (względem trendu) charakter

- addytywny lub
- multiplikatywny.

67 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Addytywne wskaźniki okresowości

Addytywny (bezwzględny) wskaźnik okresowości:

$$g_i' = \frac{\sum_{t \in n_i} (y_{it} - \hat{y}_{it})}{n_i}$$

gdzie:

- i=1,2,...,d - liczba podokresów w cyklu okresowości (np. dla danych kwartalnych o rocznym cyklu wahań d=4, dla danych miesięcznych d=12, etc.)
- n_i – liczba numerów obserwacji, które dotyczą i-tego podokresu.

Miary g_i' mają charakter absolutny i trzeba je skorygować, aby ich suma w cyklu wahań była równa 0. W tym celu definiujemy *wskaźnik korygujący*:

$$k_g = \frac{1}{d}\sum_{i=1}^{d} g_i'$$

68 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Addytywne wskaźniki okresowości

Wówczas *czyste* wskaźniki okresowości to:

$$g_i = g_i' - k_g$$

i spełniają one warunek

$$\sum_{i=1}^{d} g_i = 0$$

Multiplikatywne wskaźniki okresowości

Multiplikatywny (względny) wskaźnik okresowości (surowy):

$$O_i' = \frac{\sum_{t \in n_i} (y_{it} / \bar{y}_{it})}{n_i}$$

(oznaczenia jak w przypadku addytywnym).

Stosujemy *wskaźnik korygujący*: $k = \frac{d}{\sum_{i=1}^{d} O_i'}$

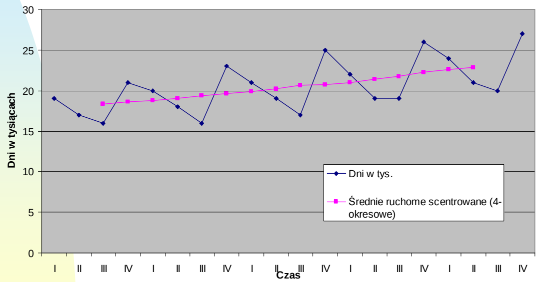
Wówczas oczyszczony wskaźnik okresowości: $O_i = O_i' k$

i zachodzi:

$$\sum_{i=1}^{d} O_i = d$$

Analiza okresowości – przykład 1

Przypadek 1: trend wyodrębniony metodą mechaniczną, tj. średnimi ruchomymi scentrowanymi.



Przykład 1 – wskaźniki bezwzględne

				Kwartały				
Lata	*Kwartały*	*Dni w tys.*	*Średnie ruchome scentrowane (4-okresowe)*	*I*	*II*	*III*	*IV*	
1991	I	19						
	II	17						
	III	16	18.4			-2.4		
	IV	21	18.6				2.4	
1992	I	20	18.8	1.3				
	II	18	19.0		-1.0			
	III	16	19.4			-3.4		
	IV	23	19.6				3.4	
1993	I	21	19.9	1.1				
	II	19	20.3		-1.3			
	III	17	20.6			-3.6		
	IV	25	20.8				4.3	
1994	I	22	21.0	1.0				
	II	19	21.4		-2.4			
	III	19	21.8			-2.8		
	IV	26	22.3				3.8	
1995	I	24	22.6	1.4				
	II	21	22.9		-1.9			
	III	20						
	IV	27						
	Suma (licznik wyrażenia we wzorze)			4.75	-6.50	-12.13	13.75	Suma kontrolna
	Bezwzględny wskaźnik okresowości			1.19	-1.63	-3.03	3.44	-0.03
	Współczynnik korygujący			-0.01				
	Skorygowany (czysty) wskaźnik [tys. h]			1.20	-1.62	-3.02	3.45	0.00

Przykład 1 – wskaźniki względne

				Kwartały				
Lata	Kwartały	Dni w tys.	Średnie ruchome scentrowane (4-okresowe)	I	II	III	IV	
1991	I	19						
	II	17						
	III	16	18.4			0.9		
	IV	21	18.6				1.1	
1992	I	20	18.8	1.1				
	II	18	19.0		0.9			
	III	16	19.4			0.8		
	IV	23	19.6				1.2	
1993	I	21	19.9	1.1				
	II	19	20.3		0.9			
	III	17	20.6			0.8		
	IV	25	20.8				1.2	
1994	I	22	21.0	1.0				
	II	19	21.4		0.9			
	III	19	21.8			0.9		
	IV	26	22.3				1.2	
1995	I	24	22.6	1.1				
	II	21	22.9		0.9			
	III	20						
	IV	27						
			Suma (licznik wyrażenia we wzorze)	4.23	3.69	3.39	4.67	Suma kontrolna
			Względny wskaźnik okresowości	1.06	0.92	0.85	1.17	3.9979
			Współczynnik korygujący	1.0005				
			Skorygowany (czysty) wskaźnik	1.058	0.924	0.849	1.169	4.0000

Przykład 1 – interpretacja

W pierwszym kwartale każdego badanego roku, tylko i wyłącznie na skutek działania czynnika okresowości, liczba dni nie przepracowanych z powodu choroby jest wyższa od poziomu zjawiska określonego przez trend średnio o 5.8%, w drugim i trzecim kwartale niższa średnio o 7.6% i 15.1%, a w czwartym wyższa średnio o 16.9% (współczynniki względne).

Współczynniki bezwzględne podają skalę tych odchyleń w liczbach bezwzględnych.

Analiza okresowości – przykład 2

Przypadek 2: trend wyodrębniony metodą analityczną, tj. metodą regresji liniowej.

Czas	Lata	Kwartały	Dni w tys.					Prosta regresji
1	1991	I	19		Współczynniki równania			17,51
2		II	17		prostej regresji y=at+b			17,83
3		III	16		a	0,31		18,14
4		IV	21		b	17,20		18,46
5	1992	I	20					18,77
6		II	18					19,09
7		III	16					19,40
8		IV	23					19,71
9	1993	I	21					20,03
10		II	19					20,34
11		III	17					20,66
12		IV	25					20,97
13	1994	I	22					21,29
14		II	19					21,60
15		III	19					21,91
16		IV	26					22,23
17	1995	I	24					22,54
18		II	21					22,86
19		III	20					23,17
20		IV	27					23,49

Przykład 2 – wskaźniki bezwzględne

					Kwartały					
Czas	Lata	Kwartały	Dni w tys.	Prosta regresji	I	II	III	IV		
1	1991	I	19	17.5	1.5					
2		II	17	17.8		-0.8				
3		III	16	18.1			-2.1			
4		IV	21	18.4				2.6	a	0.31
5	1992	I	20	18.8	1.3				b	17.2
6		II	18	19.1		-1.1				
7		III	16	19.4			-3.4			
8		IV	23	19.7				3.3		
9	1993	I	21	20.0	1.0					
10		II	19	20.3		-1.3				
11		III	17	20.6			-3.6			
12		IV	25	20.9				4.1		
13	1994	I	22	21.2	0.8					
14		II	19	21.5		-2.5				
15		III	19	21.9			-2.9			
16		IV	26	22.2				3.8		
17	1995	I	24	22.5	1.5					
18		II	21	22.8		-1.8				
19		III	20	23.1			-3.1			
20		IV	27	23.4				3.6		
				Suma (licznik wyrażenia we wzorze)	6.05	-7.50	-15.05	17.40	Suma kontrolna	
				Bezwzględny wskaźnik okresowości	1.21	-1.50	-3.01	3.48	0.18	
				Współczynnik korygujący	0.05					
				Skorygowany (czysty) wskaźnik [tys. h]	1.17	-1.55	-3.06	3.44	0.00	

Przykład 2 – wskaźniki względne

					Kwartały					
Czas	Lata	Kwartały	Dni w tys.	Prosta regresji	I	II	III	IV		
1	1991	I	19	17.5	1.1					
2		II	17	17.8		1.0				
3		III	16	18.1			0.9			
4		IV	21	18.4				1.1	a	0.31
5	1992	I	20	18.8	1.1				b	17.2
6		II	18	19.1		0.9				
7		III	16	19.4			0.8			
8		IV	23	19.7				1.2		
9	1993	I	21	20.0	1.1					
10		II	19	20.3		0.9				
11		III	17	20.6			0.8			
12		IV	25	20.9				1.2		
13	1994	I	22	21.2	1.0					
14		II	19	21.5		0.9				
15		III	19	21.9			0.9			
16		IV	26	22.2				1.2		
17	1995	I	24	22.5	1.1					
18		II	21	22.8		0.9				
19		III	20	23.1			0.9			
20		IV	27	23.4				1.2		
				Suma (licznik wyrażenia we wzorze)	5.31	4.64	4.27	5.83	Suma kontrolna	
				Względny wskaźnik okresowości	1.06	0.93	0.85	1.17	4.01	
				Współczynnik korygujący	1.00					
				Skorygowany (czysty) wskaźnik	1.06	0.93	0.85	1.16	4.00	

Porównanie wyników przykładów

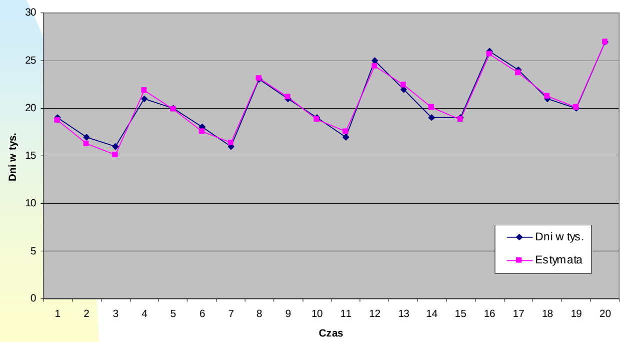
		Kwartał			
		I	II	III	IV
Metoda	Wskaźnik				
mechaniczna	bezwzględny	1.20	-1.62	-3.02	3.45
	względny	1.06	0.92	0.85	1.17
analityczna	bezwzględny	1.17	-1.55	-3.06	3.44
	względny	1.06	0.93	0.85	1.16

Konkluzja: miary okresowości uzyskane metodą mechaniczną i metodą analityczną nie są identyczne, ale bardzo zbliżone.

Wybór modelu

Ponieważ amplitudy zmian okresowych są względnie stałe, należałoby wybrać model addytywny.

Porównanie przebiegu rzeczywistego z modelem



Analiza sezonowości - Autokorelacja

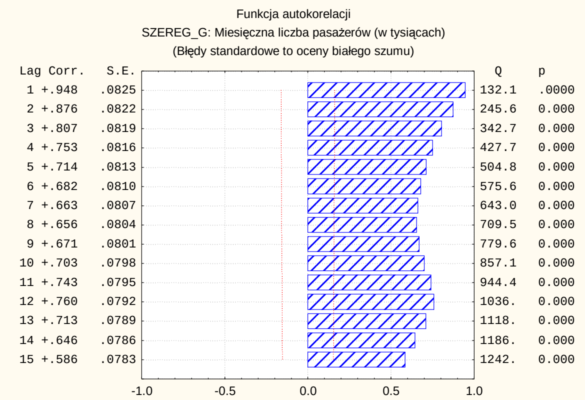
Cel: pomaga wykryć długość okresu wahań cyklicznych

Opisuje zależność przebiegu od niego samego, ściślej: zależność korelacyjną między *i*-tym elementem szeregu a *i*k*-tym elementem szeregu (dla różnych rzędów/opóźnień *k*).

$$\rho(X(t),k)=\frac{\sum_t (X(t)-\bar{X})(X(t+k)-\bar{X})}{\sigma^2}$$

Autokorelacja - przykład

Funkcja autokorelacji dla serii G



Przebieg wykładu

- Zbieranie i przygotowywanie danych
- Analiza szeregów czasowych
- Dekompozycja szeregu czasowego
- Prognozowanie
- Prognozowanie w przedsiębiorstwie
- Uwagi końcowe
- Studium przypadku

Prognozowanie

Motto: Prognozowanie to sztuka przewidywania przyszłości ...
i uzasadniania, dlaczego owe przewidywania się nie sprawdzają

Forecasting is the art of saying what will happen, and then explaining why it didn't
[Chatfield 1986]

Definicja (cechy) prognozy

- prognoza jest formułowana z wykorzystaniem dorobku nauki,
- prognoza jest stwierdzeniem odnoszącym się do określonej przyszłości,
- prognoza jest stwierdzeniem weryfikowalnym empirycznie,
- prognoza nie jest stwierdzeniem stanowczym (jest niepewna), ale jest stwierdzeniem akceptowanym.

Za [Cieślak 1993]

Celowość konstruowania prognoz

Analogia do praw mechaniki Newtona (!):

- prawidłowości obserwowane w danych są tym wyraźniejsze, im silniejsze są powiązania zjawisk (procesów, podmiotów) je generujących,
- zależności obecne w świecie rzeczywistym charakteryzują się często występowaniem pewnej inercji,

(dotyczy m.in.. prognozowania gospodarczego)

Założenia

- zaobserwowany trend nie zmieni się co do kształtu i siły działania w okresie przyszłym,
- wahania przypadkowe nie zakłócą znacząco zaobserwowanego trendu.

Wnioski:

- Spełnienie powyższych założeń ma większą szansę powodzenia dla okresów leżących bliżej ostatniego okresu badanego aniżeli dla okresów bardziej odległych.
- Niezbędna wiedza dziedzinowa, czyli znajomość charakteru zjawiska, przyczyn (np. ekonomicznych czy pozaekonomicznych) warunkujących jego rozwój.

Etapy postępowania prognostycznego

- Zgromadzenie materiału empirycznego (ze źródeł wewnętrznych i zewnętrznych).
- Przetworzenie materiału empirycznego.
- Przyjęcie reguły prognostycznej.
- Ustalenie horyzontu prognozy.
- Ocena jakości skonstruowanej prognozy.
  - ... świecenie oczami przed przełożonymi lub zleceniodawcą, gdy prognoza się nie sprawdza ;-)

Błąd prognozy *ex ante*

Błąd prognozy *ex ante* jest to dokonana w chwili budowy prognozy ocena różnicy między rzeczywistą wartością zmiennej Y w momencie/okresie t>n (n – numer ostatniej znanej obserwacji) a wyznaczoną prognozą. Ocenia *wiarygodność prognozy*.

Cechy:

- estymata,
- opinie sformułowane na jego podstawie da się zweryfikować po upływie czasu, do którego prognoza się odnosi,
- daje się wyznaczyć tylko dla prognoz ilościowych,
- nie wszystkie metody pozwalają na jego wyznaczenie.

Błąd prognozy *ex post*

Błąd prognozy *ex post* to inaczej *trafność* prognozy.

Rodzaje:

- Bezwzględny błąd prognozy *ex post*

$$q_t = y_t - y_t^*$$

- Względny błąd prognozy *ex post*

$$\Psi_t = \frac{y_t - y_t^*}{y_t} 100\%$$

- Średni błąd absolutny:

$$MSE = \frac{1}{n}\sum_t |z_t|$$

- Średni procentowy błąd absolutny:

$$MAPE = \frac{MSE}{y_t'} \cdot 100$$

89 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Metody prognozowania

90 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Rodzaje modeli szeregów czasowych dla prognozowania

Dokonuje się podziału ze względu na (domniemaną) postać SC na:

- modele ze stałym poziomem zmiennej prognozowanej,
- modele z trendem,
- modele z wahaniami okresowymi (sezonowymi i/lub cyklicznymi),
- + kombinacje.

91 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Modele ze stałym poziomem zmiennej prognozowanej

92 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Modele prognozowania naiwne

- oparte na bardzo prostych przesłankach dotyczących przeszłości, i
- zakładające, iż nie wystąpią zmiany w dotychczasowym sposobie oddziaływania czynników określających wartości zmiennej prognozowanej,
- umożliwiają budowanie jedynie prognoz krótkookresowych,

Najprostsza wersja (oparta na modelu tzw. *błądzenia losowego* znanym ze statystyki):

$$y_t^* = y_{t-1}$$

Badania pokazują [Makridakis 1989], że w niektórych zastosowaniach prognozy naiwne sprawdzają się bardzo dobrze (np. rynek papierów wartościowych) (!).

Modele średniej ruchomej (prostej)

Wykorzystywane zarówno do wygładzania, jak i prognozowania.

Założenie: poziom wartości zmiennej prognozowanej jest prawie stały w rozpatrywanym okresie (niewielkie odchylenia losowe, brak sezonowych i cyklicznych).

Idea: wartość zmiennej prognozowanej w następnym okresie jest równa średniej arytmetycznej z k ostatnich wartości tej zmiennej:

$$y_t^* = \frac{1}{k}\sum_{i=t-k}^{t-1} y_i$$

gdzie: k - stała wygładzania.

Problem: jak dobrać *k* ?

Dobór stałej wygładzania *k*

Można oprzeć na badaniu średniokwadratowego błędu prognozy *ex post:*

$$s^* = \left[\frac{1}{n-k}\sum_{t=k+1}^{n}\left(y_t - y_t^*\right)^2\right]$$

Idea: wybierz tę wartość *k*, która minimalizuje estymatę błędu s*.

Wada modelu średniej ruchomej: wszystkie obserwacje z rozpatrywanego „okna historii” partycypują w tym samym stopniu w budowaniu prognozy.

Idea: Nadawać mniejsze wagi obserwacjom starszym => postarzanie informacji.

Model średniej ruchomej ważonej

Prognoza zbudowana na podstawie ważonej średniej ruchomej z odpowiednio dobranymi wagami:

$$y_t^* = \sum_{i=t-k}^{t-1} y_i w_{i-t+k+1}$$
$$\forall i \in \langle 1, k-1 \rangle : \; w_i < w_{i+1}$$
$$w_1 > 0, \; w_k \le 1$$
$$\sum_{i=1}^{k} w_i = 1$$

gdzie:

- w_i – waga nadana przez prognostę.

Problem: jak dobrać wagi w_i ?

Model wygładzania wykładniczego

Idea: model średniej ruchomej ważonej z wagami dobranymi wg *prawa wykładniczego*.

Rozważamy tzw. *prosty* model wyrównywania wykładniczego.

Założenia: prawie stały poziom zmiennej i wahań przypadkowych.

Formuła:

$$y_t^* = \alpha y_{t-1} + (1-\alpha) y_{t-1}^*$$
$$\alpha \in (0,1]$$

gdzie:

- α – *parametr wygładzania*, waga nadana ostatniej (najnowszej) obserwacji zmiennej prognozowanej.

97 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Model wygładzania wykładniczego

Inna reprezentacja:

$$y_t^* = y_{t-1}^* + \alpha\left(y_{t-1} - y_{t-1}^*\right)$$

Jest to zatem w sumie prognoza naiwna korygowana ostatnio popełnionym błędem bezwzględnym *ex post*.

Problem: jak dobrać α ?

Odpowiedź: można estymować podobnie jak w przypadku stałej wygładzania.

$$s^* = \left[\frac{1}{n-k}\sum_{t=k+1}^{n}\left(y_t - y_t^*\right)^2\right]$$

98 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Modele z trendem

99 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Model z trendem – postać ogólna

Model addytywny

$$y_t = f(t) + \varsigma_t$$

Model multiplikatywny

$$y_t = f(t)\varsigma_t$$

gdzie:

- f(t) – funkcja trendu,
- ζ_t - zmienna losowa (wahania przypadkowe).

100 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Stosowane modele szeregów z trendem

- analityczne,
- oparte na funkcjach segmentowych,
- model Holta,
- autoregresyjne,
- podwójnej średniej ruchomej.

Modele analityczne

Kluczowy problem: ustalenie postaci analitycznej funkcji trendu (na podstawie przesłanek wynikających z mechanizmu rozwojowego zmiennej prognozowanej, czyli pewnej *wiedzy dziedzinowej*).

Modele przydatne w praktyce można podzielić na wykorzystujące:

- funkcje o stałych przyrostach (model liniowy),
- funkcje o rosnących przyrostach,
- funkcje o malejących przyrostach.

(Zasadniczo chodzi tu o wartość bezwzględną przyrostu)

Etap 1: Wybór funkcji trendu

Wyboru funkcji trendu dokonujemy na podstawie przesłanek

- empirycznych (analiza danych),
- dedukcyjnych (wiedza dziedzinowa).

Im bardziej złożona funkcja trendu, tym lepiej powinna być uzasadniona od strony dedukcyjnej, tj. wynikać z głębszych, merytorycznych przesłanek dotyczących mechanizmu rozwojowego zjawiska.

(Pewien związek z zasadą brzytwy Ockhama)

Funkcje o rosnących przyrostach

- funkcja wykładnicza

$$y_t = \exp(\alpha + \tilde{\beta} t),\ \ \tilde{\beta} > 0, \quad \text{lub}\ \ y_t = \alpha\beta^t,\ \beta < 1$$

gdzie: β - stopa wzrostu.

- wielomian stopnia drugiego (parabola)

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2,\ \alpha_2 > 0$$

- funkcja potęgowa

$$y_t = \alpha t^\beta,\ \beta > 1$$

Przykład: przedsiębiorstwo wprowadza na rynek nowy produkt, którego sprzedaż wzrasta (przez jakiś czas) „nieograniczenie”.

Uwaga! Stosować tylko do prognoz krótkookresowych!

Funkcje o malejących przyrostach

- funkcja logarytmiczna

$$y_t = \alpha + \beta \ln t,\ \beta > 0$$

- funkcja potęgowa

$$y_t = \alpha t^{\beta},\ \beta \in (0,1)$$

- wielomian (parabola) odwrotnościowy

$$y_t = \alpha_0 + \alpha_1 t^{-1} + \alpha_2 t^{-2},\ \alpha_2 < 0$$

- wielomian (parabola)

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2,\ \alpha_2 < 0$$

105 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Funkcje o malejących przyrostach

Gdy ponadto wiemy, że wzrost zdąża do pewnego poziomu:

- funkcja liniowo-odwrotnościowa

$$y_t = \alpha + \frac{\beta}{t},\ \beta < 0$$

- funkcja ilorazowa

$$y_t = \frac{\alpha t}{\beta + t},\ \alpha, \beta > 0$$

Przykład: względne nasycenie rynku.

Zaleta: mniejsze błędy (ryzyko) niż funkcje o rosnących przyrostach.

106 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Funkcja logistyczna

Szczególne istotna w zastosowaniach ekonomicznych, bo modeluje tzw. *krzywą życia produktu*.

$$y_t = \frac{\alpha}{1 + \beta e^{-\delta t}}$$

Trzy fazy:

- wprowadzenie produktu na rynek,
- przyspieszone i malejące tempo wzrostu popytu na produkt,
- nasycenie rynku, spadek popytu.

107 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Etap 2: Estymowanie parametrów funkcji trendu

Najczęściej: metoda najmniejszych kwadratów (MNK).

[Nie dotyczy funkcji logistycznej (zawiera nieliniowy związek pomiędzy parametrami a zmiennymi); wymaga innych metod, np. Hotellinga, Marquardta.]

Ocena dopasowania:

- odchylenie standardowe składnika resztowego (mianowane),
- współczynnik zmienności losowej (nie mianowany),
- współczynnik determinacji,
- skorygowany współczynnik determinacji.

108 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Prognozowanie z trendem - przykład

Stawianie prognozy na podstawie trendu wyznaczonego metodą analityczną i wyznaczonych absolutnych wskaźników okresowości.

						Kwartały				
Czas	Lata	Kwartały	Dni w tys.	Prosta regresji	Prognoza	I	II	III	IV	
14		II	19	21.5			-2.5			
15		III	19	21.9				-2.9		
16		IV	26	22.2					3.8	
17	1995	I	24	22.5		1.5				
18		II	21	22.8			-1.8			
19		III	20	23.1				-3.1		
20		IV	27	23.4					3.6	
21	1996	I		23.7						
22		II		24.0						
23		III		24.3						
24		IV		24.6	28.1					
a	0.31	uma (licznik wyrażenia we wzorze)				6.05	-7.50	-15.05	17.40	Suma kor
b	17.2	ezwzględny wskaźnik okresowości				1.21	-1.50	-3.01	3.48	0.18
		Współczynnik korygujący				0.05				
	Skorygowany (czysty) wskaźnik [tys. h]					1.17	-1.55	-3.06	3.44	0.00

109 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Modele autoregresyjne

Przesłanka: wiele zjawisk charakteryzuje się pewną bezwładnością (opóźnieniem).

Np. w sprzedaży: tzw. *zasada echa*: popyt restytucyjny na trwałe dobra konsumpcyjne jest echem popytu z lat wcześniejszych.

Postać ogólna:

$$y_t = f\left(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, \zeta_t\right)$$

W praktyce najczęściej

- liniowa

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \zeta_t$$

- logarytmiczno-liniowa

$$\ln y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \zeta_t$$

110 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Modele szeregów czasowych z wahaniami okresowymi

111 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Metody

Dla przebieg z wahaniami okresowymi

- sezonowymi („łatwiejsze”),
  - metoda wskaźników,
  - analiza harmoniczna,
- cyklicznymi,
  - podejścia wieloetapowe, najczęściej oparte na wskaźnikach.

112 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Modele ARMA i ARIMA

Bardzo ogólna klasa modeli szeregów czasowych, tzw. modele autoregresji i średniej ruchomej.

Podstawa: zjawisko autokorelacji, tj. korelacji wartości zmiennej prognozowanej z wartościami tej samej zmiennej opóźnionymi w czasie.

Wyróżnia się trzy rodzaje:

- modele autoregresji,
- modele średniej ruchomej,
- modele mieszane autoregresji i średniej ruchomej.

ARMA = Autoregressive moving average

ARIMA = Autoregressive integrated moving average

Szacowanie jakości prognozy

Estymowanie błędu prognozy

Średni błąd prognozy $S_p(Y_T)$:

$$S_p(Y_T) = S(z_t) \cdot \sqrt{\frac{(T-\bar{t})^2}{\sum_{t=0}^{n-1}(t-\bar{t})^2} + \frac{1}{n} + 1}$$

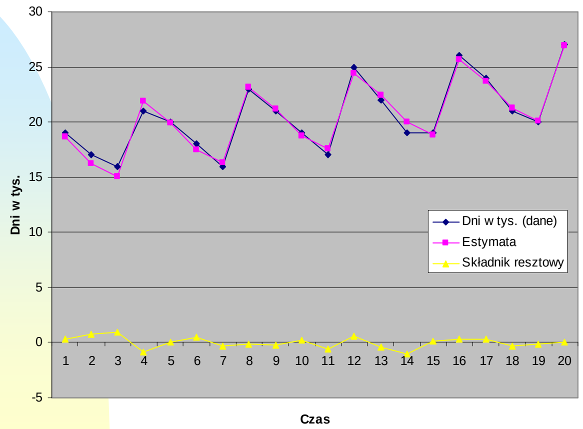
Gdzie:

- T – okres (moment), na który prognozujemy, T=n,n+1,...
- $S(z_t)$ – odchylenie standardowe składnika resztowego
- t – numeracja okresów w empirycznym szeregu czasowym
- t^{-} - średnia arytmetyczna numerów okresów,
- N – liczba badanych okresów (liczebność próby)

Estymowanie błędu prognozy – przykład

Czas	Lata	Kwartały	Dni w tys.	Prosta regresji	Bezwzględny wskaźnik okresowości	Estymata	Składnik resztowy	Wartość bezwzględna składnika resztowego	
1	1991	I	19	17,51	1,17	18,68	0,32	0,32	Współczyn
2		II	17	17,83	-1,55	16,28	0,72	0,72	prostej regr
3		III	16	18,14	-3,06	15,08	0,92	0,92	a
4		IV	21	18,46	3,44	21,90	-0,90	0,90	b
5	1992	I	20	18,77	1,17	19,94	0,06	0,06	
6		II	18	19,09	-1,55	17,54	0,46	0,46	
7		III	16	19,40	-3,06	16,34	-0,34	0,34	
8		IV	23	19,71	3,44	23,15	-0,15	0,15	
9	1993	I	21	20,03	1,17	21,20	-0,20	0,20	
10		II	19	20,34	-1,55	18,79	0,21	0,21	
11		III	17	20,66	-3,06	17,60	-0,60	0,60	
12		IV	25	20,97	3,44	24,41	0,59	0,59	
13	1994	I	22	21,29	1,17	22,46	-0,46	0,46	
14		II	19	21,60	-1,55	20,05	-1,05	1,05	
15		III	19	21,91	-3,06	18,85	0,15	0,15	
16		IV	26	22,23	3,44	25,67	0,33	0,33	
17	1995	I	24	22,54	1,17	23,71	0,29	0,29	
18		II	21	22,86	-1,55	21,31	-0,31	0,31	
19		III	20	23,17	-3,06	20,11	-0,11	0,11	
20		IV	27	23,49	3,44	26,93	0,07	0,07	
					Suma wartości bezwzględnych reszt			8,22	
							MSE	0,41	

Estymowanie błędu prognozy – przykład



Estymowanie błędu – interpretacja

Prognozowana liczba dni nie przepracowanych z powodu choroby dla IV kwartału 1996

- wynosi 28.1 tys. dni,
- z dokładnością 0.41 dni
  (czyli 0.41/28.1=1.45% dla tego okresu)

Rozkład odchyleń losowych modelu

Badania rozkładu odchyleń losowych dotyczą:

- symetrii (czy liczby odchyleń dodatnich i ujemnych są takie same),
- losowości (test serii),
- niezależności (dotyczy niezależności następujących po sobie reszt),
- nieobciążoności (czy wartość oczekiwana reszty jest równa 0),
- normalności (np. przy użyciu percentyli).

Z każdym z powyższych badań związany jest pewien test statystyczny. Dobry model powinien dawać negatywny wynik wszystkich powyższych testów.

Inne zagadnienia związane z prognozowaniem

Prognozy ilościowe i jakościowe

- Prognozy ilościowe - dotyczą przyszłej wielkości zjawiska.
- Prognozy jakościowe - dotyczą „zachowania” zjawiska w przyszłości (np. charakterystyki trendu, nasycenia, załamania, etc.); zazwyczaj oparte na sądach eksperta (-ów), modele „myślowe”.

Szczególnym przypadkiem prognozowania jakościowego (istotnym w zastosowaniach ekonomicznych) jest *prognozowanie ostrzegawcze*.

121 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Prognozowanie ostrzegawcze

Zadaniem prognozy ostrzegawczej jest dostarczenie *na czas* obserwacji o ewentualnej przyszłej niekorzystnej zmianie *kierunku rozwoju* czy *natężenia* badanego zjawiska (Siedlecka 1996)

W szczególności: Przypuszczenie, że w przyszłym momencie T_0 stan analizowanego zjawiska będzie niższy niż w momencie T_0-1.

		Rzeczywistość	
		Tak	Nie
Prognoza spadku	Tak		
	Nie		

122 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Prognozy punktowe i przedziałowe

- punktowe – prognoza jest pojedynczą wartością liczbową dla każdego okresu (momentu) *t*,
- przedziałowe – określamy przedział poziomu zjawiska dla każdego okresu (momentu) *t* (metodami estymacji parametrów).

123 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Prognoza statystyczna

Wielkość $x_p(t)$ nazywa się prognozą statystyczną nie znanej wartości X(t) szeregu czasowego dla $t \in \langle n+1, T\rangle$, jeżeli dla pewnych ε i η zachodzi:

$$P\{|X(t)-x_p(t)|<\varepsilon\}>1-\eta$$

- ε - *wiarygodność* prognozy,
- 1-η - (bliskie 1) *dokładność* (precyzja) prognozy.

Jest to tzw. warunek *dopuszczalności* prognozy.

124 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

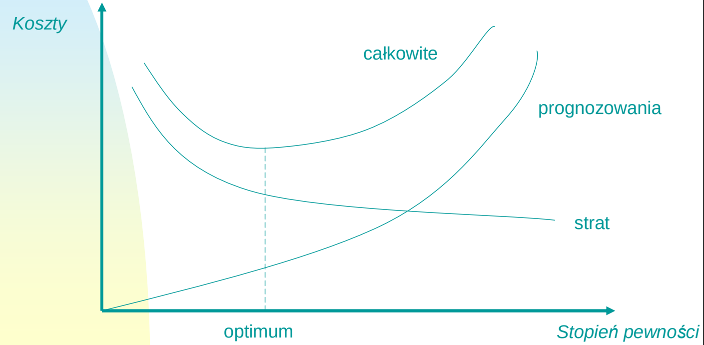
Koszty postępowania prognostycznego

Można podzielić na:

- koszty zebrania danych,
- koszty przechowywania danych,
- koszty przetwarzania danych.

Koszty prognozowania a koszty strat

Większe wydatki na przygotowanie prognoz => mniejszy stopień niepewności w procesie decyzyjnym i, w konsekwencji, mniejsze koszty (błędów).



Horyzont prognozy - definicja „naiwna”

Horyzont prognozy (horyzont czasowy) to ostatni moment (okresem), dla którego budujemy prognozę.

Okres prognozy długość przedziału pomiędzy ostatnim momentem (okresem), dla którego znane są wartości analizowanych przebiegów (ostatnim momentem, dla którego znane są dane obserwowane, „historyczne”), a horyzontem prognozy.

Tradycyjnie wyróżnia się prognozy:

- krótkoterminowe (< 3 miesiące),
- średnioterminowe (3 miesiące do 2 lat), i
- długoterminowe (> 2 lat).

(w kontekście prognozowania w przedsiębiorstwie)

Horyzont prognozy a typy prognozy

- Krótkoterminowa – zakładamy, że w badanym zjawisku zajdą tylko zmiany ilościowe.
- Średnioterminowa – w badanym zjawisku zajdą zmiany ilościowe i niewielkie zmiany jakościowe.
- Długoterminowa – w badanym zjawisku mogą wystąpić zarówno zmiany ilościowe, jak i jakościowe.

Horyzont prognozy - definicja „realistyczna”

Horyzont prognozy to

$$T = \max_t \{\forall \tau = 1..t : \delta_\tau \le \delta\}$$

gdzie:

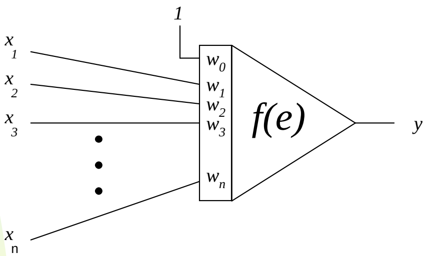
- δ dopuszczalny błąd prognozy,
- δ_τ błąd prognozy dla (przyszłego) okresu (momentu) .

129 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Sztuczne sieci neuronowe w prognozowaniu

130 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Model sztucznego neuronu

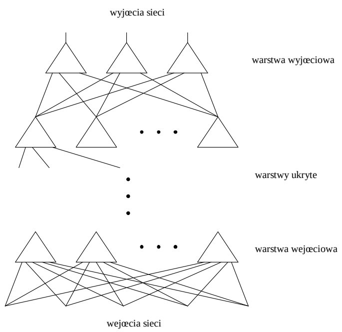


- *n wejść x_i* (*synapsy*) opatrzonych *wagami w_i* (wektor wag **w** i wektor wejść **x**)
- waga synapsy decyduje o jej ważności

131 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Sieć warstwowa

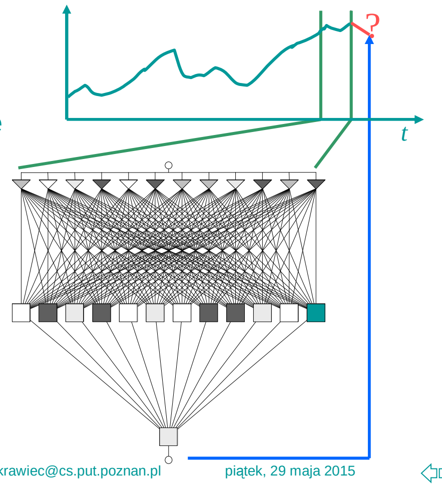
- neurony ułożone warstwami
- połączenia tylko pomiędzy kolejnymi warstwami
- jednokierunkowa (brak sprzężeń zwrotnych)
- neurony nieliniowe => *perceptron*
- działanie sieci: propagacja sygnału od warstwy wejściowej do wyjściowej



132 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Zarys idei

Prognozowanie
- ilościowe
- jakościowe



Inne zagadnienia związane z tematem

Zagadnienia teoretyczne:
- Podejścia hybrydowe: jednoczesne stosowanie wielu metod reprezentujących różne „filozofie”.
- Prognozowanie zbiorów szeregów i analiza zależności pomiędzy szeregami.
- Sterowanie procesami (co muszę podać na wejście, żeby otrzymać pożądaną wartość wyjścia).
- Analiza ryzyka.
- Wizualizacja.

Przebieg wykładu

- Zbieranie i przygotowywanie danych
- Analiza szeregów czasowych
- Dekompozycja szeregu czasowego
- Prognozowanie
- Prognozowanie w przedsiębiorstwie
- Uwagi końcowe
- Studium przypadku

Prognozowanie w przedsiębiorstwie

Prognozowanie w przedsiębiorstwie

Otoczenie marketingowe przedsiębiorstwa – zespół czynników zewnętrznych bezpośrednio lub pośrednio wpływających na jego działania (Dittman 1998)

otoczenie = mikrootoczenie + makrootoczenie

Mikrootoczenie – elementy bezpośrednio wpływające na działanie przedsiębiorstwa (dostawcy, konkurenci, pośrednicy marketingowi, nabywcy, inne podmioty)

Makrootoczenie – wpływ pośredni (czynniki demograficzne, ekonomiczne, społeczno-kulturowe, naturalne, technologiczne, polityczno prawne, etc). Cecha wspólna: przedsiębiorstwo ma na nie znikomy wpływ.

137 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Prognozowanie w przedsiębiorstwie...

... może być realizowane w warunkach:

- pewności (rzadko),
- ryzyka (znajomość [rozkładów] prawdopodobieństw powiązanych z wariantami),
- niepewności (brak znajomości rozkładów prawdopodobieństw),
- niepełnej informacji (nieznajomości wszystkich wariantów/kryteriów)

138 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Po co prognozować w przedsiębiorstwie?

Potrzeba prognozowania w przedsiębiorstwie wynika z:

- niepewności przyszłości,
- opóźnienia w czasie między momentem podjęcia decyzji a wynikłymi z niej skutkami.

Procedury prognostyczne są obecnie częścią systemu wspomagania decyzji większości systemów informacji marketingowej.

Badania pokazują, że ok.. 90% przedsiębiorstw w Europie Zachodniej stosuje prognozowanie krótkoterminowe.

139 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

Niestety ...

... prognozy zjawisk gospodarczych, nawet przy lepszym niż obecnie poznaniu ich mechanizmów, nie osiągną nigdy stopnia pewności prognoz zjawisk fizycznych (Dittman 1998, s. 18)

140 © krzysztof.krawiec@cs.put.poznan.pl piątek, 29 maja 2015

„Szczoteczka pocieszenia” albo *Schandenfreude*

Przykłady wyjątkowo nietrafnych prognoz:

- liczba koni w Paryżu (koniec XIX wieku),
- oszacowanie liczby samochodów we Francji: prognoza 2 mln (1945), rzeczywistość 14 mln (1970),
- oszacowanie wielkości potencjalnego rynku kserokopiarek przez IBM: prognoza 5 tys. (1959), rzeczywistość 200 tys. (1970),

Metody prognozowania sprzedaży

- ilościowe
  - modele szeregów czasowych,
  - modele ekonometryczne,
  - modele analogowe,
  - modele zmiennych wiodących,
  - modele analizy kohortowej,
  - testy rynkowe,
- jakościowe
  - opinie sprzedawców,
  - opinie kierownictwa,
  - opinie ekspertów,
  - badania intencji nabywców.

(Dittman 1998)

Przebieg wykładu

- Zbieranie i przygotowywanie danych
- Analiza szeregów czasowych
- Dekompozycja szeregu czasowego
- Prognozowanie
- Prognozowanie w przedsiębiorstwie
- Uwagi końcowe
- Studium przypadku

Narzędzia i pakiety programistyczne

Pakiety analizy danych (przydatne do analiz off-line przeprowadzanych raz na jakiś czas):

- Statgraphics
- SPSS
- R
- QS
- STATISTICA
- SCA Statistical System

Rozwiązania dedykowane (tam gdzie wymagane jest prognozowanie on-line)

- => studium przypadku

Literatura

Kassyk-Rokicka, H. Statystyka nie jest trudna. Tom I: Mierniki statystyczne. Polskie Wydawnictwo Ekonomiczne, Warszawa, 1998.

Siedlecka, U. Prognozowanie ostrzegawcze w gospodarce. Państwowe Wydawnictwo Ekonomiczne, Warszawa, 1996.

Dittman, P., Metody prognozowania sprzedazy w przedsiebiorstwie. Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław, 1998.

Box, G.E.P., Jenkins, G.M. Analiza szeregow czasowych. Państwowe Wydawnictwo Naukowe, Warszawa, 1983.

Gately, E. Sieci neuronowe. Prognozowanie finansowe i projektowanie systemów transakcyjnych. WIG-Press Warszawa, 1999.

Rybinski, K. Analiza techniczna. Wydawnictwo AWA-press, Warszawa, 1995.

Cytowania:

Cieślak, M. Prognozowanie gospodarcze, 1993.

Ostasiewicz S., Metody dyskryminacyjne w prognozowaniu dyskretnym, Ossolineum, Wrocław 1989.