



POLSKA AKADEMIA NAUK
Instytut Badań Systemowych

**ZARZĄDZANIE
RYZYSKIEM
INWESTYCYJNYM**
wybrane zagadnienia

**Maciej Krawczak,
Antoni Miklewski,
Andrzej Jakubowski,
Piotr Konieczny**





ZARZĄDZANIE RYZYKIEM INWESTYCYJNYM
wybrane zagadnienia

Publikację opiniowali do druku:

Prof. dr hab. Bogdan KRAWIEC

Doc. dr hab. Leszek ZAREMBA

Publikacja finansowana przez
KOMITET BADAŃ NAUKOWYCH w ramach projektu
badawczego Nr 1 1H02D 003 14 nt. „Zarządzanie ryzykiem
cenowym banku: krótkoterminowe prognozy cen rynkowych”

Copyright © by Instytut Badań Systemowych PAN
Warszawa 2000

ISBN 83-85847-52-9

ISSN 0208-8029

Semia

Bibl. podręcznik



444 30



CZEŚĆ V.

ANALIZA I PREDYKCJA SZEREGÓW CZASOWYCH

Spis treści rozdziału:

1.	Wprowadzenie.....	211
2.	Teoria stóp procentowych.....	213
	2.1. Stopa procentowa.....	213
	2.2. Stopy procentowe banku centralnego.....	214
	2.3. Ryzyko stopy procentowej.....	215
	2.4. Immunizacja portfeli inwestycyjnych.....	216
3.	STRUKTURA CZASOWA STÓP PROCENTOWYCH.....	217
4.	LINIOWE MODELE SZEREGÓW CZASOWYCH.....	219
	4.1. ARMA.....	219
	4.2. Dopasowanie liniowego modelu szeregu czasowego.....	223
	4.3. Słabości modeli liniowych.....	224
5.	OPÓŹNIENIA I TEORIA INFORMACJI.....	228
	5.1. Rekonstrukcja przestrzeni stanu.....	229
	5.2. Analiza właściwości.....	230
6.	MODELOWANIE.....	235
	6.1. Modele lokalne.....	235
	6.2. Modele globalne.....	236
7.	ZAGADNIENIA MODELOWANIA SZEREGÓW CZASOWYCH.....	238
	7.1. Przewidywanie przedziałów błędów.....	238
	7.2. Predykcje iterowane i bezpośrednie.....	243
	7.3. Przejście sieci.....	244
	7.4. Analiza sieci.....	245
8.	PROGNOZOWANIE.....	248
	8.1. Prognozowanie bez przedziału błędów.....	248
	8.2. Predykcja z przedziałami błędów.....	249
9.	PODSUMOWANIE.....	251
	LITERATURA.....	252



1. WPROWADZENIE

Szeregi czasowe są podstawą do scharakteryzowania obserwowanego złożonego zjawiska i do predykcji przyszłych stanów tego zjawiska. Poprzez obserwowane zjawisko rozumiemy tutaj złożony system, którego stan (lub stany) mogą być obserwowane w postaci szeregu czasowego (lub szeregów czasowych). Do takich systemów możemy zaliczyć systemy ekonomiczne, a obserwowanymi stanami takiego systemu są stopy procentowe.

Predykcja oraz chęć (albo często potrzeba) zrozumienia przeszłości skłania nas do poszukiwania praw, które wyjaśniają istotę zachowania obserwowanego zjawiska. W przypadku, gdy są znane deterministyczne równania opisujące zjawisko to rozwiązanie tych równań pozwala określić stan zjawiska na podstawie warunków początkowych. W przypadku, gdy nie są znane równania opisujące obserwowane zjawisko, a jesteśmy zainteresowani w predykcji stanów przyszłych, musimy określić zarówno reguły rządzące zjawiskiem (model) jak i aktualny stan. W tym rozdziale ograniczymy się do przypadku drugiego, tzn. gdy nie są znane równania opisujące interesujące nas zjawisko, a poszukiwane reguły opisujące dynamikę tego zjawiska muszą być uzyskane z przebiegu historii tego zjawiska. Dodatkowo, teoria informacji pozwala wyjaśnić pewne cechy charakterystyczne obserwowanego zjawiska. Metody sztucznej inteligencji, takie jak sztuczne sieci neuronowe pozwalają na konstrukcję modelu - typu black box - symulującego zjawisko, zaś uczenie maszynowe uczenia pozwalają na opis zjawiska w postaci skończonej liczby reguł w postaci JEŻELI ... TO ... W obu przypadkach zależy nam na narzędziach pozwalających wyjaśnić obserwowane zjawisko.

Analiza szeregów czasowych ma trzy zasadnicze cele:

- prognozę,
- modelowanie,
- analizę właściwości.

Celem prognozowania (często też używany jest termin predykcji) jest "dość" dokładne przewidywanie zmiany stanu systemu dla "stosunkowo" krótkiego horyzontu.

Celem modelowania jest znalezienie opisu zmian stanu systemu w długim horyzoncie czasowym. Bardzo często opis zjawiska służący do prognozowania może zasadniczo różnić się od opisu zjawiska służącego do modelowania. Np. równania opisujące krótkoterminowe zmiany systemu z reguły nie opisują właściwie zmiany systemu dla długich horyzontów czasowych i odwrotnie - modele opisujące zmiany systemu w długich przedziałach czasowych nie nadają się do krótkoterminowych prognoz.

Trzeci cel analizy szeregów czasowych, nazwany tutaj analizą właściwości, służy do określenia fundamentalnych właściwości, takich jak np. liczba stopni swobody systemu.

Do lat dwudziestych naszego wieku prognozowanie było zwykłą ekstrapolacją. Za początek nowoczesnych metod prognozowania można przyjąć drugą połowę lat dwudziestych kiedy to Yule (1927) zaproponował metodę autoregresji do prognozowania liczby plam na słońcu. W modelu Yule'a prognoza przyszłej liczby plam była ważoną sumą poprzednio zaobserwowanych ilości plam na słońcu. Przez pół wieku liniowy model Yule'a uzupełniony przez szum był podstawową metodą używaną do predykcji szeregów czasowych.

Oczywiście taki prosty liniowy model musi mieć wady, mianowicie nie jest w stanie modelować dynamiki złożonych systemów.

Pod koniec lat siedemdziesiątych naszego wieku w dziedzinie predykcji szeregów czasowych zostały zanotowane dwa znaczące osiągnięcia, związane z rozwojem technologii komputerowej pozwalającej na konstrukcję złożonych obliczeniowo algorytmów. Pierwszym koniecznym odnotowaniem osiągnięciem była tzw. rekonstrukcja przestrzeni stanów - podobna do idei obserwowalności stanów w teorii systemów dynamicznych. Rekonstrukcja przestrzeni stanów pozwala określić czy szereg czasowy jest opisem systemu deterministycznego, czy stochastycznego. Drugie bardzo ważne osiągnięcie to rozwój techniki zwanej "uczeniem maszynowym" - jest to znana już technika sztucznej inteligencji - a w konsekwencji rozwój sztucznych sieci neuronowych. Obecnie uczenie maszynowe jak i sztuczne sieci neuronowe są bardzo znanymi i wypróbowanymi narzędziami sztucznej inteligencji i często służą to tzw. eksploracji danych, czyli wnioskowaniu na podstawie danych. Obie wspomniane techniki charakteryzują się olbrzymimi możliwościami modelowania systemów.

W rozdziale przedstawimy, w części pierwszej, krótkie wprowadzenie do zagadnień teorii stóp procentowych. W części drugiej, także w charakterze wprowadzenia, przedstawione są zagadnienia analizy podstawowych czynników wpływających na kształtowanie się rynkowych stóp procentowych. Następnie, w części trzeciej, przedstawione są w zarysie liniowe modele szeregów czasowych. Problem rekonstrukcji przestrzeni stanów dla systemów dynamicznych z punktu widzenia fizyki i teorii informacji jest przedstawiony w części czwartej, natomiast w części piątej jest pokazane jak te teorie mogą być wykorzystane do scharakteryzowania szeregów czasowych. W części szóstej przedstawiony jest przegląd technik służących do predykcji szeregów czasowych, poczynając od modeli lokalnych a kończąc na modelach opartych na sieciach neuronowych. W części siódmej dyskutowane są modele sieci neuronowych pod kątem błędu

predykcji jak również pod kątem predykcji zmian gęstości prawdopodobieństwa. Część ósma poświęcona jest prognozowaniu w szeregach czasowych, zaś w końcowej części omawiane są przyszłe kierunki rozwoju badań związanych z predykcją. Rozdział ten zawiera także dwa dodatki, dodatek pierwszy poświęcony jest krótkiemu opisowi nowego algorytmu uczenia sieci neuronowych, zaś dodatek drugi omawia idee uczenia maszynowego pozwalającego generować reguły.

2. TEORIA STÓP PROCENTOWYCH

Pojęcie rynkowej stopy procentowej rozpatrywanej dla danego horyzontu czasowego wiąże się bezpośrednio z wszelkiego rodzaju inwestycjami dokonywanymi na różnego rodzaju rynkach; np. na rynku finansowym. Każda inwestycja wiąże się z rezygnacją wydatków w danej chwili na rzecz przyszłych korzyści, korzyści jednak obciążonych pewnym ryzykiem. W pracy Jajugi (1996) podkreślono, że takie właśnie ujęcie procesu inwestycyjnego uwypukla wszystkie najważniejsze cechy tego co rozumiemy pod pojęciem inwestowania. Efekty z inwestowania są spodziewane po pewnym określonym czasie, który musi upłynąć, aby określona inwestycja przyniosła pewne zyski. Nieodłącznym elementem wszelkiego rodzaju inwestowania jest ryzyko, które nie może być pomijane i z którym musi się liczyć każdy inwestor, gdyż przewidzenie przyszłości z dużą dokładnością jest zupełnie niemożliwe.

2.1. Stopa procentowa

Z przedstawionych powyżej cech procesu inwestycyjnego (Hirschleifera, 1965) wynika bezpośrednio, że wartość pieniądza jest różna w różnym czasie. Przykładowo, w chwili obecnej 1 mln zł może być dużo więcej warte od 1,5 mln zł ale po upływie określonego czasu, mimo, że ta druga kwota nominalnie jest o 50% większa od pierwszej. Pieniądz "leżący", czyli nie zainwestowany. Przyczyną są wszelkiego rodzaju procesy inflacyjne wpływające na obniżenie siły nabywczej pieniądza.

Procesy inflacyjne mają duży wpływ na wartość pieniądza, ale nie są one jedynymi składnikami wpływającymi na wielkość oraz zmiany stóp procentowych obowiązujących na danym rynku. Nawet w przypadku zerowej inflacji (Sierpińska i Jachna, 1994; Jajuga, 1996) można wyróżnić kilka zasadniczych przyczyn zmienności wartości pieniądza w czasie:

- bieżąca konsumpcja – duża część każdego społeczeństwa preferuje bieżącą konsumpcję na ustalonym poziomie niż odłożoną w czasie konsumpcji na wyższym poziomie,
- możliwość inwestowania – tzn. odpowiednio zainwestowana kwota pieniędzy (np. złożona w banku, przeznaczona na zakup obligacji, akcji, walut obcych itp.) może w przyszłości wartość tej kwoty (nie uwzględniając ryzyka),
- ryzyko – każda inwestycja jest jednak obciążona pewnym ryzykiem, z reguły im większe są spodziewane profity tym większe jest ryzyko nie uzyskania tych profitów,
- płynność finansowa – każde inwestowanie jest związane z zamrożeniem określonych środków finansowych, a tym samym zmniejszeniem bieżącej płynności finansowej.

Podsumowując, każdy inwestor inwestując (odkładając w czasie uzyskanie zysków) oczekuje pewną premię; ta premia jest stopa (rate of return) z podejmowanej inwestycji, czyli stopa procentowa (interest rate). Podstawowe dwa składniki stopy procentowej to “cena czasu” i “cena ryzyka” (Jajuga, 1996).

2.2. Stopy procentowe banku centralnego

Należy odróżnić pojęcie rynkowej stopy procentowej (tj. stopy procentowej obowiązującej dla danego okresu na rynku pieniężnym) od stóp procentowych ustalanych przez bank centralny (np. stopy kredytu redyskontowego, lombardowego). Stopy procentowe ustalone przez bank centralny mają charakter normatywny są pewnymi parametrami sterującymi, za pomocą których steruje (w pewnym sensie) na stan rynku pieniężnego w danym kraju.

Na rynku pieniężnym występuje cały złożony proces wzajemnych oddziaływań pomiędzy poziomem stopy redyskontowej i lombardowej banku centralnego, a poziomem rynkowych stóp procentowych, obowiązujących dla zadanych okresów zapadalności podjętych zobowiązań finansowych. Bank centralny dysponuje szeregiem instrumentów umożliwiających aktywną ingerencję w procesy występujące na rynku pieniężnym.

Każdy inwestor może posiłkować się kredytem zaciągany w bankach. Taki kredyt jest nie tyle nabyciem praw własności do wypożyczonego kapitału, lecz uzyskaniem prawa do dysponowania tym kapitałem w określonym czasie na określonych warunkach. Właśnie zbycie prawa do dysponowania określonym kapitałem na określonych warunkach

musi ileś kosztować, te koszty stanowią dochód, jaki właściciel kapitału otrzymuje za zbycie tego prawa, a jednocześnie stanowią cenę kapitału pożyczkowego i jest nazywany jest *procentem*. W praktyce, w transakcjach kredytowych cenę pożyczki ustala się jako stosunek procentu do sumy wypożyczonego kapitału; stosunek ten nosi właśnie nazwę *stopy procentowej*.

2.3 Ryzyko stopy procentowej.

Wysokość procentu od kapitału jest określana na początku rozpatrywanego kontraktu dotyczącego wypożyczenia na określony czas tego kapitału i jest ona ustalana na podstawie obowiązujących w danej chwili rynkowych stóp procentowych. Ryzyko stopy procentowej wynika z samego faktu, że jedne podmioty finansują swoją działalność z pożyczonych środków od innych podmiotów, np. banków. W sytuacji wzrostu stóp procentowych, pożyczkobiorcy uzyskują niższe wyniki od przewidywanych. To z kolei może ujemnie wpływać na wyniki finansowe uzyskiwane przez podmioty udostępniające te środki.

W teorii finansów często rozróżnia się ryzyko stopy procentowej od ryzyka inflacji. Ryzyko inflacji, zwane też ryzykiem siły nabywczej pieniądza, określane jest często jako określona niepewność co do siły nabywczej przyszłych dochodów i łączy się je z określonymi oczekiwaniami inflacyjnymi. Można stwierdzić, że zmienność oczekiwań inflacyjnych ze strony podstawowych podmiotów rynku finansowego (a tym samym ryzyko inflacji) wywołuje ryzyko nieoczekiwanych zmian rynkowych stóp procentowych; np. rentowności do wykupu rocznych bonów skarbowych. Ryzyko inflacji oraz ryzyko rynkowych stóp procentowych są więc ze sobą ściśle powiązane. Ryzyko inflacji jest w tym przypadku jednym z podstawowych źródeł ryzyka stopy procentowej. W krajach rozwijających się, wpływ ryzyka inflacji na ryzyko rynkowych stóp procentowych jest na tyle silny, że często ryzyka te są często ze sobą utożsamiane.

Można wykazać (Jakubowski, 1996), że w warunkach rosnących oczekiwań inflacyjnych rentowność do wykupu czysto dyskontowych obligacji skarbowych wzrasta. Ale rentowność do wykupu obligacji czysto dyskontowych jest właśnie miernikiem rynkowych stóp procentowych *spot*. Tak więc w przypadku, gdy oczekiwana stopa inflacji rośnie, rosną również rynkowe stopy procentowe. Wzrost inflacji nie jest jednak warunkiem koniecznym wzrostu tych stóp. Często bowiem zachodzi sytuacja, że rynkowe stopy procentowe wzrastają w warunkach stałej inflacji. Może być to na przykład spowodowane nagłym wzrostem popytu na kapitał pożyczkowy, wywołany pewnymi czynnikami zewnętrznymi,

oddziaływującymi (w warunkach stałych oczekiwań inflacyjnych) na dany rynek finansowy.

Warto również podkreślić, że rozpatrywane powyżej ryzyko inflacji silnie oddziałuje również na inne - poza ryzykiem stóp procentowych - kategorie ryzyka inwestycyjnego; na przykład - na *ryzyko walutowe*, *ryzyko rynku akcji* itp.

2.4. Immunizacja portfeli inwestycyjnych

We współczesnej teorii inżynierii finansowej istnieje wiele metod zabezpieczania się przed ryzykiem stopy procentowej takich, jak dywersyfikacja przedsięwzięć inwestycyjnych lub stosowanie pochodnych instrumentów finansowych, tj. kontraktów terminowych i opcji dotyczących stóp procentowych.

Oddzielna klasa metod, jakie są w tym zakresie stosowane, to tzw. immunizacja portfeli inwestycyjnych ze względu na ryzyko stóp procentowych. W szczególności, w odniesieniu do rynku obligacji, pod pojęciem immunizacji rozumiemy takie zaprojektowanie udziałów wartościowych poszczególnych obligacji (o różnych terminach wykupu) wchodzących w skład analizowanego portfela, aby wartość globalna tego portfela była jak najmniej wrażliwa na nieoczekiwane zmiany rynkowych stóp procentowych.

Zagadnienie to rozpatruje się przy zadanym horyzoncie inwestycyjnym wynikającym z terminu płatności przyszłych zobowiązań finansowych. W najprostszych modelach immunizacyjnych zakłada się, że w przyszłości występować będzie pojedyncze zobowiązanie. Natomiast w modelach bardziej złożonych podstawowym problemem jest dopasowanie strumienia przyszłych dochodów wynikających z faktu posiadania określonego portfela obligacji (płatności odsetkowe) ze strumieniem przyszłych zobowiązań, rozpatrywanych w dyskretnych chwilach czasowych. Zagadnienie immunizacji nie ma na ogół jednoznacznego rozwiązania - istnieje wiele (lub nieskończenie wiele) portfeli umożliwiających dopasowanie przyszłych dochodów do przyszłych zobowiązań. Umożliwia to dodatkowo sformułowanie pewnej funkcji celu - np. maksymalizacja zysku lub minimalizacja kosztu utworzenia określonego portfela obligacji. Zagadnienie immunizacji staje się w rozpatrywanym przypadku zagadnieniem optymalizacji, rozwiązywanym za pomocą jednej z wielu technik programowania matematycznego. W zagadnieniu tym problem immunizacji portfela formułuje się w postaci określonego zbioru ograniczeń.

Zagadnienie immunizacji portfela obligacji jest pewnym szczególnym (ale ważnym) przypadkiem bardziej ogólnej problematyki zarządzania portfelowego obligacjami w warunkach ryzyka stóp procentowych.

Podstawowe etapy rozwoju teorii immunizacji wyznaczają prace Macaulay'a (1938) - wprowadzenie pojęcia okresowości obligacji, Redingtona (1952) - pierwsze użycie terminu "immunizacja" w odniesieniu do zagadnień inwestycyjnych, Coxa, Ingersolla, Rossa (1971) - model stochastyczny dynamiki zmian struktury czasowej stóp procentowych, Fishera, Weila (1977) - uogólnienie pojęcia okresowości obligacji, Bierwaga (1983, 1987) - koncepcja "okna okresowości", Brennana i Schwartza (1983) - dwuwskaznikowe modele dynamiki zmian stóp procentowych, oraz prace Garbade'a (1986, 1989), Littermana, Scheinkmana (1991) i Dahla (1993) - wprowadzające koncepcję czynnikowej okresowości i czynnikowej wypukłości obligacji oraz czynnikowy model zmian struktury czasowej stóp procentowych.

Ostatnie lata to burzliwy rozwój zarówno prac teoretycznych jak i badań empirycznych w zakresie immunizacji portfela obligacji ze względu na ryzyko stóp procentowych. Najnowsze wyniki z tego zakresu są publikowane w wielu czasopismach naukowych z teorii finansów; m.in. w *Journal of Finance* i *Journal of Fixed Income Securities*.

3. STRUKTURA CZASOWA STÓP PROCENTOWYCH

Przedmiotem rozważań niniejszej części jest analiza podstawowych czynników wpływających na kształtowanie się rynkowych stóp procentowych charakteryzujących się - w zależności od rozpatrywanego horyzontu inwestycyjnego - określoną strukturą czasową. Identyfikacja struktury czasowej w chwili bieżącej, jak również analiza dynamiki zmian tej struktury w przyszłości ma ogromne znaczenie z punktu widzenia wyceny tzw. wartości bieżącej oraz wartości przyszłej wszelkiego rodzaju przedsięwzięć inwestycyjnych. Struktura czasowa stóp procentowych wpływa na:

- wartości instrumentów finansowych funkcjonujących na rynku pieniężnym (bony skarbowe, bony pieniężne NBP, papiery komercyjne, weksle),
- rynek kapitałowy (obligacje, akcje, kontrakty terminowe forward, futures, opcje),
- wartość bieżącą netto (NPV) projektów inwestycyjnych przedsiębiorstw,
- wartość bieżącą oraz przyszłą płatności dokonywanych w ramach funduszy powierniczych, emerytalnych i ubezpieczeniowych.

Znajomość obecnej i przyszłej postaci struktury czasowej stóp procentowych jest również niezbędna dla sektora bankowego. Określenie

właściwej strategii dotyczącej udzielania kredytów krótko-, średnio- oraz długoterminowych (a w tym - odpowiednich proporcji określających portfel kredytowy banku), a także polityki przyjmowania depozytów terminowych - jest niemożliwe bez identyfikacji obowiązujących na danym rynku tzw. stóp procentowych *spot*. Odrębnym zagadnieniem jest prognozowanie zachowania się tych stóp procentowych w przyszłości.

W ogólnym przypadku odpowiednio precyzyjna identyfikacja struktury czasowej obowiązujących w danej chwili rynkowych stóp procentowych, właściwa interpretacja kształtu tej struktury oraz generowanie na tej podstawie prognoz co do zachowania się tych stóp w przyszłości jest zagadnieniem niezmiernie trudnym. Wynika to z bardzo dużej nieliniowości systemów finansowych i z niedoskonałości narzędzi używanych do predykcji szeregów czasowych.

Zarówno modelowanie matematyczne jak również modelowanie-symulacja przy wykorzystaniu narzędzi sztucznej inteligencji wymagają wyznaczenia pewnej liczby parametrów. Do wyznaczania tych parametrów niezbędna jest duża liczba danych historycznych. W zasadzie rynek amerykański dysponuje najpełniejszymi danymi, na którym nastąpił w ostatnim dwudziestolecu dynamiczny rozwój rynku obligacji (w tym tzw. czysto-dyskontowych obligacji skarbowych o okresach zapadalności do 30 lat). Takie dane umożliwiają identyfikację struktury czasowej stóp procentowych *spot* dla odpowiednio długiego horyzontu - nawet dla okresów kilkudziesięcioletnich. Dla celów identyfikacji często wykorzystuje się parametr rentowności do wykupu obligacji wielokuponowych - prowadzi to jednak do pewnych przybliżeń i niedokładności. Dodatkowa trudność przy identyfikacji polega natym, że istnieje kilka istotnie różniących się metod określania stóp procentowych *spot* na podstawie znanej rentowności obligacji wielokuponowych o różnym terminie zapadalności. Zastosowanie każdej z tych metod może prowadzić do istotnie różnych wyników.

Na rynku finansowym w Polsce - oprócz krótkoterminowych bonów skarbowych o terminach zapadalności 8 tyg., 13 tyg., 26 tyg., 39 tyg. oraz 52 tyg. (tj. 1 rok), które są pewną formą obligacji czysto-dyskontowych - istnieją jeszcze obligacje 2- i 5-letnie o stałym kuponie oraz obligacje skarbowe 1-, 3- i 10-letnie o zmiennym kuponie (tzw. obligacje indeksowane). Obligacje te są przedmiotem transakcji zarówno na rynku pierwotnym, jak i na rynku wtórnym. Należy podkreślić, że obligacje o zmiennym kuponie nie nadają się dla celów identyfikacji stóp procentowych *spot*. Na rynku polskim możemy próbować określać strukturę czasową stóp procentowych dla horyzontu czasowego nie dłuższego niż 5 lat, tak wszelkie prognozy stóp procentowych w Polsce dokonywane dla okresów dłuższych niż 5 lat nie mają sensu.

4. LINIOWE MODELE SZEREGÓW CZASOWYCH

Liniovne modele szeregów czasowych są intuicyjnie łatwe do zrozumienia oraz są łatwe do zastosowania. Jednak modele liniowe zupełnie nie nadają się do opisu trochę bardziej złożonych systemów. W tej części zostaną krótko przedstawione podstawowe cechy oraz przyczyny ograniczeń zastosowania modeli liniowych.

Historycznymi pozycjami literatury dotyczącymi modeli liniowych są prace Kołmogorowa z 1941 roku i Wienera z 1947 roku, współczesną, ale już klasyczną, pozycją literatury jest książka Boxa i Jenkinsa z 1976 roku.

4.1. ARMA

W pierwszym etapie przedstawimy ideę modeli liniowych, a następnie zadanie dopasowania modelu do rozpatrywanego systemu. Wprowadzimy wejścia zewnętrzne do modelu (modele średniej ruchomej - moving average models), a następnie tzw. pamięć modelu tzn. modele auto-regresyjne (autoregressive models).

4.1.1. Cechy modeli liniowych

Modele średniej ruchomej (Moving Average (MA) models)

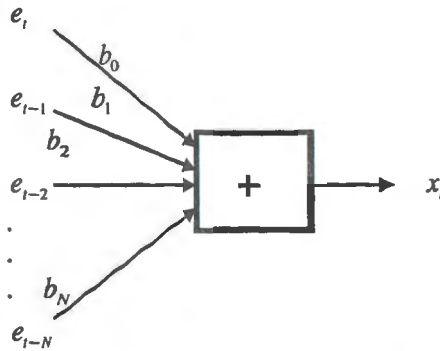
Założmy, że jest dany ciąg zewnętrznych wejść $\{e_t\}$, które oddziałują na obserwowany szereg czasowy $\{x_t\}$. Następnie zakładając liniowość systemu oraz obowiązywanie zasady *przyczynowości* (tzn. że na aktualną wartość x ma wpływ zarówno aktualna wartość e_t jak i N poprzednich wartości założonego wejścia e_{t-1}, \dots, e_{t-N}) możemy przytoczyć następującą zależność między wejściem i wyjściem modelu,

$$x_t = \sum_{n=0}^N b_n e_{t-n} = b_0 e_t + b_1 e_{t-1} + \dots + b_N e_{t-N}. \quad (1)$$

Zależność (1) można przedstawić w postaci prostego liniowego systemu - liniowego elementu przetwarzającego - Rys.1.

W teorii systemów równanie (1) opisuje filtr, tzn. generator nowego szeregu czasowego x na podstawie szeregu e przez liniowy filtr, którego współczynnikami są b_0, \dots, b_N . Specjaliści od ekonometrii lub statystycy nazywają równanie (1) modelem średniej ruchomej rzędu N (Moving Average model) $MA(N)$. Fizyczne znaczenie równania (1) jest następujące, filtr który jest uśrednia kilku - kilkunastu ostatnich wartości wejść. Wśród

inżynierów jest znane pojęcie impulsowego filtra o skończonej odpowiedzi (Finite Impulse Response) ze względu na fakt, że gdy zanika wejście e to po N krokach zanika również wyjście filtra x .



Rys. 1. Model średniej ruchomej (MA).

Własności nowego wyjściowego szeregu czasowego x zależą od własności wejściowego szeregu czasowego e . Zadaniem naszym jest określenie takiego systemu, który byłby niezależny od określonego ciągu wejść. Dla systemów liniowych odpowiedź filtra jest niezależna od wejścia.

Przedstawimy teraz trzy różne, ale jednocześnie równoważne, postacie modelu MA, a mianowicie określone:

- w dziedzinie czasowej - jako impulsowa odpowiedź filtra,
- w dziedzinie częstotliwości - jako gęstość widmowa,
- w kategoriach jego współczynników autokorelacji.

W pierwszym przypadku zakłada się, że wejście przybiera wartości różne od zera dla punktu t_0 , a w pozostałych punktach wejście jest zerowe.

Odpowiedź czasowa na wejściowy impuls jest dana przez współczynniki b równania (1); tzn. dla każdego kroku impuls wejściowy przesuwa się, aż dla $t = N$ wyjście systemu jest równe zero. Ciąg współczynników b_N, \dots, b_0 jest zatem odpowiedzią impulsową systemu. Odpowiedź na konkretne wejście może być traktowana jako superpozycja odpowiedzi z odpowiednimi opóźnieniami. Funkcja przejścia modelu dokładnie opisuje system liniowy, tzn. taki system, w którym obowiązuje zasada superpozycji i wyjście jest określone poprzez odpowiedź impulsową oraz wejście.

Czasami jest wygodniej rozważać filtr w dziedzinie częstotliwości. Jeżeli wejściem modelu typu MA jest impuls (płaska gęstość widmowa), to dyskretna transformata Fouriera wyjścia jest równa

$$\sum_{n=0}^N b_n \exp(-i2\pi n f). \quad (2)$$

Gęstość widmowa jest dana przez kwadrat amplitudy wyrażenia:

$$|b_0 + b_1 e^{-i2\pi 1 f} + \dots + b_N e^{-i2\pi N f}|^2. \quad (3)$$

Trzecia postać modelu MA, która przedstawia te same informacje, jest opisywana w kategoriach współczynników autokorelacji, tzn. w kategoriach wartości średniej $\mu = \langle x_t \rangle$ i wariancji $\sigma^2 + \langle (x_t - \mu)^2 \rangle$, w postaci

$$\rho_\tau \equiv \frac{1}{\sigma^2} \langle (x_t - \mu)(x_{t-\tau} - \mu) \rangle. \quad (4)$$

Nawias $\langle \rangle$ oznacza wartość oczekiwaną, często oznaczaną przez $E\{ \}$. Współczynniki autokorelacji określają jak dużo dwie realizacje szeregu czasowego, oddalone od siebie o τ kroków, różnią się od siebie w sensie wartości średniej. W przypadku, gdy wejście systemu jest procesem stochastycznym o wartościach nieskorelowanych wzajemnie ze sobą, tzn. $\langle e_i e_j \rangle = 0$ dla $i \neq j$, współczynniki autokorelacji są określone następującym wyrażeniem

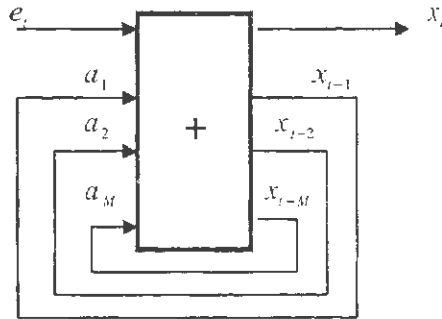
$$\rho_\tau = \begin{cases} \frac{1}{\sum_{n=0}^N b_n^2} \sum_{n=\tau}^N b_n b_{n-|\tau|} & |\tau| \leq N \\ 0 & |\tau| > N \end{cases}. \quad (5)$$

Modele autoregresji (AutoRegressive) AR

Z punktu widzenia teorii systemów filtry MA pracują w pętli otwartej, tzn. bez sprzężenia zwrotnego, dlatego sygnał wejściowy jest przenoszony bezpośrednio na wyjście modelu. Poprzez wprowadzenie do modelu sprzężenia zwrotnego można otrzymać model dynamiczny

$$x_t = \sum_{m=1}^M a_m x_{t-m} + e_t. \quad (6)$$

Taki model jest nazywany w literaturze modelem autoregresyjnym rzędu M (AR(M)), Rys. 2. W zależności od zastosowań człon e_t może oznaczać albo wejście, które może być w sposób zamierzony zmieniane, lub też może to być szum. W przypadku, gdy e_t jest białym szumem to współczynniki autokorelacji szeregu wyjściowego x można wyrazić przez współczynniki modelu.



Rys. 2. Model autoregresji.

Ze względu na wyprowadzenie sprzężenia zwrotnego każdy współczynnik autokorelacji jest rozwiązaniem odpowiedniego układu równań liniowych. Poprzez pomnożenie równania (6) przez $x_{t-\tau}$, a następnie wyznaczenie wartości oczekiwanej i normalizację (Box - Jenkins, 1976), współczynniki autokorelacji modelu AR można wyznaczyć rozwiązując układ równań liniowych, zwany równaniami Yula-Walkera:

$$\rho_t = \sum_{m=1}^M a_m \rho_{t-m}, \quad \tau > 0. \tag{7}$$

W przeciwieństwie do modelu MA współczynniki autokorelacji nie muszą równać się zero po M krokach. Wyznaczając transformaty Fouriera równania (6) można otrzymać następującą zależność:

$$\text{wyjscie} = \frac{\text{wejscie}}{1 - \sum_{m=1}^M a_m \exp(-i2\pi mf)}, \tag{8}$$

zaś gęstość widmowa wyjścia jest dana jako:

$$\frac{1}{\left|1 - a_1 e^{-i2\pi 1/f} - a_2 e^{-i2\pi 2/f} - \dots - a_M e^{-i2\pi M/f}\right|^2}. \quad (9)$$

W celu wygenerowania określonej realizacji szeregu muszą być określone warunki początkowe, które z reguły przyjmuje się jako pierwsze M wartości szeregu x .

Modele ARMA (AutoRegressive Moving Average)

Następnym krokiem może być połączenie modeli AR i MA w jeden model, który jest nazywany ARMA(M,N)

$$x_t = \sum_{m=1}^M a_m x_{t-m} + \sum_{n=0}^N b_n e_{t-n}. \quad (10)$$

Zastosowanie przekształcenia z (Oppenheim-Schafer, 1989), które opisuje dyskretną transformatę Fouriera,

$$X(z) \equiv \sum_{t=-\infty}^{\infty} x_t z^t, \quad (11)$$

pozwała uzyskać ciekawe wyniki. Przekształcenie z , na jednostkowym kole $z = \exp(-i2\pi f)$, ogranicza się do dyskretnej transformaty Fouriera, natomiast na zewnątrz jednostkowego koła przekształcenie z określa stopień zbieżności szeregu. Przekształcenie z wyjścia modelu ARMA daje następujące wyrażenie:

$$X(z) = A(z)X(z) + B(z)E(z) = \frac{B(z)}{1 - A(z)} E(z), \quad (12)$$

bez uwzględniania warunków początkowych.

Przekształcenie z zastosowane do szeregu czasowego generowanego przez równanie (10) może być zdekomponowane na dwa człony, pierwszy to funkcja rzeczywista, drugi - człon zależny od wejścia. Liczba biegunów i zer określa liczbę stopni swobody systemu.

4.2. Dopasowanie liniowego modelu szeregu czasowego

Dopasowanie współczynników

Układ liniowych równań Yule'a-Walkera (równanie (7)) pozwala na wyrażenie współczynników autokorelacji szeregu czasowego w kategoriach

współczynników modelu AR. Z drugiej strony równanie (7) pozwala na estymację współczynników modelu AR(M) na podstawie struktury obserwowanego szeregu. Estymacja współczynników jest właśnie zadaniem regresji, tzn. wyrażeniem następczej (przyszłej) wartości jako funkcji M poprzednich wartości; tzn. wyrażenie x_t jako funkcji $\{x_{t-1}, \dots, x_{t-M}\}$.

Zadanie to można sformułować jako zadanie minimalizacji kwadratu błędu, tzn. parametry (współczynniki) są określane w taki sposób, że tworzy się kwadraty różnicy między wyjściem modelu i obserwowaną wielkością i sumuje dla każdego kroku, a następnie taką funkcję błędu minimalizuje się ze względu na poszukiwane współczynniki. W zasadzie nie ma możliwości przedstawienia w jawnej prostej postaci wyrażenia na współczynniki modelu MA czy też ARMA na podstawie obserwowanych wielkości szeregu czasowego. Ale oczywiście istnieją dostatecznie proste metody numeryczne pozwalające w sposób rekurencyjny wyznaczyć poszukiwane współczynniki (Box-Jenkins, 1976; Press, 1992).

Oczywiście należy zdawać sobie sprawę z tego, że rzeczywisty system rzadko generuje sygnały, które mogą być opisane modelem (10), model (10) należy raczej traktować jako liniową aproksymację rzeczywistego systemu.

Wybór modelu i rzędu modelu

Do tej pory rozważaliśmy problem wyznaczania współczynników modelu ARMA, charakteryzowanego także przez rząd modelu (M, N). Właściwy dobór rzędu modelu ARMA jest bardzo ważnym odrębnym zagadnieniem. Zagadnienie to w ogólności nie ma jednoznacznego rozwiązania, tak w sensie doboru wartości współczynników jak i doboru liczby poszukiwanych współczynników modelu. Istnieje zasada, że zwiększenie rzędu modelu powoduje zmniejszenie błędu dopasowania. Istnieje kilka sposobów doboru właściwego rzędu modelu (Akaike, 1974), jednak głównym ich założeniem jest liniowość modelu oraz szczególnie rozkład szumu. Problem doboru rzędu modelu i wartości współczynników jest o wiele istotniejszym i trudniejszym zagadnieniem.

4.3. Słabości modeli liniowych

Jak już pokazaliśmy dla modelu ARMA trzy różne reprezentacje, tzn. poprzez współczynniki, gęstość widmową i współczynniki autokorelacji przechowują dokładnie te same informacje dotyczące liniowego systemu z nieskorelowanym białym szumem. W ten sposób, jeśli gęstość widmowa dobrze charakteryzuje cechy szeregu czasowego, to model typu ARMA będzie oczywiście dobrą reprezentacją. Prostota ta zawodzi jednak, gdy w

systemie występuje choćby prosta nieliniowość, która może doprowadzić do bardzo złożonej gęstości widmowej.

Dwa różne szeregi czasowe mogą mieć bardzo zbliżoną szerokość pasma widma, ale te szeregi czasowe mogą charakteryzować się zupełnie odmiennymi cechami. Dzieje się tak np. gdy porównujemy jeden szereg czasowy opisujący liniowy system zakłócony szumem i drugi nieliniowy system deterministyczny (bez szumu) o małej liczbie stopni swobody. Jest to właśnie przykład przypadków, do których nie mogą być stosowane operatory liniowe.

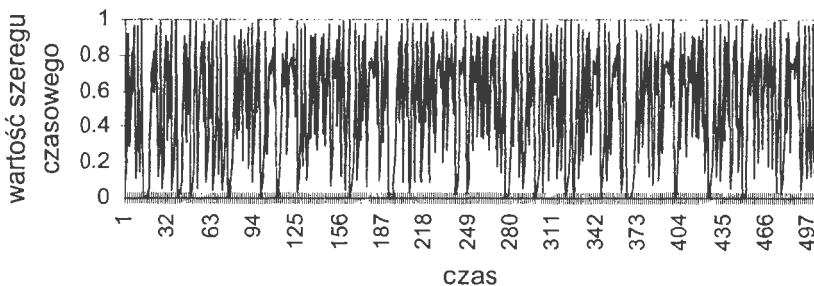
Rozważmy teraz dwa nieliniowe przykłady:

- Jako pierwszy przykład został przytoczony podany w 1947 r. przez Ulama i von Neumanna (1947) (Stanisław Ulam był wibitnym polskim matematykiem ze słynnej Szkoły Lwowskiej). W tym przykładzie model jest następujący, mianowicie wartość następna w szeregu czasowym jest określana z wartości bieżącej wykorzystując prostą aproksymację kwadratową

$$x_t = \lambda x_{t-1}(1 - x_{t-1}). \quad (13)$$

Ten model został spopularyzowany jako model rozwoju populacji, często w literaturze wspomniany jako “prosty model charakteryzujący złożoną dynamikę systemu”.

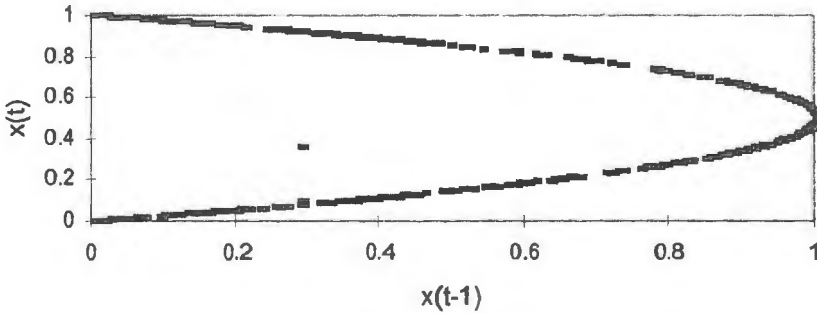
Warto zwrócić uwagę na fakt, że nawet takie proste równanie dynamiki systemu generuje bardzo złożony szereg czasowy, Rys. 3 dla 500 punktów, przyjmując $\lambda = 4$.



Rys. 3. Szereg czasowy generowany przez równanie (13).

Kwadratowe równanie (13) opisuje dynamikę wielu systemów, np. dynamikę przepływu cieczy, dynamikę w reaktorach chemicznych czy też dynamikę rynków finansowych. Równanie to jest nazywane

odzworowaniem logistycznym lub też odzworowaniem kwadratowym. Mimo, że równanie (13) generuje złożony szereg czasowy, to tzw. *dziwny atraktor* (Peitgen i in., 1996) takiego dynamicznego systemu, którego kształt przedstawia Rys. 4.



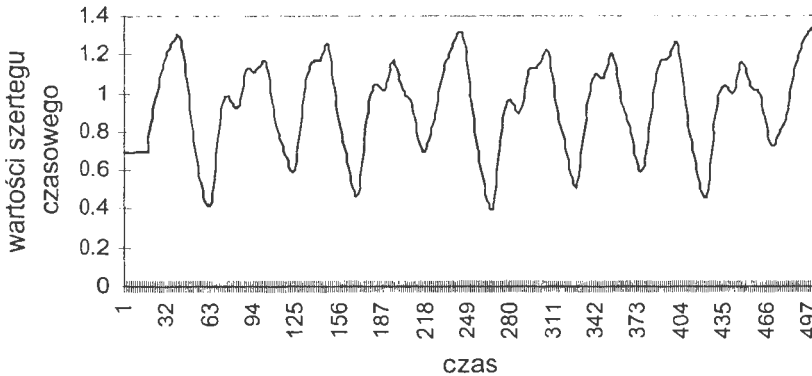
Rys. 4. Dziwny atraktor równania (13).

Wartość szeregu w modelu x_t w chwili t w sposób deterministyczny zależy od wartości poprzedniej x_{t-1} ; parametr λ jest odpowiedzialny za jakościowe aspekty modelu (parametr ten wpływa na "rozwarcie" paraboli) opisując tzw. chaos deterministyczny dla bardzo małych wartości. Przykładowo, dla $\lambda = 4$ w każdej iteracji albo jest tracony albo jest dodawany jeden bit informacji. Rozważając zależność x_t od x_{t-1} , każda wartość x_t może być określona przez dwie różne wartości x_{t-1} . Te dwie różne wartości są równoważne. Jeżeli znamy przed iteracją położenie punktu x_{t-1} z dokładnością do pewnej wartości parametru ε , to (w średniej) znajomość x_t jest określona z dokładnością 2ε . Taka wykładnicza zmiana dokładności określenia wartości x_t związana z każdą iteracją klasyfikuje zachowanie modelu jako deterministyczny chaos (Schuster, 1995).

- Drugi przykład jest równie prosty, rozważmy teraz szereg czasowy generowany przez równanie Mackey'a-Glassa (1977):

$$x(t+1) = bx(t) + a \frac{x(t-s)}{1+x^c(t-s)}. \quad (14)$$

Dla przykładowych parametrów: $a = 0.2$, $b = 0.9$, $c = 10$, $s = 18$ oraz warunków początkowych $x(0) = x(1) = \dots = x(18) = 0.7$ przebieg czasowy szeregu przedstawiony jest na Rys. 5.



Rys. 5. Pierwsze 500 punktów równania Mackey'a-Glassa.

Dziwny atraktor równania Mackey'a-Glassa przedstawiony jest na Rys. 6.

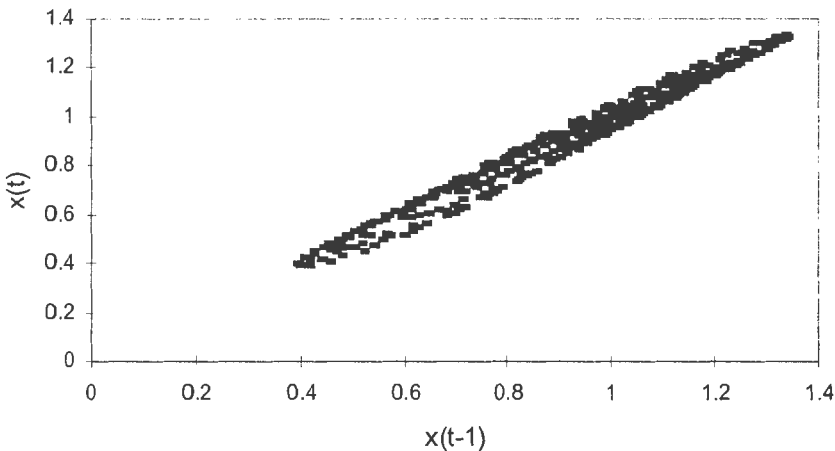


Fig. 6. Dziwny atraktor równania Mackey'a-Glassa.

Oba przedstawione systemy są całkowicie deterministyczne, tzn. rozwój tych systemów jest całkowicie określony przez warunki początkowe x_0 (odpowiada to szerokopasmowej gęstości widmowej). W przypadku

modelu ARMA występowanie szerokopasmowego składnika gęstości widmowej musi pochodzić od zewnętrznego szumu oddziałującego na wejścia systemu, jednak w rozpatrywanych tutaj przypadkach dwóch jednowymiarowych systemów odpowiada to paraboli i elipsy. W systemach deterministycznych właśnie nieliniowości są najbardziej interesujące i charakteryzujące dynamikę systemów, w takich przypadkach założenie o liniowości systemu i rozpatrywanie modeli liniowych jest bardzo ograniczone.

W literaturze teorii systemów rozważane są m.in. dwa rodzaje systemów, mianowicie *silne* i *słabe* modele. Modele silne to układ kilku równań z kilkoma parametrami, opisujących obserwowane zjawisko z pewnym "nadmiarem". W modelach słabych występuje bardzo mało założeń, a brak wiedzy o obserwowanym zjawisku jest kompensowany wprowadzaniem dużej ilości parametrów, których znaczenie trudno jest określić, a które należy wyznaczyć.

5. OPÓŹNIENIA I TEORIA INFORMACJI

Zagadnienie występowania opóźnień jest bardzo istotne w systemach, w modelowaniu zjawisk. Najpierw omówimy powiązanie rekonstruowanej przestrzeni stanów z wewnętrznymi stopniami swobody, a następnie przedstawimy pokrótce pewne elementy teorii informacji, które mogą być użyteczne przy analizowaniu szeregów czasowych.

Idea Yule'a dotycząca predykcji w szeregach czasowych polega na tym, że przyszła wartość szeregu jest określona przez wartości szeregu bezpośrednio ją poprzedzające. Model ARMA opisany równaniem (10) można przedstawić w następującej postaci

$$x_t = a \cdot x_{t-1} + b \cdot e_t, \quad (15)$$

gdzie

$$x_t = (x_t, x_{t-1}, \dots, x_{t-(d-1)}) \text{ i } a = (a_1, a_2, \dots, a_d).$$

W przedstawionych rozważaniach notacja jest trochę zmieniona, tzn. poprzednio używane było oznaczenie M jako rząd modelu typu AR, a teraz oznaczyliśmy rząd tego modelu przez d i jest on rozważany jako opóźnienie. W teorii informacji znany jest fakt wpływu opóźnienia na dynamikę systemu, wpływ ten został zasygnalizowany przez Ruelle'a (1980), a sam fakt został udowodniony przez Takensa (1981) oraz potem uściślony przez

Sauera, Yorka i Casdagliego (1991). Idea stosowania wektorów opóźnień, wystarczającej długości, nie jest jedynie sposobem opisu systemu liniowego, ale jest również użyteczna do wyekspozowania pełniejszej geometrycznej struktury systemu nieliniowego. Analizując szeregi czasowe chcielibyśmy wyciągnąć możliwie jak najwięcej informacji o obserwowanym systemie, ale bez uprzednich założeń o postaci równań, zakładając jedynie, że dane generowane są przez system dynamiczny, którego dynamikę może opisywać układ nieznanych równań różniczkowych.

5.1. Rekonstrukcja przestrzeni stanu

W analizie i predykcji szeregów czasowych rozważa się cztery różne przestrzenie:

1. *przestrzeń stanu* - jest to przestrzeń układu równań opisujących system; przestrzeń ta precyzuje maksymalną liczbę stopni swobody systemu, w przypadku np. systemu opisywanego przez równania różniczkowe cząstkowe typu Naviera-Stokesa przestrzeń konfiguracji ma wymiar nieskończony, a tym samym liczba stopni swobody systemu jest nieskończenie duża,
2. *przestrzeń rozwiązań* - jest to z reguły podprzestrzeń przestrzeni stanu; ze względu na fakt, że często nie wszystkie stopnie swobody są uaktywnione lub pewne stopnie swobody są skorelowane, wymiar przestrzeni rozwiązań może być znacznie mniejszy niż wymiar przestrzeni stanu; wymiar przestrzeni rozwiązań oznacza liczbę parametrów, które jednoznacznie określają system; często zdarza się, że dynamikę systemu opisuje układ równań różniczkowych cząstkowych, który może być zastąpiony układem równań różniczkowych zwyczajnych (Lorenz, 1963); redukcja wymiarowości systemu jest bardzo charakterystyczna dla systemów dysypatywnych (rozproszonych): rozproszenie systemu często może być zmniejszone do podprzestrzeni przestrzeni stanu,
3. *przestrzeń obserwacji* - jest to przestrzeń z reguły ograniczona jest do jednego wymiaru; zdarza się bardzo często, że jesteśmy w stanie obserwować pewien system jako skalarną funkcję zmiennych określających przestrzeń stanu; przykładowo to mogą być właśnie obserwowana stopa procentowa, poziom bezrobocia czy też inflacja,
4. *przestrzeń zrekonstruowana* - jest to przestrzeń otrzymana ze skalarnej obserwacji systemu po uwzględnieniu obserwacji historycznych (opóźnień).

Rozpatrując pewne zjawisko reprezentowane przez pewien system mamy do dyspozycji obserwowany szereg czasowy i nie posiadamy żadnej

innej informacji o obserwowanym systemie i o szeregu czasowym, a naszym zadaniem jest określenie jakie informacje można uzyskać o systemie i o dynamice tego systemu.

Założmy, że y jest wektorem stanu w przestrzeni rozwiązań, niech równanie $\frac{dy}{dt} = f(y)$ określa dynamikę systemu i niech obserwowana wielkość szeregu czasowego będzie określona jako $x_t = x(y(t))$.

Następnie oznaczamy opóźnienie przez τ i wymiar przez d , zaś wektor opóźnień możemy zdefiniować w następujący sposób

$$\text{wektor opóźnień: } x_t = (x_t, x_{t-\tau}, \dots, x_{t-(d-1)\tau}). \quad (16)$$

Zależność między x i y ma cechy tzw. *zanurzenia* (embedding), co w języku matematycznym oznacza, że x zależy przynajmniej od kilku współrzędnych wektora y , zaś pozostałe współrzędne wektora y są sprzężone i jako całość oddziałują na x . Dowód tego faktu jest podzielony na dwie części: w pierwszej części wykazuje się, że linearyzacja jest zawsze poprawna lokalnie, a w drugiej części - globalnie (Gershenfeld, 1989).

Problem wykorzystania opóźnień i zanurzenia jest stosunkowo nowym podejściem do analizy szeregów czasowych i różni się od tradycyjnego podejścia, a mianowicie:

- gwarantuje uzyskanie dokładnej informacji o zachowaniu się nieobserwowanych aktualnie stopni swobody systemu,
- pozwala zapomnieć o ścisłych założeniach stochastycznych,
- pozwala rozważać tylko wielkości, które są istotne, czyli niezmiennicze.

5.2. Analiza właściwości

Analiza właściwości szeregów czasowych opiera się na rozważaniach dotyczących używanych w teorii informacji miar redundancji i entropii systemu, których podstawą jest opis liczby bitów, jakie nowy pomiar daje o systemie. Początkowe rozważania dotyczą samego szeregu czasowego, w oderwaniu od obserwowanego systemu, zaś w następnym etapie należy przeprowadzić rozważania dotyczące prognozowania i oszacowania błędu.

Złożone szeregi czasowe (o skomplikowanym przebiegu) mogą być generowane zarówno przez stosunkowo proste (jak było pokazane wcześniej), jak i bardzo złożone systemy. Rozważymy metody analizy właściwości szeregów czasowych, które mogą być używane do określenia istotnych cech systemu (cech poza obserwowanymi wielkościami)

przydatnych do analizy, modelowania i do prognoz. Posłużymy się elementami teorii informacji w celu wprowadzenia podstawowej miary predykcji dla szeregów czasowych. Początki metod analizy szeregów czasowych sięgają prac Schustera z 1898 roku. W przypadku prostych liniowych systemów tradycyjna analiza spektralna jest przydatnym i dostatecznym narzędziem, ale jednocześnie inne systemy nieliniowe mogą mieć bardzo podobne pasma gęstości widmowej.

Związki między teorią informacji i twierdzeniem o ergodyczności są dość dobrze znane, ale Shaw (1981) połączył dynamikę systemów dyssypatywnych z teorią informacji, zaś Fraser (1989) używając teorii informacji podał sposób wyznaczania optymalnej wielkości opóźnienia.

Załóżmy, że szereg czasowy $x(t)$ został zdyskretyzowany na N części. Jeżeli n_T oznacza liczbę obserwowanych punktów i poszczególne wartości szeregu powtarzają się n_x razy, to prawdopodobieństwo estymacji tej wartości wynosi $p_1(x) = \frac{n_x}{n_T}$ (indeks l oznacza rozważanie jednowymiarowego rozkładu). Entropia takiego rozkładu jest określona przy pomocy tego prawdopodobieństwa w następujący sposób:

$$H_1(N) = -\sum_{x=1}^N p_1(x) \log_2 p_1(x). \quad (17)$$

Rozważmy zatem dwa skrajne przypadki. W pierwszym zakładamy, że istnieje tylko pojedyncza wartość x i prawdopodobieństwo p występowania tej wartości jest równe 1 , $\log_2 p_1(x) = 0$. O pozostałych wartościach zakładamy, że występują z prawdopodobieństwem równym 0 . W tym przypadku następny stan jest określony z prawdopodobieństwem 1 i nazywany jest on przypadkiem *największego uporządkowania*. W drugim przypadku rozważmy wszystkie N wartości tak samo prawdopodobnych. W tym przypadku entropia przyjmuje maksymalną wartość równą $H_1(N) = \log_2 N$ i jest on nazywany przypadkiem *największego nieporządku* - ponieważ każdy stan jest tak samo prawdopodobny. Oczywiście wszystkie pozostałe przypadki znajdują się pomiędzy tymi dwoma ekstremalnymi. Rozważmy teraz przypadek, w którym M stanów szeregu czasowego jest tak samo prawdopodobnych. Ze wzrostem N wartość entropii najpierw rośnie odpowiednio do $\log_2 N$ (ze względu na fakt, że pojawienie się wszystkich wartości jest jednakowo prawdopodobne) aż do asymptotycznego osiągnięcia wartości $\log_2 M$. W przypadku bardziej złożonego rozkładu prawdopodobieństwa wartość entropii będzie rosła do wartości $D_1 \log_2 N$,

gdzie $D_1 \leq 1$. Współczynnik D_1 nazywany jest wymiarem informacji (Hentschel i Procaccia, 1983).

Prawdopodobieństwo konkretnego wektora opóźnień w przestrzeni d -wymiarowej $x_t = (x_t, x_{t-\tau}, \dots, x_{t-(d-1)\tau})$ jest szacowane w podobny sposób

$p_d(x) = \frac{n_x}{n_\tau}$. Całkowita entropia jest określana następującym wyrażeniem

$$H_d(\tau, N) = - \sum_{x_t=1}^N \dots \sum_{x_{t-(d-1)\tau}=1}^N \dots p_d(x_t, x_{t-\tau}, \dots, x_{t-(d-1)\tau}) \log_2 p_d(x_t, x_{t-\tau}, \dots, x_{t-(d-1)\tau}). \quad (18)$$

Wyrażenie określa średnią liczbę bitów potrzebnych do opisu szeregu czasowego.

Innym terminem znanym w teorii informacji, który znajduje zastosowanie w analizie szeregów czasowych, jest *informacja wzajemna*. Informacja wzajemna między dwiema próbkami jest określana jako różnica między ich entropią całkowitą a sumą ich skalarnych entropii:

$$\begin{aligned} I_2(\tau, N) &= - \sum_{x_t=1}^N p_1(x_t) \log_2 p_1(x_t) - \sum_{x_{t-\tau}=1}^N p_1(x_{t-\tau}) \log_2 p_1(x_{t-\tau}) \\ &+ \sum_{x_t=1}^N \sum_{x_{t-\tau}=1}^N p_2(x_t, x_{t-\tau}) \log_2 p_1(x_t, x_{t-\tau}) \quad (19) \\ &= 2H_1(\tau, N) - H_2(\tau, N) \end{aligned}$$

Jeżeli próbki są statystycznie niezależne

$$p_2(x_t, x_{t-\tau}) \equiv p_1(x_t) p_1(x_{t-\tau})$$

to informacja wzajemna staje się równa zero co oznacza, że nie można wyciągnąć żadnej informacji o drugiej próbce znając pierwszą. Jednak, gdy pierwsza próbka określa całkowicie drugą próbkę, tzn. $H_1 = H_2$, wówczas wspólna informacja jest równa entropii skalarniej $I_2 = H_1$. Dla pozostałych przypadków informacja wzajemna wyznacza w bitach stopień określenia jednej zmiennej poprzez znajomość drugiej.

Informacja wzajemna może być uogólniona na przypadek wielowymiarowy przez:

- połączoną informację wzajemną

$$I_d(\tau, N) = dH_1(\tau, N) - H_d(\tau, N) \quad (20)$$

- lub przez nadmiarową informację wzajemną

$$R_d(\tau, N) = H_1(\tau, N) + H_{d-1}(\tau, N) - H_d(\tau, N). \quad (21)$$

Nadmiarowość określa średnią liczbę bitów obserwacji, która może być określona na podstawie $d-1$ poprzednich obserwacji. Zależność pomiędzy połączoną informacją wzajemną i nadmiarowością określa następujący wzór

$$R_d = I_d - I_{d-1}.$$

W przypadku systemów opisywanych układem równań różniczkowych stany przeszłe jednoznacznie określają stany przyszłe systemu. Jeżeli liczba d jest dużo mniejsza od minimalnego wymiaru zanurzenia, to liczba $d-1$ określająca poprzednie obserwacje nie jest w stanie określić następnej obserwacji i w takim przypadku wartość nadmiarowości dąży do zera:

$$\begin{aligned} P_d(x_t, x_{t-\tau}, \dots, x_{t-(d-1)\tau}) &= P_1(x_t) P_{d-1}(x_{t-\tau}, \dots, x_{t-(d-1)\tau}) \\ \Rightarrow \\ H_d &= H_1 + H_{d-1} \\ \Rightarrow \\ R_d &= 0 \end{aligned} \quad (22)$$

W przypadku zaś, gdy d jest dużo większe niż wymagany wymiar zanurzenia, to nowa obserwacja staje się nadmiarowa, wręcz zbędna:

$$\begin{aligned} P_d(x_t, x_{t-\tau}, \dots, x_{t-(d-1)\tau}) &= P_{d-1}(x_{t-\tau}, \dots, x_{t-(d-1)\tau}) \\ \Rightarrow \\ H_d &= H_{d-1} \\ \Rightarrow \\ R_d &= H_1 \end{aligned} \quad (23)$$

Można pokazać, że minimalna wartość nadmiarowości jest równa minimalnemu wymiarowi zanurzenia, tzn. takiej najmniejszej wymiarowości przestrzeni euklidesowej, w której rozwiązania systemu (trajektorie) nie przecinają się.

Wykładniki Lapunova i horyzont prognozowania

Źródło entropii, zwanej też entropią Kołmogorowa-Sinaia, oznaczane jest przez $h(\tau, N)$ i jest zdefiniowane jako asymptotyczny wskaźnik wzrostu informacji:

$$h(\tau, N) = \lim_{N \rightarrow \infty} \lim_{d \rightarrow \infty} H_d(\tau, N) - H_{d-1}(\tau, N). \tag{24}$$

Pojęcie źródła entropii odnosi się do dodatnich wykładników Lapunowa (Ruelle, Eckmann, 1985):

$$h(\tau) = \tau h(1) = \tau \sum_i \lambda_i^+ \tag{25}$$

Wykładniki Lapunowa λ_i są wartościami własnymi lokalnej linearyzacji dynamiki systemu, określają one średni wskaźnik odchylenia osi głównej od pobliskiej trajektorii. Wykładniki Lapunowa można wyznaczyć z jakobianu, lub też poprzez wykorzystanie sztucznych sieci neuronowych, bądź poprzez analizę trajektorii (Brown, Bryant, Abarbanel, 1991).

W przypadku, gdy wymiar zanurzenia d jest dostatecznie duży, to nadmiarowość jest właśnie różnicą pomiędzy entropią skalarną i estymowanym źródłem informacji:

$$R_d(\tau, N) \approx H_1(\tau, N) - h(\tau, N). \tag{26}$$

Dla małych opóźnień

$$\begin{aligned} H_{d-1}(0, N) &= H_d(\tau, N) \\ \Rightarrow \\ R_d(0, N) &= H_1(N) \end{aligned} \tag{27}$$

zaś dla dużych opóźnień

$$\begin{aligned} \lim_{\tau \rightarrow \infty} H_{d-1}(\tau, N) &= dH_1(\tau, N) \\ \Rightarrow \\ R_d(\infty, N) &= 0 \end{aligned} \tag{28}$$

Wartość parametru τ , dla którego nadmiarowość znika szacowuje granicę przewidywalności zachowania się systemu, tzn. horyzont przewidywalności. Horyzont ten jest bardzo krótki, jeśli d jest mniejsze niż minimalny wymiar zanurzenia. Gdy jednak wartość parametru d jest większa niż wymiar zanurzenia, wówczas nadmiarowość będzie się zmniejszać znacznie wolniej, a nachylenie dla małych τ będzie źródłem entropii:

$$R_d(\tau, N) = H_1(N) - \tau h(1). \quad (29)$$

W ten równania (18) i (21) pozwalają na oszacowanie (Bonnländer, Weigend, 1994):

- rozwiązania,
- minimalnego wymiaru zanurzenia,
- źródła entropii,
- horyzontu prognozowania,
- liczby stopni swobody,
- wskaźnika zapominania warunków początkowych przez system.

6. MODELOWANIE

Obserwacja systemu dynamicznego opisanego układem deterministycznych równań różniczkowych pozwala, na podstawie obserwowanych wielkości, tak dobrać wielkość opóźnienia (związanego z zanurzeniem), aby odtworzyć liczbę stopni swobody tego systemu. Wartości odtworzonych zmiennych pozwalają uzyskać dokładne prognozy. Przedstawimy teraz procedurę uzyskiwania dokładnych prognoz w szeregach czasowych generowanych przez deterministyczny system. W tym celu zostaną wykorzystane poprzednio przedstawione narzędzia oparte na teorii informacji.

Przedstawimy pokrótce różne podejścia wyznaczające najlepsze punkty w modelach lokalnych i w modelach globalnych.

6.1. Modele lokalne

Prawdopodobnie najprostszym rozwiązaniem zagadnienia prognozowania jest utworzenie bazy danych z historycznych ciągów $\{(x_{t-1}, \dots, x_{t-d})\}$ o długości d - ciągi te tworzą *wzorce wejściowe* i odpowiadające tym ciągom stany następne $\{x_t\}$. Rozpatrywanie nowego ciągu o długości d polega na przeszukaniu bazy danych i znalezieniu wzorca, który jest "najbliższy"

rozpatrywanemu, po czym przyjmuje się jako prognozę odpowiedni stan następny. Znalezienie *najbliższego wzorca* wymaga przechowywania wszystkich zarejestrowanych wzorców w pamięci. W przypadku sztucznych sieci neuronowych *wiedza* o systemie przechowywana jest w wagach połączeń między neuronami, a poszczególne wzorce mogą być odtworzone dopiero po procesie *uczenia* sieci.

Teraz rozważmy obserwacje, na które nałożony jest szum. Poprawienie jakości prognozy można uzyskać uśredniając po sąsiednich punktach. Występuje tutaj tzw. *dylemat odchylenia wariancji* (Geman, Bienenstock, Doursat, 1992). Wzięcie bowiem zbyt małej liczby sąsiadujących punktów do uśredniania powoduje, że prognoza jest ciągle obciążona szumem (duża wariancja, małe odchylenie). Wzięcie, natomiast, zbyt dużej liczby punktów sąsiadujących do uśredniania pozwala uzyskać bardzo stabilną prognozę (wariancja może dążyć do zera, a odchylenie jest duże).

Tak jak poprzednie, tak i ten model wykorzystuje informacje lokalne. Liniowy model do predykcji szeregów czasowych został zaproponowany przez Farmera i Sidorowicha (1987) posiada dwa zasadnicze parametry, mianowicie parametr określający liczbę sąsiadów k oraz liczbę opóźnień zanurzania d .

W ogólności modele liniowe dają całkiem dobre rezultaty w prognozowaniu szeregów czasowych. Przyczyną tego faktu jest to, że każda podprzestrzeń jest lokalnie liniowa.

6.2. Modele globalne

W poprzednio rozważanych modelach lokalnych wszystkie obserwowane wielkości musiały być przechowywane w pamięci (komputera), gdyż wielkości te służyły do prognozowania, a dodatkowo do predykcji każdej nowej wielkości w każdym następnym kroku był tworzony nowy model. Rozważmy teraz modele globalne, które charakteryzują się tym, że tylko raz jest konstruowana (modelowana) powierzchnia zanurzona w przestrzeni opóźnień, a potem odczytuje się prognozę z tej powierzchni.

W pracach Yule'a (1927) zaproponowano, aby ta powierzchnia była zwykłą hiperpłaszczyzną. W końcu lat siedemdziesiątych zostały zaproponowane pierwsze prace osłabiające takie założenie dla modeli typu AR. Mianowicie Granger i Anderson (1980) wprowadzili iteracje drugiego rzędu między mierzonymi wejściami x_t, x_{t-1} . Taki model został nazwany modelem bilinearnym. W tym samym czasie Tong i Lim (1980) rozdzielili przestrzeń wejść na dwie podprzestrzenie i rozważali dwa modele typu AR. Modele takie zostały nazwane progowymi modelami autoregresyjnymi

(Threshold AutoRegressive models). W ostatnich latach popularne są także inne modele np.: MARS (Multivariate Adaptive Regression Splines) wprowadzone przez Friedmana (1991). Tutaj jednak rozpatrzmy zupełnie inne koncepcyjnie podejście oparte na sztucznych sieciach neuronowych.

Sieci neuronowe są często używane do klasyfikacji, tzn. dany jest zbiór wzorców wprowadzanych jako wejścia do sieci, a zadanie polega na przyporządkowaniu każdego wejścia do odpowiedniej klasy. Inne bardzo ważne zastosowanie sieci neuronowych to regresja nieliniowa, w tym zastosowaniu zadanie polega na znalezieniu gładkiej interpolacji pomiędzy punktami. W rozważanych sieciach neuronowych informacja jest podawana na wejścia sieci. W zastosowaniu sieci neuronowych do szeregów czasowych informacja o *obrazach* i historii tych obrazów - wzorców jest podawana cyklicznie. Początki wykorzystania sieci neuronowych do prognoz w szeregach czasowych sięgają początkowych lat wiedzy o sztucznych sieciach neuronowych. W 1964 roku Hu wykorzystał rezultaty prac Widrowa, a mianowicie jego liniową adaptacyjną sieć do prognozy pogody. Następną era w sztucznych sieciach neuronowych jest okres rozwoju gradientowego algorytmu uczenia sieci, zwanego algorytmem wstecznej propagacji błędu uczenia (backpropagation). Algorytm ten pozwala na uczenie nieliniowych sieci neuronowych. Lapades i Farber (1987) zastosowali taką sieć i tego typu algorytm do uczenia sieci opisującej zależność pomiędzy wyjściem sieci (następny punkt w szeregu czasowym) a wejściem sieci (czyli sekwencją poprzednich punktów szeregu czasowego). Do eksperymentu użyli oni szereg czasowy generowany przez generator komputerowy. Następnie Weigend, Huberman i Rumelhart (1990) określili pewne warunki, które powinna spełniać sieć do prognoz rzeczywistych szeregów czasowych.

Podobnie jak w modelach stochastycznych (ARMA), także sieć neuronowa musi "znać" historię całego szeregu czasowego. Najprostszym sposobem wprowadzenia historii szeregu czasowego do sieci neuronowej jest wprowadzanie ciągów danych (o odpowiednie długości - opóźnienie) na wejście sieci neuronowej. Warto wspomnieć, że sieć neuronowa bez ukrytej warstwy neuronów jest równoważna modelowi AR. Znane są sieci, w których wprowadzono opóźnienie (Lang, Waibel, Hinton, 1990) oraz inne modyfikacje sieci, których celem jest wprowadzenie złożonych wag (połączeń między neuronami), a to wszystko, aby w połączeniach zawrzeć jak najwięcej informacji, np. prace: Cuna (1989), Wana (1994).

Algorytm uczenia wielowarstwowych sieci neuronowych - algorytm wstecznej propagacji jest algorytmem gradientowym i wyznacza jedynie lokalne minima błędu uczenia. W zależności od zainicjowanych wartości wag połączeń sieci otrzymuje się różne rozwiązania. W pracach Krawczaka (1994, 1998, 1999a, 2000a) przedstawiono nowy algorytm uczenia tej klasy

sieci neuronowych. W największym skrócie idea nowego algorytmu polega na potraktowaniu procesu uczenia sieci jako szczególny przypadek zagadnienia sterowania optymalnego. W tych pracach wykorzystano ideę programowania dynamicznego i aproksymacji funkcji zwrotnej (poprzez zastosowanie metody kontynuacji).

7. ZAGADNIENIA MODELOWANIA SZEREGÓW CZASOWYCH ZA POMOCĄ SZTUCZNYCH SIECI NEURONOWYCH

W tej części pokażemy dokładny sposób otrzymywania przedziałów błędu predykcji poprzez wprowadzenie dodatkowego wyjścia sieci. Następnie wprowadzamy reprezentację, która pozwala przewidywać zmianę gęstości prawdopodobieństwa. Porównujemy predykcję bezpośrednią i otrzymaną iteracyjnie. Rozważamy także pewne sposoby uniknięcia przeuczenia sieci (zbyt dokładne dopasowanie) i kończymy przykładem na to jak analiza nauczonej sieci neuronowej może pomóc w scharakteryzowaniu szeregu czasowego.

7.1. Przewidywanie przedziałów błędów

Dotychczas rozważaliśmy jak prognozować następne stany szeregu czasowego znając stany poprzednie. Często pożądane jest i ważne, aby również oceniać pewność prognozowania.

Przed rozwiązaniem tego problemu przedstawimy trzy założenia wynikające z minimalizacji sumy kwadratów błędów w metodzie największej wiarygodności:

1. Wartości błędów w różnych punktach czasu są wzajemnie niezależne. Statystyczna niezależność zakłada, że prawdopodobieństwo zajścia takich dwóch zdarzeń jest równe iloczynowi indywidualnych prawdopodobieństw. Po obustronnym zlogarytmowaniu, otrzymujemy z prawej strony sumę logarytmowanych prawdopodobieństw. Tak więc sumowanie błędów zakłada statystyczną niezależność mierzonych błędów.
2. Błędy mają rozkład Gaussa. Błąd definiujemy jako różnicę między wartością rzeczywistą a wartością wynikającą z prognozy (predykcji). Wartość prognozowana jest deterministyczną funkcją sygnału wejściowego (funkcja ta jest określona przez parametry sieci). Przy zadanym wejściu (a przy określonych wagach oznacza to także, że jest dane wyjście (prognoza)), zakładamy teraz, że dodajemy do tego wyjścia (idealnego) pewien szum (zniekształcenie sygnału) o rozkładzie Gaussa w celu wygenerowania rzeczywistej (obserwowanej) wartości

sygnału. Prawdopodobieństwo rozkładu wartości obserwowanych błędów ma rozkład Gaussa, przy zadanej predykcji nazywane jest czasami błędem modelu lub *całkowitym rozkładem warunkowym* (warunkowy ponieważ zależy bezpośrednio od predykcji i pośrednio od wejścia). Dokonując logarytmowania tego rozkładu otrzymujemy kwadrat różnicy. Kwadrat różnicy można interpretować jako kwadrat błędu.

3. Błędy mają rozkłady identyczne, tzn. zakładamy, że rozmiar przedziału błędów nie zmienia się w przestrzeni stanu. (Błędy sumuje się z tą samą wagą dla każdego punktu danych).

Postaramy się teraz wprowadzić metodę estymacji przedziałów błędów. Przyjmijmy pewną parametryczną postać całkowitego rozkładu warunkowego (założenie zostanie później osłabione). Będziemy estymować lokalne przedziały błędów poprzez szukanie takich parametrów, które maksymalizują wiarygodność modelu dla rozważanych danych. Przechodząc do szczegółów, jeżeli stosujemy wyjście sieci y do aproksymacji funkcji $f(x)$ to zakładamy, że pożądane wyjście d (wartości obserwowanych) można zmodelować jako

$$d(x) = f(x) + n(x)$$

gdzie $n(x)$ jest szumem zgodnym z założonym wcześniej modelem rozkładu błędów. Gdy estymacja średniej tego rozkładu $y(x)$ jest funkcją wejścia, to wariancja σ^2 może także zmieniać się w wejściowej przestrzeni x . Natomiast, gdy poziom szumu zmienia się w przestrzeni wejścia (tzn. $\sigma^2(x)$ zależy od x i nie jest stałe), to nie tylko chcemy, aby sieć nauczyła się i zapamiętała funkcję wyjściową $y(x)$, ale również chcemy nauczyć sieć odwzorowywać funkcję $v(x)$, która estymuje wariancję $\sigma^2(x)$ tego rozkładu.

Estymację $\sigma^2(x)$ możemy uzyskać poprzez dodanie dodatkowego wyjścia v , które oblicza $v(x)$. Ponieważ $\sigma^2(x)$ musi być dodatnie, wybieramy wykładniczą funkcję aktywacji dla $v(x)$, która w naturalny sposób zapewnia te ograniczenie

$$v(x) = \exp \left[\sum_k w_{vk} h_k(x) + \beta \right] \quad (30)$$

gdzie β jest dodatkowym wejściem neuronu (bias), w_{vk} jest wagą między ukrytymi neuronami k i v .

Tak więc sieć ma dwa wyjścia. Dla jednego z nich wzorcem jest d , dla drugiego, możemy wzorzec efektywnie otrzymać przez maksymalizowanie wiarygodności naszej sieci przy założonych danych. Zakładając statystyczną niezależność błędów (jest to pierwsze założenie powyżej) równoważnie możemy minimalizować ujemny logarytm wiarygodności lub koszt

$$C = -\sum_i \log p(d_i | \mathbf{x}_i, N) \quad (31)$$

Końcowa postać kosztu C zawiera tylko estymację $y(x_i)$ średniej założonego błędu modelu jako funkcję wejścia, więc zakłada się, że wariancja jest stała, oraz stałe są człony otrzymane po zróżnicowaniu (Rumelhart, Durbin, Golden, Chauvin, 1994).

Następnie zakładamy że wariancja zależy od wejścia, i z założenia pozostawione są powyższe człony w wyrażeniu na koszt C . Przy dowolnej architekturze sieci i dowolnym modelu błędu można wyprowadzić poprawne, zależne od zmian wag równania, które w procesie uczenia sieci będą powodować zmniejszanie się gradientu.

W celu zilustrowania tego rozważmy przypadek kiedy odchylenia obserwowanych wartości od średniej podlegają rozkładowi Gaussa, a ich wariancja równa jest $v(x)$, tzn. jest spełnione drugie założenie. Założenie normalności rozkładów błędu wokół $y=f(x)$ odpowiada całkowitemu, warunkowemu rozkładowi prawdopodobieństwa o postaci

$$p(d_i | x_i, N) = \frac{1}{\sqrt{2\pi v(x_i)}} \exp\left\{-\frac{[d_i - y(x_i)]^2}{2v(x_i)}\right\}, \quad (32)$$

gdzie wyjście sieci $y(x_i) \approx \mu(x_i)$ estymuje średnią, a $v(x_i) \approx \sigma^2(x_i)$ estymuje wariancję. Logarytmując (z minusem) otrzymamy monotonicznie ujemny logarytm wiarygodności

$$-\log p(d_i | x_i, N) = \frac{\log 2\pi v(x_i)}{2} + \frac{[d_i - y(x_i)]^2}{2v(x_i)}. \quad (33)$$

Pozbywając się stałych członów, sumując poprzez wszystkie wzorce otrzymujemy całkowity koszt:

$$C = \sum_i \frac{1}{2} \left(\frac{-[d_i - y(x_i)]^2}{v(x_i)} + \log[v(x_i)] \right). \quad (34)$$

W tym miejscu należy poczynić założenie dotyczące funkcji przejścia neuronów w sieci w celu sformułowania równań zależnych od zmian wag. W tym celu wybieramy liniową funkcję aktywacji dla neuronów wyjściowych y , następnie tangens hiperboliczny jako funkcję aktywacji dla neuronów w warstwie ukrytej i wykładniczą funkcję aktywacji dla neuronów v . Teraz możemy obliczyć pochodne funkcji kosztu C względem wag sieci. Zmiana wag w odniesieniu do neuronów y i v daje

$$\Delta w_{yj} = \eta \frac{1}{v(x_i)} [d_i - y(x_i)] h_j(x_i) \quad (35)$$

$$\Delta w_{vk} = \eta \frac{1}{2v(x_i)} \left\{ [d_i - y(x_i)]^2 - v(x_i) \right\} h_k(x_i) \quad (36)$$

gdzie η jest współczynnikiem uczenia sieci. Równania zależne od zmian wag dla wag nie połączonych z wyjściem można wyprowadzić w taki sam sposób jak w standardowej metodzie backpropagation. Zauważmy, że równanie (36) odpowiada uczeniu oddzielnej sieci aproksymującej funkcję $v(x)$. Zauważmy także, że jeżeli $v(x_i)$ jest stałe to równania (35)-(36) redukują się do znanej, w standardowej metodzie backpropagation, postaci z funkcją kosztu jako sumą kwadratów błęd.

Wariancja $v(x_i)$, z uwzględnieniem wzorca i , powoduje zróżnicowanie wag dla każdego wzorca. Człon $\frac{1}{v(x)}$ w otrzymanych równaniach może być interpretowany jako postać ważonej regresji, zmniejszając efektywny współczynnik uczenia w obszarach o wysokim szumie.

Jeżeli człon ważonej regresji ma znaczący wpływ na zmniejszanie się gradientu, to częściej osiągane są minima lokalne: sieć dopasowuje się do pierwszej statystycznej własności, którą może napotkać na drodze szukania małych błędów, nie biorąc pod uwagę innych wzorców będących *dużymi błędami*. W celu odrzucenia wcześniej ważonych, różnych wzorców (które mogą mieć za podstawę niedokładne wartości $v(x_i)$ zanim funkcja $f(x)$ zgrubnie będzie aproksymowana przez $y(x)$), proces uczenia dzielimy na trzy fazy:

- **I faza (średnia)**

Losowy podział dostępnych danych na dwie połowy, na zbiory A i B . Uczenie warunkowej wartości oczekiwanej $y(x)$ wykorzystując zbiór A jako zbiór uczący. W tej fazie stosujemy prosty gradient malejący na prostej funkcji kosztu błędu kwadratowego, tj. równania (35)-(36) bez członów $\frac{1}{v(x)}$. W celu zapobiegnięcia przeuczenia sieci, uważa się, że uczenie jest zakończone gdy osiąga się minimum błędu kwadratowego dla testującego zbioru B .

- **II faza (wariancja)**

Dołączając warstwę ukrytych neuronów połączonych zarówno do wejść jak i do neuronów ukrytych sieci z fazy I. Nie zmieniając wag uzyskanych w fazie I należy nauczyć neurony związane z v w celu predykcji kwadratów błędów, i ponownie zastosować gradient prosty tak jak w fazie I. Zbiorem uczącym jest w tej fazie zbiór B , ze zbiorem A jako zbiorem testującym - zastosowanie zbioru A w tej fazie do uczenia przy przeuczeniu sieci w fazie I mogłoby doprowadzić do błędnego oszacowania $v(x)$. Aby uniknąć tego zjawiska zamieniamy właśnie zbiory danych. Początkowa wartość β dla neuronu v jest wartością logarytmu naturalnego średniego błędu kwadratowego (z fazy I) dla zbioru B . Faza II kończy się, gdy kwadrat błędu zbioru A przekracza zadany poziom lub zaczyna wzrastać.

- **Faza III (ważona regresja)**

W tej fazie ponownie dzielimy zbiór danych na dwie nowe połowy A' i B' . Otrzymane wartości wag połączeń między neuronami potraktujemy jako wartości początkowe w nowej iteracji w celu zminimalizowania pełnej funkcji kosztu C stosując równania (35)-(36) dla zbioru A' . Nauczanie uważa się za zakończone, gdy C osiągnie swoje minimum na zbiorze B' . Zauważmy, że człony $\frac{1}{v(x)}$ w równaniach (35)-(36) zawierają postać regresji ważonej, skupiając się na obszarach, gdzie estymowane $v(x)$ jest właściwe.

Podsumowując, wyszliśmy od zasady największej wiarygodności i doszliśmy do przedziałów błędów lokalnych. W każdym przykładzie musimy wybrać specyficzny model błędu. Podejście tutaj przedstawione obejmuje dowolny rozkład z parametrem położenia (średnia warunkowa) i parametrem skali (lokalny przedział błędu). Przedstawione jest również przejście od regresji do klasyfikacji. Umożliwia to ilościowy opis niepewności estymacji prawdopodobieństwa (wzorca należącego do pewnej klasy) poprzez dostarczenie szerokości tego rozkładu, w zależności od wzorca wejścia.

Trzy fazowe podejście w sposób istotny może zredukować prawdopodobieństwo wystąpienia minimów lokalnych w procesie uczenia sieci.

7.2. Predykcje iterowane i bezpośrednie

Rozważymy problemy związane z predykcją bezpośrednią i predykcją iterowaną z fizycznego punktu widzenia, a następnie z perspektywy sieci neuronowych.

Fizycy często przedstawiają szeregi czasowe jako sekwencje pomiarów wyjścia pewnego układu dynamicznego. Na rozróżnienie predykcji bezpośredniej i iterowanej ma wpływ wprowadzenie dwóch skal czasowych: skalę dynamiki systemu (np. wzrosty i spadki) i skalę czasu pomiarów.

Rozważmy skrajny przypadek, gdy pomiary zjawiska otrzymuje się w skali czasu znacznie szybszej niż wewnętrzna dynamika systemu. Iterowany predyktor będzie się starał wychwycić zmiany, które w istocie są szumem: system nie zmienia się znacząco w ciągu jednego kroku czasowego, zmienia się jedynie szum, ponieważ zwykle ma więcej składowych o wysokich częstościach niż sygnał. Tak więc, w tym przypadku predyktor bezpośredni będzie bardziej odpowiedni, ponieważ dłuższy czas predykcji będzie bliższy dynamice systemu. Inny przypadek to, gdy dane z systemu są pobierane zbyt rzadko i skala czasu przy predykcji iterowanej jest bliższa wewnętrznej skali czasu.

Zastosowanie do systemu predyktora iterowanego często wymaga wyższej złożoności funkcyjnej, np. wolne od szumu odwzorowanie kwadratowe, równanie (13). Dla przypadku predykcji jednokrokowych (następnie poddawanych iteracji) wystarcza model, którego zachowanie *wejście-wyjście* może być reprezentowane poprzez parabolę (jest to łatwe zadanie dla sieci neuronowej z kilkoma ukrytymi neuronami). Natomiast dla przypadku predykcji bezpośrednich o kroku h zależność staje się skomplikowanym wielomianem stopnia 2^h . Przykładowo predykcja krokowa dla $h = 10$ wymaga sieci neuronowej z liczbą ukrytych neuronów rzędu 1000.

Możemy rozróżnić trzy przypadki sieci neuronowych do predykcji (przypadki te różnią się sposobem uczenia):

- w pierwszym przypadku parametry sieci są optymalizowane ze względu na minimalizację błędu w prognozie jednokrokowej, zaś iterowanie dodawane jest później, po nauczaniu sieci,
- w drugim przypadku, celem uczenia sieci jest minimalizacja błędu iterowanej prognozy na h kroków naprzód; można to sobie wyobrazić

jako umieszczenie h kopii identycznych sieci „jedna na drugiej”, tzn. druga kopia przyjmuje wyjście z pierwszej jako jedno ze swoich wejść itd.; następnie wykonuje się dopasowanie parametrów tak, aby błąd po h iteracjach, czyli na szczycie rozłożonej sieci, był jak najmniejszy (Rumelhart, Hinton, Williams, 1986),

- trzecia sieć ma odmienne zadanie; przewiduje bezpośrednio z punktu bieżącego do punktu w przyszłości odległego o h kroków - bez żadnych pośrednich prognoz i sprzężeń.

Rozważania przedstawione powyżej odnoszą się do przypadku, w którym szum ma inną skalę czasową niż dynamika układu. Dla deterministycznych systemów chaotycznych (bez szumu) Farmer i Sidorowich (1988) dowodzą, że prognozy iterowane dają lepsze przewidywania niż bezpośrednie.

7.3. Przeuczenie sieci

Przeuczenie sieci obserwuje się podczas uczenia, gdy nauczona sieć charakteryzuje więcej cech, które źle się uogólniają, niż cech, które dobrze się uogólniają (sieć nauczyła się przypadków mocno zasumionych - nie niosących żadnej informacji). Intensywność zjawiska zależy od poziomu szumu w danych i od rozmiaru zbioru uczenia. Bardzo trudnym przypadkiem bywają finansowe szeregi czasowe.

Rozważa się trzy sposoby obróbki tego rodzaju szeregów czasowych przy prognozowaniu:

- **wczesne zatrzymanie**
najprostszy algorytm polega na użyciu sieci nadmiernie dużej z liczbą wag prawie tak dużą jak liczba punktów i zatrzymanie, gdy błąd uczenia przestaje się polepszać. Inicjalizacja sieci odbywa się z bardzo małymi wagami - wystarczająco dużymi dla uniknięcia symetrii, ale na tyle małymi by utrzymać się w zakresie liniowym dla ukrytych neuronów.
- **dodanie szumu do wejścia**
alternatywą jest dodanie szumu do każdego wejścia w każdej iteracji algorytmu backpropagation (Bishop, 1994).
- **uwzględnienie złożoności sieci**
do składnika błędu w funkcji błędu uczenia dodaje się drugi składnik kary za złożoność sieci. W podejściu bayesowskim taki składnik ma interpretację a priori, tzn. odpowiada naszej wierze w rozkład, z którego uzyskano wagi, np.:

$$\sum_i \frac{(w_i / w_0)^2}{1 + (w_i / w_0)^2} \quad (37)$$

Znana jest też technika tzw. *pseodo-danych* polegająca na „podpowiadaniu” sieci spodziewanego wyjścia. Posłużył się on danymi finansowymi z jednego kraju do „podpowiedzi” dla predykcji dla innego kraju. W trakcie uczenia funkcje błędu uczenia były przełączane ze zwykłej funkcji błędu na funkcję błędu z „podpowiedzią”.

Inną metodą (Weigend, 1992) jest dostarczanie dodatkowych informacji do sieci, np. poprzez wprowadzanie na wejście sieci informacji o znaku i/lub wartości bezwzględnych zmian stóp procentowych.

7.4. Analiza sieci

W części 2.1 przedstawiliśmy, że liniowy system zależny od czasu jest w pełni scharakteryzowany przez swoją transformatę Fouriera (lub równoważnie przez swoje współczynniki ARMA lub funkcję autokorelacyjną). Następnie pokazaliśmy, jak to można uogólnić dla przypadku systemów nieliniowych. W części 3.2 opisaliśmy niezależne od modelu własności punktów obserwowanych bezpośrednio w przestrzeni zanurzonej, a w części 4.1 pośrednią analizę szeregu czasowego poprzez obserwację zachowania błędów spowodowanych zmianą parametrów lokalnych modeli liniowych. Obecnie podamy możliwości nauczonej sieci neuronowej do analizy szeregów czasowych. Zajmiemy się „wyciąganiem” charakterystycznych własności takich jak: minimalny wymiar zanurzenia, estymację wykładników Lapunova, ocenę „wielkości” nieliniowości procesu (systemu) - związanej z stopniem entropii systemu liniowego.

Wymiary

Początkowo, gdy był używany algorytm backpropagation, zmieniano liczbę neuronów w warstwie ukrytej i wykreślano błąd nauczania (lub testowania) jako funkcji liczby elementów ukrytych. Zwykle błąd ten początkowo spada, następnie osiąga minimum; a potem rośnie. Dla osiągniętej minimalnej wartości błędu nauczania sieci (lub błędu testowania nauczonej sieci) liczba ukrytych neuronów może być traktowana jako rodzaj miary stopni swobody systemu. W podobny sposób, zmieniając liczbę wejściowych elementów, można określić wymiar zanurzenia.

Określone wymiary są silnie zależne od funkcji aktywacji algorytmu optymalizującego. Widmo wartości własnych macierzy kowariancji aktywacji ukrytych neuronów jest obliczane jako funkcja czasu uczenia. Kowariancja

$$C_{ij} = \langle (f_i - \bar{f}_i)(f_j - \bar{f}_j) \rangle \quad (38)$$

opisuje dwupunktowe oddziaływanie pomiędzy aktywacjami dwóch ukrytych neuronów i oraz j ($\bar{f}_i = \langle f_i \rangle$ jest średnią aktywacją ukrytego neuronu i). Liczba znacząco dużych wartości własnych macierzy kowariancji może służyć jako miara efektywnego wymiaru przestrzeni ukrytych neuronów.

Wykładniki Lapunova

Wciąż trudno jest estymować wykładniki Lapunova (zdefiniowane w części 5.2) z krótkich zapisów zaszumionych systemów. Nadzieja jest w tym, że jeśli sieć nauczy się wiarygodnie emulować takie systemy, to wykładniki Lapunova mogą być znalezione przez sieć. Weigend i in. (1990) podają przykład w jaki sposób szybkość dywergencji procesu chaotycznego może być estymowana poprzez analizę błędów „out-of-sample” jako funkcję czasu predykcji. Alternatywny sposób to obliczenie z sieci Jakobianu i uśrednienie go po wszystkich punktach pomiarowych w celu otrzymania *globalnych* wykładników Lapunova (Gencay i Dechert, 1992; Nychka i in., 1992).

Często współczynniki Lapunova silnie zależą od lokalizacji w przestrzeni stanów. W takich przypadkach, uśrednianie po atraktorze zamazuje informacje potencjalnie interesujące.

Nieliniowość

Wykresy DVS analizują błąd w funkcji nieliniowości modelu (mniejsze sąsiedztwo \Rightarrow więcej nieliniowości). Idea porównywania błędu sieci „out-of-sample” z błędem „out-of-sample” modelu liniowego jest dobra również dla sieci neuronowych.

Ponadto, możemy wykorzystać własność sieci do scharakteryzowania ilości nieliniowości poprzez badanie rozkładu aktywacji f sigmoidalnych ukrytych neuronów. Stosunek kwadratowego członu rozwinięcia Taylora sigmoidy do jego członu liniowego, czyli stosunek drugiej pochodnej do pierwszej, może być wyrażony przez wielkości sieciowe:

$$\frac{|f''(\xi)|}{|f'(\xi)|} = a|1 - 2f| \quad (39)$$

gdzie f' i f'' oznaczają tu pierwszą i drugą pochodną sigmoidalnej funkcji aktywacji

$$f(\xi_h^{(i)}) = \frac{1}{1 + \exp(-a\xi_h^{(i)})} = \frac{1}{2} \left(1 + \tanh \frac{a}{2} \xi_h^{(i)} \right) \quad (40)$$

α oznacza nachylenie sigmoidy, a

$$\xi_h^{(r)} = \sum_{i=1}^d w_{hi} x_i^{(r)} + b_h$$

oznacza wejście "net" sieci. Weigend i in. (1990) podają rozkład tej statystyki dla kilku szeregów czasowych, przedstawiając różne stopnie nieliniowości dla różnych szeregów czasowych.

Po ustaleniu istotności nieliniowości dla danego problemu, pojawiają się następujące pytania: Gdzie występują nieliniowości? Do czego one są używane? Są takie szeregi czasowe, w których cała nieliniowość sieci mogła być usunięta przez „preprocessing” każdego wejścia z osobna, innymi słowy sieć mogła być zredukowana do liniowej superpozycji indywidualnie przetransformowanych wejść:

$$\text{wyjście } y = \sum_i w_i f_k(x_i) \quad (41)$$

Takie nieliniowości nazywane są nieliniowościami pierwszego rodzaju, i mogą być usunięte przez wstępną obróbkę nazywaną *nieliniowościami wstępnej obróbki*. Istnieją jednak również przypadki, gdy oddziaływania między wejściami są naprawdę istotne. Takie nieliniowości nazywane są nieliniowościami drugiego rodzaju lub *nieliniowościami interakcji*.

Rozróżnienie nieliniowości drugiego rodzaju od nierówności pierwszego rodzaju może być dokonana w następujący sposób: obliczmy hesjan - macierz drugich pochodnych wyjścia sieci względem wejść i oraz j

$$\frac{\partial^2 y}{\partial x_i \partial x_j} \quad (42)$$

i następnie obliczmy to dla danych punktów pomiarowych (gęstość empiryczna). Jeżeli, z jednej strony, wielkość członów poza przekątną (tzn., suma ich kwadratów) jest pomijalna w porównaniu z członami z głównej przekątnej, to sieć stosuje tylko wstępną obróbkę nieliniowości. Jeżeli wielkość członów pozadiagonalnych jest duża, to sieć rzeczywiście wykorzystuje interakcje pomiędzy zmiennymi wejściowymi.

W rozdziale 5.2 ustaliliśmy, że stopień źródłowej entropii szeregu jest dolnym ograniczeniem dla każdego modelu. Dla modelu liniowego stopień entropii jest dany przez (Cover i Thomas, 1991):

$$\frac{\log 2\pi e}{2} + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log S(\lambda) d\lambda \quad (43)$$

gdzie $S(\lambda)$ jest gęstością widma mocy; tj. transformatą Fouriera funkcji autokorelacji (w części 4.1 przedstawiona jest dyskusja dlaczego model liniowy jest całkowicie opisany przez swoje widmo, lub równoważnie przez funkcję autokorelacji).

8. PROGNOZOWANIE

W tej części omawiamy istotność właściwego otrzymywania predyktora. Najpierw wymieniamy pewne standartowe miary wykonania, następnie dyskutujemy niepewność wprowadzoną przez podział dostępnego zbioru danych na podzbiory: treningowy, przekrojowego sprawdzania i testowy.

8.1. Prognozowanie bez przedziału błędów

Podstawowa miara dokładności predykcji używa wyłącznie wartości \hat{x}_t otrzymywanych z predykcji (w odróżnieniu od obserwowanych wartości x_t). Definiujemy *znormalizowany błąd średniokwadratowy*:

$$E_{NMS} = \frac{\sum_{t \in T} (\text{obserwacja}_t - \text{predykcja}_t)^2}{\sum_{t \in T} (\text{obserwacja}_t - \bar{\text{średnia}}_T)^2} \approx \frac{1}{\hat{\sigma}_T^2} \frac{1}{N} \sum_{t \in T} (x_t - \hat{x}_t)^2 \quad (44)$$

gdzie $t = 1, \dots, N$ oznacza punkty w zbiorze testowym T , a $\bar{\text{średnia}}_T$ i $\hat{\sigma}_T^2$ oznaczają odpowiednio średnią z próby i wariancję obserwowanych wartości w T . Z definicji widać, że prosta prognoza średniej całkowitej (występująca w mianowniku) daje wartość $E_{NMS} = 1$.

Dogodnie jest odnieść E_{NMS} do współczynnika korelacji R pomiędzy predykcją a wartością celu. Niech

$$x_t = \rho \hat{x}_t + \varepsilon_t,$$

ε_t jest błędem prognozy.

Dla wariancji otrzymamy

$$\sigma_x^2 = \rho\sigma_{\hat{x}}^2 + \sigma_\varepsilon^2.$$

Zakładając, że wariancje x_t i \hat{x}_t są równe otrzymujemy

$$1 - \rho^2 = \frac{\sigma_\varepsilon^2}{\sigma_x^2} \quad (45)$$

Jeśli średnie x_t i \hat{x}_t są również zbliżone, to ρ może być przybliżone przez współczynnik korelacji pomiędzy prognozą i wzorcem.

Wykreślając $(1-R^2)$ i E_{NMS} w tej samej skali (np. jako funkcję czasu uczenia niezbędnego do monitorowania przeuczenia sieci), E_{NMS} jest ograniczony od dołu przez $(1-R^2)$.

Dla danych zbliżonych do przypadkowego błędzenia, dobrą miarą określającą jakość predykcji jest stosunek błędów kwadratowych:

$$\frac{\sum_{t \in T} (cel_t - predykcja_t)^2}{\sum_{t \in T} (cel_t - predykcja_{t-1})^2} \quad (46)$$

W mianowniku użyte są ostatnie zaobserwowane wartości jako predykcje. Wartość tego stosunku powyżej 1.0 odpowiada zatem predykcji gorszej od błędzenia przypadkowego, zaś poniżej 1.0 - ulepszeniu w stosunku do błędzenia przypadkowego. Jeżeli występują istotne autokorelacje, zamiast silnej hipotezy zerowej można spróbować wykorzystać model AR niskiego rzędu.

Często można otrzymać więcej informacji o błędach niż o ich wartościach średnich. Rozkład błędów posortowanych według wielkości ułatwia rozróżnienie prognoz o tym samym średnim błędzie, ale bardzo różnym rozkładzie. W innych podstawowych testach wykreśla się błędy predykcji względem prawdziwych (bądź przewidywanych) wartości. Rozkład taki powinien być płaski gdy błędy mają rozkład Gaussa, a proporcjonalny do średniej gdy błędy mają rozkład Poissona. Interesujące może być wykreślenie współczynników autokorelacji reszt względem liczby training epochs i sprawdzenie jak to się ma do błędu związanego z przeuczeniem sieci.

8.2. Predykcja z przedziałami błędów

Jeżeli dostępne są przedziały błędów predykcji \hat{x}_k to możemy zastosować jako miarę predykcji wiarygodność obserwowanych danych,

dyponując przewidywanymi wartościami i przedziałami błędów predykcji. W każdym przypadku trzeba założyć model błędów. Często stosowane założenia o niezależnych błędach gaussowskich mogą być niewłaściwe, ale dają prosty wynik. Prawdopodobieństwo zaobserwowania danego punktu x_t , dane jest przez scałkowanie rozkładu Gaussa po jednostkowym przedziale (odpowiadającym błędowi zaokrąglenia jednego bitu w konwersji analogowo-cyfrowej) scentrowanym w x_t :

$$p(x_t | \hat{x}_t, \hat{\sigma}_t) = \frac{1}{\sqrt{2\pi\hat{\sigma}_t^2}} \int_{x_t-0.5}^{x_t+0.5} \exp\left(-\frac{(\xi - \hat{x}_t)^2}{2\hat{\sigma}_t^2}\right) d\xi. \quad (47)$$

Jeżeli błąd predykcji jest duży, to obliczone prawdopodobieństwo będzie relatywnie małe, prawie niezależne od wartości punktu z predykcji. Jeżeli błąd predykcji jest mały i punkt z predykcji jest blisko obserwowanej wartości, to prawdopodobieństwo będzie duże, ale jeżeli błąd predykcji jest mały i wartość przewidywana odległa od rzeczywistej, to prawdopodobieństwo będzie bardzo małe. Przy założeniu błędów niezależnych, wiarygodność całego testu dla pomiarów przy danych wartościach prognozowanych i przedziałach błędu predykcji, wynosi:

$$p(D|M) = \prod_{i=1}^N p(x_i | \hat{x}_i, \hat{\sigma}_i). \quad (48)$$

W końcu, po zlogarytmowaniu (51) możemy zdefiniować *ujemny, średni logarytm wiarygodności*

$$-\frac{1}{N} \sum_{i=1}^N \log p(x_i | \hat{x}_i, \hat{\sigma}_i) \quad (49)$$

Standartowe podejście do wyboru modelu polega na podziale danych na podzbiory: treningowy, weryfikujący, testowy. Ponieważ nie ma jednego sposobu podziału danych jedynym sposobem jest metoda prób i błędów do doboru większości parametrów sieci (liczba ukrytych warstw, liczba neuronów w każdej warstwie, początkowe wagi, itd., Weigend i LeBaron, 1994). W ten sposób można wykonać histogramy błędów uczenia. Z porównania histogramów sieci nauczanych dla różnych podziałów wynika, że losowość w podziale danych generuje znacznie większą zmienność niż losowość w inicjalizacji sieci i jej architekturze. Zauważmy, że sieć jest w stanie wyjaśnić około 50% zmienności danych. Jednak wyniki są podobne do wyników uzyskanych przy zastosowaniu modeli liniowych (przy

wykorzystaniu tych samych danych (uczenie + weryfikacja). Jedną z możliwych przyczyn takiej sytuacji jest brak nieliniowości, a inną to niestacjonarność nieliniowości, które zmieniają się w czasie, i mogą być uśrednione przez metodę *bootstrap* (Weigend i LeBaron, 1994).

9. PODSUMOWANIE

Przedstawiliśmy przegląd wyników, które zdają się tworzyć stały postęp w przewyżnianiu niewiedzy w analizie i predykcji szeregów czasowych. Najlepszym modelem do predykcji jest model, który wymaga jak najmniej informacji do jego opisu. Właśnie celem *Teorii Informacji Algorytmicznej* Kolmogorowa (1965). Klasyczna teoria informacji, opisana w rozdziale 5.2, mierzy informację w odniesieniu do rozkładu prawdopodobieństwa zespołu obserwacji strumienia symboli. W algorytmicznej teorii informacji, informacja jest mierzona w pojedynczym strumieniu symboli poprzez liczbę bitów potrzebnych dla specyfikacji najkrótszego algorytmu, który może je wygenerować. Doprowadziła ona do znacznego ulepszenia wyników uzyskanych przez Gödela (1931) i Turinga (1936) (za Chaitin, 1990). Jednakże nie wiadomo czy istnieje uniwersalny algorytm znajdujący najkrótszy program generujący obserwowaną sekwencję.

Chociaż występują głębokie teoretyczne ograniczenia w predykcji szeregów czasowych, to jednak uzyskiwane są wyniki na tyle przydatne, że są szeroko wykorzystywane w instytucjach kapitałowych. W tym rozdziale pominieliśmy wiele ważnych problemów szeregów czasowych, między innymi:

- **Budowę modeli parametrycznych**

Dla systemu liniowego możliwa jest identyfikacja wewnętrznej funkcji przejścia niezależnej od wejść zewnętrznych, a dla systemów nieliniowych takie rozdzielanie jest niemożliwe. Sparаметryzowana sieć pozwala scharakteryzować system w kategorii jego jakościowego zachowania (Kevrekidis i in., 1993).

- **Sterowanie systemów nieliniowych**

Problem sterowania systemami nieliniowymi polega na tym, aby obserwacje i dostęp do niektórych wejść systemu nieliniowego mogłyby być użyte do zbudowania modelu, który z kolei mógłby być zastosowany poprowadzenia systemu w kierunku pożądanego stanu (Miller, Sutton, Werbos, 1990; White i Sofge, 1992).

Warto zwrócić uwagę na kilka zagadnień związanych z wykorzystaniem sieci neuronowych do predykcji szeregów czasowych:

- **Czyszczenie danych**

W rozdziale w zasadzie zakładaliśmy, że na wejściach nie ma szumu. W przypadku zaszumionych szeregów można zastosować filtr: sieć uczy się przewidywać środkową wartość okna przesuwanego się wzdłuż szeregu czasowego od kilku poprzedzających i od kilku następujących wartości.

- **Szeregi czasowe wielu zmiennych i niestacjonarne**

Jak dotąd rozpatrywane były jedynie szeregi jednej zmiennej (jedna zmienna w funkcji czasu). Twierdzenie Takensa (5.10) stwierdza, że przeszłe wartości pojedynczej obserwacji są wystarczające do rekonstrukcji lokalnych wartości rozwiązania - o ile rozważamy system dynamiczny wolny od szumu. W rzeczywistych systemach taka sytuacja ma miejsce raczej rzadko. Uwzględnienie większej ilości informacji może spowodować, że problem okaże się bardziej stacjonarny, a w wielu przypadkach istnieją zbiory zmiennych, które występują grupowo. W literaturze ekonometrycznej własność ta nazywana jest kointegracją (Engle i Granger, 1987).

- **Sieci rekurencyjne**

Praca ta nie objęła sieci rekurencyjnych. Jedną z istotnych przewag, jaką mają one nad sieciami *feed-forward*, jest zdolność odnalezienia stanu w obecności szumu, podobnie jak w ukrytych modelach Markowa (Mozer, 1994).

W pracach Krawczaka (1999a, 1999b, 2000a, 2000b) przedstawiających nowy algorytm uczenia sieci neuronowych dopuszcza się istnienie sprzężeń zwrotnych w sieci typu *feed-forward* i dodatkowo algorytm ten wyznacza minimum globalne uczenia sieci. Wydaje się, że tego typu sieć może znaleźć zastosowanie przy predykcji finansowych szeregów czasowych.

Literatura

1. Akaike, H. (1974) A New Look at the Statistical Model Identification. *IEEE Trans. Auto. Control*, 19: 716-723.
2. Bierwag, G. (1987) *Duration Analysis - Managing Interest Rate Risk*. Ballinger Publishing Company, Cambridge, Mass.
3. Bishop, C. M. (1994) Training with Noise is equivalent to Tikhonov Regularization. *Neural Computation*.
4. Bonnländer, B. V., and Weigend A. S. (1994) Selecting Input Variables Using Kernel Density Estimation. In *Proceedings of the International*

- Symposium on Artificial Neural Networks (ISANN'94)*, Tainan, Taiwan, R. O. C.
5. Box, G. E. P., and F. M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*, 2nd ed. Oakland, CA: Holden-Day.
 6. Brennan M., Schwartz E. (1983): Duration, Bond Pricing and Portfolio Management. W: *Innovations in Bond Portfolio Management*, Bierwag G., Kaufman G., Toevs (Eds.), Greenwich, Conn., JAI Press.
 7. Brown, R., P. Bryant, and H. D. I. Abarbanel. 1991. "Computing the Lyapunov Spectrum of a Dynamical System from an Observed Time Series." *Phys. Rev. A* **43**: 2787-2806.
 8. Chaitin, G. J. 1990. *Information, Randomness & Incompleteness*. Series in Computer Science, Vol. 8, 2nd ed. Singapore: World-Scientific.
 9. Cover, T. M., and J. A. Thomas. 1991. *Elements of Information Theory*. New York: John Wiley.
 10. Cox J., Ingersoll J., Ross S. (1981): A Re-examination of Traditional Hypotheses about the Term Structure of Interest Rates. *Journal of Finance*, September 1981, pp. 769-799.
 11. Dahl H. (1993): A Flexible Approach to Interest-Rate Risk Management. In: Zenios S.A. (Ed.), *Financial Optimization*. Cambridge University Press, Cambridge.
 12. Engle, R. F., and C. W. J. Granger. 1987. "Cointegration and Error-Correcting Representation, Estimation and Testing." *Econometrica* **55**: 251-276.
 13. Farmer, J. D., and J. J. Sidorowich. 1988. "Exploiting Chaos to Predict the Future and Reduce Noise." *Evolution, Learning, and Cognition*, edited by Y. C. Lee. Singapore: World Scientific.
 14. Fisher L., Weil R.L. (1977): Coping with the Risk of Market-Rate Fluctuations - Returns to Bondholders from Naive and Optimal Strategies. *Journal of Business*, October 1977, pp. 408-431.
 15. Fraser, A. M. 1989. "Information and Entropy in Strange Attractors." *IEEE Trans. Info. Theory* **IT-35**: 245-262.
 16. Friedman, J. H. 1991. "Multivariate Adaptive Regression Splines." *Ann. Stat.* **19**: 1-142.
 17. Garbade K. (1986): „*Modes of Fluctuations in Bond Yields - an Analysis of Principal Components*”. Technical Report, Bankers Trust Company, Money Market Center, New York, June 1986.
 18. Garbade K. (1989): „*Polynomial Representations of the Yield Curve and its Modes of Fluctuations*”. Bankers Trust Company, Money Market Center, New York, No. 53, July 1989.
 19. Geman, S., E. Bienenstock, and R. Doursat. 1992. Neural Networks and the Bias/Variance Dilemma." *Neural Computation* **5**: 1-58.

20. Gershenfeld, N. A. 1989. "An Experimentalist's Introduction to the Observation of Dynamical Systems." In *Directions in Chaos*, edited by B.-L. Hao, Vol. 2, 310-384.
21. Gershenfeld, N. A., and A. S. Weigend. 1994. "The Future of Time Series: Learning and Understanding." In *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld, 1-70. Reading, MA: Addison-Wesley.
22. Gödel, K. 1931. "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme, I." *Monatshefte für Mathematik und Physik* **38**:173- 198. An English translation of this paper is found in *On Formally Undecidable Propositions* by K. Gödel (New York: Basic Books, 1962).
23. Granger, C. W. J., and T. Teräsvirta. 1993. *Modelling Nonlinear Economic Relationships*. Oxford, UK: Oxford University Press.
24. Hentschel, H. G. E., and I. Procaccia. 1983. "The Infinite Number of Generalized Dimensions of Fractals and Strange Attractors." *Physica D* **8**: 435-444.
25. Hirschleifer J. (1965): Investment decisions under uncertainty - choice theoretic approach. *The Quarterly Journal of Economics*, Vol. 74, pp. 509-536.
26. Hu, M. J. C. 1964. "Application of the Adaline System to Weather Forecasting." E. E. Degree Thesis. Technical Report 6775-1, Stanford Electronic Laboratories, Stanford University.
27. Jajuga K., Jajuga T. (1996): *Inwestycje - instrumenty finansowe, ryzyko finansowe, inżynieria finansowa*. PWN, Warszawa.
28. Jakubowski, A. 1996 *Modelowanie struktury czasowej stóp procentowych*. Projekt badawczy KBN Nr PB 536/HO2/96/10 - G 37, IBS PAN, Warszawa.
29. Kellison S.G. (1991): *The Theory of Interest*. Irwin, Homewood.
30. Kevrekidis, I. G., R. Rico-Martinez, R. E. Ecke, R. M. Farber, and A. S. Lapedes. 1993. "Global Bifurcations in Rayleigh-Benard Convection." Los Alamos preprint, LA-UR-93-2922, May 1993.
31. Kolmogorov, A. 1941. "Interpolation und Extrapolation von stationären zufälligen Folgen." *Bull. Acad. Sci. (Nauk)* **5**: 3-14. U.S.S.R., Ser. Math.
32. Kolmogorov, A. N. 1965. "Three Approaches to the Quantitative Definition of Information." *Prob. Infor. Trans.* **1**: 4-7.
33. Krawczak M.: Backpropagation versus dynamic programming approach for neural networks learning. In: *Emerging Knowledge Engineering and Connectionist-Based Information Systems*, Dunedin, New Zealand, 22-23 November, 1999, Univ. of Otago, ss. 1057-1062.

34. Krawczak M.: Dynamic learning for feedforward neural networks. In: Proceedings of the conference: ICONIP'99 ANZIIS'99&ANNES'99 &ACNN'99, Perth, Western Australia, 16-20 November, 1999, ss. 33-36.
35. Krawczak, M. 2000a. Backpropagation versus dynamic programming approach. *BULLETIN OF POLISH ACADEMY OF SCIENCES, SERIA TECHNICAL SCIENCES*. Vol. 48, No. 2
36. Krawczak, M. 2000b. Feedforward Neural Networks Learning by Continuation Method. IIZUKA 2000, Fukuoka, Japonia.
37. Krawczak, M. 2000c. Neural Networks Learning by Homotopy Method. ICONIP 2000, Taejin, Korea.
38. Krawczak, M., K. Mizukami, 1994. The control theory approach to perceptron learning process. *44 Conference of IEE of Japan*, Okayama.
39. Lang, K. J., A. H. Waibel, and G. E. Hinton. 1990. "A Time-Delay Neural Network Architecture for Isolated Word Recognition." *Neural Networks* 3: 23-43.
40. Lapedes, A., and R. Farber.1987. "Nonlinear Signal Processing Using Neural Networks." Technical Report No. LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM.
41. le Cun, Y.1989. "Generalization and Network Design Strategies." In *Connectionism in Perspective*, edited by R. Pfeifer, Z. Schreter, F. Fogelman, and L. Steels. Amsterdam: North Holland.
42. Litterman R., Scheinkman J. (1991): Common Factors Affecting Bond Returns. *Journal of Fixed Income*, June 1991, pp. 54-61.
43. Lorenz, E. N. 1963. "Deterministic Non-Periodic Flow." *J. Atmos. Sci.* 20: 130- 141.
44. Macaulay F.R, (1938): *Some Theoretical Problems Suggested by the Movement of Interest Rates, Bond Yields and Stock Prices in the United States since 1856*. Columbia University Press, National Bureau of Economic Research, New York.
45. M. Mackey and L. Glass (1977):Oscillations and chaos in physiological control systems. *Science*, 197,287.
46. Miller, W. T., R. S. Sutton, and P. J. Werbos, eds. 1990. *Neural Networks for Control*. Cambridge, MA: MIT Press.
47. Morrison, F. 1996. *Sztuka modelowania układów dynamicznych*. PWN. Warszawa.
48. Mozer, M. C. 1994. "Neural Net Architectures for Temporal Sequence Processing." In *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld, 243-264. Reading, MA: Addison-Wesley.
49. Nychka, D., S. Ellner, D. McCaffrey, and A. R. Gallant. 1992. "Finding Chaos in Noisy Systems." *J. Roy. Stat. Soc. B* 54(2): 399-426.

50. Oppenheim, A. V., and R. W. Schaffer. 1989. *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice Hall.
51. Peitgen, H.-O., Jurgens, H., i Saupe, D. 1996. Granice chaosu - fraktale, cz. 1, cz. 2. PWN. Warszawa.
52. Peters E.E. (1997): *Teoria Chaosu a Rynki Kapitałowe*. WIG Press, Warszawa.
53. Peters E.E. (1994): *Fractal Market Analysis*. Wiley, New York.
54. Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1992. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge: Cambridge University Press.
55. Redington F.M. (1952): Review of the Principle of Life Office Valuations. *Journal of The Institute of Actuaries*, 1952, vol. 78, pp. 286-340.
56. Ruelle, D., and J. P. Eckmann. 1985. "Ergodic Theory of Chaos and Strange Attractors." *Rev. Mod. Phys.* **57**: 617-656.
57. Rumelhart, D. E., R. Durbin, R. Golden, and Y. Chauvin. 1994. "Backpropagation: The Basic Theory." In *Backpropagation: Theory, Architectures, and Applications*, edited by Y. Chauvin and D. E. Rumelhart. Hillsdale, NJ: Lawrence Erlbaum Associates. Also in: In *Mathematical Perspective on Neural Networks*, edited by P. Smolensky, M. C. Mozer, and D. E. Rumelhart. Hillsdale, NJ: Lawrence Erlbaum Associates.
58. Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. „Learning Internal Representations by Error Propagation." In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume I Foundations*, edited by D. E. Rumelhart and J. L. McClelland, 318-362. Cambridge, MA: MIT Press/Bradford Books.
59. Sauer, T., J. A. Yorke, and M. Casdagli. 1991. "Embedology." *J. Stat. Phys.* **65**(3/4): 579-616.
60. Shaw, R. S. 1981. "Strange Attractors, Chaotic Behavior and Information Flow." *Z. Naturforsch.* **36A**: 80-112.
61. Sierpińska M., Jachna T. (1994): *Ocena przedsiębiorstwa według standardów światowych*. PWN, Warszawa, Wyd. 2.
62. Schuster, H.G. 1995. *Chaos deterministyczny*. PWN. Warszawa.
63. Takens, F. 1981. "Detecting Strange Attractors in Turbulence." In *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young. Lecture Notes in Mathematics, Vol. 898, 336-381. Warwick, 1980. Berlin: Springer-Verlag.
64. Tong, H., and K. S. Lim. 1980. "Threshold Autoregression, Limit Cycles and Cyclical Data." *J. Roy. Stat. Soc. B* **42**: 245-292.

65. Traczyk, W. (red) 1995. Problemy sztucznej inteligencji. Wiedza i Życie. Warszawa.
66. Ulam, S. M, and J. v. Neumann. 1947. *Bulletin Amer. Math. Soc.* **53**: 1120.
67. Wan, E. A. 1994. "Times Series Prediction Using a Connectionist Network with Internal Delay Lines." In *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A. S. Weigend and N. A. Gershenfeld, 195- 217. Reading, MA: Addison-Wesley.
68. Weigend, A. S., B. A. Huberman, and D. E. Rumelhart. 1990. Predicting the Future: A Connectionist Approach." *International Journal of Neural Systems* 1:193-209.
69. Weigend, A., B. A. Huberman, and D. E. Rumelhart. 1992. "Predicting Sunspots and Exchange Rates with Connectionist Networks." In *Nonlinear Modeling and Forecasting*, edited by M. Casdagli and S. Eubank. 395-432. Redwood City, CA: Addison-Wesley.
70. Weigend, A. S., and B. LeBaron. 1994. "Evaluating Neural Network Predictors by Bootstrapping." In *Proceedings of International Conference on Neural Information Processing (ICONIP'94)*, Seul, Korea. Also Technical Report CU-CS-725-94, Computer Science Department, University of Colorado at Boulder.
71. Weron, A., i Eron, R. 1998. Inżynieria finansowa. WNT. Warszawa.
72. Yule, G. 1927. "On a Method of Investigating Periodicity in Disturbed Series with Special Reference to Wolfer's Sunspot Numbers." *Phil. Trans. Roy. Soc. London* **A 226**: 267-298.





Zarządzanie ryzykiem inwestycji ,

Maciej Krawczak,
Antoni Miklewski,
Andrzej Jakubowski,
Piotr Konieczny

Rozwój rynków finansowych doprowadził do sytuacji, w której jednym z najistotniejszych celów w banku staje się zarządzanie ryzykiem cenowym. Analizując ryzyko cenowe banku, działającego w warunkach polskich, wydaje się, że najważniejsze stają się krótkoterminowe prognozy cen rynkowych (np. stóp procentowych, kursów walutowych itd.). Bezpieczne zarządzanie ryzykiem cenowym oraz krótkoterminowe prognozowanie cen stają się w Polsce jednymi z najważniejszych elementów polityki banków.

W książce przedstawiono:

- modele struktury terminowej stóp procentowych,
- zagadnienia zarządzania portfelem obligacji,
- metodę value at risk,
- procesy stochastyczne w modelowaniu i prognozowaniu szeregów czasowych,
- analizę i predykcję szeregów czasowych z wykorzystaniem elementów sztucznej inteligencji i teorii informacji.

ISBN 83-85847-52-9

W celu uzyskania bliższych informacji i zakupu dodatkowych egzemplarzy prosimy o kontakt z Instytutem Badań Systemowych PAN
ul. Newelska 6, 01-447 Warszawa
tel. 837-35-78 w. 241 e-mail: bibliote@ibspan.waw.pl