

SYSTEM UCZENIA GŁĘBOKIEGO DLA ELIMINACJI SZUMÓW Z WYKORZYSTANIEM DYWERGENCJI ALPHA

Ryszard Szupiluk  <https://orcid.org/0000-0002-4416-7131>

Kolegium Analiz Ekonomicznych
Szkola Główna Handlowa w Warszawie
e-mail: rszupi@sgh.waw.pl

Streszczenie: W niniejszym artykule przedstawimy system eliminacji szumów z wielowymiarowych szeregów czasowych. Elementami tego systemu będą bloki: dekompozycji PCA/EVD, grupowania oparty na funkcji dywergencji alpha, dekompozycji odwrotnej PCA oraz filtracji indywidualnej. W ramach etapu grupowania przedyskutowane będą typowe problemy tego zadania, motywujące do poszukiwania nowych rozwiązań. Zaprezentujemy także eksperymenty potwierdzające słuszność opracowanej koncepcji.

Słowa kluczowe: grupowanie, filtracja szumów, eliminacja szumów, funkcje dywergencji, finansowe szeregi czasowe, uczenie głębokie

JEL classification: C02, C50

WPROWADZENIE

Eliminacja szumów i zakłóceń jest jednym z typowych zadań analizy danych. Motywowane to jest licznymi praktycznymi zastosowaniami jak: identyfikacja trendów ekonomicznych, separacja sygnałów medycznych czy rekonstrukcja sygnałów mowy. Stosowane są tu różne techniki bazujące m.in. na wygładzaniu średnimi ruchomymi, zastosowaniu transformacji Fouriera, czy rozkładzie szeregów na komponenty o określonych charakterystykach [Shiryayev 1999, Hamilton 1994, Vaseghi 1997, Richter et al. 2022].

W niniejszym opracowaniu przedstawimy system eliminacji szumów, który ze względu na swoją złożoność i charakter stosowanych metod, może być określony jako system uczenia głębokiego. Jednym z jego elementów jest analiza składowych głównych (PCA-ang. Principal Component Analysis), co można

<https://doi.org/10.22630/MIBE.2023.24.4.19>

bezpośrednio powiązać z dekompozycją względem składowych własnych (EVD – ang. Eigenvalue Value Decoposition), stąd będziemy stosować akronim PCA/EVD [Jolliffe 2002]. Prezentowana koncepcja eliminacji szumów zakłada podejście wielowymiarowe, w którym zestaw szeregów czasowych poddajemy dekompozycji PCA/EVD otrzymując ukryte komponenty. Zakładamy przy tym, że niektóre z nich są szumami, których identyfikacja a następnie eliminacja pozwoli na usunięcie wspólnych zakłóceń o niejako fizycznym charakterze, tj. takich gdzie występuje pewien obiektywny/odrębny składnik (lub składniki) identyczny w kilku szeregach. Jest to podejście nawiązujące do metod ślepej separacji sygnałów [Comon i Jutten 2010]. Kluczowym elementem jest ocena i wybór komponentów, które można uznać za szumy. Rozważymy tą kwestię jako problem grupowania.

Zagadnienie grupowania, określane także jako klasteryzacja lub segmentacja, należy do podstawowych zagadnień uczenia maszynowego [Mitchell 1997, Richter et al. 2022]. Mimo ideowej prostoty niektórych rozwiązań, a także możliwych wygodnych praktycznych przypadków, w ogólnym problemie grupowania występuje wiele ograniczeń i niejednoznaczności istotnie wpływających na jakość otrzymywanych rozwiązań. Stąd zastosowanie konkretnych algorytmów powinno uwzględniać specyfikę danego problemu, zwłaszcza charakterystykę danych [Everitt 2011].

W prezentowanym systemie zastosujemy algorytm grupowania oparty na funkcji dywergencji alpha [Amari 1985], ze względu na jej właściwości możliwe do wykorzystania w przypadku danych o charakterystyce szumowej lub posiadających istotny składnik losowy. Należy przy tym zaznaczyć, że ta metoda grupowania (lub oceny podobieństwa) może także stanowić samodzielne podejście w identyfikacji szumów.

WIELOWARSTWOWY SYSTEM ELIMINACJI SZUMÓW

Założmy, że analizowane szeregi czasowe, generowane są w pewnym wspólnym środowisku, jak np. rynki finansowe, a w związku z tym, składają się zarówno z niepożądanych szumów, zakłóceń lub krótkoterminowych fluktuacji własnych, jak i użytecznych składników wspólnych, jak trendy i cykle średnio i długoterminowe. Naszym celem jest w pierwszym etapie wydzielenie z bazowych szeregów czasowych wspólnych komponentów szumowych a następnie ich eliminacja oraz dokonanie transformacji odwrotnej. Dla tak uzyskanych „oczyszczonych” szeregów czasowych możliwa jest przejście do etapu eliminacji szumów indywidualnych, jedną ze standardowych metod filtracji. System eliminacji szumów, przedstawiony na rysunku 1, składa się z następujących warstw.

Wejście. Zestaw szeregów czasowych \mathbf{x}_i , $i=1,\dots,N$. zbieramy w jednej zmiennej wielowymiarowej $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_N]$.

Warstwa 1. Dekompozycja PCA/EVD. Zmienną \mathbf{X} poddajemy transformacji PCA otrzymując składowe główne $\mathbf{Y}=[y_1, y_2, \dots, y_N]$. Można przyjąć $\mathbf{Y}=\mathbf{U}^T\mathbf{X}$, gdzie \mathbf{U} jest macierzą kolumnowych wektorów własnych macierzy kowariancji \mathbf{X} otrzymanych z rozkładu EVD $\mathbf{X}\mathbf{X}^T/N=\mathbf{U}\mathbf{S}\mathbf{U}^T$, macierz \mathbf{S} oznacza diagonalną macierz wartości własnych rozkładu.

Warstwa 2. Grupowanie i identyfikacja szumu. Otrzymane składowe główne grupujemy, wydzielając grupę komponentów związanych z zaszumieniem lub zakłóceniami. Ten aspekt omówimy dokładniej w dalszej części artykułu.

Warstwa 3. Eliminacja szumów. Grupę komponentów szumowych eliminujemy. W praktyce procesowej oznacza to zastąpienie danego sygnału zerami lub przyjęcie w transformacji odwrotnej odpowiedniej kolumny zerowej.

Warstwa 4. Transformacja odwrotna. Dokonujemy transformacji odwrotnej do transformacji z warstwy pierwszej, otrzymując szeregi czasowe pozbawione komponentów szumowych.

Warstwa 5. Eliminacja szumów indywidualnych. Może się odbywać różnymi technikami, w tym zwykła średnią kroczącą.

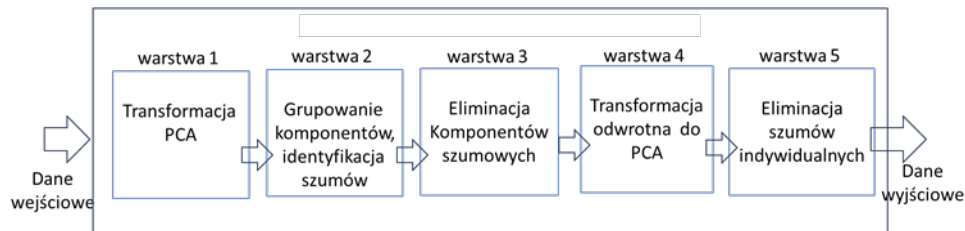
Wyjście. Otrzymujemy sygnały przefiltrowane.

Prezentowany układ, określa pewien ogólny schemat, który może być w szerokim zakresie modyfikowany lub dopasowywany do konkretnej sytuacji. W powyższej procedurze, w zasadzie jednym istotnym wymogiem/założeniem jest odwracalność macierzy transformacji. Przykładowo, w przypadku pracy z danymi typu Biga Data (m.in. duży wolumen, duża zmienność) wskazanym może być skorzystanie z adaptacyjnych algorytmów PCA [Oja 1992]. Należy przy tym zauważyć, iż w prezentowanym podejściu, bazując na EVD, nie musimy czynić żadnych dodatkowych założeń co do charakterystyki analizowanych szeregów bazowych, jak również odnośnie charakterystyki komponentów w nich zawartych. W naszym przypadku, przyjmujemy, że nie są dla nas istotne założenia związane z istnieniem statystyk drugiego rzędu, stacjonarnością lub ergodycznością sygnałów. Oczywiście można uwzględniać te kwestie w miarę potrzeb.

Zadania warstwy piątej, nie będą szerzej omawiane w niniejszej pracy, gdyż mogą być wykonane standardowymi technikami licznie ujętymi w literaturze [Therrien 1992, Oppenheim i Schafer 2014, Richter 2022].

Należy jednak zaznaczyć, iż eliminacja szumów indywidualnych powinna się odbywać po wcześniejszym usunięciu szumów wspólnych i zadania warstwy piątej nie powinny być wykonywane wcześniej. Zmiana tej kolejności oznaczać będzie konwolucyjne zmieszanie szumu wspólnego i indywidualnego danego sygnału z jego komponentami użytecznymi. Skuteczne wydzielenie takich komponentów jest zadaniem dekonwolucji wielowymiarowej, będącej nadal aktualnym wyzwaniem badawczym [Levin et al. 2009].

Rysunek 1. System uczenia głębokiego eliminacji szumów



Źródło: opracowanie własne

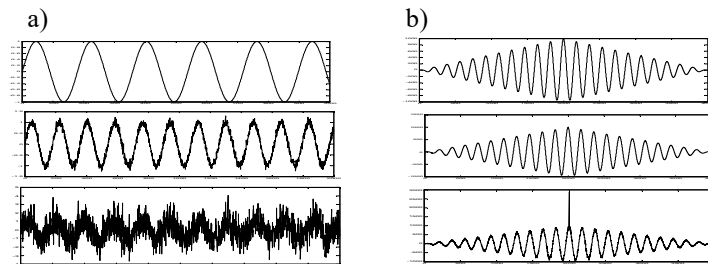
W przypadku zmiany postaci dekompozycji mogą się pojawić dodatkowe założenia. Niemniej, można zauważyć, że w instrumentalnej optyce uczenia maszynowego, istotność przyjętych założeń warunkowana jest specyfiką użycia danej metody w konkretnej sytuacji.

Jedną z modyfikacji może być przyjęcie innych dekompozycji wielowymiarowych jak: analiza składowych niezależnych (ICA – ang. Independent Component Analysis), nieujemna faktoryzacja macierzy (NMF – ang. Nonnegative Matrix Factorization) lub nieliniowe PCA [Comon i Jutten 2010]. Obecny wybór PCA/EVD motywowany jest doskonale zbadanymi jej właściwościami teoretycznymi oraz potwierdzonymi licznymi praktycznymi zastosowaniami.

PROBLEM GRUPOWANIA DANYCH SZUMOWYCH

Zasadniczym elementem prezentowanej koncepcji jest ocena komponentów uzyskanych z PCA/EVD, czyli rozróżnienie między komponentami użytecznymi a szumowymi. W przypadku finansowych szeregów czasowych, takimi jak np. logarytmiczne stopy zwrotu, sytuacja może być o tyle kłopotliwa, że komponenty szumowe mogą mieć charakterystykę daleką od czystych białych szumów losowych [Shiryaev 1999]. W efekcie ocena na bazie funkcji autokorelacji może nie być zbyt użyteczna, funkcja ta nie jest też wysoce czuła na szum, co pokazuje rysunek 1, gdzie dopiero stosunkowo duży poziom szumu jest widoczny w postaci piku na funkcji autokorelacji ($SNR \approx -2$ dB). Dlatego raczej będziemy oczekiwać, że komponenty zakłócające są w pewnym stopniu podobne do białych szumów, niż nimi faktycznie są. Stąd wyznaczenie komponentów szumowych potraktujemy jako zadanie przypisania do grupy o charakterze szumowym niż stricto spełniające warunki białego szumu.

Rysunek 2. Zaszumione sygnały a) i odpowiadające im funkcje autokorelacji b), gdzie zerowe opóźnienie odpowiada obserwacji środkowej.



Źródło: opracowanie własne

Zagadnienie grupowania można przedstawić jako identyfikację w posiadanym zbiorze: danych, obiektów, obserwacji lub wektorów do siebie podobnych lub sobie bliskich. Przynależność do danej grupy bazuje zwykle na wybranym matematycznym kryterium podobieństwa lub odległości. Można wyróżnić takie klasy metod grupowania jak: metody oparte na metrykach, grupowanie hierarchiczne lub metody oparte na rozkładach [Hennig et al. 2015, Richter et al. 2022]. Możliwe jest także grupowanie w oparciu o modele danych i ich parametry, a także grupowanie uwzględniające kontekst praktycznych zastosowań. W szerszym rozumieniu grupowanie można określić jako identyfikację wewnętrznej struktury wielowymiarowych danych [Maharaj et al. 2019, Meila 2003].

W każdym z wymienionych podejść występują algorytmy w różnych wariantach i lokalnych modyfikacjach. Jednak, mimo, że zasadniczą formułą implementacji oraz realizacji algorytmów grupowania jest uczenie maszynowe, gdzie użytkownik metody musi zwykle określić ile grup (klastrow) ma uzyskać, lub zaakceptować regułę to określającą, co można potraktować jako tzw. hiperparametr [Agrawal et al. 2005, Amorim i Hennig 2015]. Określanie pewnych cech odnoszących się do modelu, procesu uczenia czy otrzymanych wyników jest typowe w całokształcie metod uczenia maszynowego. Jednak w przypadku grupowania danych wielowymiarowych mamy z reguły znacznie ograniczone możliwości oceny słuszności przyjętych założeń lub sposobu parametryzacji algorytmów, a w konsekwencji, trudności weryfikacji czy kontroli uzyskanych rezultatów [Hüllermeier 2010]. Nie tylko różne metody, ale także te same, tylko inaczej sparametryzowane (np. k-średnich z różnymi metrykami lub początkowymi centrami) mogą prowadzić do całkowicie odmiennych wyników [Celebi et al. 2013].

W przypadku szeregów czasowych o charakterze szumowym lub ze znaczącym komponentem szumowym, problemem może być sposób oceny ich podobieństwa na bazie charakterystyk indywidualnych bądź wzajemnych relacji opartych na standardowych metrykach lub statystykach drugiego rzędu. Z tego

względu, zastosujemy funkcję dywergencji alpha jako metodę oceny podobieństwa, wykorzystując jej interesujące właściwości analityczne, w szczególności cechę asymetrii miary podobieństwa [Amari i Nagaoka 2000].

FUNKCJA DYWERGENCJI ALPHA

Dywergencja $D(\mathbf{y}||\mathbf{z})$, to funkcja dwuargumentowa, określona na nieujemnych zmiennych z i y , która spełnia warunek $D(\mathbf{y}||\mathbf{z}) \geq 0$, gdzie $D(\mathbf{y}||\mathbf{z}) = 0$ wtedy i tylko wtedy, gdy $\mathbf{y} = \mathbf{z}$. Dywergencja nie musi spełniać nierówności trójkąta $D(\mathbf{y}||\mathbf{z}) \leq D(\mathbf{y}||\mathbf{x}) + D(\mathbf{x}||\mathbf{z})$. Dywergencje są z reguły niesymetryczne $D(\mathbf{y}||\mathbf{x}) \neq D(\mathbf{x}||\mathbf{y})$. Dla niektórych dywergencji konieczny jest warunek sumowania się wartości zmiennych z i y do jedności. Dywergencje mogą być zdefiniowane dla wielkości ciągłych jak i dyskretnych. Obecnie funkcje dywergencji stosowane są do oceny podobieństwa (lub jego braku) między rozkładami oraz/lub nieujemnymi zmiennymi, wektorami, macierzami lub funkcjami [Amari 1985, Cichocki et al. 2009, Csiszar 1974].

Dywergencja, może być też przyjmowana lub interpretowana, jako miara quasi-odległości. Miary dywergencji, uzyskiwane mogą być zarówno na drodze aksjomatycznej, tzn. jako charakterystyki o pewnych interesujących właściwościach matematycznych [Csiszar 2008], jak i wynikać z badania praktycznych problemów [Cichocki et al. 2009].

Większość z dość licznej listy dywergencji można zaklasyfikować lub uznać za szczególne przypadki kilku zasadniczych rodzin. Dwie szerokie klasy stanowią dywergencje Ciszar oraz Bregman'a.

Dywergencje Ciszar, zwane także f -dywergencjami, wyrażają się ogólną formułą [Csiszar 1974]:

$$D_c(\mathbf{y} || \mathbf{z}) = \sum_{i=1}^N \left(z_i f \left(\frac{y_i}{z_i} \right) \right), \quad (1)$$

gdzie elementy wektorów \mathbf{x} oraz \mathbf{y} spełniają warunki $x_i \geq 0, y_i \geq 0$ zaś $f(u)$, $u = \frac{y}{z}$ jest funkcją wypukłą w $(0, \infty)$ oraz ciągłą w punkcie 0, N oznacza ilość elementów wektorów \mathbf{y} oraz \mathbf{z} . Przyjmuje się, że jeśli funkcja dywergencji (1) spełnia dodatkowe warunki, takie że, $f(1) = 0$ oraz jest ona ściśle wypukła w 1, to może ona być miarą podobieństwa.

Dywergencje Bregmana stanowią szeroką klasę określaną przez [Bregman 1967, Dhillon i Tropp 2007]:

$$D_\phi(\mathbf{y} || \mathbf{z}) = \sum_{t=1}^N \left(\phi(y_t) - \phi(z_t) - \phi'(z_t)(y_t - z_t) \right), \quad (2)$$

gdzie $\phi(u)$ jest ściśle wypukłą funkcją klasy C^1 (wszędzie różniczkowalną) posiadającą ciągłą pochodną $\phi'(u)$.

W dalszych rozważaniach skoncentrujemy się na dywergencji alpha, która jest klasą dywergencji należącą zarówno do rodziny dywergencji Bregmana jak i f-dywergencji Ciszar'a [Amari 1985, Amari i Nagaoka 2000]. Dywergencja ma postać:

$$D_A^\alpha(\mathbf{y}||\mathbf{z}) = \sum_{i=1}^N \frac{y_i^\alpha z_i^{1-\alpha} + (\alpha - 1)z_i - \alpha y_i}{\alpha(\alpha - 1)}. \quad (3)$$

Dla parametru $\alpha=2$ stanowi odległość Pearson'a, dla $\alpha=0.5$ odległość Hellingera, dla $\alpha=-1$ odległość chi-kwadrat Neyman'a. Dla $\alpha \rightarrow 1$ otrzymuje się uogólnioną dywergencję Kullback'a-Leibler'a: $D_A^{\alpha \rightarrow 1}(\mathbf{y}||\mathbf{z}) = D_{GKL}(\mathbf{y}||\mathbf{z})$ zaś dla $\alpha \rightarrow 0$ otrzymuje się dualną uogólnioną dywergencję Kullback'a-Leibler'a: $D_A^{\alpha \rightarrow 0}(\mathbf{y}||\mathbf{z}) = D_{DGKL}(\mathbf{y}||\mathbf{z})$ [Cichocki et al. 2009]. Alpha dywergencja Amarięgo może zostać otrzymana z dywergencji Ciszar'a za pomocą nieliniowości:

$$f(u) = \frac{u(u^{\alpha-1} - 1)}{\alpha^2 - \alpha} + \frac{1-u}{\alpha}. \quad (4)$$

Należy tu zaznaczyć, że dywergencja alpha nie wymaga sumowania wartości zmiennych do jedności. My jednak będziemy dokonywać wstępnego przetworzenia danych tak, aby ten warunek został spełniony.

GRUPOWANIE Z FUNKCJĄ DYWERGENCJI ALPHA

W zadaniu grupowania z wykorzystaniem dywergencji alpha, wykorzystamy jej cechę niesymetryczności wraz z typową charakterystyką szumów losowych, jaką jest ich samopodobieństwo [Mandelbrot 1997]. Możemy oczekiwać, że te dwie właściwości skutkować będą otrzymywaniem relatywnie niedużych wartości symetrycznych dla „czystych” szumów $D^\alpha(\mathbf{x}||\mathbf{y})=D^\alpha(\mathbf{y}||\mathbf{x})$, i wyników niesymetrycznych $D^\alpha(\mathbf{x}||\mathbf{y}) \neq D^\alpha(\mathbf{y}||\mathbf{x})$ o większych wartościach dla danych o mniejszej zawartości szumu. Cały algorytm grupowania można przedstawić następująco:

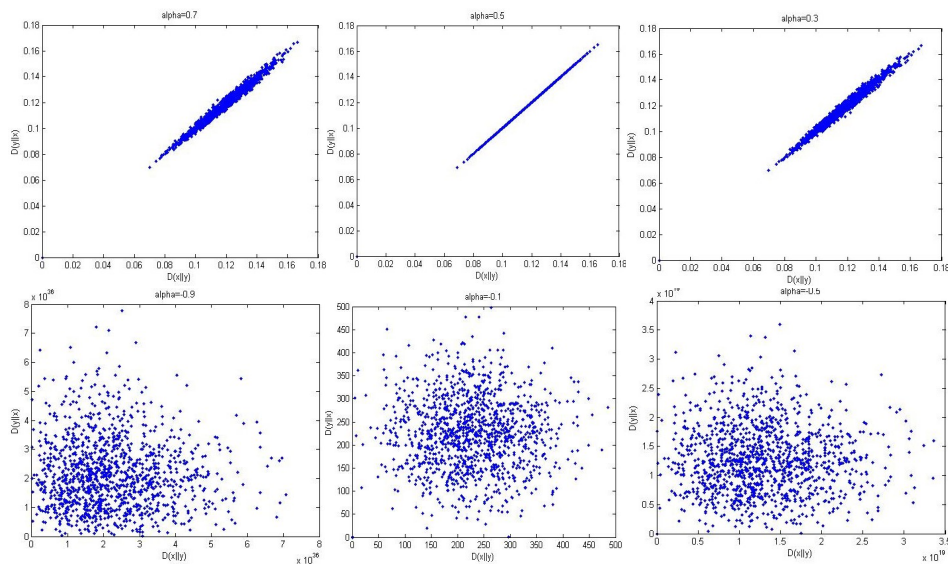
1. Dla komponentów \mathbf{Y} dokonaj wstępnego przygotowania danych, postaci (zapis Matlaba) $\mathbf{y}_i = \mathbf{y}_i - \min(\mathbf{y}_i)$ oraz $\mathbf{y}_i = \mathbf{y}_i / (\sum(\mathbf{y}_i) + \exp(-100))$.
2. Dla wszystkich par komponentów otrzymanych z PCA oraz przyjętego α obliczamy dywergencje $D^\alpha(\mathbf{y}_i||\mathbf{y}_j)$ oraz $D^\alpha(\mathbf{y}_j||\mathbf{y}_i)$.
3. Znajdujemy takie pary komponentów, że $D^\alpha(\mathbf{y}_i||\mathbf{y}_j) \approx D^\alpha(\mathbf{y}_j||\mathbf{y}_i) \approx 0$.
4. Weryfikujemy otrzymane wyniki przy różnych wartościach α .

Typowy zakres α należy do przedziału $(0,1)$, zmiana tego parametru w tym zakresie nie powinna zasadniczo zmieniać wyników grupowania. Można także oprzeć się na uśrednionym pomiarze wartości dywergencji przy różnych wartościach α .

BADANIE PRAKTYCZNE

Obecnie przedstawimy powyższą procedurę na przykładzie symulowanych szeregów czasowych. Na początek sprawdzimy słuszność koncepcji oceny szumów (sygnałów losowych) dla wygenerowanych szeregów z rozkładu gaussowskiego ($x=\text{randn}(50,1000)$ w zapisie Matlaba). Rysunek 3 przedstawia rozrzut punktów we współrzędnych $D(y|x)$, $D(x|y)$. Celem badania jest test efektu symetryczności oraz niedużych wartości miar dywergencji dla sygnałów losowych. Badanie przeprowadzono dla parametrów α : 0.7; 0.5, 0.3, -0.9, -0.1, -0.5. Jak widać, dla parametrów dodatnich uzyskujemy efekty zgodne z oczekiwaniami. Ujemne parametry α wykazują w zasadzie całkowity brak zależności co do kierunku i wartości, w zakresie tych parametrów metoda nie działa. Jest to o tyle znamienne, że przy $\alpha \rightarrow -1$, alpha dywergencja zbiega do popularnej miary chi-kwadrat Neyman'a.

Rysunek 3. Wartości $D(y|x)$ względem $D(x|y)$ dla sygnałów losowych (szumy gaussowskie) dla parametrów α : 0.7, 0.5, 0.3 w górnym wierszu oraz -0.9, -0.1, -0.5 w dolnym

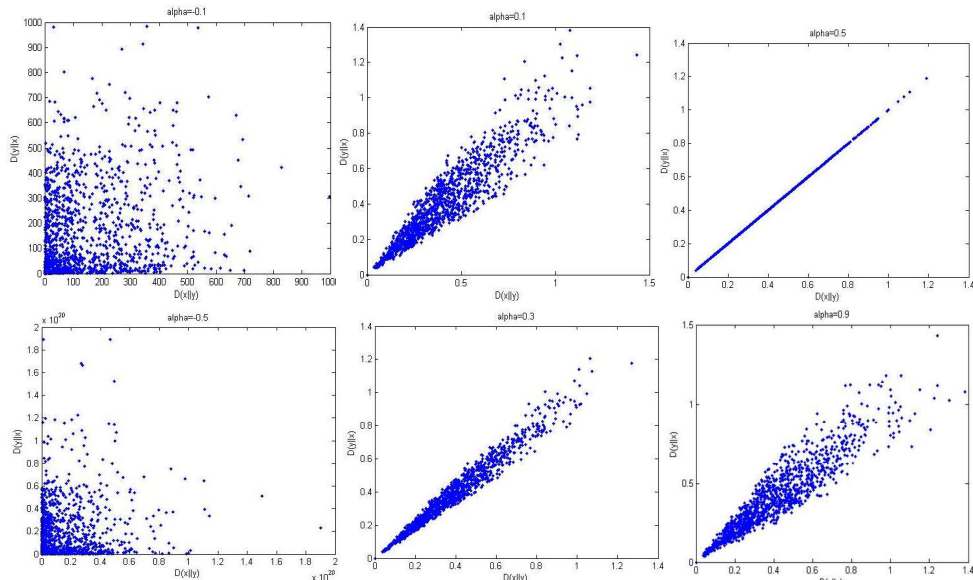


Źródło: opracowanie własne

Kolejne badanie przeprowadzimy dla sygnałów symulujących giełdowe szeregi czasowe, które wygenerowaliśmy jako $x = \text{abs}(\text{cums}(\text{randn}(50,1000)))$ w zapisie Matlab.

Rysunek 4 przedstawia rozrzut punktów we współczesnych $D(y||x)$, $D(x||y)$ dla parametrów α : -0.3, -0.1, 0.1, 0.5, 0.3, 0.9. Ponownie ujemne parametry α wykazują brak zależności co do kierunku i wartości miar dywergencji α , natomiast dla dodatnich α mamy wyniki zgodne z oczekiwaniami. Widzimy zróżnicowanie wartości zarówno jeżeli chodzi o położenie wokół linii symetrii $D(y||x) = D(x||y)$, jak i zróżnicowanie względem przyjmowanych wielkości.

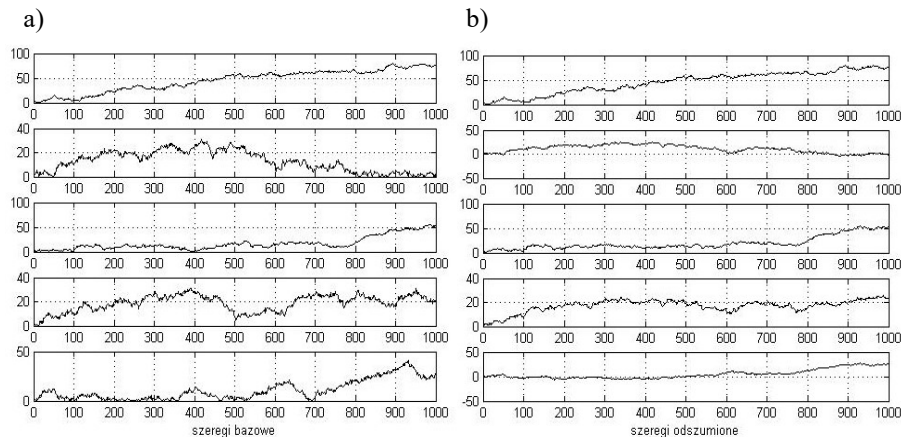
Rysunek 4. Wartości $D(y||x)$ względem $D(x||y)$ dla sygnałów losowych (szumy gaussowskie) dla parametrów α : -0.1, 0.1, 0.5 w górnym wierszu oraz -0.3, 0.3, 0.9 w dolnym



Źródło: opracowanie własne

Te pary punktów, które leżą blisko punktu (0,0) oraz w bliskiej odległości od linii symetrii, możemy uznać za szumowe. Słuszność takiego podejścia prezentuje badanie wykonane na symulowanych finansowych szeregach czasowych, z których usunięto punkty uznane za szumowe. Jak widać, usunięcie tych komponentów pozwoliło na wygładzenie szeregów czasowych, przy zachowaniu głównych trendów. Oznacza to, że w istocie zostały usunięte szumy i krótkoterminowe fluktuacje.

Rysunek 5. Działanie systemu eliminacji szumów a) oryginalne bazowe szeregi czasowe, b) szeregi odszumione



Źródło: opracowanie własne

PODSUMOWANIE

Przedstawiony system uczenia głębokiego do eliminacji szumów określa pewien ogólny schemat postępowania, którego poszczególne elementy, mogą być w szerokim stopniu modyfikowane. Dotyczy to zwłaszcza postaci dekompozycji wielowymiarowej. Wybór podejścia opartego na PCA/EVD, związany był, z jednej strony, z pożądanymi właściwościami matematycznymi tej metody, a z drugiej, z jej popularnością teoretyczną i praktyczną. Pozwoliło to poświęcić zasadniczą uwagę aspektowi identyfikacji komponentów szumowych, co wykonano w ramach problemu grupowania.

Koncepcja grupowania oparta na dywergencji alpha, znalazła potwierdzenie w przeprowadzonych eksperymentach. Metoda ta może być także rozważana jako samodzielne podejście analityczne, niezależne od prezentowanego systemu filtracji. Jednak wykorzystanie dywergencji alpha wydaje się zasadne do przypadków, w których występują szumy lub sygnały w jakimś stopniu o charakterze losowym. Dywergencja ta jest także stosunkowo obszerne opracowana teoretycznie. Niemniej, zastosowanie innych dywergencji, z których wiele jest mało zbadanych, otwiera szanse na uzyskanie nowych ciekawych wyników.

Ze względu na ograniczony charakter niniejszego opracowania pominieliśmy omawianie ostatniej warstwy naszego systemu jaką jest filtracja indywidualnych szeregów czasowych. Jest to jednak temat bardzo dobrze opisany i na chwilę obecną przyjmujemy, że stosujemy standardowe metody dla tego problemu. Nie wykluczamy jednak, wprowadzenia w przyszłości pewnych nowych rozwiązań także w tej warstwie.

BIBLIOGRAFIA

- Agrawal R., Gehrke J., Gunopulos D., Raghavan P. (2005) Automatic Subspace Clustering of High Dimensional Data, *Data Mining and Knowledge Discovery*, 11, 5-33.
- Amari S. (2012) *Differential-Geometrical Methods in Statistics*. Springer Verlag.
- Amari S. and Nagaoka H. (2000) *Methods of Information Geometry*. Oxford University Press, New York.
- Amorim R. and Hennig C. (2015) Recovering the Number of Clusters in Data Sets with Noise Features using Feature Rescaling Factors. *Information Sciences*, 324, 126-145.
- Bregman L (1967) The Relaxation Method of Finding a Common Point of Convex Sets and its Application to the Solution of Problems in Convex Programming. *Comp. Math. Phys., USSR*, 7, 200-217.
- Celebi M., Kingravi H., Vela P. (2013) A Comparative Study of Efficient Initialization Methods for the k-Means Clustering Algorithm. *Expert Systems with Applications*, 40(1), 200-210.
- Cichocki A., Zdunek R., Phan A.-H., Amari S. (2009) *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis*. John Wiley.
- Comon P., Jutten Ch. (2010) *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic Press.
- Csiszar I. (1974) Information Measures: A Critical Survey. [In:] *Transactions of the 7th Prague Conference*, 83-86.
- Csiszar I. (2008) Axiomatic Characterizations of Information Measures. *Entropy*, 10, 261-273.
- Dhillon I. S., Tropp J. A. (2007) Matrix Nearness Problems with Bregman Divergences. *SIAM Journal on Matrix Analysis and Applications*, 29(4), 1120-1146.
- Everitt B. (2011) *Cluster Analysis*. Chichester, West Sussex, U.K, Wiley.
- Hamilton J. D. (1994) *Time Series Analysis*. Princeton, Princeton University Press.
- Hennig Ch., Meila M., Murtagh F. (2015) *Handbook of Cluster Analysis*. CRC Press.
- Hüllermeier E. (2010) Uncertainty in Clustering and Classification. [In:] Deshpande A., Hunter A. (eds) *Scalable Uncertainty Management. SUM 2010. Lecture Notes in Computer Science()*, 6379, 16-19. Springer, Berlin, Heidelberg.
https://doi.org/10.1007/978-3-642-15951-0_6
- Jolliffe I. T. (2002) *Principal Component Analysis*. Springer Verlag.
- Levin A., Weiss Y., Durand F. and Freeman W.T. (2009) Understanding and Evaluating Blind Deconvolution Algorithms. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, Miami, 20-25 June 2009, 1964-1971.
- Mandelbrot B. B. (1997) Multifractals and 1/f Noise. *Wild Self-Affinity in Physics (1963–1976)*, Springer-Verlag. <https://doi.org/10.1007/978-1-4612-2150-0>
- Maharaj E., D'Urso P., Caiado J., (2019) *Time Series Clustering and Classification*, CRC Press, Boca Raton.
- Meila M. (2003) Comparing Clusterings by the Variation of Information, *Learning Theory and Kernel Machines. Lecture Notes in Computer Science*, 2777, 173-187.
- Mitchell T. (1997) *Machine Learning*. McGraw-Hill, Boston.

- Oja E. (1992) Principal Components, Minor Components and Linear Neural Networks. *Neural Networks*, 5, 927-935.
- Oppenheim R., Schaffer A. (2014) *Discrete - Time Signal Processing*, Pearson India.
- Richter A., Sheuli P., Kępuska V., Silaghi M. (2022) *Signal Processing and Machine Learning with Applications*. Springer Cham.
- Shiryaev A. N. (1999) *Essentials of Stochastic Finance: Facts, Models, Theory*. Singapore, World Scientific.
- Therrien, C. W. (1992) *Discrete Random Signals and Statistical Signal Processing*. Prentice Hall, New Jersey.
- Vaseghi S. V. (1997) *Advanced Signal Processing and Digital Noise Reduction*. John Wiley and Sons, Chichester.

DEEP LEARNING SYSTEM FOR NOISE ELIMINATION USING ALPHA DIVERGENCE

Abstract: In this article, we present a system for noise eliminating from multivariate time series. The system contain the following layers: PCA/EVD decomposition, grouping based on the Alpha divergence function, reverse decomposition to PCA, and individual filtration. During the grouping stage, typical problems of this task will be discussed, motivating the search for new solutions. We will also present experiments confirming the validity of the developed concept.

Keywords: clustering, noise filtering, divergence functions, financial time series, deep learning

JEL classification: C02, C50