METODY ILOŚCIOWE W BADANIACH EKONOMICZNYCH
Tom XXIII/2, 2022, s. 59 – 66

IDENTYFIKACJA UKRYTYCH KOMPONENTÓW W SZEREGACH CZASOWYCH METODAMI NIEUJEMNEJ FAKTORYZACJI MACIERZY

Ryszard Szupiluk https://orcid.org/0000-0002-4416-7131
Kolegium Analiz Ekonomicznych
Szkoła Główna Handlowa w Warszawie
e-mail: rszupi@sgh.waw.pl

Streszczenie: W niniejszym artykule przedstawimy zastosowanie metod nieujemnej faktoryzacji macierzy do identyfikacji ukrytych komponentów zawartych w ekonomicznych szeregach czasowych. Wyprowadzony zostanie nowy algorytm nieujemnej faktoryzacji macierzy oparty na dywergencji Fermiego-Diraca. Wykorzystując uzyskane ukryte komponenty dokonamy eliminacji szumów z szeregów czasowych reprezentujących wyniki predykcji. Całą koncepcję przetestujemy w problemie prognozy obciążenia systemu elektroenergetycznego.

Słowa kluczowe: nieujemna faktoryzacja macierzy, identyfikacja ukrytych komponentów, ślepa separacja, predykcja

JEL classification: C02, C50

WSTĘP

Jednym z istotnych zagadnień analizy danych empirycznych jest poszukiwanie interesujących i użytecznych metod dekompozycji oraz faktoryzacji macierzy. W obszarze ekonomii i finansów motywacją dla tych poszukiwań jest dążenie do odkrycia, identyfikacji lub rekonstrukcji niejawnych komponentów (sygnałów, szeregów czasowych) zawartych w danych rynkowych lub biznesowych. Z praktycznego punktu widzenia znajomość takich komponentów może być wykorzystana w zadaniach eliminacji szumów, redukcji wpływu niepożądanych zakłóceń, estymacji trendów lub poprawie jakości predykcji [Szupiluk 2014].

https://doi.org/10.22630/MIBE.2022.23.2.6

W niniejszym artykule przedstawimy zastosowanie algorytmu nieujemnej faktoryzacji macierzy (ang. Nonnegative Matrix Factorization - NMF) w dekompozycji szeregów czasowych, a następnie wykorzystanie uzyskanych komponentów w autorskim systemie poprawy predykcji. Wyprowadzony zostanie nowy algorytm nieujemnej faktoryzacji bazujący na dywergencji Fermiego-Diraca należącej do klasy dywergencji Bergmana [Cichocki et al. 2009].

Metoda NMF adresowana jest do danych nieujemnych, co w przypadku wartości rynkowych jak akcje lub indeksy giełdowe, oznacza raczej analizę wartości bazowych niż wartości zmian tych instrumentów, choć przy odpowiedniej transformacji także jest to możliwe. Ma to szczególne znaczenie dla szybkiej i bezpośredniej interpretacji sytuacji rynkowej w przypadku podejmowania decyzji inwestycyjnych, opartych na analizie technicznej. automatycznych giełdowych systemach inwestycyjnych lub systemach predykcyjnych rynku energii.

NIEUJEMNA FAKTORYZACJA MACIERZY

Problem nieujemnej faktoryzacji macierzy można opisać jako przedstawienie danej macierzy **X** iloczynem dwóch macierzy nieujemnych **A**, **S** [Berry et al. 2009]

$$\mathbf{X} \cong \mathbf{AS}, \tag{1}$$

gdzie elementy macierzy **A** są nieujemne $a_{ij} \geq 0$, co zapisywane jest także jako $\mathbf{A} \geq 0$, analogiczny warunek nieujemności dotyczy macierzy **S**.

W praktyce zastosowań często zagadnienie NMF przedstawia się w kategoriach problemu ślepej separacji sygnałów/źródeł (ang. Blind Signal /Source Separation – ang. BSS) [Comon, Jutten 2010]. W takiej interpretacji **X** oznacza nieujemną macierz danych obserwowanych, macierz **S** zawiera wiersze stanowiące ukryte komponenty zawarte w danych obserwowanych, zaś **A** oznacza system mieszający. Zakłada się przy tym, że liczba wierszy macierzy **X** jest nie mniejsza od liczby wierszy macierzy **S**.

W problemie BSS bazując na samych obserwacjach staramy się wyodrębnić nieznane macierze **A** oraz **S**, przy czym w wielu przypadka oczekuje się, że identyfikacja ukrytych komponentów (wiersze **S**) jest jednoznaczna z wydzieleniem pewnych konkretnych fizycznych sygnałów jak sygnały mowy, medyczne lub geologiczne. W tak postawionym problemie oczekuje się, że niezależnie od użytych metod BSS wraz z odpowiadającymi im założeniami, otrzyma się podobny „fizyczny" efekt, np. konkretną wypowiedź wyłowioną z gwaru rozmów. W przypadku danych ekonomicznych, choć otrzymane komponenty zawarte w S stanowią zwykle raczej pewną bazę analityczną, to możemy poszukiwać lub oczekiwać wyników także o niejako fizycznym charakterze takich jak trendy lub szumy i zakłócenia.

Nieujemna faktoryzacja macierzy **X=AS,** podobnie jak inne metody BSS, charakteryzuje się naturalną niejednoznacznością rozwiązań, ze względu na

$\mathbf{ADD}^{-1}\mathbf{X}$, gdzie dowolna odwracalna macierz **D** odpowiada za przeskalowanie i permutację sygnałów separowanych w stosunku do źródłowych. Oznacza to, że poszukując obiektywnych fizycznych sygnałów akceptujemy jako satysfakcjonujące rozwiązanie w postaci sygnałów przeskalowanych oraz występujących w zmienionej kolejności do źródłowych.

Rozwiązanie problemu NMF wiąże się z przyjęciem kryterium jakości faktoryzacji, które jednocześnie pełni rolę funkcji celu. Funkcja celu [Cichocki et al. 2009]

$$J = L(\mathbf{X}, \mathbf{AS}), \tag{2}$$

określa różnicę, podobieństwo lub odmienność między macierzą **X** oraz jej faktoryzacją **AS**. Przyjmowane obecnie funkcje celu opierają się zwykle na funkcjach dywergencji $J = D(\mathbf{X}, \mathbf{AS})$.

Dywergencją $D(y \| z)$, określana jest funkcja dwóch nieujemnych zmiennych z i y, która spełnia warunek $D(y \| z) \geq 0$, gdzie $D(y \| z) = 0$, gdy $y = z$ [Csiszar 1974]. Dywergencja, nie musi spełniać nierówności trójkąta $D(y \| z) \leq D(y \| x) + D(x \| z)$, oraz nie musi być dla niej spełniony warunek symetryczności, tj. zwykle $D(y \| x) \neq D(x \| y)$. Dla niektórych dywergencji dochodzą dodatkowe warunki, np. sumowania się wartości zmiennych z i y do jedności. Jedną z szerszych rodzin dywergencji stanowią dywergencje Bregmana definiowane jako [Bregman 1967],

$$D_{\phi}(\mathbf{y} \| \mathbf{z}) = \sum_{i=1}^{N} \Big(\phi(y_i) - \phi(z_i)) - \phi^{'}(z_i)(y_i - z_i) \Big), \tag{3}$$

gdzie $\phi(u)$ jest ściśle wypukłą funkcją, wszędzie różniczkowalną, posiadającą ciągłą pochodną $\phi^{'}(u)$ [Dhillon i Tropp 2005] . Dla problemu NMF (3) można przedstawić w postaci

$$D_{\phi}(\mathbf{X} \| \mathbf{AS}) = \sum_{i=1}^{m} \sum_{t=1}^{N} \Big(\phi(x_{it}) - \phi([\mathbf{AS}]_{it})) - \phi^{'}(x_{it})(x_{it} - [\mathbf{AS}]_{it}) \Big). \tag{4}$$

Dla nieliniowości $\phi(u) = u\ln(u) + (1-u)\ln(1-u)$, $u \in [0,1]$, otrzymuje się dywergencję Fermiego-Diraca postaci [Cichocki et al.2009]

$$D_{FD}(\mathbf{y} \| \mathbf{z}) = \sum_{i=1}^{N} \left(y_i \ln\left(\frac{y_i}{z_i} \right) + (1 - y_i) \ln\left(\frac{1 - y_i}{1 - z_i} \right) \right), \tag{5}$$

dla $y_i, z_i \in (0,1)$.

W praktyce wykorzystanie dywergencji może wiązać się z koniecznością przekształcenia analizowanych danych do przedziału (0,1).

ALGORYTM UCZENIA NMF

Multiplikatywny algorytm NMF dla dywergencji Bregmana o naprzemiennych iteracjach. odpowiednio dla macierzy A ma postać [Dhillon i Sra 2005]

$$\mathbf{A} \leftarrow \mathbf{A}.\frac{(\phi^{"}(\mathbf{AS}).*\mathbf{X})\mathbf{S}^T)}{(\phi^{"}(\mathbf{AS}).*(\mathbf{AS}))\mathbf{S}^T}, \tag{6}$$

a dla macierzy **S**

$$\mathbf{S} \leftarrow \mathbf{S}.\frac{\mathbf{A}^T(\phi^{"}(\mathbf{AS}).*\mathbf{X})}{\mathbf{A}^T(\phi^{"}(\mathbf{AS}).*(\mathbf{AS}))}, \tag{7}$$

gdzie .*, oznacza mnożenie element przez element.

W przypadku przyjęcia funkcji celu w postaci dywergencji Fermiego-Diraca otrzymujemy drugą pochodną funkcji $\phi(u)$ w postaci $\phi^{"}(u)=\left(1-u^2\right)^{-1}$, co prowadzi do następującego algorytmu estymacji macierzy **A** oraz **S**

$$\mathbf{A} \leftarrow \mathbf{A}.\frac{(\mathbf{1}./(\mathbf{1}-(\mathbf{AS}).\hat{}2).*\mathbf{X})\mathbf{S}^T)}{(\mathbf{1}./(\mathbf{1}-(\mathbf{AS}).\hat{}2).*(\mathbf{AS}))\mathbf{S}^T}, \tag{8}$$

oraz

$$\mathbf{S} \leftarrow \mathbf{S}.\frac{\mathbf{A}^T(\mathbf{1}./(\mathbf{1}-(\mathbf{AS}).\hat{}2).*\mathbf{X})}{\mathbf{A}^T(\mathbf{1}./(\mathbf{1}-(\mathbf{AS}).\hat{}2).*(\mathbf{AS}))}, \tag{9}$$

Uzyskany algorytm (8)-(9) jest jednym z wielu wariantów możliwych do uzyskania na bazie funkcji celu w postaci dywergencji Fermiego-Diraca. Postacie te zależne są od przyjętej metody minimalizacji funkcji celu.

WIELOMODELOWY SYSTEM POPRAWY JAKOŚCI PREDYKCJI

Model nieujemnej faktoryzacji macierzy (1) można zinterpretować w kategoriach systemu poprawy predykcji. Przyjmijmy, że wiersze macierzy $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_n]^T$ zawierają wyniki prognoz z różnych modeli. Załóżmy następnie, że są one kombinacją liniową **X=AS** ukrytych komponentów odpowiadających za poprawne wyniki prognoz, jak również komponentów odpowiedzialnych za błędy. Zatem możemy zapisać, że macierz

$$\mathbf{S} = [\mathbf{s}_1,...,\mathbf{s}_n]^T = [\hat{\mathbf{s}}_1,...,\hat{\mathbf{s}}_p, \tilde{\mathbf{s}}_{p+1},...., \tilde{\mathbf{s}}_{p+q}]^T, \tag{10}$$

składa się z komponentów konstruktywnych $\hat{\mathbf{s}}_i$ oraz destrukcyjnych $\hat{\mathbf{s}}_j$, $n = p + q$. Identyfikując macierz **A** oraz **S**, a następnie eliminując komponenty destrukcyjne (czyli przyjmując $\hat{\mathbf{s}}_j = \mathbf{0}$) powinniśmy uzyskać poprawę jakości predykcji

$$\hat{\mathbf{X}} = \mathbf{A}[\hat{\mathbf{s}}_1,...,\hat{\mathbf{s}}_p,\mathbf{0}_{p+1},....,\mathbf{0}_n]^T, \tag{11}$$

gdzie $\hat{\mathbf{X}}$ oznacza macierz której wiersze zawierają poprawione prognozy [Szupiluk et al. 2007]. Ten sam efekt zostanie osiągnięty w wyniku zastąpienia w macierzy mieszającej $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2,..., \mathbf{a}_n]$ odpowiednich kolumn wektorami zer

$$\hat{\mathbf{A}} = [\mathbf{a}_1,...,\mathbf{a}_p,\mathbf{0}_{p+1},\mathbf{0}_{p+2},...,\mathbf{0}_n]. \tag{12}$$

Prowadzi to do analogicznej do (11) formuły na poprawione wyniki prognoz

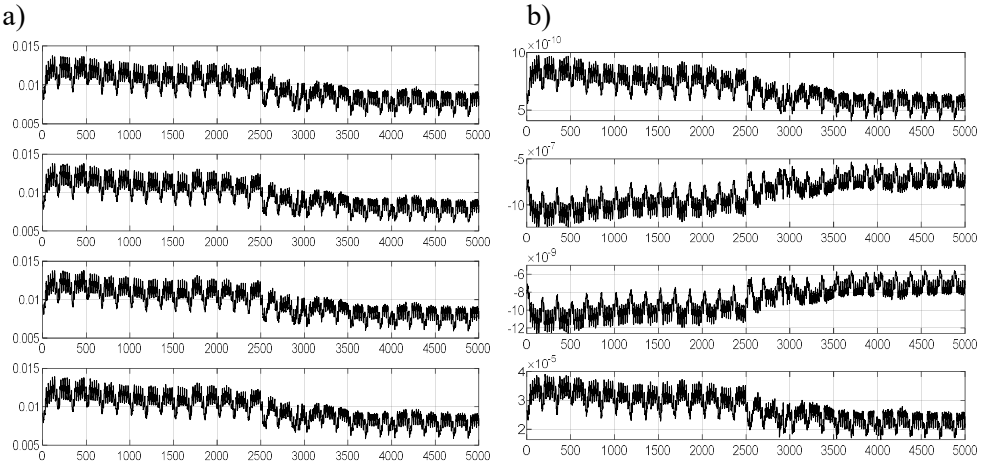
$$\hat{\mathbf{X}} = \hat{\mathbf{A}}\mathbf{S}. \tag{13}$$

Identyfikacja macierzy **A** oraz **S**, intepretowanych w kategoriach problemu BSS jako macierz mieszająca oraz macierz sygnałów źródłowych (ukrytych komponentów) zostanie dokonana metodą nieujemnej faktoryzacji macierzy. Należy zauważyć, że osiągnięcie celu stawianego zadaniom BSS oznacza eliminację komponentów zakłócających o niejako fizycznym charakterze, np. szumów w danych. Ich eliminacja powinna skutkować poprawą prognozy na ocenianą wg dowolnego kryterium matematycznego. Jednak w praktyce problemów ekonomicznych, uzyskanie czystych fizycznych komponentów może być trudne do osiągniecia, gdyż trudno zakładać, iż określone dane jak np. szeregi zużycia energii elektrycznej, są efektem kombinacji pewnej ograniczonej liczby sygnałów źródłowych.

EKSPERYMENT PRAKTYCZNY

Obecnie przedstawimy zastosowanie opisanej wyżej metody do poprawy predykcji zużycia energii elektrycznej. Predykcja godzinnego obciążenia jest dokonywana z wyprzedzeniem 24 godzin. Dane objaśniające obejmują informacje o godzinnym zużyciu energii w ciągu ostatnich 24 godzin, jak również dane kalendarzowe typu: dzień, miesiąc. Dla tak postawionego problemu, opracowano cztery modele neuronowe MLP o 12, 18, 24, 27 neuronach w warstwie ukrytej. Na Rysunku 1 widzimy wyniki z modeli predykcyjnych oraz ukryte komponenty otrzymane po nieujemnej faktoryzacji algorytmem opartym na dywergencji Fermiego-Diraca. Tabela 1 zawiera wyniki predykcji uzyskane dla poszczególnych modeli oraz wyniki po usunięciu poszczególnych komponentów ukrytych.

Rysunek 1. Wyniki modeli predykcyjnych a) oraz ukryte komponenty b)



Źródło: opracowanie własne

Tabela 1. Przykładowy zestaw danych

MSE 10^{-3}	MLP12	MLP18	MLP24	MLP27
Modele podstawowe	8,164	**7,782**	7,891	8,097
Uśnięty 1 komponent	97,586	7,507	7,991	**7,404**
Uśnięty 2 komponent	8,164	98,588	7,891	8,097
Uśnięty 3 komponent	8,164	7,516	97,334	7,389
Uśnięty 4 komponent	7,688	7,782	7,575	99,588

Źródło: opracowanie własne

Najlepsze wyniki predykcji uzyskano, po usunięciu 2 komponentu ukrytego, uzyskano zmniejszeni błędu MSE, czyli poprawę jakości predykcji, o ok.4.8%. Uzyskane wyniki należy traktować jako poglądowe potwierdzenie poprawności działania przedstawionej koncepcji. Natomiast rzeczywisty stopień poprawy, możliwy do uzyskania może się zmienić w przypadku zwiększenia liczby modeli predykcyjnych oraz testu eliminacji nie tylko pojedynczych komponentów ale także ich kombinacji.

Jednocześnie należy wspomnieć o pewnych słabościach powyższej metody. W wyniku badań zaobserwowano stosunkowo dużą (np. w porównania z metodą ICA) niestabilność działania uzyskanego algorytmu. Choć od strony formalne jego działanie jest stosunkowo skuteczne, tj. uzyskujemy nieujemną faktoryzację macierzy, zaś możliwa do uzyskania poprawa wyników stosunkowo znaczącą, to jednocześnie jednak zaobserwowano znaczą wrażliwość na warunki początkowe.

PODSUMOWANIE

W niniejszej pracy zaprezentowano nowy algorytm nieujemnej faktoryzacji macierzy oparty na dywergencji Fermiego-Diraca. Został on następnie zastosowany w systemie poprawy wyników predykcji w podejściu wielomodelowym, co można jednocześnie zaklasyfikować do kategorii agregacji modeli. Przeprowadzone eksperymenty potwierdziły zasadność przyjętych koncepcji. Jednocześnie jednak ujawniły pewne trudności tego podejścia. Niestabilność i wrażliwość na warunki początkowe uzyskanego algorytmu, skłaniają do refleksji nad jakością algorytmu w jego obecnej postaci w praktycznych zastosowaniach. Poprawę efektywności można upatrywać m.in. w odmiennych sposobach optymalizacji funkcji celu.

Jednak głównym celem obecnej pracy było uzyskanie nowego algorytmu NMF i potwierdzenie skuteczności jego działania w aspekcie formalnym, czyli wykazanie, że podstawowe zadanie nieujemnej faktoryzacji jest wykonywalne, i może być zastosowane w realnych zadaniach. Dostosowanie wyprowadzonego algorytmu do praktycznego działania w ramach stabilnego systemu poprawy predykcji jest zasadniczo odrębnym zadaniem, adresowanym do możliwych kolejnych badań.

BIBLIOGRAFIA

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.

Berry M., Browne M., Langville A., Pauca P., Plemmons R. (2007) Algorithms and Applications for Approximate Nonnegative Matrix Factorization. Computational Statistics & Data Analysis, 52(1), 155-173.

Cichocki A., Zdunek R., Phan A.-H., Amari S. (2009) Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis. John Wiley.

Csiszar I. (1974) Information Measures: A Critical Survey. Transactions of the 7th Prague Conference, 83-86.

Comon P., Jutten Ch. (2010) Handbook of Blind Source Separation: Independent Component Analysis and Applications. Academic Press.

Dhillon I. S., Sra S. (2005) Generalized Nonnegative Matrix Approximations with Bregman Divergences. Advances in Neural Information Processing Systems, 283-290.

Dhillon I. S., Tropp J. A. (2007) Matrix Nearness Problems with Bregman Divergences. SIAM Journal on Matrix Analysis and Applications, 29(4), 1120-1146.

Gillis N. (2020) Nonnegative Matrix Factorization. SIAM, Philadelphia.

Lee D. D., Seung H. S. (1999) Learning the Parts of Objects by Non-Negative Matrix Factorization, Nature, 401, 788-791.

Szupiluk R. (2014) Dekompozycje wielowymiarowe w agregacji predykcyjnych modeli Data Mining. Szkoła Główna Handlowa. Oficyna Wydawnicza.

Szupiluk R., Wojewnik P., Ząbkowski T. (2007) Smooth Component Analysis as Ensemble Method for Prediction Improvement. Lecture Notes in Computer Science, Springer Berlin-Heidelberg, 4666, 277-284.

LATENT COMPONENT IDENTIFICATION IN TIME SERIES BY NONNEGATIVE MATRIX FACTORIZATION

Abstract: In this article, we present the use of non-negative matrix factorization methods to identify latent components contained in economic time series. A new non-negative matrix factorization algorithm based on Fermi-Dirac divergence is derived. Using the obtained latent components, we eliminate noise from the time series representing the prediction results. We test the entire concept in the problem of power system load prediction.

Keywords: non-negative matrix factorization, latent components identification, blind separation, prediction

JEL classification: C02, C50