METODY ILOŚCIOWE W BADANIACH EKONOMICZNYCH
Tom XXIV/2, 2023, s. 79 – 91

ANALIZA SKŁADOWYCH NIEZALEŻNYCH NA RYNKACH FINANSOWYCH – MOŻLIWOŚCI I OGRANICZENIA

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 problematykę zastosowań analizy składowych niezależnych na rynku finansowym w kontekście algorytmicznej i statystycznej charakterystyki tej metody. Wskażemy, że specyfika uczenia maszynowego oraz problemowy kontekst ślepej separacji, w jakim osadzona jest analiza składowych niezależnych, ma zasadniczy wpływ na możliwości i ograniczenia interpretacji statystycznej uzyskanych wyników. Przedstawimy także propozycję algorytmu, bardziej zorientowanego na spełnienie warunku niezależności, niż algorytmy ukierunkowane na separację.

Słowa kluczowe: analiza składowych niezależnych, ślepa separacja, uczenie maszynowe, analiza finansowa

JEL classification: C02, C50

WPROWADZENIE

Analiza składowych niezależnych (ICA – ang. Independent Component Analysis) należy do popularnych metod analizy wielowymiarowej [Salazar i Vergara 2018]. Jej zasadniczym celem jest wydzielenie, ze zbioru obserwowanych sygnałów, składowych/komponentów statystycznie niezależnych [Hyvarinen et al. 2001]. Znaczenie tej metody wiąże się w dużej mierze z zastosowaniami w problemie ślepej separacji sygnałów (BSS – ang. blind singal/source separation), dla którego ICA stanowiła pierwsze efektywne rozwiązanie i nadal jest główną techniką. Z tego względu, w początkowym etapie rozwoju zagadnienia te były zasadniczo z sobą utożsamiane [Jutten, Hérault 1991; Bell, Sejnowski 1995]. Obecnie ICA stanowi samodzielne podejście w analizie danych, które może być rozważane w kategoriach transformacji, reprezentacji i dekompozycji danych.

https://doi.org/10.22630/MIBE.2023.24.2.6

Jedną z charakterystycznych cech ICA jest jej zasadniczo algorytmiczny charakter, w związku z czym, zaliczana jest do metod uczenia maszynowego, może być także intepretowana i rozważana w kategoriach sieci neuronowych. Związki te były szczególnie silne w początkowym okresie rozwoju ICA [Hyvarinen, Oja 1996], jednak są obecne także w aktualnych rozwiązaniach [Haykin 2009].

Skuteczność rozwiązań ICA w obszarze ślepej separacji, niekiedy jedynych możliwych, zaowocowała licznymi zastosowaniami w obszarze medycyny, telekomunikacji czy akustyki [Comon, Jutten 2010]. Na tym tle, jej obecność w obszarze ekonomii, w szczególności rynków finansowych, generujących dane w kategoriach big data, wygląda stosunkowo skromnie, choć oczywiście jest obecne [Szupiluk 2014; Lassance et al. 2022]. Zastosowania ICA w obszarach ryzyka, analizy portfelowej, estymacji trendów, wydają się nadal raczej badawczymi propozycjami, niż ogólnie przyjętymi metodami rozwiązań. Jedną z przyczyn tego stanu rzeczy może stanowić algorytmiczna i statystyczna specyfika rozwiązań ICA.

W niniejszym artykule przeprowadzimy dyskusję właściwości analizy składowych niezależnych oraz możliwości i ograniczenia zastosowań ICA na rynkach finansowych. W tym celu prześledzimy proces wyprowadzenia jednego z najbardziej popularnych algorytmów jakim jest ICA Natural Gradient, a następnie odniesiemy się do szerszego kontekstu innych algorytmów ICA.

ANALIZA SKŁADOWYCH NIEZALEŻNYCH

Analiza składowych niezależnych powstała w związku z pracami nad problemem ślepej separacji (BSS) i w takim kontekście jest zwykle przedstawiana [Comon 1994; Cardoso1998]. W problemie BSS, staramy się odtworzyć sygnały źródłowe zmieszane w pewnym systemie. Zarówno system jak i sygnały źródłowe są nieznane, zaś identyfikacja odbywa się tylko na podstawie zmieszanych danych, co wiąże się z założeniem pewnego modelu generującego. Typowym modelem generującym jest liniowy model statyczny postaci

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k), \tag{1}$$

gdzie: $\mathbf{A} \in R^{N \times M}$, $\mathbf{x}(k) = [x_1(k),...,x_N(k)]^T$ jest wektorem sygnałów obserwowanych, zaś $\mathbf{s}(k) = [s_1(k),...,s_M(k)]^T$ źródłowych, k oznacza indeks czasu lub numer obserwacji. W celu znalezienia $\mathbf{x}$, poszukuje się takiego systemu separującego, reprezentowanego przez macierz $\mathbf{W}$, że dla sygnałów separowanych $\mathbf{y}(k) = [y_1(k),...,y_M(k)]^T$ zachodzi

$$\mathbf{y}(k) = \mathbf{W}\mathbf{x}(k) = \mathbf{W}\mathbf{A}\mathbf{s}(k) = \mathbf{P}\mathbf{D}\mathbf{s}(k), \tag{2}$$

gdzie: **P** - macierz permutacji określająca kolejność estymowanych sygnałów, **D** - diagonalna macierz skalująca. Równanie (2) oznacza, że estymowane sygnały **y** mogą być przeskalowane oraz w innej kolejności uporządkowane niż **s**. Jest to nieunikniona konsekwencja faktu, iż przyjmując *N*=*M*, dla dowolnej odwracalnej macierzy $\mathbf{E} \in R^{N \times N}$ zachodzi $\boldsymbol{x} = \boldsymbol{AEE}^{-1}\boldsymbol{s}(k)$. W ramach ICA przyjmuje się zwykle następujące założenia [Cardoso 1999; Hyvarinen et al. 2001]:

a) kolumny macierzy **A** są liniowo niezależne,
b) liczba komponentów (sygnałów) obserwowanych nie może być mniejsza od liczby niezależnych komponentów źródłowych, $M \le N$,
c) sygnały modelowane są jako zmienne losowe lub stochastyczne białe szumy, przy czym co najwyżej jeden z sygnałów ma rozkład gaussowski.

Celem ICA jest wydzielenie z wielowymiarowych danych obserwowanych sygnałów (komponentów, składowych) niezależnych, co oznacza, że spełniona będzie równość

$$p_1(y_1)p_2(y_2)...p_M(y_M) = p_{1...M}(y_1, y_2, ..., y_M)\,, \tag{3}$$

gdzie $p_i(y_i)$ oznacza funkcję gęstości prawdopodobieństwa sygnału y_i (rozkład brzegowy), zaś $p_{1...M}(y_1, y_2, ..., y_M)$ jest rozkładem łącznym $y_1, y_2, ..., y_M$.

Rozwiązaniem tak postawionego zadania można dokonać na wiele sposób, jednym z nich jest popularny algorytm ICA Natural Gradient. Punktem wyjścia jest tu zdefiniowanie funkcji celu w postaci wzajemnej informacji, której minimalizacja ze względu na macierz **W** ma prowadzić do spełnienia (3) [Amari et al. 1999]. Mamy więc

$$\mathbf{W}_{opt} = \min_{\mathbf{W}} I(\mathbf{y}) = \min_{\mathbf{W}} \int_{-\infty}^{+\infty} p(y_1, ..., y_M) \log \frac{p(y_1, ..., y_M)}{\prod_{i}^{M} p_i(y_i)} dy_1 ... dy_M\,. \tag{4}$$

Przekształcając $I(\boldsymbol{x})$ otrzymuje się

$$I(\mathbf{x}) = -H(\mathbf{x}) - \log(|\det(\mathbf{W})|) - \sum_{i=1}^{M} H_i(y_i)\ , \tag{5}$$

gdzie $H_i(y_i)$ oznacza entropie poszczególnych sygnałów estymowanych zaś $H(\mathbf{x})$ oznacza entropię zmiennej **x,** która nie zależy od **W,** więc czynnik ten może zostać pominięty. Wartość bezwzględną (5) zastępuje się wygodniejszą analitycznie formą $\mathbf{WW}^{\mathrm{T}}$. Daje to funkcję celu postaci

$$L(\mathbf{W}) = -\log(\det(\mathbf{WW}^T)) - \sum_{i=1}^{M} H_i(y_i)\ . \tag{6}$$

Minimalizacja funkcji (6) metodą Gradientu Naturalnego prowadzi do następującego algorytmu [Amari 1998; Cichocki, Amari 2002]:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k)\left[\mathbf{I} - E\left\{\mathbf{f}(\mathbf{y})\mathbf{y}^T\right\}\right]\mathbf{W}(k) \ , \tag{7}$$

gdzie E{.} oznacza wartość oczekiwaną, zaś nieliniowa kowariancja $\mathbf{R}_{fy} = E\left\{\mathbf{f}(\mathbf{y})\mathbf{y}^T\right\}$ zawiera wektor funkcji nieliniowych $\mathbf{f}(\mathbf{y}) = [f_1(y_1),...,f_M(y_M)]^T$ o elementach

$$f_i(y_i) = -\frac{\partial \log(p_i(y_i))}{\partial y_i} . \tag{8}$$

Optymalna postać nieliniowości wyrażona przez (8) zakłada znajomość rozkładów prawdopodobieństwa poszczególnych sygnałów, a jest to *a priori* nieznane. Z tego względu, stosowane są estymowane modele parametryczne rozkładów, modele adaptacyjne lub z góry określa się postać nieliniowości aby eksplorować statystyki określonego rzędu [Karvanen et al. 2002]. Stosowane są także reguły heurystyczne, np. można także zauważyć, że formuła (8) prowadzi do nieliniowości szybciej rosnących od funkcji liniowej dla rozkładów bardziej stromych niż gaussowski oraz wolniej rosnących dla rozkładów wolniej rosnących, Stąd dobór nieliniowości może odbywać się zgodnie z regułą [Cichocki et al. 1997]

$$f_i(y_i) = \begin{cases} y_i^3 & dla \ \ \mathrm{kurt_N}(y_i) > 0, \\ \tanh(y_i) & dla \ \ \mathrm{kurt_N}(y_i) < 0, \\ y_i & dla \ \ \mathrm{kurt_N}(y_i) = 0. \end{cases} \tag{9}$$

gdzie $\mathrm{kurt_N}(y) = E\{y^4\}/E^2\{y^2\} - 3$ oznacza znormalizowaną kurtozę. Kolejną kwestią jest występujący w (7) współczynnik uczenia. Jego dobór jest kwestią kluczową dla efektywnego działania algorytmu. Typowe postacie zakładają wartości zmienne w czasie, zwykle zanikające do zera. Choć możliwe są także złożone systemy w postaci filtrów górno- i dolnoprzepustowych [Cichocki, Amari 2002].

Można zauważyć, ze występująca w (7) nieliniowa macierz kowariancji $\boldsymbol{R}_{fy} = E\{\boldsymbol{f}(\boldsymbol{y})\boldsymbol{y}^T\}$ jest szczególnym przypadkiem bardziej ogólnej postaci $\boldsymbol{R}_{fg} = E\{\boldsymbol{f}(\boldsymbol{y})\boldsymbol{g}(\boldsymbol{y})^T\}$, której zastosowanie prowadzi do algorytmu [Cichocki, Unbehauen 1996]

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k)\left[\mathbf{I} - E\left\{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})\right\}\right]\mathbf{W}(k) \ , \tag{10}$$

gdzie wektor funkcji nieliniowych $\mathbf{g}(\mathbf{y}) = [g_1(y_1),\ldots,g_M(y_M)]^T$ składa się zwykle z funkcji $g(y_i)$ dualnych do $f(y_i)$. Algorytm (10) można potraktować jako rozwinięcie całej koncepcji, choć powstał on w sposób heurystyczny wcześniej, niż wyprowadzony z teorii informacji algorytm Natural Gradient.

Podobne struktury widzimy także w algorytmach wyprowadzanych z innych zasad, w tym algorytm EASI oparty na funkcji największej wiarygodności [Cardoso, Laheld 1996]

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k)\left[\mathbf{I} - E\left\{\mathbf{y}\mathbf{y}^T\right\} - E\left\{\mathbf{f}(\mathbf{y})\mathbf{y}^T\right\} + E\left\{\mathbf{y}\mathbf{f}^T(\mathbf{y})\right\}\right]\mathbf{W}(k). \tag{11}$$

lub wyprowadzony z maksymalizacji negentropii jako miary niegaussowości (a to jako miary niezależności) algorytm FastICA Hyvarinen, Oja 2000]

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k)\mathbf{R}\left[E\left\{\mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k)\right\} - diag\left(E\{f(y_i)y_i\}\right)\right]\mathbf{W}(k) \ , \tag{12}$$

gdzie $\mathbf{R} = diag\left(1/E\{f(y_i)y_i\} - E\{f'(y_i)\}\right)$ oznacza macierz diagonalną.

Można zauważyć, że mimo odmiennych kryteriów początkowych, uzyskane postacie algorytmów ICA mają stosunkowo podobną strukturę, która może być poddawana dalszym modyfikacjom. Jednym z nich jest przyjęcie wersji on-line dla $\boldsymbol{R}_{fy} = E\{\boldsymbol{f}(\boldsymbol{y})\boldsymbol{y}^T\}$.

PROBLEMATYKA ZASTOSOWAŃ ICA W EKONOMII I FINANSACH

Zastosowanie ICA w ekonomii, finansach lub zarzadzaniu wiąże się zwykle z przyjęciem podejścia separacyjnego lub analitycznego, choć oczywiście możliwe jest także łączenie tych optyk. W odróżnieniu jednak, od takich metod wielowymiarowych jak analiza składowych głównych lub analiza czynnikowa, na otrzymane wyniki i interpretacji istotny wpływ ma algorytmiczny charakter analizy składowych niezależnych.

Traktując ICA jako metodę rozwiązania problemu ślepej separacji, mamy do czynienia z przypadkiem zasadniczo instrumentalnym. Dąży się tu zwykle do osiągnięcia konkretnego wymiernego rezultatu, jakim są źródłowe komponenty użyteczne lub intepretowane jako praktycznie fizyczne sygnały. Przykładem może tu być wydzielenie z wyników predykcji finansowych szumów i zakłóceń [Szupiluk 2014]. W takim podejściu kryterium przesądzającym o zasadności użycia ICA jest poprawa prognozy. Wszelkie dodatkowe rozważania o matematycznej charakterystyce otrzymanych komponentów, stopniu ich niezależności, spełnieniu założeń związanych z danymi, systemem je generującym czy ogólnie metodą ICA mogą być nieistotne lub zbędne. Decyduje skuteczność - w szczególności w systemach predykcyjnych opartych na uczeniu maszynowym. Osiągnięcie pozytywnych rezultatów wiąże się tu zasadniczo z testem różnych algorytmów ICA, w poszukiwaniu najlepszego w danym przypadku bez względu na jego specyfikę.

Należy jednak zaznaczyć, że separacja szeregów finansowych jak indeksy giełdowe, kursy akcji, ceny walut lub towarów - w poszukiwaniu składowych związanych z jednoznacznym zjawiskiem rynkowym (sygnał źródłowy) jest praktycznie niewykonalna, ze względu na trudności ze spełnieniem najbardziej podstawowych założeń ICA/BSS, jak mniejsza sygnałów źródłowych niż obserwowanych (czyli poddawanych separacji). Sytuacja, iż wybrany do analizy zestaw instrumentów finansowych jest efektem odziaływania nie większej liczby fundamentalnych czynników niż liczba tych instrumentów, wydaje się sytuacją wysoce nieprawdopodobną, o ile w ogóle możliwą.

Także założenie o liniowym modelu generującym, jest w ogólnym przypadku mało realistyczne. Wiele informacji dociera i oddziałuje na określone instrumenty z opóźnieniami lub impulsowo (nieliniowo), co przemawia za uwzględniającymi je modelami dynamicznymi i nieliniowymi. Przyjęcie dynamicznego modelu mieszającego, oznacza konieczność zastosowania także dynamicznego separującego (estymującego składowe niezależne), co skutkuje, że za akceptowane rozwiązania uznane będą także przefiltrowane sygnały źródłowe. Z kolei przyjęcie nieliniowego modelu mieszającego oznacza w ogólnym przypadku brak jednoznaczności uzyskiwanych wyników (możliwe są różne zestawy składowych niezależnych). W przypadku nieliniowego modelu dynamicznego te niedogodności sumują się. W praktyce znacząco ogranicza to pole separacyjnych zastosowań modeli nieliniowych oraz dynamicznych.

Także uwzględnienie szumu addytywnego czyli przyjęcie modelu $\mathbf{x}=\mathbf{A}\mathbf{s}+\mathbf{v}$, gdzie $\mathbf{v}$ oznacza szum addytywny, zasadniczo uniemożliwia w większości skuteczną separację. Wiąże się to z faktem, iż w przypadku ogólnym nie istnieje tu liniowa transformacja pozwalająca na odwrócenie procesu mieszania i odtworzenie sygnałów źródłowych.

Stosunkowo odmienna sytuacja występuje w przypadku kiedy zastosowanie ICA ma mieć charakter analityczny. Taką sytuację mamy w przypadku estymacji trendów rynkowych, jako bezpośrednich wyników działania metody ICA. Zasadniczą kwestią jest tu charakterystyka otrzymanych wyników.

ICA JAKO REPREZENTACJA ANALITYCZNA

Przyjęcie metody ICA jako metody prowadzącej do nowej reprezentacji analitycznej uwalania nas od interpretacji wyników w kategoriach separacji rynkowych „fizycznych źródeł". Pojawiają się jednak kwestie związane z odpowiedzią na pytanie, jaką charakterystykę matematyczną posiadają komponenty otrzymane w wyniku działania algorytmów ICA. Oznacza to analizę poszczególnych etapów ICA prowadzących do ich powstania.

W odróżnieniu od separacji fizycznych sygnałów źródłowych, w przypadku poszukiwania interesującej reprezentacji analitycznej zastosowania modeli dynamicznych na gruncie finansowym wydają się całkiem możliwe. W takim przypadku, przyjęcie/założenie określonego modelu nie przesądza natury zjawiska lub zależności między danymi służy wyłącznie do opracowania algorytmu numerycznego. Nie wpływa także zasadniczo na ideę wyprowadzenia algorytmu ICA. Nie jest także problemem w zastosowaniach finansowych założenie, iż separowane sygnały mają być białymi szumami, oznacza to wyłącznie, że w ICA nie jest brana pod uwagę struktura czasowa analizowanych danych, natomiast w niczym to nie przeszkadza takie dane wykorzystywać. Jak widać znaczenie poszczególnych założeń w ICA może mieć różną wagę, co jest jedną z cech charakterystycznych dla metod uczenia maszynowego.

Natomiast kluczowym elementem staje się istota algorytmów ICA i ich interpretacji, czyli pojęcie niezależności. Choć spektrum kryteriów oceny niezależności statystycznej sygnałów, jest stosunkowo szerokie i obejmuje m.in. takie podejścia jak: minimalizacja wzajemnej informacji, maksymalizacja entropii, minimalizacja kurtozy, maksymalizacja funkcji największej wiarygodności, w większości przypadków, przy nieco głębszym wglądzie, widzimy, że prowadzą one do, stosunkowo podobnych, lub zasadniczo tych samych formuł [Cichocki, Amari 2002; Comon, Jutten 2010]. Wiąże to się z reguły, z aproksymacją pojęć wykorzystujących rozkłady (funkcje gęstości prawdopodobieństwa) „wygodnymi” statystykami. Najczęściej są to statystyki czwartego rzędu wyrażane różnymi sposobami m.in. momentami, *L*-momentami lub kumulatami. Przykładem może być aproksymacja funkcji gęstości prawdopodobieństwa szeregami Edegwortha lub Gram-Chalier, gdzie przyjmuje się, że ograniczenie do statystyk czwartego rzędu w większości przypadków daje wystarczającą dokładność przybliżenia [Hyvarinen et al. 2001].

Jednak optymalizacja funkcji celu może wiązać się nie tylko z formalnymi przekształceniami aproksymującymi, ale także krokami heurystycznymi. W przypadku algorytmu Natural Gradient, jako kryterium oceny statystycznej niezależności przyjmuje się wzajemną informację, która jest optymalizowana jedynie częściowo. Jak wskazano wyżej, eliminowany jest składnik związany z entropią sygnałów źródłowych, ponieważ system ICA nie ma wpływu na tą część funkcji celu. Kolejnym zabiegiem, jest eliminacja wartości bezwzględnej z funkcji celu. W rezultacie, oddalamy się, w stopniu najczęściej trudnym do określenia, od spełnienia warunku niezależności.

Należy zaznaczyć, że algorytmy (7)-(12) nie stanowią wersji ostatecznych lub zamkniętych. Na gruncie praktycznego problemu BSS dozwolona jest w zasadzie dowolna modyfikacja prowadząca do większej efektywności separacji, zaś pojęcie niezależności staje się pewnym użytecznym kryterium a nie głównym celem działania algorytmu. Przykładowo poprawa własności numerycznych algorytmu (11) prowadzi do Normalized EASI [Cardoso, Laheld 1996]

$$\mathbf{W}(k+1)=\left(\mathbf{I}-\eta\frac{\mathbf{y}\mathbf{y}^T-\mathbf{I}}{1+\eta\mathbf{y}^T\mathbf{y}}-\frac{\eta}{2}\frac{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})-\mathbf{g}(\mathbf{y})\mathbf{f}^T(\mathbf{y})}{1+\eta\left|\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})\right|}\right)\mathbf{W}(k). \tag{13}$$

Kolejnym elementem, kluczowym dla efektywnego działania algorytmu ICA (jako metody separacji) jest współczynnik uczenia, a jego prosta postać (9) może być rozwijana do [Douglas, Cichocki 1997]

$$\mu(k)=\mu(k-1)-\rho(k)\left[\frac{m-1}{1+\mu(k-1)}+\frac{1-\mathbf{y}^{\mathrm{T}}(k-1)\mathbf{f}(\mathbf{y}(k-1))}{1+\mu(k-1)\left[1-\mathbf{y}^{\mathrm{T}}(k-1)\mathbf{f}(\mathbf{y}(k-1))\right]}-\right.$$
$$\left.-\mathbf{y}^{\mathrm{T}}(k)\mathbf{f}(\mathbf{y}(k))+\mathbf{f}^{\mathrm{T}}(\mathbf{y}(k))\mathbf{f}(\mathbf{y}(k-1))\mathbf{y}^{\mathrm{T}}(k-1)\mathbf{y}(k)\right]. \quad (14)$$

Z zanikającym do zera współczynnikiem uczenia wiąże się jeszcze kwestia punktu stopu. Algorytm (7) można uznać za nauczony jeżeli $\mathbf{W}(k+1)=\mathbf{W}(k)$ co oczekujemy, że zachodzi jeżeli $\mathbf{R}_{fy}=E\{\mathbf{f}(\mathbf{y})\mathbf{y}^{T}\}=\mathbf{I}$, a co w interpretacji statystycznej oznaczałoby nieliniową dekorelację. Jednak w praktyce warunek $\mathbf{W}(k+1)=\mathbf{W}(k)$ wiąże się raczej z wymuszonym przyjęciem $\mu(k)=0$.

Wszystkie te zabiegi są dlatego uzasadnione, że zasadniczym zadaniem ICA jest estymowanie nie tyle składowych niezależnych, co separacja rzeczywistych sygnałów. Efektem końcowym jest rekonstrukcja źródeł, a nie test niezależności uzyskanych komponentów. Zaś separacja może się dokonać nie tylko w oparciu o regułę niezależności, ale także na bazie innych pojęć jak gładkość, nieujemność czy niestacjonarność.

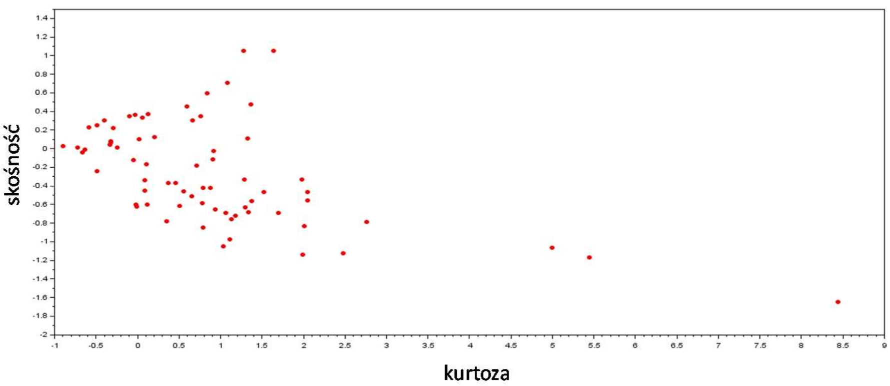
W efekcie jeżeli weźmiemy pod uwagę przekształcenia i redukcje początkowej funkcji celu, wraz z modyfikowanymi heurystycznie algorytmami uczenia, w których współczynniki uczenia oraz dobór funkcji nieliniowych realizowany jest w złożonych systemach, to pojawia się pytanie o naturę otrzymanych komponentów. Czy eksplorujemy wyłącznie pojęcie niezależności i w jakim stopniu, czy znaczenie mają już także inne charakterystyki prowadzące do separacji? Wydaje się, że nie sposób jednoznacznie odpowiedzieć na to pytanie, choć elementem, który wyraźnie występuje w finalnych, implementacyjnych wersjach algorytmów, są statystyki czwartego rzędu, a w szczególności kurtoza.

KURTOZA W ICA A RYNKI FINANSOWE

Pojęciem szczególnie często pojawiającym się w kontekście algorytmów ICA jest kurtoza, która stanowi nie tylko funkcję aproksymującą entropię lub negentropię, ale jest także samodzielną podstawą wyprowadzania algorytmów [Douglas, Kung 1998; Girolami, Fyfe 1996]. Kurtoza stosowana jest także w doborze nieliniowości. Przyjmując, że w wielu przypadkach algorytmy ICA ograniczają się do eksploracji kurtozy lub ogólnej statystyki czwartego rzędu, otrzymujemy narzędzie analityczne stosunkowo wygodne w interpretacji zarówno statystycznej jak i ekonomicznej oraz łatwe w zakresie weryfikacji wyników. Nadal jednak należy mieć na uwadze, że ICA jest metodą uczenia maszynowego, gdzie jej instrumentalna natura powoduje, że poszczególne pojęcia bardzo szybko mogą być modyfikowane i rozwijanie przy zachowaniu tych samych etykiet. W efekcie, w obszarze ICA mamy do czynienia zarówno z różnymi postaciami samej kurtozy jak i licznymi jej modyfikacjami i generalizacjami [Lambert 1996; Miettinen et al. 2015].

W niektórych algorytmach analizy składowych niezależnych owa natura statystyk czwartego rzędu jest wprost sygnalizowana w nazwie jak w algorytmie FOBI-Fourth Order Blind Identification [Cardoso 1999]. Jednak nawet ograniczenie „niezależności” do statystyk czwartego rzędu, z rynkowego punktu widzenia, może być problematyczne, ze względu na rzeczywistą charakterystyką danych finansowych. Po pierwsze, jest kwestią otwartą czy dany szereg czasowy ma w ogóle wyższe momenty [Peters 1996]. Przyjęcie, że duża część danych na rynku finansowym ma rozkłady nieskończenie podzielne oznaczałoby zasadniczą barierę w zastosowaniu ICA zgodnie z jej założeniami. Po drugie, jeżeli jednak przyjmiemy, że statystki te istnieją lub ograniczymy się do empirycznych charakterystyk z próby, to z reguły otrzymamy wykres wartości kurtozy prezentowany na rysunku 1.

Rysunek 1. Kurtoza i skośność obserwowana dla różnych podokresów indeksu WIG 20 z lat 2011-2015



Źródło: opracowanie własne

Rysunek 1 przedstawia wartości znormalizowanej kurtozy oraz skośności dla losowo wybranych przedziałów danych reprezentujących stopy zwrotu, z indeksu WIG 20 z okresu 2011-2015, a więc okresu gospodarczo stosunkowo stabilnego, widzimy jednak, że obydwie statystyki są dalekie od stabilności. W takim wypadku, trudno oczekiwać stabilności od algorytmów je stosujących. Brak takiej stabilności w przypadku separacji sygnałów technicznych najczęściej rozwiązuje się wersję algorytmów on-line, jednak w przypadku takiego podejścia dla danych finansowych, pojawia się problem ich analitycznej interpretacji gdy estymujemy sygnał ze zmiennej w czasie kurtozy.

NOWE POSTACIE ALGORYTMU ICA

W świetle poważnych rozważań wydaje się, że w przypadku zastosowań typowych algorytmów ICA dla danych finansowych stosunkowo trudno pogodzić skuteczność separacji z wymogiem i interpretacją ich niezależności, dominuje optyka separacyjna.

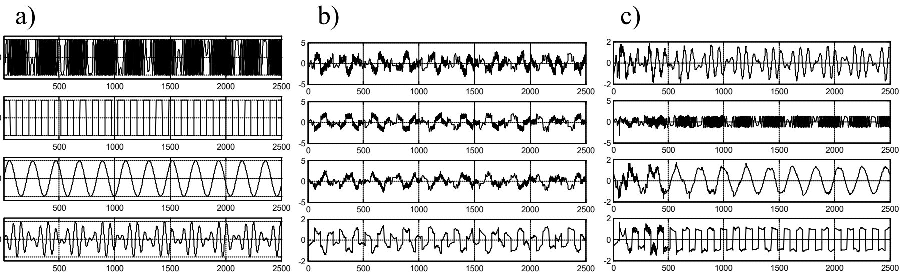
Natomiast, możliwa jest modyfikacja algorytmów ICA wychodząca bardziej naprzeciw warunkowi niezależności statystycznej, przy zachowaniu pewnych właściwości separacyjnych. Obecnie zaproponujemy metodę, która zakłada wprost eksplorację statystyk wyższych rzędów niż czwarty.

$$\begin{aligned}\mathbf{W}(k+1) &= \mathbf{W}(k) + \left(\mu_1(k)\left[\mathbf{I} - E\{\mathbf{y}^{\cdot 3}\mathbf{y}^T\}\right] + \mu_2(k)\left[\mathbf{I} - E\{\mathbf{y}^{\cdot 4}\mathbf{y}^T\}\right]\right)\mathbf{W}(k) + \\ &+ \left(\mu_1(k)\left[\mathbf{I} - E\{\mathbf{y}^{\cdot 5}\mathbf{y}^T\}\right] + \mu_2(k)\left[\mathbf{I} - E\{\mathbf{y}^{\cdot 7}\mathbf{y}^T\}\right]\right)\mathbf{W}(k).\end{aligned} \tag{15}$$

gdzie $\boldsymbol{y}^{\cdot k} = \left[y_1^k, y_2^k \ldots, y_M^k\right]^T$.

W algorytmie (15) wykorzystaliśmy podstawową ideę struktury algorytmów ICA rozszerzając ją na nieliniowe dekorelacje względem statystyk wyższych rzędów niż czwarty. Nawiązuje to do bezpośredniej idei niezależności statystycznej, wg której nie powinny w takim przypadku występować korelacje nieliniowe dowolnego rzędu. Jednak choć podstawowa idea (15) wydaje się stosunkowo prosta to występujące oddzielne współczynniki uczenia dla poszczególnych $\mathbf{R}_{y^k y} E\{\mathbf{y}^{\cdot k}\mathbf{y}^T\}$, czynią to podejście wymagającym implementacyjnie. Aby potwierdzić skuteczność działania (15), wykonaliśmy eksperyment z separacją stosunkowo prostych sygnałów, łatwych w ocenie jakości separacji, co prezentuje rysunek 2.

Rysunek 2. Sygnały a) źródłowe, b) zmieszane, b) separowane



Źródło: opracowanie własne

Choć dokładniejsze badania tego algorytmu wychodzą poza zakres niniejszej pracy, to już pobieżne badania pozwalają na sformułowanie paru uwag. Algorytm dla założonego rzędu statystyk działa poprawnie, tzn. następuje zadekorelowanie wzgl. tych rzędów. Jednak zwiększanie rzędu eksplorowanych statystyk nie przekłada się na automatyczną poprawę jakości separacji.

PODSUMOWANIE

Analiza składowych niezależnych jest metodą, która powstała w kontekście problemu ślepej separacji stanowiąc główny sposób jego rozwiązania. Z punktu widzenia implementacyjnego, należy do metod nienadzorowanego uczenia maszynowego, przy czym istnieje wiele algorytmów ICA o różnych właściwościach. Zawarte w nazwie pojęcie niezależności odnosi się do idei i kryteriów jakie zastosowano aby ową separację osiągnąć, nie stanowi jednak w istocie celu nadrzędnego ICA, jest nim bowiem efektywna separacja sygnałów źródłowych. W rzeczywistości, korzystając z gotowych procedur, często trudno stwierdzić jakimi w istocie charakterystykami statystycznymi charakteryzują się uzyskane komponenty. Z dużą dozą pewności są one nieliniowo zdekorelowane, w większości przypadków z użyciem statystki czwartego rzędu. Choć mogą to być także inne zależności.

W efekcie, jeżeli staramy się wśród danych finansowych identyfikować określone sygnały źródłowe, to ICA może być dobrym wyborem o ile możliwe są do spełnienia jej podstawowe założenia, co w rzeczywistych przypadkach może być trudne lub zgoła niemożliwe. W szczególności relacja między ilością źródeł a obserwacji wydaje się trudną do przekroczenia barierą,

W przypadku reprezentacji analitycznej danych ekonomicznych, interpretacja statystyczna powinna być dokonana na bazie specyfikacji konkretnego algorytmu ICA. Kwestia zastosowań praktycznych w tym kontekście, jest bowiem ściśle związana z charakterystyką uzyskanych składowych. Natomiast jako efektu działania ICA, raczej nie powinniśmy oczekiwać otrzymania komponentów w pełni niezależnych

Pewnym rozwiązaniem jest samodzielna implementacja algorytmów, gdzie możemy kontrolować zarówno kryteria ich działania i techniczne aspekty implementacji. Jak wskazano w zaproponowanej autorskiej wersji algorytmu ICA, można też w ten sposób nieco bardziej zbliżyć się do spełnienia warunku niezależności.

BIBLIOGRAFIA

Amari S. (1998) Natural Gradient Works Efficiently in Learning. Neural Computation, 10, 271-276 .

Amari S., Cichocki A. (1998) Adaptive Blind Signal Processing - Neural Network Approaches. Proceedings of the IEEE, 86(10), 2026-2048.

Amari S., Cichocki A., Yang H. (1999) Unsupervised Adaptive Filtering, chapter Blind Signal Separation and Extraction - Neural and Information Theoretic Approaches. John Wiley.

Bell A. J., Sejnowski T. J. (1995) An Information-Maximization Approach to Blind Separation and Blind Deconvolution. Neural Computation, 7(6), 1129-1159. https://doi.org/10.1162/neco.1995.7.6.1129

Cardoso J., Laheld B. (1996) Equivariant Adaptive Source Separation. IEEE Trans. Signal Processing, 44(12), 3017-3030.

Cardoso J. (1998) Blind Signal Separation: Statistical Principles. Proceedings of the IEEE, 86(10), 2009-2025.

Cardoso J. (1999) High-Order Contrasts for Independent Component Analysis. Neural Computation, 11(1), 157-192.

Cichocki A., Unbehauen R. (1996) Robust Neural Networks with On-Line Learning for Blind Identification and Blind Separation of Sources. IEEE Trans. Circuits and Systems I: Fundamentals Theory and Applications, 43(11), 894-906.

Cichocki A., Sabala I., Choi S., Orsier B., Szupiluk R. (1997) Self Adaptive Independent Component Analysis for Sub-Gaussian and Super-Gaussian Mixtures with Unknown Number of Sources and Additive Noise. In Proc. Int. Symposium on Nonlinear Theory and its Applications, NOLTA-97, 731-734.

Cichocki A., Amari S. (2002) Adaptive Blind Signal and Image Processing. John Wiley, Chichester.

Comon P. (1994) Independent Component Analysis, a New Concept ? Signal Processing, Elsevier, 36(3), 287-314.

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

Douglas S., Cichocki A. (1997) On-Line Step Size Selection for Training Adaptive Systems, IEEE Signal Processing Mag., 14(6), 45-46.

Douglas S., Kung S. (1998): Kuicnet Algorithms for Blind Deconvolution. Proc. IEEE Workshop on Neural Networks for Signal Processing, Cambridge, UK.

Girolami M., Fyfe C. (1996) Negentropy and Kurtosis as Projection Pursuit Indices Provide Generalized ICA Algorithms. Advances in Neural Information Processing Systems, NIPS'96 Workshop, Snowmaas.

Haykin S. (2009) Neural Networks and Learning Machines. Pearson Education, New Jersey.

Hyvarinen A., Oja E. (1996) Simple Neuron Models for Independent Component Analysis. Int. Journal of Neural Systems, 7(6), 671-687.

Hyvarinen A., Oja E. (2000) Independent Component Analysis: Algorithms and Applications. Neural Networks, 13(4-5), 411-430.

Hyvarinen A., Karhunen J., Oja E. (2001) Independent Component Analysis. Wiley John, New York.

Jutten C., Hérault J. (1991) Blind Separation of Sources, Part 1: An Adaptive Algorithm Based on Neuromimetic Architecture. Signal Processing, 24(1), 1-10.

Karvanen, J., Eriksson, J. & Koivunen, V. (2002) Adaptive Score Functions for Maximum Likelihood ICA. The Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology, 32, 83-92. https://doi.org/10.1023/A:1016367418778

Lambert R. (1996) Multi-Channel Blind Deconvolution: FIR Matrix Algebra and Separation of Multi-Path Mixtures. Elec. Eng. Univ. of Southern California.

Lassance N., DeMiguel V., Vrins F. (2022) Optimal Portfolio Diversification via Independent Component Analysis. Operations Research, 70(1), 55-72.

Miettinen J., Taskinen S., Nordhausen K., Oja H. (2015) Fourth Moments and Independent Component Analysis. Statistical Science, 30(3), 372-390.

Peters E. (1996) Fractal Market Analysis. John Wiley&Son.

Puuronen J., Hyvärinen A. (2015) Independent Component Analysis with an Inverse Problem Motivated Penalty Term. 2015 International Joint Conference on Neural Networks.

Salazar A., Vergara L. (2018) Independent Component Analysis (ICA): Algorithms, Applications and Ambiguities. Nova Science Publishers.

Szupiluk R. (2014) Dekompozycje wielowymiarowe w agregacji predykcyjnych modeli data mining. Szkoła Główna Handlowa, Oficyna Wydawnicza, Warszawa.

ANALYSIS OF INDEPENDENT COMPONENTS IN FINANCIAL MARKETS – POSSIBILITIES AND LIMITATIONS

Abstract: In this article, we will present the application of independent components analysis for the financial market in the context of the algorithmic and statistical characteristics of this method. We will point out that the specificity of machine learning and the problematic context of blind separation in which the analysis of independent components is embedded have a significant impact on the possibilities and limitations of the statistical interpretation of the obtained results. We will also present a novel algorithm that is more focused on meeting the independence condition than algorithms dedicated for separation.

Keywords: independent components analysis, blind signal separation, machine learning, financial analysis

JEL classification: C02, C50