

## FAKTOROVÁ ANALÝZA BINÁRNÍCH PROMĚNNÝCH POMOCÍ NEURONOVÉ SÍTĚ HOPFIELDOVA TYPU

DUŠAN HŮSEK, ALEXANDER A. FROLOV, HANA ŘEZANKOVÁ, VÁCLAV SNÁŠEL

**ABSTRACT.** The problem of binary factorization of complex patterns in recurrent Hopfield-like neural network was studied by means of computer simulation. The network ability to perform a factorization was analyzed depending on the number and sparseness of factors mixed in presented patterns. Binary factorization in sparsely encoded Hopfield-like neural network is treated as efficient statistical method and as a functional model of hippocampal CA3 field.

**Резюме.** С помощью компьютерного моделирования изучалась задача бинарной факторизации сложных образов в рекуррентной нейронной сети Хопфилда. Анализировалась способность сети осуществлять факторизацию в зависимости от числа и разреженности факторов смешанных в подаваемых образах. Бинарная факторизация в нейронной сети Хопфилдовского типа с разреженным кодированием рассматривается как эффективный статистический метод и функциональная модель поля CA3 гиппокампа.

### 1. PROBLÉM INFORMAČNÍ REDUNDANCE

Význam pojmu redundance jistě každý intuitivně chápe. V každodenním životě se setkáváme v čase a prostoru s opakujícími se regulárními informačními strukturami a ukazuje se, že tyto pravidelnosti mají pro jednotlivé jedince, a nejen to, dokonce pro jednotlivé živočišné druhy tendenci být univerzální. Tyto pravidelnosti vytvářejí společný informační prostor a reprezentují určitý lexikon vlastností, který je jejich zárukou. Univerzalita tohoto lexikonu je dána existencí současně se vyskytujícími vlastností, které tvoří invariantní shluky, v dalším textu označované jako faktory. Shlukování vlastností je dáno evolučním nárůstem komplexnosti problému za předpokladu, že entropie prostředí, které obklopuje člověka, je podstatně větší než entropie prostředí, které by bylo uspořádáno jen náhodně. Idea evoluční strukturalizace může být použita pro různou úroveň pohledu, počínaje od genetické úrovně až po makroskopickou úroveň organických látek. Pro odstranění informační redundance v mozku se k uchování informací využívá „umění“ strukturalizace, mechanismu, který se implicitně zakódoval ve struktuře mozku prostředky přirozeného výběru. Každý přicházející vzor tedy může být vyjádřen jako nelineární kombinace skutečně existujících invariantních faktorů. Budeme uvažovat pouze prostorovou redundanci, tj. redundanci způsobenou společně se vyskytujícími několika faktory v jednom vzoru. Marr se zabýval s touto problematikou ve svých článcích (1970, 1971). Nastínil zde možnost efektivního uchování redundantní informace, pokud by určité skupiny vlastností, které se vyskytují společně, byly ze vstupního vzoru vyextrahovány a přidány do databáze primitiv (zkušeností) v mozku jako nová entita (koncept). S využitím

---

2000 *Mathematics Subject Classification.* 68T05.

*Klíčová slova.* Neuronové sítě, Hopfieldova síť, faktorová analýza, binární faktorizace, řídké kódování, model paměti, hippocampus.

Tato práce byla vypracována za podpory grantu GA ČR 201/01/1192 a 201/00/1031.

tohoto slovníku pak mozek později interpretuje a zaznamenává své zkušenosti. Marr také uvádí, že realizace úlohy pamatování je blízká procesu shlukování. Problém výše zmíněného efektivního uchování informací by mohl být realizován pomocí vhodné organizace neuronové sítě a speciálních přirozených pravidel a dynamiky učení, o čemž bude pojednáno později.

## 2. STATISTICKÉ HLEDISKO

Jak bylo ukázáno v předchozí kapitole, odstranění informační redundance je hlavním zdrojem informační komprese a tudíž je i nutnou podmínkou pro efektivní uchování informací. Eliminace redundance též umožňuje odhalovat skrytou strukturu, která není zřejmá ze zdrojových dat. Tento často se vyskytující se problém je konvenčně řešen pomocí různých statistických algoritmů. Odhalování skryté struktury spočívá ve zjištění faktorů (shluků) a v přiřazení proměnných k těmto faktorům. Jde o problém shlukování proměnných.

Základním algoritmem pro eliminaci redundance je *analýza hlavních komponent* známá pod zkratkou PCA (principal component analysis), která umožňuje extrahovat hlavní vztahy ve vícerozměrných datech. Obvyklým způsobem nalezení hlavních komponent v datové množině je výpočet vlastních čísel korelační matice. Pro kategoriální data může být použita analýza hlavních komponent s optimálním škálováním. Ve srovnání s lineární PCA umožňuje kategoriální (nelineární) PCA, aby proměnné existovaly na různých škálách měření. Vážená forma PCA je označována jako *korespondenční analýza*. Je vhodná pro kategoriální data, která mohou být též zadána pomocí četností. Pro řešení problémů musí být použita vícenásobná korespondenční analýza, která je někdy označována jako *analýza homogeneity*.

Nejčastěji používanou metodou pro shlukování kvantitativních proměnných je *faktorová analýza*. Tato technika předpokládá lineární vztahy mezi proměnnými. Příslušnost proměnné k určitému faktoru není pevná a mohou být nalezeny překrývající se shluky. Speciální technikou pro shlukování (jak proměnných tak i statistických jednotek) je *shluková analýza*. Při tomto postupu můžeme rozlišit různé algoritmy. Metody k-průměrů se snaží identifikovat relativně homogenní skupiny založené na vybraných charakteristikách. Jiné metody jsou označovány jako hierarchické. Jsou založeny na maticích nepodobnosti, resp. podobnosti, které obsahují příslušné charakteristiky pro všechny páry sledovaných proměnných. Dále existují překrývající se (mlhavé) shlukovací algoritmy. Matice vzdáleností je také základem pro další techniku – *vícerozměrné škálování* (MDS – *multidimensional scaling*). Tato metoda je označována jako alternativní k faktorové analýze. Při MDS uživatel může analyzovat nejen korelační matice, ale také jiné matice podobnosti, resp. nepodobnosti.

Pro některé metody obsažené ve statistických programových systémech existují určitá omezení. Například pro optimální škálování prováděné v SPSS musí být kategorie kódovány od hodnoty 1 a při vícerozměrném škálování systém SPSS nedokáže analyzovat více než 100 proměnných.

V poslední době bylo vyvinuto mnoho algoritmů založených na neuronových sítích, jejichž cílem je eliminovat redundanci jak v lineárním tak v nelineárním případě (analýza nezávislých komponent, síť PCA). Většina z nich využívá uměle konstruovaná pravidla učení, víceúrovňové sítě s nelineární transformační funkcí nebo převádějí nelineární případ na lineární, což je umožněno pomocí křížových korelací vstupních vzorů. V tomto článku navrhneme metodu faktorové analýzy pro dostatečně nelineární případ, který vychází z pravidla učení Hebbova typu a dynamiky

neuronové sítě Hopfieldova typu. Tento konceptuální mechanismus se navíc jeví jako pravděpodobná varianta průběhu paměťových procesů v mozku.

### 3. FAKTORIZACE

Faktorizaci definujeme jako dekompozici komplexního vektoru signálu na množinu jednoduchých faktorů založených na korelacích mezi komponentami. V případě binární faktorizace je komplexní vektor signálu (vzoru) vyjádřen jako logický součet vážených binárních faktorů:  $X^0 = \bigvee_{l=1}^L \alpha_l \mathbf{f}^l$ . Tento případ faktorizace umožňuje její interpretaci v terminologii atraktoru neuronových sítí s binární aktivitou. Hlavní myšlenka spočívá v tom, že se síť může snadno učit křížové korelace, které jsou obsaženy ve vstupním komplexním vzoru při použití Hebbova pravidla učení. Při použití Hebbova pravidla učení se vytváří matice vztahů ve formě kovarianční matice pro množinu naučených vzorů. Skupiny neuronů, které mají tendenci být současně v excitovaném stavu (reprezentují společný faktor), budou více korelovat a odpovídající síla vztahu, reprezentovaná vahou u příslušné vazby, bude vyšší ve srovnání s neurony, které patří k různým faktorům. Tudiž každá skupina neuronů, která tvoří faktor, může odpovídat atraktoru dynamiky sítě. Tento příspěvek je věnován zkoumání podmínek, za kterých se v řídicí kódované neuronové síti Hopfieldova typu vytváří atraktory odpovídající jednotlivým faktorům s využitím Monte Carlo počítačové simulace chování dané neuronové sítě.

### 4. POPIS MODELU

Detailní teoretická a výpočetní analýza řídicí kódovaných neuronových sítí Hopfieldova typu je popsána např. v [2], [5], [6], [12]. Na rozdíl od těchto prací jsme trénovali síť množinou komplexních vzorů (binární superpozice faktorů). V etapě učení tak byla plně propojená síť  $N$  binárních neuronů trénována množinou  $M$  vzorů ve tvaru  $X^m = \bigvee_{l=1}^L \alpha_l^m \mathbf{f}^l$ , kde  $\mathbf{f}^l \in B_p^N$  (součin  $pN$  je udržován na konstantní hodnotě),  $L$  je počet faktorů a  $\alpha_l^m \in B_{p_f}^L$  jsou faktorová skóre.<sup>1</sup> Jak faktory, tak faktorová skóre, byly statisticky nezávislé. Dále  $p$  a  $p_f$  jsou mírou řídkosti (poměr počtu aktivních prvků k celkovému počtu prvků) kódování faktorů vzhledem k neuronům a vzorů vzhledem k faktorům. V limitě  $p_f \rightarrow 0$  se vzory stávají čistými faktory, které odpovídají obvyklému Hopfieldovu případu.

Váhová incidenční matice  $J$  byla vytvořena použitím korelačního Hebbova pravidla:

$$\mathbf{J}_{ij} = \frac{1}{A} \sum_{m=1}^M (X_i^m - q\{X^m\})(X_j^m - q\{X^m\}) \quad i \neq j, \quad \mathbf{J}_{ii} = 0,$$

kde  $q\{X^m\} = \sum_{i=1}^N X_i^m / N$  je celková aktivita vzoru.

Ve druhé fázi aktivní dynamiky, byla síť po prezentaci vstupního vzoru ponechána v relaxačním iterativním procesu, dokud nedospěla do nějakého atraktoru. Aktivní dynamika sítě je určena synchronní dynamickou rovnicí pro aktivitu  $X$  v čase  $t + 1$ :

$$X_i(t + 1) = \Theta(h_i(t) - T(t)), \quad X_i(0) = f_i^l, \quad i = 1, \dots, N$$

kde

---

<sup>1</sup> $B_p^N = \{\mathbf{X} | X_i \in \{0, 1\}, P\{X_i = 1\} = p \forall i = 1, \dots, N\}$ .

$$h_i(t) = \sum_{j=1}^N \mathbf{J}_{ij} X_j(t)$$

je postsynaptický potenciál,  $\Theta$  je skoková funkce a  $T(t)$  je aktivační práh. Tento práh  $T(t)$  je vybírán v každém časovém kroku tak, že aby výsledná aktivita sítě setrvala na konstantní úrovni  $p$ . V každém kroku  $n = pN$  jsou pak vybírání „vítězi“ (neurony s největším postsynaptickým potenciálem) a pouze tyto jsou pak v následující kroku aktivní. Tato procedura zabezpečuje, že atraktory budou pouze bodové nebo cyklické délky dva. U cyklického atraktoru byl za stabilní vzor (bodový atraktor) brán první vzor v popisované etapě. Ke zjištění, zda daný je faktor je atraktorem, byl jako počáteční stav sítě brán čistý faktor.

## 5. MÍRY

Těsnost mezi faktory popisovanými v předchozí kapitole a konečným vzorem byla měřena pomocí veličiny překrytí:

$$m^f = m(X^{in}, X^f) = \frac{1}{N} \sum_{i=1}^N X_i^{in} X_i^f.$$

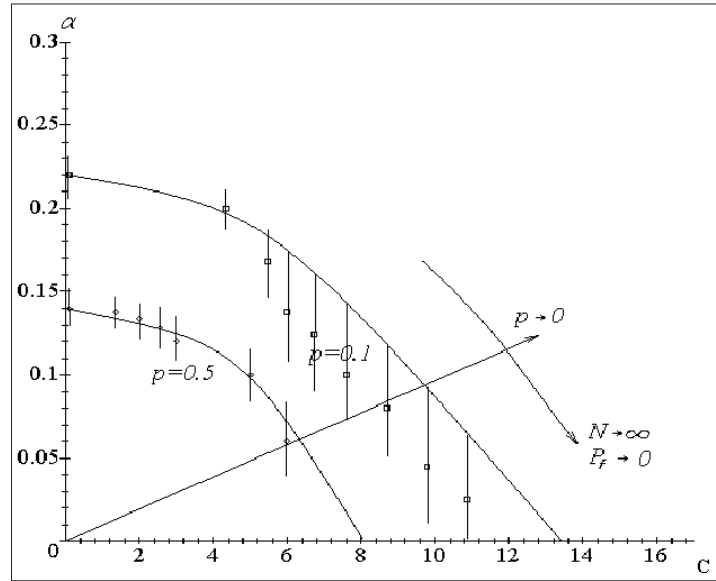
Jako míru relativní informační zátěže používáme veličinu  $\alpha \approx Lh(p)/N$ , kde  $h(p)$  je Shannonova funkce. Informační kapacita sítě je  $\alpha_{cr}$ , což je maximální  $\alpha$ , pro které ještě existují stabilní stavy v sousedství naučených faktorů. Průměrný počet aktivních faktorů ve vzoru  $C = Lp_f$  jsme použili jako míru komplexnosti problému. Nejjednodušší případ  $C = 1$  odpovídá standardnímu Hopfieldovu modelu.

## 6. VÝSLEDKY SIMULACE

Abychom analyzovali závislost  $\alpha_{cr}$  na parametrech sítě, bylo počítačově simulováno chování modelu neuronové sítě. Výpočty byly provedeny pro  $N$  v rozsahu od 200 do 4000, pro  $p = 0, 5; 0, 1$  a  $0,02$ . Program generoval náhodné faktory, které pak byly náhodně smíchány (s ohledem na  $p$  a  $p_f$ ) do množiny  $M$  vzorů. Síť pak byla touto množinou vzorů trénována. Nakonec bylo testováno, zda se k příslušným faktorům vytvořily v síti odpovídající atraktory. Rozdělení konečných překrytí  $m^f$  je bimodální se dvěma mody – s prvním pro  $m^f \approx 1$  a druhým pro  $m^f \approx 0$ , které odpovídají stabilizaci sítě v pravých a nepravých atraktorech. Prahová hodnota  $m_{thr}^f$  používaná pro separaci byla určena jako minimum v případě symetrického rozdělení mezi dvěma mody. Pravděpodobnost existence stabilních atraktorů v sousedství faktorů byla odhadována pomocí pravděpodobnosti, že  $m^f$  patří ke „skutečnému“ (pravému) modu.

Pro každou velikost sítě pravděpodobnost  $P$  rychle klesá, jestliže vzrůstá zátěž sítě a jak  $N$  vzrůstá, křivka funkce  $P(\alpha)$  se blíží skokové funkci. Bod, ve kterém se objevuje pokles, je označen jako  $\alpha_{cr}$ . Závislost  $\alpha_{cr}$  na  $C$  je znázorněna na obrázku 1.

Pro každé  $p$  křivka v souřadnicích  $\alpha, C$  odděluje fázi faktorizace od fáze, kde k faktorizaci nedochází z důvodů přetížení sítě. S rostoucími hodnotami  $C$  se křivka blíží první souřadnici a lze odhadnout hodnotu  $C_{max}$  – maximální komplexnost vzorů, pro které je ještě možná faktorizace. S rostoucí řídkostí kódování ( $p$  klesá) vzrůstá  $C_{max}$ .

Obr. 1 Kritické křivky v rovině  $\alpha, C$ .

## 7. ZÁVĚR

Před třiceti lety navrhl Marr konceptuální model hippocampu, jehož funkční idea je blízká naší představě o faktorizaci. Marr předpokládal, že hippocamp hraje roli zpracovatele komplexního vstupního vzoru s následným převodem do množiny klasifikačních jednotek. Úloha dočasného uchování informace byla přiřazena poli CA3 v mozku, která díky extenzivnímu systému rekurentních kolaterálů je přirozeným autoasociátorem. Předpokládali jsme, že by vzhledem ke zde prezentované schopnosti autoasociátoru provádět faktorizaci mohlo toto pole provádět dekompozici komplexní informace do elementárních faktorů. Počítačová simulace dokázala platnost této myšlenky. Rekurentní neuronové sítě se podařilo extrahovat faktory ze struktury komplexních vzorů, přičemž bylo při učení použito Hebbovo pravidlo (založené na pozorování chování neuronů v živých organismech). Náš výzkum se prozatím nedotkl dalších parametrů důležitých pro praktické nasazení této metody nelineární faktorizace – nevěnovali jsme se zatím struktuře oblastí atrakce, podílu nepravých atraktorů, jejich eliminaci atd. Zodpovězení těchto otázek bude předmětem dalšího výzkumu. Faktorizace je pravděpodobně závislá na absolutní komplexnosti vzorů. Tím se dostáváme k myšlence zvětšit parametry sítě tak, aby byly blízké hodnotám reálným u živých organismů ( $N \approx 10^5$ ,  $p \approx 0,04$ ,  $P\{\mathbf{J}_{ij} \neq 0\} \approx 0,02$ ) a použít biologicky pravděpodobnější neuronovou dynamiku k odhadu kritických makro-parametrů, tak aby mohly být porovnány s jejich analogiemi v experimentu chování.

## LITERATURA

- [1] Amit, D.J., Gutfreund, H. & Sompolinsky, H. (1987): Statistical mechanics of neural networks near saturation. *Ann. Phys.*, **173**, pp. 30-67.
- [2] Amari, S. & Maginu, K. (1988): Statistical neurodynamics of associative memory. *Neural Networks*, **1**, pp. 63-73.
- [3] Amari, S. (1989): Characteristics of sparsely encoded associative memory. *Neural Networks*, **2**, pp. 451-457.

- [4] Buzsaki, G. (1996): Hippocampo-neocortical dialogue. *Cerebral Cortex*, **6**, pp. 81-92.
- [5] Frolov, A.A. & Muraviev, I.P. (1993): Informational characteristics of neural networks capable of associative learning based on Hebbian plasticity. *Network*, **4**, pp. 495-536.
- [6] Frolov, A.A., Husek, D. & Muraviev, I.P. (1997): Informational capacity and recall quality in sparsely encoded Hopfield-like neural network: Analytical approaches and computer simulation. *Neural Networks*, **10**, pp. 845-855.
- [7] Gifi, A. (1990): *Nonlinear Multivariate Analysis*. New York: John Wiley & Sons.
- [8] Hopfield, J.J. (1982): Neural network and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, **79**, pp. 2544-2548.
- [9] Marr, D. (1970): A theory of cerebral neocortex. *Proc R Soc Lond, B*, **176**, pp. 161-234.
- [10] Marr, D. (1971): Simple memory. A theory of archicortex. *Phil Trans R Soc Lond, B*, **262**, pp. 24-81.
- [11] Mc Donald, R.P. (1985): *Factor Analysis and Related Methods*. New Jersey: Lawrence Erlbaum Associates, Publisher.
- [12] Perez-Vicente, C.J. & Amit, D. (1989): Optimized network for sparsely encoded patterns. *J. of Physics A: Math. Gen.*, **22**, pp. 559-569.

ÚSTAV INFORMATIKY AV ČR, POD VODÁRENSKOU VĚŽÍ 2, 182 07 PRAHA  
E-MAIL: [husek@cs.cas.cz](mailto:husek@cs.cas.cz)

INSTITUTE OF HIGHER NERVOUS ACTIVITY, RAS, MOSKVA, RUSSIA

VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE

VYSOKÁ ŠKOLA BÁŇSKÁ - TU OSTRAVA