

VYUŽITÍ SAMOORGANIZUJÍCÍ SE NEURONOVÉ SÍTĚ PRO HODNOCENÍ CHOVÁNÍ SPOTŘEBITELE

J. Weinlichová, J. Fejfar

Došlo: 3. září 2010

Abstract

WEINLICHOVÁ, J., FEJFAR, J.: *Usage of self-organizing neural networks in evaluation of consumer behaviour.* Acta univ. agric. et silvic. Mendel. Brun., 2010, LVIII, No. 6, pp. 625–632

This article deals with evaluation of consumer data by Artificial Intelligence methods. In methodical part there are described learning algorithms for Kohonen maps on the principle of supervised learning, unsupervised learning and semi-supervised learning. The principles of supervised learning and unsupervised learning are compared. On base of binding conditions of these principles there is pointed out an advantage of semi-supervised learning. Three algorithms are described for the semi-supervised learning: label propagation, self-training and co-training. Especially usage of co-training in Kohonen map learning seems to be promising point of other research.

In concrete application of Kohonen neural network on consumer's expense the unsupervised learning method has been chosen – the self-organization. So the features of data are evaluated by clustering method called Kohonen maps. These input data represents consumer expenses of households in countries of European union and are characterised by 12-dimension vector according to commodity classification. The data are evaluated in several years, so we can see their distribution, similarity or dissimilarity and also their evolution. In the article we discuss other usage of this method for this type of data and also comparison of our results with results reached by hierarchical cluster analysis.

behaviour of consumers, neural networks, Kohonen map, semi-supervised learning, label propagation, self-training, co-training

Využití neuronových sítí v ekonomických aplikacích je již dnes reálnou možností. Jsou využívány například pro podporu investičního rozhodování nebo jiných rozhodovacích procesů, jejichž cílem je dosažení minima, a v optimalizačních problémech, např. čerpání úvěrů nebo zpracování surovin ve výrobním procesu. Zkoumá se jejich využití ve všech oblastech podnikání, v marketingu, v odhadech rizik, finančních oblastech, managementu i výrobě. Zajímavou oblastí využití neuronových sítí je obchodní politika firmy a její strategický management. Hlavní využití neuronových sítí se v současnosti orientuje na oblast predikce, klasifikace a optimalizace. Při správném nasazení neuronových sítí např. při predikci finančních ukazatelů a na základě získaných údajů při vhodném manažerském

rozhodnutí může firma získat tolik důležitou strategickou výhodu. Další velkou oblastí s možností aplikace neuronových sítí jsou klasifikační problémy. Různé klasifikace také slouží jako podklad pro manažerská rozhodování, takový příklad uvádějí např. (Konečný, Trenz, Svobodová; 2010). Obecně se však dnes v praxi pro klasifikace většiny problémů využívají převážně statistické metody, regresní analýzy a metody rozhodovacích stromů. Nejčastěji používanou metodou je statistická metoda shlukové analýzy.

Cílem této práce je využití principu samoorganizace v neuronových sítích také jako ekvivalentu ke shlukové analýze a nalezení nových možností využití různých principů učení, jejichž nasazení není příliš prozkoumáno.

METODY

Kohonenovy mapy jsou založeny na principu soutěžní strategie učení (Competitive learning). To znamená, že neurony výstupní vrstvy – reprezentanti (codebook vectors) mezi sebou soutěží a aktivní je pouze vítěz. Ostatní neurony nemají žádný výstupní signál.

Následující podkapitoly uvádějí přehled učicích metod aplikovatelných na samoorganizující se neuronovou síť. Je zde uveden jejich princip, jednotlivé varianty a případné souvislosti s jinými metodami strojového učení. Na základě uvedených principů a omezení je vybrána učicí metoda použitelná v úloze hodnocení chování spotřebitele.

Učení bez učitele – shlukování

Úkolem je aproximovat hustotu pravděpodobnosti skutečných vstupních vektorů $x \in \mathfrak{X}$ pomocí konečného počtu takzvaných „reprezentantů“ (codebook vectors) (Šíma, 1996). Každému vstupnímu vektoru (jenž reprezentuje zkoumaný objekt) potom odpovídá jeden reprezentant. Přitom jeden reprezentant může odpovídat více zkoumaným objektům – zastupuje jejich skupiny (shluky).

Vektorová kvantizace

Pomocí vzdálenosti vstupního vektoru a vektoru reprezentanta můžeme určit, jak moc si objekt a reprezentant odpovídají. Zpočátku jsou reprezentanti rozmístěni náhodně, nebo mohou pravidelně vyplňovat celý vstupní prostor. Tento princip implementuje „batch training algorithm“ Kohonenovy mapy (Vesanto, 2000). Můžeme tedy pro první vstupní vektor vybrat reprezentanta, který je mu nejbližší a určit jeho vzdálenost jako

$$c = \arg \min_{l=1, \dots, h} \{ \|x - w_l\| \}, \quad (1)$$

kde x je vstupní vektor a w_l je l -tý reprezentant.

Nalezení vhodných reprezentantů pro všechny vstupní vektory tedy znamená minimalizace chyby dané vzorcem

$$E = \int \|x - w_c\|^2 p(x) dx, \quad (2)$$

kde w_c je nejbližší reprezentant.

Obyčejně neznáme hustotu pravděpodobnosti, ale máme k dispozici množinu vstupních vektorů – trénovací množinu. Potom je minimalizace chyby dána vzorcem

$$E = \frac{1}{k} \sum_{l=1}^k \|x^{(l)} - w_c\|^2, \quad (3)$$

kde $x^{(l)}$ je vstupní vektor trénovací množiny (Šíma, 1996).

Samoorganizující se síť – Kohonenovo učení a algoritmus k-means

Myšlenku uvedenou v předchozí kapitole můžeme realizovat pomocí jednoduché samoorganizující se sítě učené algoritmem Kohonenova učení.

Topologie – Síť je tvořena dvěma vrstvami, které jsou úplně propojené. Vstupní vrstva slouží k distribuci vstupních vektorů $x \in \mathfrak{X}$. Počet jejich neuronů tedy odpovídá velikosti vstupního vektoru. Druhá – kompetiční vrstva je tvořena reprezentanty. Počet reprezentantů ve druhé vrstvě určuje počet tříd, do nichž se bude klasifikovat. Jejich umístění ve vstupním prostoru je dáno hodnotami vah $w_j = (w_{j1}, \dots, w_{jn})$. Podle vzorce

$$y_j^* = \begin{cases} 1 & j = \arg \min_{l=1, \dots, h} \{ \|x^{(t)} - w_l\| \} \\ 0 & \text{jinak.} \end{cases} \quad (4)$$

můžeme spočítat výstup každého neuronu. Platí tedy, že vstupem jsou libovolná reálná čísla a výstupem je 1 u nejbližšího reprezentanta – neuronu a 0 u všech ostatních. Tento princip se nazývá vítěz bere vše (winner takes all). V přírodě je pozorována také tzv. Laterální inhibice, pomocí níž vítězný neuron oslabuje ostatní sousední neurony. Navíc každý vítězný neuron změní své váhy směrem k právě předloženému vstupnímu objektu podle vzorce

$$w_{ji}^{(t)} = \begin{cases} w_{ji}^{(t-1)} + \theta (x_i^{(t-1)} - w_{ji}^{(t-1)}) & j = \arg \min_{l=1, \dots, h} \{ \|x^{(t)} - w_l\| \} \\ w_{ji} & \text{jinak.} \end{cases} \quad (5)$$

kde parametr θ určuje velikost změny vah a obvykle se v průběhu učení zmenšuje (Šíma, 1996). Toto je analogií lidského učení. V mládí každé zaznamenané události podstatně mění zkušenost, kdežto ve stáří se zkušenost už v podstatě nemění. Takto se vektor vah reprezentanta dostává do středu shluků vstupních objektů.

Tento algoritmus je velmi podobný algoritmu k-means, jenž také hledá středy shluků. Řešení mohou být stejná v závislosti na nastavení parametru θ (Šíma, 1996).

Kohonenova samoorganizační mapa – Self-organizing map

U samoorganizujících se sítí je základem schopnost odlišit podobné od různého (Mařík, 2003). Mapují mnohazměrný vstupní prostor do dvourozměrného výstupního prostoru při pokud možno co nejvěrnějším zachování topologických vztahů. Jsou použitelné v mnoha oblastech. Cílem Kohonenových map je zobrazení vícerozměrných dat do dvourozměrného prostoru tak, aby byly podobné položky z mnohazměrného prostoru umístěny v mapě poblíž sebe. Kohonenovy mapy se skládají z jednotek uspořádaných do dvourozměrného vizualizačního prostoru. Obvykle se jednotky uspořádávají do pravoúhlé, či šestiúhelníkové mřížky. Každé jednotce je přiřazen modelový vektor z mnohazměrného vstupního prostoru. Vstupní položka (charakterizována svým vektorem) je zobrazena do jednotky s nejvíce podobným modelovým vektorem. Kohonenova mapa může být náhodně inicializována – náhodné vektory ve vstupním prostoru jsou přiřazeny každému modelovému vektoru. Alternativu představuje např. inicializace na základě dvou hlavních představitelů dat (Kohonen, 2001). Po inicializaci se iterativně střídají dva

kroky až do dosažení konvergence. Prvním krokem je nalezení jednotky s nejvíce podobným modelovým vektorem (best matching unit) pro každou jednotku vstupního prostoru (dat). Ve druhém kroku se modelové vektory upraví tak, aby lépe odpovídaly datům ze vstupního prostoru s tou podmínkou, že sousední jednotky reprezentují podobné jedince vstupního prostoru. Sousednost jednotek je definována funkcí sousednosti, jejíž hodnota se s každou iterací snižuje. Pro formalizaci základního algoritmu učení Kohonenovy mapy definujeme matici dat $\mathbf{D}$, matici modelových vektorů $\mathbf{M}_t$, matici vzdáleností $\mathbf{U}$, matici sousednosti $\mathbf{N}_t$, matici rozdělení $\mathbf{P}_t$ a matici rozdělení aktivace $\mathbf{S}_t$. Matice dat $\mathbf{D}$ je velikosti $n \times d$, kde n je počet položek dat a d je počet rozměrů jejich charakteristických vektorů. Velikost matice modelových vektorů $\mathbf{M}_t$ je $m \times d$, kde m je počet jednotek Kohonenovy mapy. Hodnoty $\mathbf{M}_t$ se aktualizují s každou iterací. Matice čtvercové vzdálenosti $\mathbf{U}$ o rozměru $m \times m$ určuje vzdálenosti mezi jednotkami mapy. Matice sousednosti může být definována například funkcí:

$$N_t = e^{-U/2r_t^2}, \quad (6)$$

kde r_t definuje rádius sousednosti $\mathbf{N}_t$ a s každou iterací se monotónně snižuje. Matice sousednosti je symetrická, o rozměru $m \times m$. Má vysoké hodnoty umístěné na diagonále a reprezentuje vliv, jaký mají jednotky mezi sebou navzájem. Řídká matice (speciální třída matic, jejichž většina prvků se rovná nule) $\mathbf{P}_t$ se spočítá z daných matic $\mathbf{D}$ a $\mathbf{M}_t$ následovně:

$$P_t(i, j) = \begin{cases} 1, & \text{pokud je jednotka } j \text{ BMU} \\ 0 & \text{jindy} \end{cases} \quad (7)$$

Matice rozdělení aktivace $\mathbf{S}_t$ definuje odezvu každé jednotky na každou položku dat během iterace t a vypočítá se následovně:

$$S_t = P_t \cdot N_t, \quad (8)$$

Na konci každého cyklu se spočítají nové modelové vektory následující způsobem:

$$M_{t+1} = S_t^* \cdot D, \quad (9)$$

kde S_t^* udává matici rozložení aktivace, která byla normalizována tak, aby suma všech řádků v každém sloupci byla 1 kromě těch jednotek, ke kterým není přiřazena žádná položka dat.

Kohonenova mapa má dva hlavní parametry, jedním je velikost mapy, druhým konečný rádius sousednosti. Větší mapy dávají v zobrazování větší rozložení, ale nevýhodou je větší výpočetní složitost. Konečný rádius sousednosti udává hladkost zobrazení a měl by být nastaven vzhledem k šumu v datech. Počet iterací by měl být alespoň pětsetkrát větší, než je počet jednotek v síti (Šíma, 1996).

Nevýhoda Kohonenovy mapy spočívá zejména v nemožnosti použití některé dobře definované nákladové funkce, což ztěžuje nastavení parametrů modelu a zajištění konvergence. Není zde také žádná záruka, že samoorganizace bude fungovat, je-

likož záleží na volbě parametrů pro každou množinu dat (Bishop, 2006).

Učení s učitelem – Learning Vector Quantization

Kohonenovu síť lze použít i pro učení s učitelem. Tento přístup využívá ohodnocená (labeled) data a obecně se jeho použitím dosahuje lepších výsledků klasifikace než u učení bez učitele – shlukování. Postupuje se podle následujícího principu:

1. učení bez učitele,
2. označení výstupních neuronů kategoriemi,
3. doučení sítí jedním z algoritmů LVQ.

Semi-supervised learning – učení s částečným dohledem

Principem těchto algoritmů je využití výhod učení s učitelem (lepší výsledky klasifikace) při shlukování neoznačených dat. Je potřeba získat malou množinu ohodnocených (labeled) dat, jejímž použitím můžeme pomoci učicímu algoritmu bez učitele dosáhnout výsledků srovnatelných s klasifikací s učitelem.

Label propagation

Známým semi-supervised algoritmem je Label propagation (Zhu, 2002). Tato metoda umožňuje šíření značek (labels) ve shlucích nalezených pomocí učení bez učitele. Použití algoritmu label-propagation pro Kohonenovu mapu nalezneme v (Herrman, 2007). Další možností je použití algoritmů Self-training a Co-training popsaných v (Steven, 2007).

Algoritmus Self-training

Algoritmus Self-training funguje podle následujícího schématu:

```

Procedure selfTrain ( $L_0$ ,  $U$ )
 $L_0$  is labeled data,  $U$  is unlabeled data
 $c \leftarrow \text{train}(L_0)$ 
loop until stopping criterion is met
   $L \leftarrow L_0 + \text{select}(\text{label}(U, c))$ 
   $c \leftarrow \text{train}(L)$ 
end loop
return  $c$ .

```

Klasifikátor se naučí na množině trénovacích dat (obyčejně malé), poté klasifikuje neoznačená (unlabeled) data. Data, u nichž je velká pravděpodobnost správné klasifikace, poté přidá do množiny ohodnocených dat. Celý postup se opakuje.

Algoritmus Co-training

Algoritmus Co-training funguje následujícím způsobem:

```

Procedure coTrain ( $L$ ,  $U$ )
 $L$  is labeled data,  $U$  is unlabeled data
 $P \leftarrow \text{random selection from } U$ 
loop until stopping criterion is met
   $f_1 \leftarrow \text{train}(\text{view}_1(L))$ 
   $f_2 \leftarrow \text{train}(\text{view}_2(L))$ 
   $L \leftarrow L + \text{select}(\text{label}(P, f_1)) + \text{select}(\text{label}(P, f_2))$ 

```

Remove the labeled instances from P and replenish P from U
end loop.

Na data se zde pohlíží ze dvou „pohledů“. Na každém „pohledu“ je natrénován klasifikátor. Tyto klasifikátory si poté navzájem opravují chybné hodnocení.

Algoritmy Self-training a Co-training zatím, pokud je nám známo, nebyly použity pro Kohonenovu samoorganizující se mapu. O takové použití se v současné době snažíme.

a energie a paliva, bytové vybavení a zařízení domácnosti a opravy, zdraví, doprava, pošty a telekomunikace, rekreace a kultura, vzdělávání, stravování a ubytování, ostatní zboží a služby. Údaje o těchto komoditách jsou tříděny do listů tabulkového procesoru. Jako vstupní data byly zformovány třízmměrné matice (rok, země, výdaje v jednotlivých komoditách). Získané výsledky tak mohou postihnout jednotlivé charakteristiky v chování spotřebitelů a neklasifikovat země pouze podle celkových hodnot výdajů.

MATERIÁL

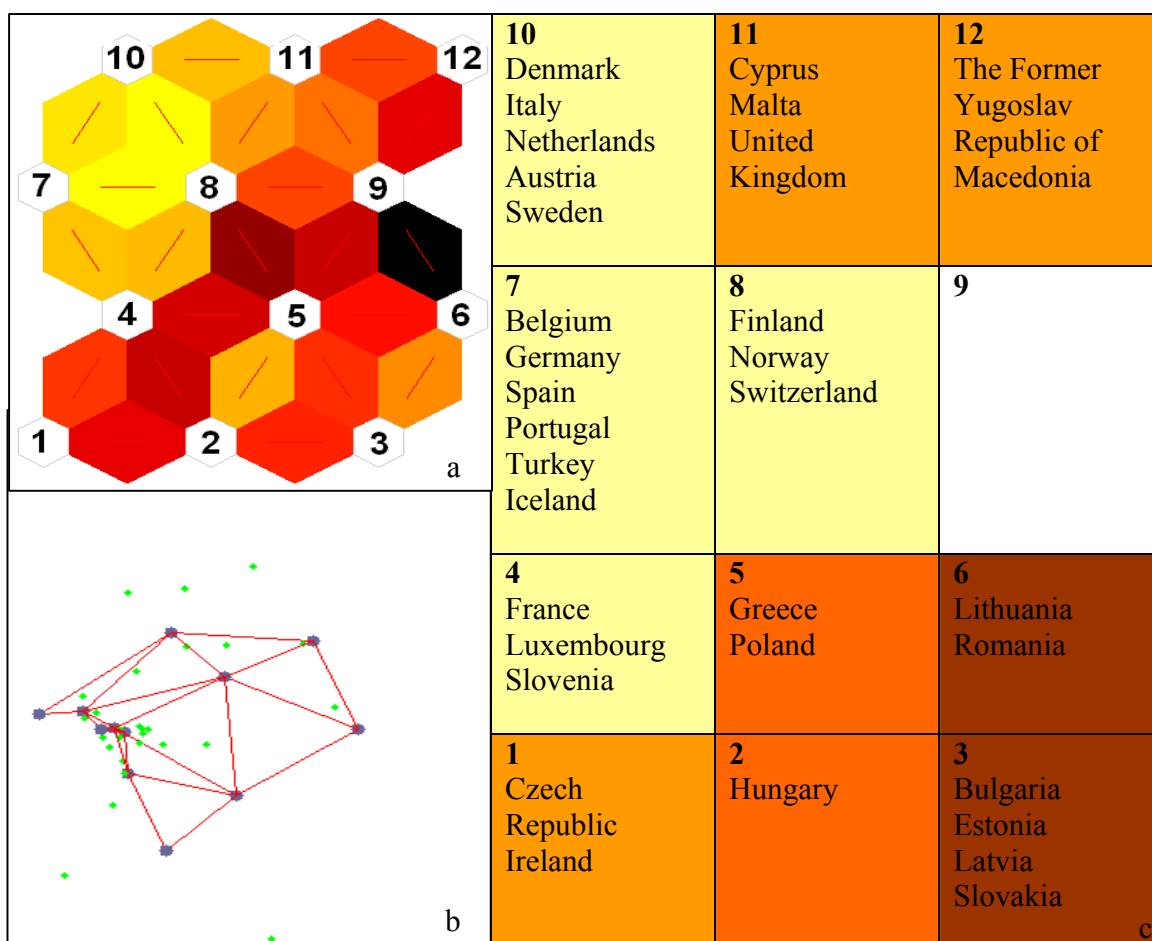
Z databáze Eurostat byly získány tabulky s výdaji domácností jednotlivých zemí za období posledních 11 let. Výdaje jsou v tabulkách dále členěny podle harmonizované komoditní klasifikace COI-COP/HICPs. Do výdajů domácností jsou tak začleněny položky: jídlo a nealkoholické nápoje, alkoholické nápoje a tabák, odívání a obuv, bydlení a voda

VÝSLEDKY A DISKUSE

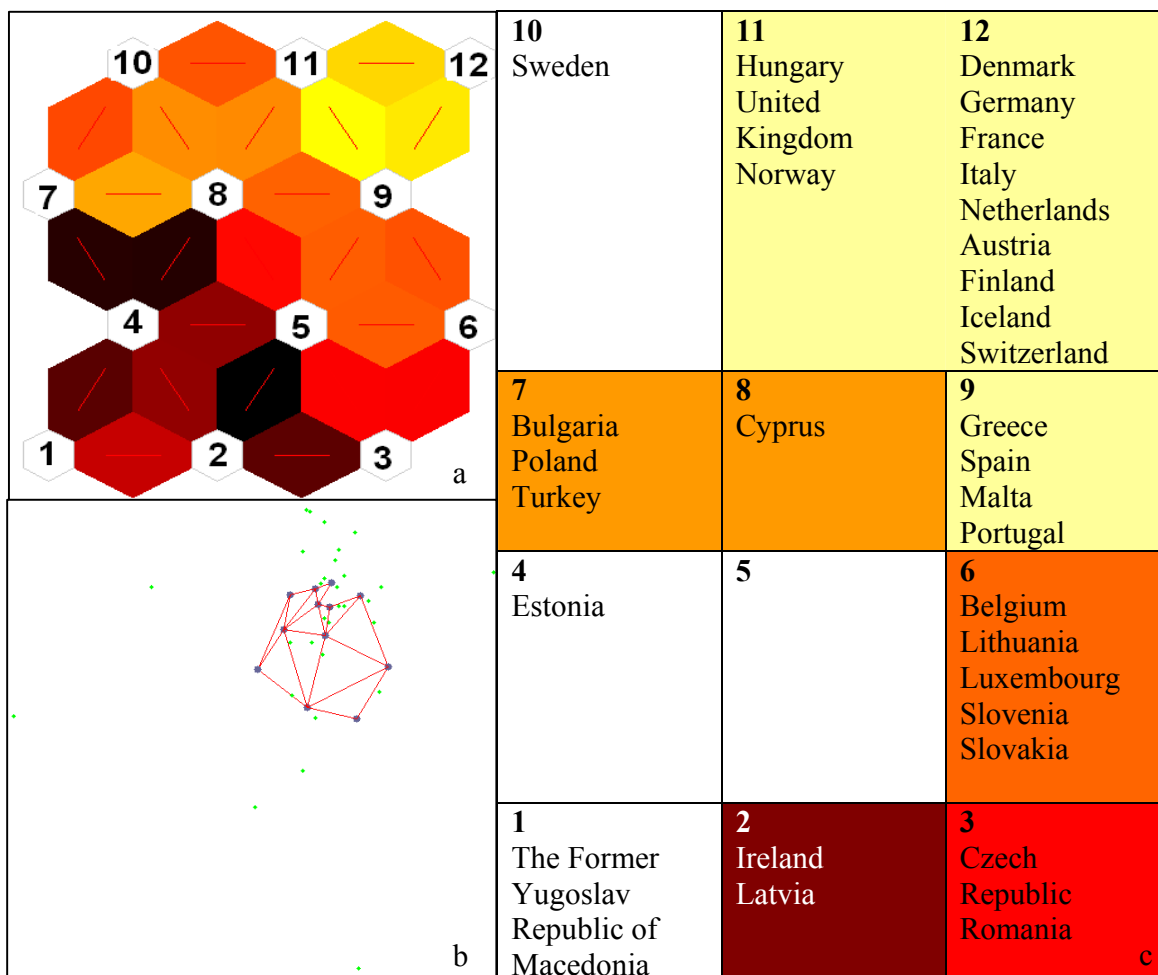
Vyhodnocování vstupních dat probíhalo v systému Matlab pomocí Neural Network Toolboxu; jako ukázka výše zmíněných metod byla zvolena Kohonenova samoorganizující mapa. Učení s učitelem nelze na těchto datech využít, jelikož není znám počet klasifikačních tříd.

I: Kohonenova mapa pro rok 2002; a) Očíslované body vyjadřují neurony, vzdálenosti nastavení vah jednotlivých neuronů jsou odlišné barvami. Čím světlejší barva, tím menší vzdálenost., b) Odpovídající samoorganizace vah neuronů do hexagonální mřížky v prostoru vah, c) Zařazení států do skupin dle reakcí neuronů. Státy světlých barev jsou si blízké, zatímco s státy s tmavou barvou jsou od nich velmi odlišné.

I: Kohonen map for year 2002; a) Numbered points means neurons, set distances of neuron's weights are contrasted by colours. The lighter colour the shorter distance it is. b) Self-organization of neuron's weights into hexagonal grid in weight area for this case, c) Country ranking by neuron's reactions. The countries with the light colours are very close while those with dark colour are very different from them.



II: Kohonenova mapa pro rok 1999; a) Očíslované body vyjadřují neurony, vzdálenosti nastavení vah jednotlivých neuronů jsou odlišné barvami. Čím světlejší barva, tím menší vzdálenost., b) Odpovídající samoorganizace vah neuronů do hexagonální mřížky v prostoru vah, c) Zařazení států do skupin dle reakcí neuronů. Státy světlých barev jsou si blízké, zatímco státy s tmavou barvou jsou od nich velmi odlišné.
 II: Kohonen map for year 1999; a) Numbered points means neurons, set distances of neuron's weights are contrasted by colours. The lighter colour the shorter distance it is. b) Self-organization of neuron's weights into hexagonal grid in weight area for this case, c) Country ranking by neuron's reactions. The countries with the light colours are very close while those with dark colour are very different from them.



Prvním úkolem před zahájením výpočtů bylo nalezení vhodné struktury sítě a počtu neuronů pro celkové hodnoty výdajů domácností. Pro reprezentaci vztahů a rozdělení těchto dat se ukázalo vhodné hexagonální uspořádání neuronů v Kohonenově vrstvě. Rovněž počet neuronů takto uspořádaných se ukázal nejvhodnější v uspořádání 3 x 4. Menší počet neuronů nepostihoval rozložení všech vlastností vstupních dat. Při použití většího počtu neuronů zůstávaly některé neurony bez reakcí na vstupní vzory a také nijak neovlivňovaly výsledné rozložení vlastností.

Po 96 trénovacích epochách se Kohonenova síť pro rok 2002 samoorganizovala ve své hexagonální topologii tak, že pokryla většinu vstupních vzorů (viz Tab. I b). Jednotlivé neurony odpovídají určitým skupinám států, které síť klasifikovala jako státy velmi podobné ve svém spotřebitelském chování. Žlutá barva v Tab. I a udává velkou podobnost skupin, které neurony určují. Tab. reprezentující kla-

sifikace Kohonenovou mapou v Tab. I c tak zobrazuje žlutě zbarvenou skupinu silných evropských států, jejichž spotřebitelské výdaje jsou na obdobné úrovni. Jedná se o státy, jejichž výdaje jsou ve většině komodit nejvyšší. Jsou to většinou ty státy EU, které jsou v ní nejdéle. Tmavě červená barva v Tab. I a udává velkou vzdálenost dvou tříd (neuronů), černá největší. Můžeme tedy v mapě pozorovat další tři shluky. Prvním je Řecko, Polsko a Maďarsko, další shluk tvoří státy z neuronů č. 11 a 12 a posledním shlukem, který je v Tab. I c reprezentován nejtmavší barvou a tedy udává státy, které mají spotřebitelské výdaje velmi vzdálené od ostatních států, je shluk Litvy, Lotyšska, Estonska, Rumunska, Bulharska a Slovenska.

V mapě pro rok 1999 (Tab. II) můžeme pozorovat odlišné rozdíly ve spotřebních výdajích mezi státy EU. Žluté silné státy s velkými spotřebními výdaji tvoří menší skupinu. Vidíme, že v budoucnu (dle mapy pro rok 2002) se výdaje slabších oranžových

a tmavších států zvyšují a přibližují se tak silným evropským státům. Je to např. Česká republika nebo Irsko. Oproti tomu jiné státy, které byly ve svých spotřebních výdajích silnější, oslabují a začínají se vzdalovat silným státům. Tento trend se dle dat z databáze Eurostat do budoucna potvrzuje.

Zde uvedené výsledky shlukování pomocí Kohonenovy mapy můžeme porovnat s výsledky získanými nad stejnými daty aplikací metody shlukové analýzy, jejichž výsledné dendrogramy jsou uvedeny v (Turčínková, Stejskal, Stávková; 2007). Při porovnání zjistíme, že Kohonenova mapa shlukla některé státy do jiných skupin, než jaké ukazují dendrogramy. To znamená, že tato mapa klasifikovala použitá data nejen podle jejich hodnot, ale zachovala i topologické vztahy dat, což je jedna z vlastností Kohonenových map. Další vlastností je grafické vyjádření těchto vztahů, takže lze tuto metodu do budoucna využít pro grafické modelování vývoje rozložení výdajů domácností evropských států.

Z hlediska metod je použita aplikace Kohonenovy mapy, která spočívá v použití učícího algoritmu bez učitele, nejčastější. Existuje zde řada osvědčených aplikací. Využití samoorganizující se sítě je popsáno např. v článku autorů Cardosa, Moura-Pirese (2002), který se věnuje problematice segmentace portugal-

ských klientů. Ve svých výsledcích většinou zaznamenali lepší nebo stejné výsledky oproti použití standardních statistických metod. Kombinaci Kohonenovy sítě a shlukovací analýzy uvádí Serrano-Cinca (1998). Díky vizualizaci podobností mezi uvažovanými firmami identifikoval, jak velké rozdíly (podle vybraných ukazatelů) existují mezi spříteľnami ve španělských regionech jako v důležitém sektoru finančního systému.

Při použití Kohonenovy mapy můžeme najít také několik nevýhod: obecně horší výsledky algoritmů učení bez učitele oproti algoritmům učení s učitelem, nemožnost změřit „výkon“ mapy, nutnost ručně kontrolovat správnost vytvořených shluků. Oproti tomu nemusíme disponovat označenými (labeled) tréninkovými daty. Ta mohou být v mnoha oblastech skutečně nedostupná. V situacích, kde můžeme získat alespoň malou množinu takových ohodnocených dat, se nabízí využití možností algoritmů s částečným dohledem učitele (semi-supervised learning). Tyto metody jsou relativně nové a v současné době jsou předmětem výzkumu. Vývin takového algoritmu pro Kohonenovu mapu je velmi slibný a budeme se mu v příštích letech intenzivně věnovat.

SOUHRN

Tato práce popisuje využití samoorganizující se neuronové sítě pro hodnocení chování spotřebitele. V metodické části hodnotí různé metody učení samoorganizující se neuronové sítě. Na základě porovnání principu učení s učitelem, učení bez učitele a učení s částečným dohledem a jejich omezujících podmínek jsou vysvětleny výhody přístupu učení s částečným dohledem – semi-supervised learning. Při zkoumání učících algoritmů učení s částečným dohledem učitele: label propagation, self-training a co-training se ukazuje jako vhodné pro budoucí výzkum zejména využití principu co-training pro učení Kohonenovy mapy.

Pro konkrétní aplikace na hodnocení výdajů domácností států Evropské unie pro roky 1999 a 2002 byl vybrán princip učení bez učitele. Výsledkem provedené analýzy jsou Kohonenovy mapy, které znázorňují rozložení výdajů jednotlivých států. Tato grafická reprezentace dat umožňuje sledovat kromě zařazení států do několika skupin podobnosti také topologické vztahy mezi těmito daty, grafickou vzdálenost jejich skupin. Z uvedených výsledků vyplývá, že Kohonenova mapa je vhodným nástrojem pro sledování struktury výdajů zemí Evropské unie, rovněž je do budoucna vhodná pro grafické sledování vývoje spotřebitelských výdajů.

chování spotřebitele, neuronové sítě, Kohonenova mapa, učení s částečným dohledem, label propagation, self-training, co-training

SUMMARY

This article deals with usage of self-organizing neural network in evaluation of consumer behaviour. In a method part there are evaluated different methods for learning the self-organizing neural network. On base of comparison of supervised learning, unsupervised learning and semi-supervised learning and their binding conditions there is pointed out an advantage of semi-supervised learning. After exploration of learning algorithms for semi-supervised learning: label propagation, self-training and co-training, the co-training algorithm for Kohonen map learning for future research is evinced.

In concrete application in evaluation of household expenses in EU countries for years 1999 and 2002 have been chosen the principle of unsupervised learning. The results of the analysis that have been made are the Kohonen maps. They show a distribution of expenses of each country. This graphic rep-

resentation makes possible to watch not only the country ranking into several groups of similarity but also topological relationships among this data as the graphical distance of their groups. From the presented results imply that the Kohonen map is appropriate tool for observation the structure of household expenses in EU countries, likewise it is appropriate for future graphical monitoring of consumer expense evolution.

LITERATURA

- BISHOP, C. M., 2006: Pattern recognition and machine learning. NY, USA: Springer. ISBN 0-387-31073-8.
- CARDOSO, M., MOURA-PIRES, F., 2002: Neural Networks in Business: Techniques and Applications. [s.l.]: [s.n.]. ISBN 1-931777-79-9. Chapter V: Segmentation of the Portuguese Clients of Pousadas de Portugal, s. 70–88.
- HERRMAN, L., ULTSCH, A., 2007: Label Propagation For Semi-Supervised Learning In Self-Organizing Maps. Proceedings of the 6th WSOM. Dostupné z < <http://www.uni-marburg.de/fb12/datenbionik/pdf/pubs/2007/WSOMssLP> >.
- KOHONEN, T., SCHROEDER, M. R., HUANG, T. S. (ED.), 2001: Self-Organizing Maps, Secaucus, NJ, USA: Springer-Verlag New York, Inc. ISBN 3540679219.
- KONEČNÝ, V., TRENZ, O., SVOBODOVÁ, E., 2010: Classification of companies with assistance of self-learning neural networks. Agricultural Economics: Zemědělská ekonomika. sv. 56, č. 2, s. 51–58. ISSN 0139-570X.
- MAŘÍK, V., ŠTĚPÁNKOVÁ, O., a kol., 2003: Umělá inteligence. Sv. 4. Praha: Academia. ISBN 80-200-1044-0.
- SERRANO-CINCA, C., 1998: From financial information to strategic groups: A self-organizing neural network approach. Journal of forecasting [online]. [cit. 2009-05-05]. Dostupný z [www: < http://apps.isiknowledge.com >](http://apps.isiknowledge.com).
- STEVEN, A., 2007: Semisupervised Learning for Computational Linguistic, Chapman & Hall/Crc Computer Science & Data Analysis. ISBN:1584885599.
- ŠÍMA, J., NERUDA, R., 1996: Teoretické otázky neuronových sítí. Matfyzpress, Praha. ISBN 80-85863-18-9. Dostupné z <<http://www2.cs.cas.cz/~sima/kniha.html>>.
- TURČÍNKOVÁ, J., STEJSKAL, L., STÁVKOVÁ, J., 2007: Chování a rozhodování spotřebitele. Brno: MSD. 102 s. ISBN 978-80-7392-013-5.
- VESANTO, J. a kol., 2000: SOM Toolbox for MATLAB 5. Helsinki University of Technology. Dostupné z <<http://www.cis.hut.fi/projects/somtoolbox>>.
- ZHU, X., GHAHRAMANI, Z., 2002: Learning from labeled and unlabeled data with label propagation. Technical Report CMU-CALD-02-107, Carnegie Mellon University.

Adresa

Ing. Jana Weinlichová, Ing. Jiří Fejfar, Ústav informatiky, Mendelova univerzita v Brně, Zemědělská 1, 613 00 Brno, Česká republika, e-mail: jiri.fejfar@mendelu.cz, jana.weinlichova@mendelu.cz

