

Porovnanie niekoľkých typov rekurentných neurónových sietí z hľadiska hĺbky pamäte

Diplomová práca

Jaroslav Ištók

Školiteľ: doc. RNDr. Martin Takáč, PhD.

1. Rekurentné neurónové siete
2. Úvod do SOM
3. Experiment so SOM
 - RecSOM
 - Activity RecSOM
 - MSOM
 - Decay MSOM
4. Úvod do SRN
5. Experiment s SRN
6. Otázky

Rekurentné neurónové siete

- Pracujú s určitým druhom historického kontextu
- Kontext je tvorený minulými stavmi a vstupmi siete a predstavuje formu pamäte.
- Výstup siete je ovplyvnený aj minulými vstupmi a nie iba aktuálnym
- Rekurentné SOM (RecSOM a MSOM)
- SRN s Elmanovou architektúrou

Pamäťová hĺbka - definícia

Hĺbka pamäti neurónovej siete vyjadruje dĺžku historického kontextu, ktorý má ešte vplyv na aktuálny výstup siete. Ak je hĺbka pamäti N , vstupy v čase väčšom ako $(T - N)$ nemajú vplyv na výstup v čase T .

Úvod do SOM

- učenie bez učiteľa
- zhlukovanie dát
- zachovanie topologických vlastností dát

Hľadanie víťaza

$$i^* = \operatorname{argmin}_i \|x - w_i\|$$

Aktualizácia váh

$$w_i(t+1) = w_i(t) + \alpha(t)h(i^*, i)[x(t) - w_i(t)]$$

Rekurentná SOM

- rozšírená o kontext
- pridané kontextové váhy

Výpočet vzdialenosti

$$d_i = (1 - \alpha) \cdot \|x(t) - w_i\|^2 + \alpha \cdot \|y(t - 1) - c_i\|^2$$

Aktualizácia váh

$$w_i(t + 1) = w_i(t) + \alpha(t)h(i^*, i)[x(t) - w_i(t)]$$

Aktualizácia kontextových váh

$$c_i(t + 1) = c_i(t) + \alpha(t)h(i^*, i)[y(t - 1) - c_i(t)]$$

Pamäťová hĺbka v rekurentnej SOM

Spôsob určovania pamäťovej hĺbky v rekurentných SOM

Každý neurón má navyše pamäťové okno, do ktorého sú ukladané vstupy, pre ktoré bol daný neurón víťazom. Neukladáme však iba vstupy, ale vždy k posledných vstupov, kde k je veľkosť posuvného okna na vstupnej sekvencii.

Pamäťová hĺbka v rekurentných SOM

Vážený priemer dĺžok najdlhších spoločných podpostupností reťazcov (určovaných od konca) v pamäťových oknách všetkých neurónov v sieti. Priemer je váhovaný počtom reťazcov uložených v jednotlivých pamäťových oknách.

Pamäťová hĺbka v rekurentných SOM

Čiže napríklad ak by pamäťové okno obsahovalo „abbdabb“ a „bbbacbb“, potom pamäťová hĺbka pre toto okno bude 2

Experiment so SOM

Čo bolo cieľom experimentu?

Nájsť optimálne hodnoty parametrov, pri ktorých daný typ siete dosahuje najvyššiu hodnotu pamäťovej hĺbky a porovnať siete s optimálnymi parametrami.

Priebeh experimentu

- Siete sme natrénovali s rôznymi kombináciami parametrov, ktoré ovplyvňujú váhu kontextu vo výpočte vzdialenosti a zloženie samotného kontextu.
- Použili sme postupne zmenšujúcu sa gausovskú funkciu susednosti
- Postupne klesajúca rýchlosť učenia
- Ako trénovacie dáta sme použili náhodne vygenerovanú postupnosť písmen **abcd**
- Na záver sme porovnali pamäťové hĺbky sietí s optimálnymi parametrami
- Vstupmi su 1-hot enkódované písmená

Použitá funkcia susednosti

$$N(i^*, i) = \exp^{-\frac{d_{\mathbb{E}}^2(i^*, i)}{\lambda^2(t)}} \quad (1)$$

RecSOM kontext, vzdialenosť a parametre

$$y(t) = [y_1(t-1), \dots, y_N(t-1)] \quad c, r \in R^N$$

$$y_i = \exp(-d_i)$$

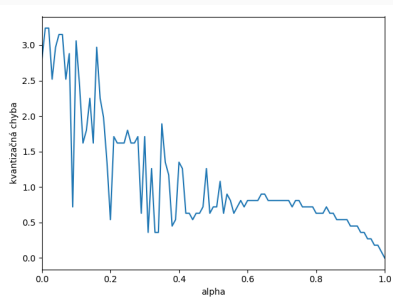
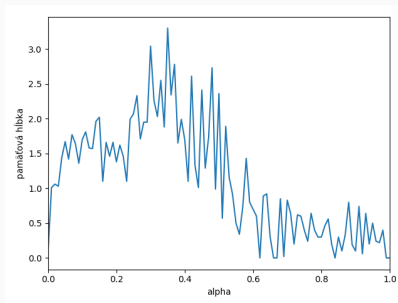
$$d_i = (1 - \alpha) \cdot \|x(t) - w_i\|^2 + \alpha \cdot \|y(t-1) - c_i\|^2 \quad c_i \in R^N$$

Pri RecSOM máme iba parameter α

Parameter	Hodnota
Hodnoty alpha	0 - 1 (krok: 0.01)
Rozmer	30x30
Počet epôch tréovania	20
Veľkosť posuvného okna	10

Tabuľka 1: Trénovacie parametre RecSOM siete

RecSOM výsledky



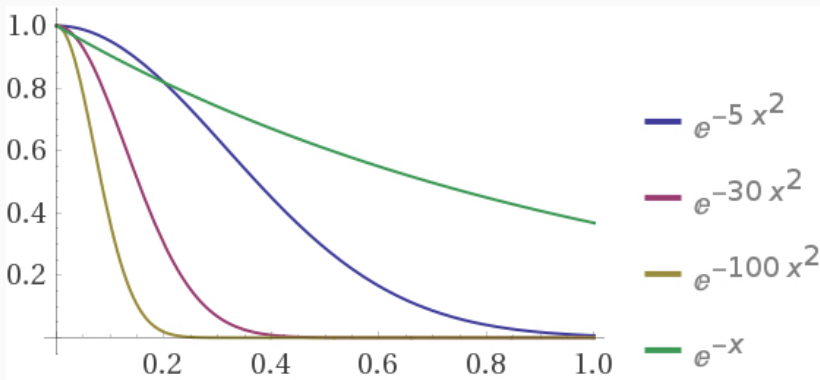
Obr. 1: Výsledky RecSOM

Activity RecSOM

Activity RecSOM

Pri RecSOM sme chceli sme ovplyvniť spôsob výpočtu kontextu. Preto sme pozmenili spôsob výpočtu aktivácie. Pridalo nám to ďalší parameter β .

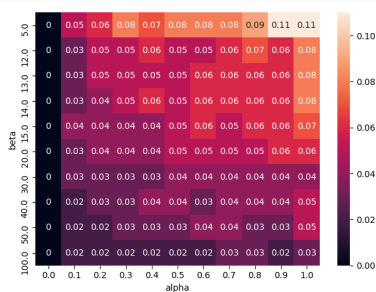
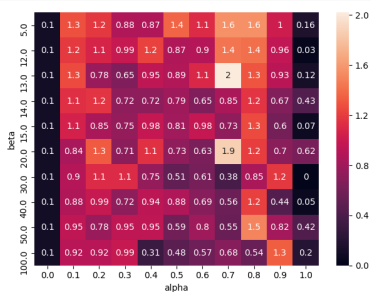
$$y_i = \exp(-\beta \cdot d_i^2)$$



Parameter	Hodnota
Hodnoty alpha	0 - 1 (krok 0.1)
Hodnoty beta	[5.0,12.0,13.0,14.0,15.0,20.0,30.0,40.0,50.0,100.0]
Veľkosť	30x30
počet epôch tréningovania	20
veľkosť posuvného okna	10

Tabuľka 2: Parametre Activity RecSOM siete

Activity RecSOM výsledky



Obr. 3: Výsledky Activity RecSOM

MSOM kontext, vzdialenosť a parametre

$$y(t) = (1 - \beta) \cdot w_{i^*}(t - 1) + \beta \cdot c_{i^*}(t - 1) \quad (2)$$

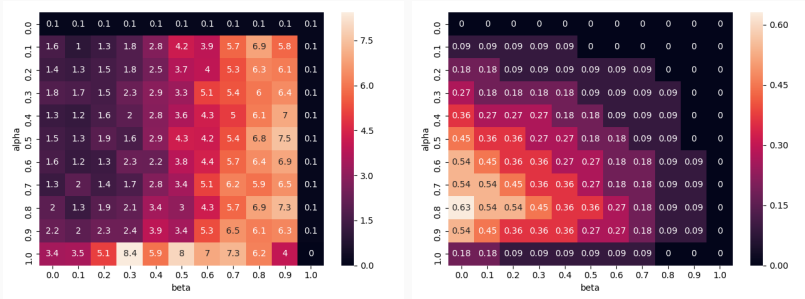
$$d_i = (1 - \alpha) \cdot \|x(t) - w_i\|^2 + \alpha \cdot \|y(t - 1) - c_i\|^2 \quad c_i \in R^N$$

Pri MSOM máme parametre α a β

Parameter	Hodnota
Hodnoty alpha	0 - 1 (krok 0.1)
Hodnoty beta	0 - 1 (krok 0.1)
Veľkosť	30x30
Počet epôch	20
Veľkosť posuvného okna	10

Tabuľka 3: Parametre MSOM siete

MSOM výsledky



Obr. 4: Výsledky MSOM

Decay MSOM kontext, vzdialenosť a parametre

$$c_t = x_t + \beta * c_{t-1} \quad (3)$$

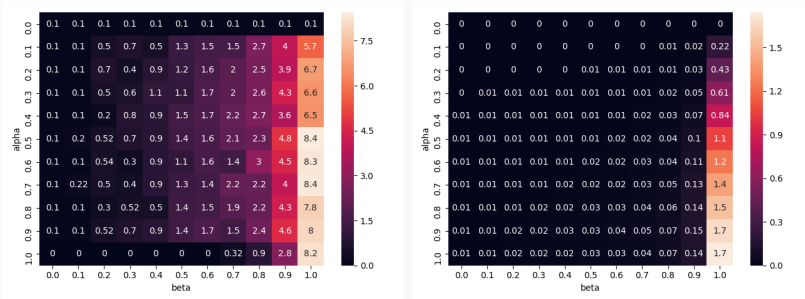
$$c = \beta^0 \cdot x_t + \beta^1 \cdot x_{t-1} + \beta^2 \cdot x_{t-2} \cdots \beta^n \cdot x_{t-n} \quad (4)$$

Modifikovaná verzia MSOM, kde sme chceli vyskúšať odlišný druh kontextu, ktorý nie je tvorený minulými stavmi siete ale samotnou históriou vstupov siete. Pri Decay MSOM máme parametre α a β

Parameter	Hodnota
Hodnoty alpha	0 - 1 (krok 0.1)
Hodnoty beta	0 - 1 (krok 0.1)
Veľkosť	30x30
Počet epôch tréningovania	20
Veľkosť posuvného okna	10

Tabuľka 4: Parametre Decaying MSOM siete

Decay MSOM výsledky



Obr. 5: Výsledky Decay MSOM

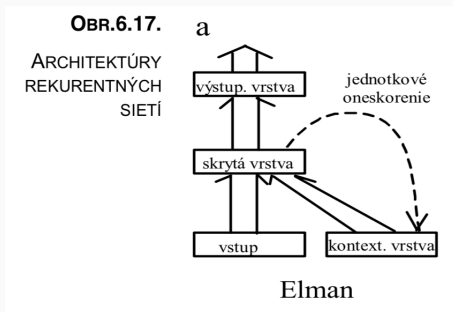
Vyhodnotenie výsledkov experimentu so SOM

- Problém s rozdielom v dimenziách medzi váhovými a kontextovými vektormi pri RecSOM a Activity RecSOM
- Vplyv zloženia kontextu na pamäťovú hĺbku
- DecaySOM
- Výpočtová náročnosť RecSOM vs MSOM

Úvod do SRN

SRN s Elmanovou architektúrou

- Trénovanie pomocou BPTT
- Skrytá vrstva
- Stavový priestor na skrytej vrstve



Obr. 6: Architektúra Elmanovej siete

Experiment s SRN

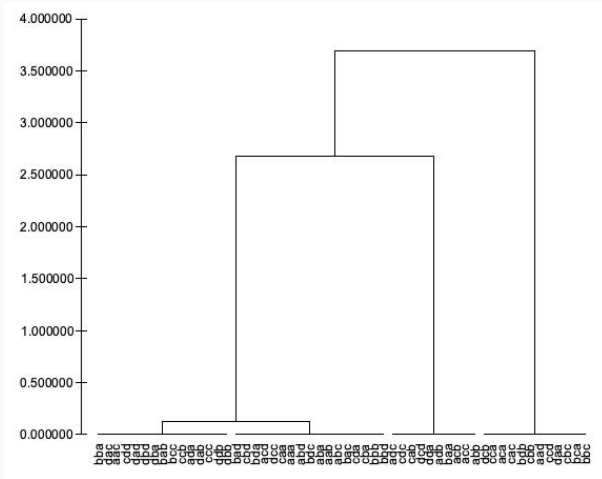
Experiment s SRN

- Hľadanie spôsobu merania pamäťovej hĺbky
- Analýza stavového priestoru
- Dendrogram

Parameter	Hodnota
veľkosť skrytej vrstvy	30
počet epôch	100
počet krokov do minulosti (T)	5

Tabuľka 5: Trénovacie parametre SRN s Elmanovou architektúrou

Dendrogram



Obr. 7: Dendrogram pre Elmanovu sieť

Otázky

Otázka

Podrobne popíšte, ako by ste použitím PCA vizualizovali v 2D na ktoré časti skrytého priestoru sa zobrazujú jednotlivé vstupné sekvencie a čo by vám to mohlo napovedať o hĺbke pamäti SRN

1. Odpoveď

Skrytý priestor v SRN je vysoko rozmerný, čiže vstupy sú ako keby zobrazované do vysoko rozmerného stavového priestoru na skrytej vrstve siete. Každý vstup spôsobí nejakú aktiváciu neurónov na skrytej vrstve a toto je jeden bod v tom priestore. Keďže body vo vysokorozmernom priestore si nevieme vizualizovať, môžeme zredukovať ich dimenzie, napríklad pomocou PCA. Keď potom vezmeme napríklad prvé dva principiálne komponenty a vykreslíme 2D scatter graf (projekciu do roviny) z týchto komponentov, tak sa nám vytvoria skupiny, kde každá skupina predstavuje vstupy, ktoré majú nejakú spoločnú vlastnosť. O pamäťovej hĺbke nám to moc nepovie, resp. sú to len 2D projekcie z veľarozmerného priestoru a nepredsievujú skutočnú blízkosť vstupov v priestore. Dendrogram nám môže o pamäťovej hĺbke povedať viac, keďže ten reprezentuje skutočné vzdialenosti jednotlivých bodov v priestore.

2. otázka

2. V časti 5.7 spomínate, že Activity RecSOM by sa dala upraviť tak, že by vracala iba Gausovskú aktivitu okolo víťaza. Ako by ste ju upravili? Očakávate, že by takto upravená RecSOM dávala z hľadiska hĺbky pamäti výsledky podobnejšie na štandardnú RecSOM, alebo na MSOM? Prečo?

2. odpoveď

Vzorec by som ponechal tak ako je pri Activity RecSOM, teda gaussian s parametrom β ktorým vieme ovplyvňovať strmosť (priebeh) tejto funkcie. Zmenila by sa hodnota vzdialenosti d . Namiesto súčtu euklidovských vzdialeností, by sme použili vzdialenosť daného neurónu od víťazného neurónu v mape (podobne ako pri funkcii susednosti), a aktivita zvyšných neurónov by závisela od ich vzdialenosti od víťazného neurónu v samotnej mape a nie vzdialenosti vo vektorovom priestore. Aktivita by nebola priamo ovplyvnená stavmi siete, ale iba samotnou pozíciou neurónu vzhľadom na víťazný neurón v danom kroku.

Podľa mňa budú výsledky takejto siete podobné výsledkom obvyčajnej RecSOM čo sa týka hĺbky pamäte, pretože kontext neobsahuje vlastnosti víťaza (ktoré sa ukázali ako dôležité z hľadiska hĺbky pamäte).

1.Otázka

Pripraviť porovnanie kvality predikcie pre výsledky zo str. 43 až 47 plus kvalitu predikcie pre 3 trénovacie sety pre Elmanovu SRN.

2.Otázka

Ako hĺbka pamäte súvisí s kvalitou predikcie? Je medzi nimi lineárny vzťah? Čiže čím väčšia hĺbka pamäte, tým lepšia predikcia? Dá sa to alebo nedá doložiť aj údajmi z literatúry?

1.Odpoved'

Pri našich implementáciach SOM nemáme žiadne predikcie (unsupervised model). Kvalitu organizácie nám reprezentuje kvantizačná chyba, ktorú sme merali a aj porovnávali s pamäťovou hĺbkou.

2.Odpoved'

Neviem to porovnať, keďže sme nenašli (neimplementovali) spôsob na odmeranie hĺbky pamäte SRN. Ako chybovú funkciu používame log loss, ktorá počas trénovania klesala až po veľmi nízke hodnoty. Bohužiaľ to nemôžem porovnať s pamäťovou hĺbkou.

Ďakujem za pozornosť