

# Semestrální práce z předmětu 36NAN neuronové sítě a neuropočítače

## Analýza spirál

Vypracoval: Jan Doležel

E-mail: [dolezj2@fel.cvut.cz](mailto:dolezj2@fel.cvut.cz)

### Úvod:

Ve spolupráci s fakultní nemocnicí v Motole jsme získali 23 na papír kreslených spirál. Tyto spirály kreslili lidé postižení parkinsonovou chorobou, ale jsou mezi nimi i referenční spirály lidí zdravých. Naším úkolem je spirály analyzovat a najít metodu klasifikace spirál dle stupně postižení.

Tento dokument spolu s dokumentem semestrální projekt – analýza spirál (dále jen 36SP) tvoří jeden celek, proto se budu odvolávat na některé věci v něm uvedené. V uvedeném dokumentu se věnuji zejména problému získání a digitalizace vstupních dat implementované v mém programu a také jedné metodě jejich klasifikace, rovněž implementované v programu. Zde uvedu různé alternativní postupy, které jsem na datech zkoušel.

### Teorie:

Metoda analýzy spirál je v současnosti velice využívaná metoda. Nyní pacienti spirály kreslí na papír tužkou a spirály hodnotí člověk (lékař). My bychom rádi této analýze naučili počítač.

Dostali jsme několik nakreslených spirál. Moje aplikace dokáže většinu z nich převést do formy posloupnosti souřadnic (x,y). Spirála může být matematicky popsána dvěma ekvivalentními způsoby – v kartézských nebo polárních souřadnicích:

$$x = a\theta \cos(\theta + c)$$

$$y = a\theta \sin(\theta + c)$$

$$r = a\theta$$

$$r = \sqrt{x^2 + y^2}$$

Tyto data pak lze transformovat, derivovat, integrovat a používat různé další matematické úpravy. Pro analýzu jsem používal statistické postupy a metody, nástroj JavaNNS a Weka.

JavaNNS je nástroj na tvorbu a učení neuronových sítí založený na knihovnách staršího SNNS, ke kterému bylo přidáno nové grafické rozhraní napsané v Javě. Tento nástroj byl vyvinut na univerzitě v Tübingenu v Německu. Existují verze pro Windows, Solaris i RedHat Linux (kvůli nativním kódům knihovny SNNS) a stáhnout si je můžete na adrese

<http://www-ra.informatik.uni-tuebingen.de/downloads/JavaNNS/>.

Nástroj Weka je velice mocný nástroj pro předpřípravu i analýzu dat. Vyvinuli ho na univerzitě ve Waikato na Novém Zélandě. Je to open-source projekt s licencí GNU/GPL a stáhnout si ho můžete na adrese

<http://www.cs.waikato.ac.nz/~ml/weka/>

ve verzích pro Windows, Mac OS X a jiné (Linux).

Mnou vytvořené nástroje byly přiloženy k dokumentu. Jejich algoritmy a použití jsou vysvětleny v druhém dokumentu (36SP).

## Experiment:

Pro analýzu počítačem se ukázalo vhodných 20 spirál. Dvě nešly převést do posloupnosti souřadnic a u jedné nevíme, zda patří člověku zdravému nebo nemocnému. Data máme od sedmi pacientů a jednoho zdravého člověka (ukázkové spirály nakreslené lékařem) a jsou kresleny levou i pravou rukou v poměru 12:8 (nemocní : zdravý).

Tyto spirály byly převedeny do posloupnosti souřadnic (x,y), čas (dva následující body jsou vzdálené 5ms) a tlak. Dva poslední parametry byly uměle přidány kvůli budoucí kompatibilitě dat s daty snímanými z tabletu. Přesný popis v 36SP.

S takto vytvořenými daty jsem důkladně analyzoval spirály podle metody dr. Pullmana (opět uvedena v 36SP). Zběžně: Tato metoda z dat vypočítává pět koeficientů  $I_1$  až  $I_5$ . Jejich význam je standardní odchylka 1. a 2. derivace, „těsnost“ spirály a procentuální podíl extrémů a inflexních bodů. Z těchto pěti koeficientů je následně vypočítána DOS (degree of severity), která by měla odpovídat závažnosti postižení na stupnici (0 až 4). Tato metoda ovšem dává nesmyslné výsledky, proto jsem se začal věnovat i metodám jiným.

V průběhu výpočtu koeficientů  $I_x$  se pro jednotlivé body vypočítávají různé parametry a výsledek výpočtu lze uložit do souboru. Přiložil jsem tedy i tyto výstupy. Jsou zabalené do souborů output-mx.zip, kde x značí, kolik bodů bylo při výpočtu použito (každý 1./2./3./4. bod). Z těchto výstupů vznikly i soubory m1 až m4 ve variantách pro JavaNNS a Weka. Obsahují vypočítané koeficienty  $I_1$  až  $I_5$  spolu s hodnocením, zda spirála patří nemocnému člověku.

Tyto soubory pak byly předloženy uvedeným nástrojům a analyzovány. Nejdříve jsem zkoušel analýzu pomocí JavaNNS. Hodnoty koeficientů vybočují z rozsahu 0 až

1, ale jelikož se neodvažuji hádat jakých rozsahů by mohli dosahovat, pracoval jsem s neurony, které byly schopny zvládnout i hodnoty mimo rozsah  $\langle 0;1 \rangle$ . Po různých experimentech s různými sítěmi, neurony a vstupními soubory, se mě podařilo vytvořit síť, která data zařadila na 100% správně (testovací množina nebyla použita, protože bylo málo dat). Síť měla pět standardních vstupů, bez skryté vrstvy a jedním výstupním neuronem s aktivační funkcí `act_perceptron`. Kučení byl použit algoritmus `backpropagation`. Jako vstup analyzovala soubor `m3`, tedy každý třetí bod na spirále.

Nakonec jsem data analyzoval v nástroji Weka za pomoci nástroje klasifikace na různých algoritmech. Většinou byl při testování používán postup `cross-validation`. Nejzajímavější výsledky uvádím v tabulce:

Použitá funkce	Procento správně klasifikovaných	Procento špatně klasifikovaných	Vstupní soubor
Logistic Function	95	5	m4
BayesNet	90	10	m1
MultilayerPerceptron	90	10	m4
Threshold Selector	90	10	m3
NaiveBayes	85	15	m2
Decision Stumps	75	25	m1
J4.8 Tree	70	30	m2

## Diskuse:

Původní záměr hodnotit pacienty podle stupně postižení nemohl být splněn, protože dostupná data nesla informaci jen o tom, zda člověk, který spirálu kreslil je parkinsonik nebo není. Mým prvním záměrem při testování spirál bylo využít síť RSOM. To se mě však z časových důvodů nepovedlo splnit, protože výpočet v síti RSOM je značně časově náročný. Jeden průchod pětiset řádkovým vektorem trvá kolem pěti minut. Jedna originální spirála se skládá z 2500 až 5000 řádků. Pro naučení se s programem pracovat, nalezení vhodných parametrů a naučení sítě na dvacet spirál jsem čas nenašel.

Použité metody se s daty vypořádali s výše uvedenými výsledky.

Zdánlivě nejlépe dopadla síť z `JavyNNS`. Její topologie byla zvolena také s ohledem na to, abych alespoň trochu ověřil relevantnost DOS. Výsledek 100% je vynikající, ale když si uvědomíme, že nebyla použita žádná testovací množina, mohla být síť jen přeučená (a pravděpodobně taky byla). Na druhou stranu oproti ostatním testovaným sítím se alespoň naučit dokázala. Ostatní buď neustále oscillovaly nebo každou spirálu hodnotili jako od parkinsonika nebo naopak každou od zdravého člověka.

Pro jednoznačnější závěr by testování pomohl mnohem větší soubor dat (tento poznatek lze aplikovat i na ostatní použité metody – dvacet spirál je opravdu málo). Při větším množství bude nejspíše nutné změnit i topologii sítě: současný výstup se

stane jedním z neuronů 1. skryté vrstvy a na výstupu bude buď funkce OR nebo perceptron backpropagation. (A ani tato síť samozřejmě nemusí vůbec fungovat.)

U výsledků z nástroje Weka již lze o různých metodách něco usoudit, jelikož byla použita i metoda validace. Logistická funkce je podobná výpočtu DOS a uvedené síti v JavaNNS. To by mohlo vypovídat o tom, že stačí u formule DOS změnit váhu jednotlivých koeficientů a mohla by dávat správné výsledky. Opět by bylo potřeba metodu vyzkoušet na větším souboru dat.

Ani ostatní uvedené metody však nedopadly zvláště špatně a na větším souboru dat by mohly získat výsledky lepší (nebo taky horší).

## Závěr:

Pro učinění nějakého konkrétního závěru je málo testovaných spirál. Bylo by potřeba sehnat další spirály. Také by pomohl tablet, ze kterého by se daly získat dynamické informace o kreslených spirálách (rychlost, tlak...)