



národní
úložiště
šedé
literatury

Nástroj pro vzdálené použití NNSU algoritmu pro separaci dat (uživatelský manuál)

Hakl, František
2013

Dostupný z <http://www.nusl.cz/ntk/nusl-170416>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 30.07.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní [nusl.cz](http://www.nusl.cz) .



Institute of Computer Science
Academy of Sciences of the Czech Republic

Nástroj pro vzdálené použití NNSU algoritmu pro separaci dat (uživatelský manuál)

František Hák

Technical report No. 1200

December 2013

Abstrakt:

Tento manuál popisuje základní použití serveru NNSU (paralelní implementace neuronové sítě s přepínacími jednotkami), který umožňuje vzdálený přístup k implementaci algoritmu NNSU a jeho pilotní použití na separování dat zaslaných na server. Účelem této volně přístupné aplikace je otestování vhodnosti separátoru na separaci uživatelských dat.

Obsahem tohoto uživatelského návodu jsou informace postačující k využívání NNSU serveru, které popisují způsob práce s daty určenými k separaci, způsob definování použité neuronové sítě, zadání výpočtu a metody hodnocení výsledné kvality separace.

Keywords:

Separace dat, neuronové sítě, paralelní výpočty

Nástroj pro vzdálené použití NNSU algoritmu pro separaci dat (uživatelský manuál)

Ing. František Hakl, CSc.
Ústav Informatiky AV ČR v. v. i.,
Pod Vodárenskou věží, 182 07 Praha
e-mail: hakl@cs.cas.cz

15. prosince 2013

Abstrakt

Tento manuál popisuje základní použití serveru NNSU (paralelní implementace neuronové sítě s přepínacími jednotkami), který umožňuje vzdálený přístup k implementaci algoritmu NNSU a jeho pilotní použití na separování dat zaslaných na server. Účelem této volně přístupné aplikace je otestování vhodnosti separátoru na separaci uživatelských dat.

Obsahem tohoto uživatelského návodu jsou informace postačující k využívání NNSU serveru, které popisují způsob práce s daty určenými k separaci, způsob definování použité neuronové sítě, zadání výpočtu a metody hodnocení výsledné kvality separace.

URL NNSU serveru

<http://www.cs.cas.cz/nnsu/>

Obsah

1	Obecné poznámky	4
2	Separace pomocí neuronové sítě s přepínacími jednotkami	6
3	Postup práce s NNSU serverem	7
4	Vytvoření uživatelského účtu	8
5	Výběr aplikační oblasti	10
6	Vytváření adresářové struktury	11
7	Zaslání učících dat na server	12
8	Zaslání testovacích dat na server	16
9	Zaslání RAW dat na server	17
10	Vytvoření neuronové sítě	18
11	Spuštění procesu učení	22
12	Ošetření nedokončené fronty	24
13	Vyhodnocení výsledků separace na datech	25
	13.1 Výsledky na testovacích datech se všemi druhy grafů	27
	13.2 Zobrazení výsledků na RAW datech	27
	13.3 Ukázka porovnání histogramů pro učící a testovací data	28
14	Doplňující nástroje a konfigurace	28

1 Obecné poznámky

Před detailním popisem používání NNSU serveru zmíníme nejdříve několik obecných poznámek.

SYNCHRONIZACE SERVERU A APLIKACE U KLIENTA

NNSU server je de facto distribuované uživatelské rozhraní sloužící k ovládní sofistikovaného separačního algoritmu. Toto rozhraní je realizováno pokud možno co nejvíce obecně, aby uživateli poskytlo dostatečné prostředky pro pilotní testy vhodnosti separace jeho dat pomocí NNSU. Tato obecnost je mimo jiné dána možností vytváření libovolné struktury adresářů na straně serveru a zaslání dat do těchto adresářů. Současně, odezva algoritmu je závislá na velikosti neuronové sítě použité k výpočtu, velikosti datových množin a výpočetní kapacity serveru, na kterém jsou výpočty prováděny. Optimalizace NNSU algoritmu je realizována pomocí genetických algoritmů, což všechno způsobuje, že jednotlivé fáze použití NNSU serveru (a zejména hlavně fáze učení separátoru na učicích datech) může být časově velmi zdoluhavé (např. hodiny výpočtů). Proto je velmi důležitá synchronizace mezi vnitřním stavem NNSU serveru a stavem www prostředí na straně klienta. Této synchronizace je dosaženo striktním dodržováním principu řízení NNSU serveru pouze přes odkazy na www ploše prohlížeče. Veškeré smysluplné akce odpovídající vnitřnímu stavu NNSU serveru jsou dostupné přes odkazy na www stránce. Z důvodu synchronizace nelze používat navigační tlačítka prohlížeče, konkrétně tlačítek "Zpět (Back)", "Dopředu (Forward)", "Znovu načíst (Reload)", stejně tak nelze používat odkazy v historii prohlížených stránek, nebo v záložkách. Jejich použití nejspíše povede k destabilizaci NNSU separátoru a ukončení komunikace mezi serverem a klientem. Po tomto přerušení je možno se znovu přihlásit k aplikaci (s možným zpožděním dle následující poznámky).

ZAMEZENÍ VÍCENÁSOBNÉHO PŘÍSTUPU

Z důvodu zamezení současného přístupu na uživatelský účet z různých IP adres i z různých instancí prohlížeče (či různých prohlížečů) v rámci jednoho počítače, je povolen v daném čase pouze jeden přístup k danému uživatelskému účtu. Toto omezení je dáno pouze v časovém rozsahu mezi přihlášením a odhlášením se z uživatelského účtu. V případě, že z nějakého důvodu nedojde k regulárnímu odhlášení se z účtu (přerušena komunikační spojnice mezi serverem a klientským počítačem, vypnutí klientského počítače) či nedostupnosti počítače, kde je uživatel přihlášen (při odchodu z práce se zapomnělo odhlášení uživatele a uživatel již není v dosahu počítače), je možné opětovné připojení až po uplynutí předem nastavené doby nečinnosti, cca 60 minut. Uživateli je tato skutečnost oznámena při pokusu o přihlášení se k aplikaci (samotný přístup na přihlašovací stránku nemění dobu nečinnosti).

JAZYK APLIKACE

Uživatelské rozhraní NNSU je dostupné pouze v anglickém jazyce. Důvodem upřednostnění anglického jazyka je snaha nabídnout tento nástroj co nejširší odborné komunitě jak ze sféry akademické, tak ze sféry průmyslové. Z důvodů omezené časové kapacity řešitelského týmu, který vyvíjel tuto aplikaci, nebylo přistoupeno k vytvoření další jazykové verze. V následujících kapitolách je proto uživatelské rozhraní demonstrováno na ukázkách, které jsou v angličtině.

PODMÍNKY PRO ULOŽENÍ DAT

Účelem NNSU serveru je poskytnout odborné veřejnosti možnost okamžitého a dostupného testování separace dat pomocí modelu neuronové sítě označené jako NNSU. Tento server není zamýšlen pro robustní a pravidelné využívání aplikacemi založenými na separaci dat, predikci časových řad či na aproximaci funkčních závislostí. Trvalé uložení dat zaslaných na server a výsledků dosažených na těchto datech není zaručeno. Doporučujeme uživatelům si numerické výsledky pro účely archivace stáhnout prostřednictvím NNSU serveru na svůj počítač.

SOFTWAREVÁ KOMPATIBILITA

Klientská část NNSU serveru byla vyvíjena a robustně testována v prohlížeči **Firefox** od společnosti Mozilla. Základní funkčnost byla otestována i pro IE, Operu a Safari. Vzhledem k rozdílnému chování těchto prohlížečů a i jejich jednotlivých verzí, není pro tyto prohlížeče garantována plná funkčnost některých kontrolních java skriptů na straně klienta.

ZŘEKnutí SE ODPOVĚDNOSTI

TATO SLUŽBA JE POSKYTOVÁNA "TAK, JAK JE", BEZ ZÁRUKY JAKÉHOKOLIV DRUHU, AŽ VÝSLOVNÉ ČI ODVOZENÉ, VČETNĚ, ALE BEZ OMEZENÍ POUZE NA, VHODNOSTI PRO URČITÝ ÚČEL A JAKÉHOKOLI POUŽITÍ VÝSLEDKŮ, DOPORUČENÍ A INFORMACÍ, VZNIKLYCH NA ZÁKLADĚ TĚTO SLUŽBY. AUTOŘI A PROVOZOVATEL TĚTO SLUŽBY V ŽÁDNÉM PŘÍPADĚ NENESOU PRÁVNÍ ANI JINOU ODPOVĚDNOST ZA JAKÝKOLI NÁROK, ŠKODU NEBO JINÝ DRUH ODPOVĚDNOSTI, VZNIKLY AŽ JIŽ NA ZÁKLADĚ SMLOUVY, PORUŠENÍM PRÁV NEBO JINAK, VYPLÝVAJÍCÍ Z, NEBO V SOUVISLOSTI, S TOUTO SLUŽBOU, POUŽITÍM TĚTO SLUŽBY, NEBO JINÝM NAKLÁDÁNÍM S TOUTO SLUŽBOU.

WARANTY DISCLOSURE

THE SERVICE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SERVICE OR THE USE OR OTHER DEALINGS WITH THE SERVICE.

DOSTUPNOST TOHOTO MANUÁLU

Tento návod je k dispozici v elektronické formě na stránkách NNSU serveru na adrese:

<http://www.cs.cas.cz/nnsu/>

2 Separace pomocí neuronové sítě s přepínacími jednotkami

Neuronová síť s přepínacími jednotkami označovaná jako NNSU (z anglické zkratky Neural Network With Switching Units) je model umělé neuronové sítě kombinující model umělé neuronové sítě známý jako Multilayer Perceptron a algoritmus označovaný jako lineární rozhodovací strom. Tyto dva modely jsou doplněny heuristickým metodami shlukové analýzy, které v prostoru separovaných dat detekují shluky sobě podobných dat, které jsou separovány nezávisle na datech v ostatních detekovaných shlucích.

Detailní informace o algoritmu NNSU jsou k dispozici v odborné literatuře, která je dostupná přímo na úvodní webové stránce NNSU separátoru. Protože primárním účelem této příručky je především poskytnout uživateli základní postupy a znalosti potřebné k použití implementovaného algoritmu pro řešení vlastních úloh separace dat, nebudeme zde dále rozebírat vlastní algoritmus NNSU, nýbrž se zaměříme na praktické postupy vedoucí k využívání serverové implementace NNSU algoritmu.

Základní scéna, ve které se odehrává separace metodami umělých neuronových sítí, je tzv. "učení s učitelem", kdy máme k dispozici data ve dvou třídách s tím, že víme, které vektory dat patří do které třídy. Pro primární separaci používáme pouze dvě třídy, pracovně je budeme označovat jako "signál" a jako "pozadí".

- Takováto data budeme označovat jako učící data, jejich znalost bude využita pro definování parametrů separátoru NNSU.
- Dále budeme rozlišovat testovací data, u kterých je, stejně jako v případě učících dat, známa příslušnost ke třídám "signál" a "pozadí". Na rozdíl od učících dat, testovací data nejsou využita pro nastavení parametrů separátoru. Porovnání výsledků separace dosažené na testovacích datech se skutečným zařazením testovacích dat do tříd slouží k verifikování robustnosti separace a detekování případného jevu přeučení dat (ang. overlearning). Současně v případě časově proměnného systému generujícího testovaná data můžeme porovnáním výsledků na testovacích datech pořízených v různých časových obdobích detekovat takové změny systému, které vedou k narušení kvality separace. V takovémto případě je na místě poslední testovaná data použít jako nová data učící, a tímto způsobem adaptovat separaci na nové podmínky, ve kterých se primární systém produkující data nachází.
- Posledním typem dat, která je nutno rozlišovat v souvislosti s používáním NNSU separátoru jsou data "ostrá", která budeme označovat jako RAW data. Tato data vznikají v režimu systému produkujícího data, ve kterém nejsme schopni říci, zda takto vzniklá data patří do třídy "signál", nebo do třídy "pozadí". Separace těchto dat je v praktickém případě hlavním cílem celého separačního procesu.

Jako příklad těchto dat mohou sloužit data získaná vyšetřením pacientů, kdy s odstupem času víme, u kterých pacientů sledovaná choroba propukla a u kterých ne. Takováto data lze použít jako data učící a testovací. Data získaná vyšetřením pacientů v současnosti jsou RAW data, jejich klasifikaci pomocí neuronové sítě můžeme podpořit rozhodnutím, které pacienty zařadit do skupiny, u které je třeba přijmout nějaké rozhodnutí o následné léčbě, a které zařadit do skupiny, u které nebude podniknuto žádné zdravotní opatření.

3 Postup práce s NNSU serverem

Využití prezentované služby používání NNSU separátoru je realizováno v následujících krocích:

1. Nejprve je třeba založit a potvrdit uživatelský účet na NNSU severu.
2. Dalším krokem je vytvoření pracovního adresáře nového projektu. NNSU server je koncipován tak, že je možno vytvářet libovolnou strukturu adresářů a podadresářů, což umožňuje přehledně strukturovat rozsáhlé experimenty s daty. Adresáře lze vytvářet v jakékoli fázi práce s NNSU serverem.
3. V daném podadresáři vytvořené adresářové struktury je nejprve nutno zaslat učící data. Učící data definují dimenzi separační úlohy, a teprve po zaslání učících dat je možno definovat neuronovou síť pro výpočet, či přidat testovací data pro verifikaci robustnosti separace, nebo přidat neznámá data (RAW data), u jejichž separace je primárním cílem celého separačního procesu.
4. Nejpřirozenějším postupem je po vložení učících dat na server definovat neuronovou síť. Po definici této neuronové sítě je možno spustit proces učení.
5. Po naučení neuronové sítě se uživateli zpřístupní grafické zobrazení výsledků k vizuálnímu vyhodnocení kvality separace. Současně s tím si uživatel může stáhnout numerické výsledky separace pro svoji vlastní specifickou analýzu separačních výsledků.
6. V průběhu další práce je možno přidávat další testovací množiny, další RAW data, a současně definovat další neuronové sítě pro naučení na učících datech.

V následujících kapitolách se seznámíme s provedením výše uvedených jednotlivých úkonů.

4 Vytvoření uživatelského účtu

NNSU server je dostupný na adrese:

<http://www.cs.cas.cz/nnsu/>

Na této adrese se objeví přihlašovací stránka NNSU serveru.



Can NNSU algorithm help to solve my problem?

NNSU server login gate:

Name:

Password:

If you forgot both, name and password, please [contact us.](#)

Go directly to the previous position

NNSU related articles:

- **High energy physics application:**
 - journals articles
 - [Methods for Multidimensional Event Classification: a Case Study Using Images from a Cherenkov Gamma-ray Telescope \(in Nuclear Instruments and Methods in Physics Research A\)](#)
 - [Application of Neural Networks to Higgs Boson Search \(in Nuclear Instruments and Methods in Physics Research A\)](#)
 - [Hadronic tau's Identification Using Artificial Neural Network, \(ATLAS Physics Communication ATL-COM-PHYS-2005-044\)](#)
 - slides, etc.
 - [Application of Neural Networks optimised by analysis Genetic Algorithms: case of ttH, H -> bb Analysis \(preliminary results\)](#)
 - [HEP\(LHC\) Data Separation Based on Subdivision of Input Space](#)
 - [NNSU and HARM Methods Support for Cuts Method.](#)
 - [Application of Neural Networks Optimized by Genetic Algorithms to Higgs Boson Search](#)
- **Time series forecasting:**
 - [Seasonal Time Series Modeling via Neural Networks With Switching Units \(PHD. thesis\)](#)
 - [Neural Network with Cooperative Switching Units with Application to Time Series Forecasting \(conference proceeding\)](#)
- **NNSU algorithm:**
 - [Optimization of Neural Networks Architectures Using Genetic Algorithm \(PHD. thesis\)](#)
 - [Parallelization of Distributed Genetically Optimized Neural Network Computations \(Master thesis, in Czech\)](#)
 - [Monitoring of a Learning Process of NNSU \(Master thesis, in Czech\)](#)

© Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague. This project was supported by grant of the TAČR under the contract No. TA01010490.

Přihlašovací stránka obsahuje několik odkazů dokumentujících použití modelu neuronových sítí s přepínacími jednotkami v různých oblastech zpracování dat, pro obecnou informaci.

Pro uživatele NNSU serveru nabízí následující tři možnosti:

- přihlášení se do aplikace, má-li uživatel již vytvořen uživatelský účet na NNSU serveru.
- přechod na stránku, na které je možno vytvořit nový uživatelský účet.
- v případě zapomenutí hesla je možno prostřednictvím této stránky zaslat dotaz na heslo, odpovídající zadanému NNSU účtu. Současně je možný přímý kontakt na správce serveru prostřednictvím e-mailu.

Na úvodní stránce je možné také zvolit možnost přejít po přihlášení přímo do uživatelského adresáře ve kterém se uživatel pohyboval naposledy při předchozím přihlášení (za předpokladu, že došlo k regulárnímu odhlášení se z aplikace).

V případě prvního použití NNSU serveru je třeba nejdříve vybrat položku vytvoření nového účtu, která zobrazí formulář pro zadávání osobních dat nového uživatele.

New NNSU account sheet:

Account name: (mandatory, a-z,A-Z,0-9, no funny local chars)

Email address: (mandatory)

Password: (mandatory, a-z,A-Z,0-9, no funny local chars)

Retype passwd: (mandatory)

Personal:

Name, family/nick name: (optional)

Institute: (optional)

Mailing address: line 1 (optional)

line 2 (optional)

line 3 (optional)

Scope of interests: (optional)

Comment (optional) (html hints):

THE SERVICE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SERVICE OR THE USE OR OTHER DEALINGS WITH THE SERVICE.

Yes, I accept.

[Make a new NNSU account](#)

[Go back](#)

Ve formuláři pro vytvoření nového uživatelského účtu jsou *povinné* položky jméno účtu, které smí obsahovat pouze velké a malé znaky **anglické** abecedy a číslice, dále platná e-mailová adresa¹ a přístupové heslo k účtu (tvořené opět pouze velkými a malými znaky **anglické** abecedy a číslicemi). Dále je možno zadat další osobní údaje, doručovací adresu a případně komentář. Povinnou akcí při vytváření účtu je odsouhlasení "Zřeknutí se odpovědnosti – Warranty Disclosure".

New NNSU user:

Account name: josef123

Email address: mob@centrum.cz

Personal:

Name, family/nick name: Josef Novák

Institute: Ustav pro výzkum dat

Mailing address: line 1

line 2

line 3

Scope of interests: data, data, data

Comment
any comment

To submit this new user account,

insert the new user password

and confirm number send to your e-mail

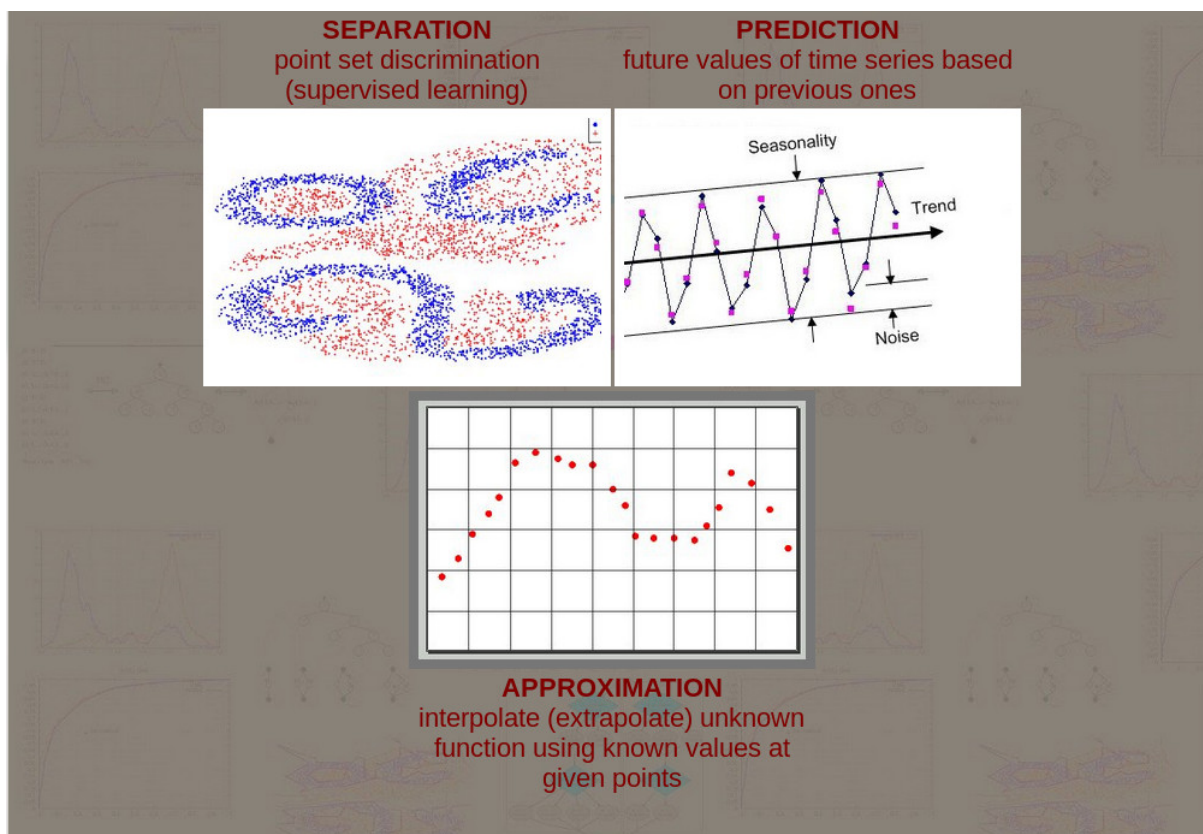
[Confirm new NNSU account](#) [Reset password](#) [Go back](#)

¹Emailová adresa nemůže být z technických důvodů v doméně cs.cas.cz.

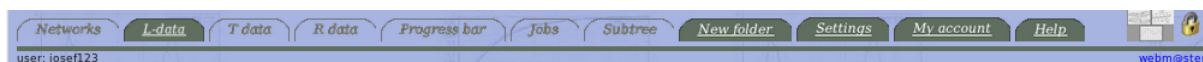
Po odeslání vyplněného formuláře aplikace vygeneruje náhodné potvrzovací číslo a toto číslo zašle na uvedenou e-mailovou adresu. Uživatel se poté zobrazí potvrzovací formulář, kde uživatel zadaným heslem k účtu a potvrzovacím číslem potvrdí vytvoření nového uživatelského účtu.



5 Výběr aplikační oblasti

Po vytvoření nového účtu se uživatel dostane na rozcestník aplikace, kde lze zvolit typ práce s neuronovou sítí s přepínacími jednotkami. Primárně je NNSU určeno pro separaci dat, ale je zjevné, že úlohy predikce časových řad a aproximace funkcí se dají do jisté míry vyjádřit jako separační problém a vice versa. Na úlohu predikovat vzrůst nebo pokles následné hodnoty v dané časové řadě se dá částečně pohlížet jako na úlohu separace, kdy separujeme vektor předchozích hodnot do dvou tříd, definovaných podle nárůstu či poklesu následující hodnoty oproti poslední známe hodnotě. Stejně tak v případě spojité pravé strany datových množin, se dá na neuronovou síť, snažící se zobrazit vektory vstupních hodnota na zadané reálné výstupní hodnoty pohlížet jako na aproximátor, aproximující funkci více proměnných do definičního oboru reálných čísel. Webová aplikace proto umožňuje vybrat jednu z těchto tří oblastí. Způsob práce pro všechny tři oblasti je zhruba totožný, rozdíl je de facto pouze ve způsobu zobrazení výsledků, který je volen tak, aby zohledňoval buď požadavek na separaci, predikci či aproximaci. V následujícím textu budeme popisovat pouze tu část, která se týká separace. Postup práce v části pro aproximaci či predikci je obdobný.



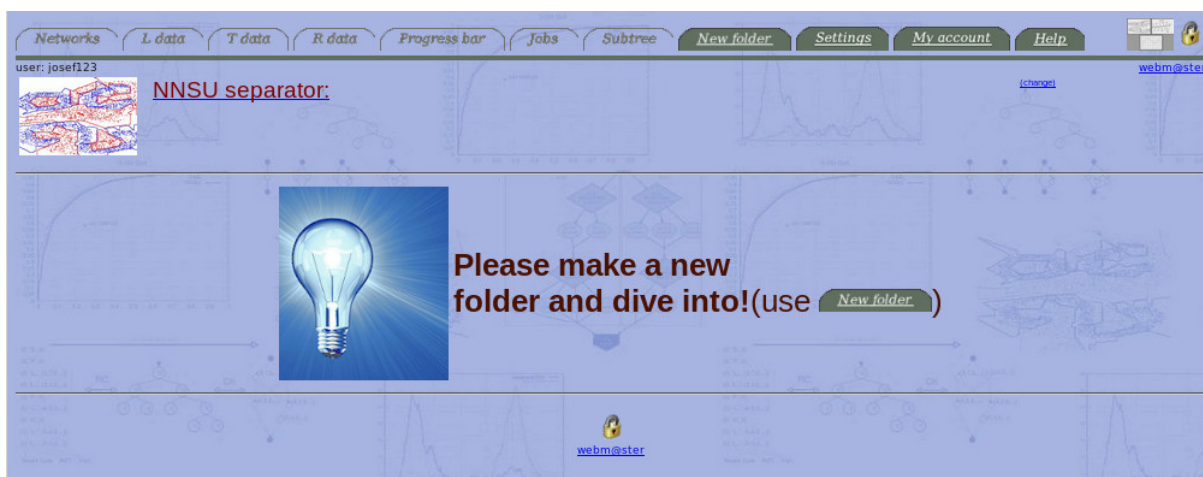
Po vstupu do příslušné oblasti je k dispozici nabídka menu, která umožňuje základní práci s aplikací. Nabídka v menu je kontextově závislá, t.j. aktivní jsou pouze ty položky v menu, které jsou v daném okamžiku smysluplné a realizovatelné.



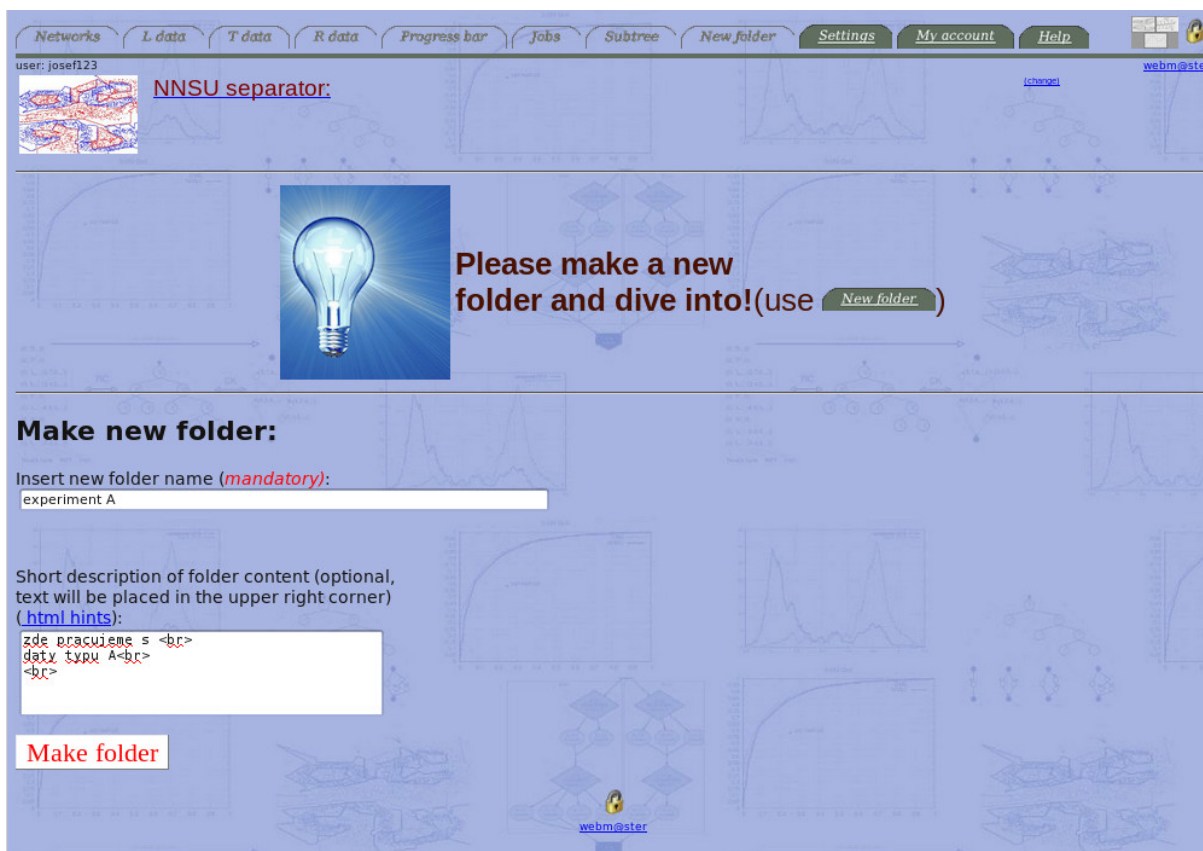
V pravé části menu je ikonka pro ukončení aplikace  a ikonka pro přímý návrat rozcestník . Akce svázané s jednotlivými položkami v menu jsou diskutovány v následujících kapitolách.

6 Vytváření adresářové struktury

Při počátku práce musíme nejdříve v nám přidělené datové oblasti vytvořit první adresář s daty.



1 K formuláři pro vytvoření adresáře se dostaneme pomocí odkazu [New folder](#).



Zobrazí se formulář, ve kterém vyplníme jméno adresáře a popis tohoto adresáře. V poli popisu lze používat jednodušší konstrukce jazyka html. Popis adresáře je možno v průběhu další

práce měnit (odkaz *change* v pravé části okna prohlížeče), např. lze měnit komentář ohledně provedených výpočtů. Jméno adresáře je nadále již neměnné.

Po zaslání formuláře pro vytvoření adresáře se zobrazí potvrzovací formulář, ve kterém můžeme zkontrolovat popis adresáře a po zadání hesla adresář vytvoříme.

7 Zaslání učicích dat na server

Po vytvoření prvního adresáře se nám zpřístupní možnost zadávat učicí data a v poli prohlížeče se zobrazí základní postup práce s daty a neuronovou sítí.

The screenshot shows the NNSU separator web interface. At the top, there is a menu bar with buttons for 'Networks', 'L-data', 'T-data', 'R-data', 'Progress bar', 'Jobs', 'Subtree', 'New folder', 'Settings', 'My account', and 'Help'. Below the menu bar, the user is identified as 'user: josef123'. The main content area features a lightbulb icon and a list of instructions:

Please make a new folder and dive into (use **New folder**),

OR

1. define learning data first (use **L-data**)
2. define network(s) and select some of them (use **Networks**)
3. define test and/or raw data and select some of them (use **T-data** and/or **R-data**)
4. run job(s) and wait ..., :- (... (see **Progress bar**)
5. browse results :-)

At the bottom of the instruction panel, there is a 'webmaster' link.

Ze všeho nejdříve je třeba zadat učící data. To z toho důvodu, že učící data definují dimenzi separační úlohy, a tato dimenze je nezbytná k vytvoření jak neuronových sítí, které v daném adresáři budou s daty pracovat, tak pro kontrolu konzistentnosti následně zadávaných testovacích dat a RAW dat s již zadanými daty učícími. Formulář pro zadání učících dat vyvoláme položkou **L-data** v menu. Učící data lze v daném adresáři zadat pouze jednou.

The screenshot shows the 'Insert learning data' section of the NNSU separator web interface. The section title is 'Insert learning data:'. Below the title, there is a paragraph explaining that learning data files are submitted here and that users cannot change previously submitted data. It also mentions that learning data are text files with a fixed number of floats per line and a desired NNSU output value at the end of each line.

Example of learning data format

0.35	458	54.21	0.023	45	-1
1.78	52	56.25	0.145	3.12	1
0.89	124	48.23	0.078	8.15	1
0.65	312	62.37	1.4	2.45	-1
0.68	258	50.5	1.35	6.84	-1
... etc. ...					

Description of learning data set (optional) ([html hints](#)):

```
demonstrační data<br>
pro učení NNSU<br>
včetně ohodnocení <br>
```

LEARN data:

demonstrační-data-LEARN-nnsu

GAIN data (optional):

demonstrační-data-LEARN-GAIN

(GAIN data are weights of learn events (rows). Multiple columns are allowed in the gain data. If omitted, the desired NNSU output values will be used to get gain plot. Use the same delimiter as in the learn data file.)

Example of corresponding gain data

-0.15	-59	-8.79
0.25	27	56.25
0.61	86	-18.78
-0.45	-156	37.24
-0.08	-134	-15.5
... etc. ...		

Ve formuláři pro zadávání učících dat jsou stručně uvedeny základní informace o použití učících dat a jsou zde uvedeny příklady, jak taková data mají vypadat. Vkládaná data musí být pouze v textové formě, na každé řádce musí být jeden datový vektor následovaný hodnotou rozdělující tyto datové vektory do dvou tříd – této hodnotě nadále budeme říkat **indikátor**.

Abychom vyšli vstříc uživatelům, jsou podporovány čtyři způsoby oddělení dat, a to pomocí znaků "mezery" – whitespaces, ":" – colon, ";" – semicolon a "," – comma. Uživatel si jeden z těchto 4 separátorů může vybrat. Dále je možno zvolit typ indikátoru, je možno volit mezi (-1, +1), (0, 1) a spojitý separátor +/- – float (předpokládáme, že hodnoty 1 či kladné float hodnoty jsou přiřazeny pozitivním datovým množinám tzv. "signálu", ale není to nutné).

Dále ve formuláři pro zadání učicích dat zadáme popis těchto dat (opět možno použít konstrukce jazyka html) a dále musíme zadat soubor s učícími daty a můžeme, ale nemusíme zadat soubor s GAIN daty. V případě nezadání tohoto souboru se za vektor vah vstupních vektorů bere sloupec indikátorů vstupních vektorů. Po odeslání vyplněného formuláře dojde k zaslání datových souborů na server do dočasného datového prostoru, kde dojde ke kontrole syntaktické správnosti zaslanych dat, t.j. konkrétně se zkontroluje zda všechny vstupní vektory na řádcích mají stejnou dimenzi (kdy oddělovač jednotlivých hodnot je dán zvoleným separačním znakem), zda každý řádek obsahuje pouze řetězce, které lze interpretovat jako float, zda počet vstupních vektorů deklarovaných indikátory jako signál je nenulový, stejně tak, že počet vektorů deklarovaných indikátory jako pozadí je také nenulový (správně zadaná úloha separace musí mít příklady pro obě dvě separované množiny). Podle zvoleného typu indikátorů se také provede kontrola, zda všechny indikátory nabývají povolených hodnot. Dojde-li k porušení některé z těchto kontrolovaných podmínek, uživateli se zobrazí informace v jakém místě zaslaneho souboru toto porušení konzistentnosti souboru učicích dat nastalo. Je-li syntaktická kontrola zaslanych dat v pořádku, zobrazí se pro zaslana data potvrzovací formulář.

Networks L data T data R data Progress bar Jobs Subtree New folder Settings My account Help

user: josef123 webmaster

NNSU separator:
[experiment A](#) (inserted at 14/01/14, 08:48)

[\[change\]](#)
**zde pracujeme s
daty typu A**

Insert data confirmation sheet:

Learning data:

File name:
demonstracni-data-LEARN-nnsu

Description:
demonstrační data
pro učení NNSU
včetně ohodnocení

First ten lines:

-6.991990e-01	2.365446e+00	2.038661e+02	1.219534e+00	1.202596e+00	6.510000e-01	1.114020e-02	3.905144e-01	4.258718e+01	1.106855e+01
-1.506336e-01	3.050116e+00	1.121747e+02	4.326431e-01	9.487013e-01	6.510000e-01	1.920608e+00	1.265683e+00	7.026385e+01	4.803064e-01
-8.765211e-02	2.997748e+00	6.415690e+01	3.567554e-01	2.200944e+00	9.010000e-01	4.954047e-01	2.230110e+00	3.271466e+01	7.107853e-01
8.541767e-01	2.940020e+00	8.280323e+01	-2.130836e+00	3.998681e-01	5.010000e-01	2.438323e+00	2.541391e+00	1.754306e+01	1.496797e+00
7.365260e-01	3.030122e+00	1.243256e+02	8.482895e-01	1.819045e+00	9.260000e-01	-1.296243e+00	8.206816e-01	3.615978e+01	9.373629e-01
-1.042895e+00	2.763246e+00	9.545007e+01	1.344408e+00	2.550441e+00	9.260000e-01	5.551000e-17	2.144902e+00	2.215951e+01	1.095666e+00
2.135763e-01	2.328650e+00	9.219188e+01	-7.551167e-01	2.554372e+00	9.010000e-01	1.563081e+00	2.638561e+00	3.481292e+01	1.134125e+00
-7.407962e-01	2.996550e+00	1.200251e+02	-1.793553e+00	1.268105e+00	9.260000e-01	-9.387599e-01	1.115296e+00	4.790099e+01	8.729717e-01
4.198343e-01	1.790258e+00	3.209485e+01	-1.083862e+00	1.987565e+00	8.510000e-01	-1.512285e+00	2.973905e+00	3.096409e+01	1.237825e+00
-3.736509e-01	1.203240e+00	7.019917e+01	-3.825090e-01	1.003051e+00	8.510000e-01	3.645265e-01	1.135208e+00	3.708669e+01	2.839391e+00

Last ten lines:

1.148888e+00	9.882097e-01	4.118485e+01	1.826642e+00	-2.364635e+00	7.828166e-02	-6.386007e-01	3.489764e-01	2.126018e+01	3.131782e+00
6.920317e-01	1.783721e+00	5.054263e+01	-1.797472e+00	2.834085e+00	1.065479e-01	9.098409e-01	-1.189575e+00	2.551075e+01	1.627656e+00
-9.376578e-01	-2.890581e+00	8.835717e+01	-1.655439e-01	2.091000e+00	1.714541e-01	1.100107e+00	-8.027026e-01	3.248019e+01	5.706570e-01
-1.172386e+00	-1.692206e+00	3.629457e+01	-1.656158e-01	-1.924447e+00	1.437700e-01	1.537532e+00	1.087007e+00	2.732647e+01	2.139015e+00
-1.122830e+00	-2.834447e+00	1.430671e+02	8.178364e-01	7.224996e-01	2.156270e-01	3.827970e-01	-6.131337e-01	4.187297e+01	2.665459e+00
-2.041852e+00	-2.596299e+00	2.555795e+01	-1.211882e+00	3.010460e+00	8.935519e-02	1.714224e-01	-5.431304e-01	2.487318e+01	1.223225e+00
-3.380812e-01	-2.765802e+00	1.601064e+02	-1.786290e-01	2.918739e+00	2.301249e-01	-5.612287e-02	7.271237e-01	1.313298e+02	9.397143e-01
7.446050e-01	1.680978e+00	3.296489e+01	-1.620753e-01	2.664957e+00	7.020013e-01	-1.113286e+00	-6.760205e-01	3.021468e+01	2.361526e+00
1.966261e-01	-2.770828e+00	4.260338e+01	-6.571499e-01	-2.715159e+00	2.270944e-01	3.188480e-01	1.478018e+00	2.490050e+01	3.063802e+00

Summary:

Data dimension = 18, size of data set = 49049, positive examples: 37039, negative examples= 12010

Average values:	-2.54e-3	7.16e-1	7.77e+1	-8.42e-3	6.53e-1	7.14e-1	1.52e-2	6.44e-1	4.39e+1	1.81e+0	7.14e-1	-4.19e+0	3
Minimal value:	-2.64e+0	-3.14e+0	1.58e+1	-3.83e+0	-3.14e+0	7.50e-2	-2.62e+0	-3.14e+0	1.50e+1	4.15e-5	7.50e-2	-1.00e+2	1
Maximal value:	2.58e+0	3.14e+0	4.32e+2	3.74e+0	3.14e+0	9.82e-1	2.65e+0	3.14e+0	3.15e+2	3.14e+0	9.81e-1	9.26e-1	4

Confirm validity of submitted data using your account password.
Remember that the learning data file **cannot be updated later.**

●●●

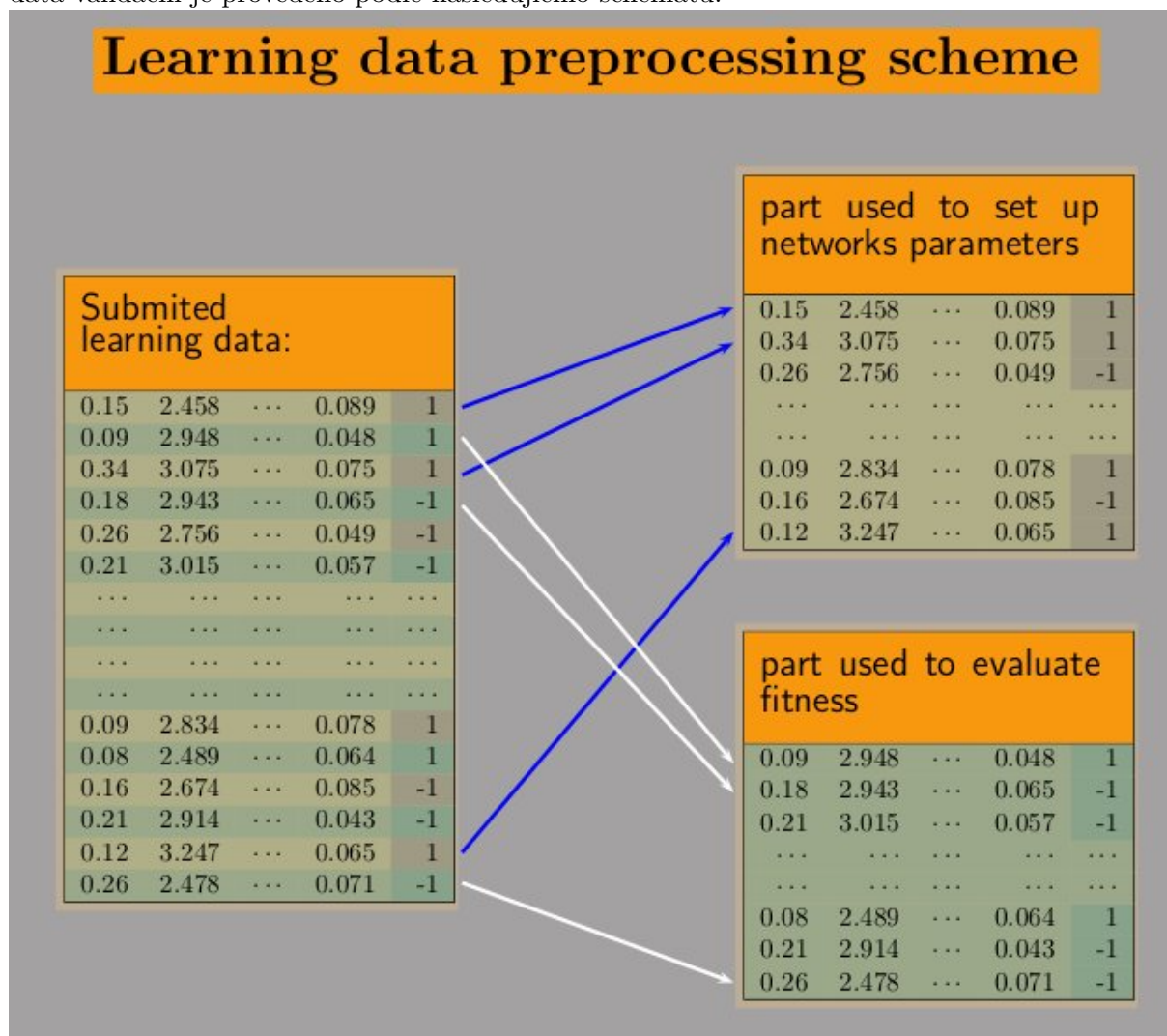
NNSU server

Verze 12/2013

14

V zobrazovacím formuláři jsou uvedeny základní informace a statistiky o zaslaných datech, které uživateli umožní kontrolu správnosti a úplnosti načtení dat na server a po zadání hesla jsou tato zasláná data zařazena do aktuálního adresáře k dalšímu zpracování.

Jak již bylo velmi stručně výše zmíněno, NNSU je algoritmus využívající genetické optimalizace. Z tohoto důvodu jsou učící data rozdělena na dvě stejně velké části, část učící a část validační. Učící část slouží k nastavení parametrů separátoru, zatímco validační část slouží k výpočtu hodnoty fitness funkce, která je využita při optimalizaci architektury NNSU sítě pomocí genetických algoritmů. Rozdělení zaslanych učících dat na vlastní vnitřní učící data a na data validační je provedeno podle následujícího schématu.



Zde je patrné, že v pořadí liché vektory dat jsou použity pro vnitřní učení jedinců v populacích genetického algoritmu a sudé vektory jsou použity jako validační množina při výpočtu fitness funkce. Tuto skutečnost je zde potřeba zmínit z toho důvodu, že statistické vlastnosti souboru učících dat by pro obě tyto podmnožiny měly být stejné. Je nezbytné k tomuto způsobu využití učících dat přihlížet při sběru či generování učícího datového souboru.

Jednou zasláná a potvrzená učící data již nelze dále měnit, jediné co lze měnit je jejich popis, kterým můžeme dokumentovat výsledky práce nad těmito daty. Po zadání učících dat odkaz [L-data](#) zpřístupní formulář pro opravu popisu dat, informaci o těchto datech, a odkaz na zasláný soubor, který si tak uživatel může zpětně downloadovat pro kontrolu.

user: josef123

NNSU separator:
[experiment A](#) (inserted at 14/01/14, 08:48)

[\[change\]](#)
zde pracujeme s daty typu A

Change description of learning data:

The learning data has been submitted already. You can change its description only.

Description of learning data set (optional) ([html hints](#)):

```
demonstrační data<br>
pro učení NNSU<br>
včetně ohodnocení <br>
```

[Change description](#)

LEARNING data:

File name: [demonstracni-data-LEARN-nnsu](#)

Unique ID: d140114101831

Description:
demonstrační data
pro učení NNSU
včetně ohodnocení

Uploaded on 14/01/14 at 10:21

First ten lines:

```
-6.991990e-01 2.365446e+00 2.038661e+02 1.219534e+00 1.202596e+00 6.510000e-01 1.114020e-02 3.905144e-01 4.258718e+01 1.106855e+00
-1.506336e-01 3.050116e+00 1.121747e+02 4.326431e-01 9.487013e-01 6.510000e-01 1.920608e+00 1.265683e+00 7.026385e+01 4.803064e-01
-8.765211e-02 2.997748e+00 6.415690e+01 3.567554e-01 2.200944e+00 9.010000e-01 4.954047e-01 2.230110e+00 3.271466e+01 7.107853e-01
8.541767e-01 2.940020e+00 8.280323e+01 -2.130836e+00 3.998681e-01 5.010000e-01 2.438323e+00 2.541391e+00 1.754306e+01 1.498797e+00
7.365260e-01 3.030122e+00 1.243256e+02 8.482895e-01 1.819045e+00 9.260000e-01 -1.296243e+00 8.206816e-01 3.615978e+01 9.373629e-01
-1.042895e+00 2.763246e+00 9.545007e+01 1.344408e+00 2.550441e+00 9.260000e-01 5.551000e-17 2.144902e+00 2.215951e+01 1.095666e+00
2.135763e-01 2.328650e+00 9.219188e+01 -7.551167e-01 2.554372e+00 9.010000e-01 1.563081e+00 2.638561e+00 3.481292e+01 1.134125e+00
-7.407962e-01 2.996550e+00 1.200251e+02 -1.793553e+00 1.268105e+00 9.260000e-01 -9.387599e-01 1.115296e+00 4.790099e+01 8.729717e-01
4.198343e-01 1.790258e+00 3.209485e+01 -1.083862e+00 1.987565e+00 8.510000e-01 -1.512285e+00 2.973905e+00 3.096409e+01 1.237825e+00
-3.736509e-01 1.203240e+00 7.019917e+01 -3.825090e-01 1.003051e+00 8.510000e-01 3.645265e-01 1.135208e+00 3.708669e+01 2.839391e+00
```

Po vložení učících dat se v daném adresáři zpřístupní položky menu pro definování neuronové sítě (sítě) [Networks](#), vkládání testovacích dat [T-data](#) a vkládání RAW dat [R-data](#). V této fázi práce můžeme pokračovat kteroukoli z nabídnutých možností.

8 Zaslání testovacích dat na server

Obdobným způsobem jako jsme zadali učící data, zadáme i testovací data. K tomu použijeme odkaz [T-data](#) v menu. Postup zadávání dat je zcela obdobný jako postup při zadávání učících dat. Ke každým testovacím datům můžeme opět zadat GAIN soubor, který obsahuje váhy jednotlivých vektorů dat pro případnou definici fitness funkce. Rozdíl oproti zadávání učících dat je pouze v tom, že testovací data svoji dimenzí (počtem čísel na řádku) musí odpovídat předem zadaným učícím datům. Dalším rozdílem je také to, že testovacích dat v daném adresáři můžeme zadat libovolně mnoho. Ke každým zadaným testovacím datům můžeme opět přiřadit popis, který lze později upravit. Dále můžeme libovolně testovací data aktivovat či deaktivovat, což znamená, že nebudou či budou zobrazována v základní ploše aplikace. Aktivací a deaktivací dat můžeme definovat úlohu k výpočtu, jak bude naznačeno později. Současně na stránce s přehledem vložených testovacích dat lze vložená data stáhnout k uživateli pro kontrolu.

Pro úplnost zde uvádíme ukázkou zadávacího a potvrzovacího formuláře pro testovací data.

Insert testing data:

You can submit testing data files here. Data are of three kinds, learning data used by a learning process only, testing data which allow to verify results on learning-independent data, and raw data used as application data.

Our testing data are text files, each line contains fixed number of floats separated by whitespace includes tabs. The last value in a row is a desired NNSU output. -1 or -1 , which indicates the separated sets.

Description of testing data set (optional) (html hints):

toto jsou testovaci data

Example of testing data format

```
0.35 458 54.21 0.023 45 -1
0.89 124 48.23 0.078 815 1
0.65 312 62.37 1.4 2.45 -1
0.68 258 50.5 1.35 6.84 -1
... etc. ...
```

Example of corresponding gain data

```
-0.15 -59 -8.79
0.25 27 56.25
0.61 86 -18.78
-0.45 -156 37.24
-0.08 -134 -15.5
... etc. ...
```

Insert testing data confirmation sheet:

Testing data:

File name: demonstracni-data-TEST-nnsu

Description: toto jsou testovaci data

First ten lines:

```
-1.190003e+00 1.439205e+00 3.368495e+01 -3.688990e+00 9.518359e-01 9.260000e-01 6.32841e-01 2.199556e+00 2.722431e+01 1.514878e+00
8.484865e-01 1.22111e-01 9.659628e-01 6.699526e-02 6.679939e+00 9.260000e-01 1.037362e-01 7.658159e+00 3.69917e+00
1.247755e+00 6.23141e-01 7.154862e+01 -1.099179e-01 2.604959e+00 9.260000e-01 1.261951e+00 2.565826e+00 6.52619e+01 1.709398e+00
1.313613e+00 6.670158e+00 9.676148e+01 1.267327e+00 5.944791e-00 9.260000e-01 1.306817e+00 2.818986e+00 2.698844e+00 1.413447e+00
1.735559e+00 6.555200e+00 1.6161279e-02 -9.819566e-01 2.663859e+00 9.030000e-01 5.949311e-01 1.140494e+00 3.810298e+01 1.132929e-01
-2.31859e-02 9.005000e+00 9.602531e+01 -1.88723e+00 1.889519e-01 9.030000e-01 -4.814769e-01 1.887651e+00 3.649912e+00 2.146405e+00
2.348197e+00 7.729690e+00 5.887626e+01 -1.037456e+00 5.314700e-01 9.260000e-01 9.150745e-01 1.889419e+00 2.312136e+01 1.819379e+00
-8.478949e-02 2.289759e+00 8.692139e+01 -1.844409e+00 6.535428e-01 6.510000e-01 -2.229785e-01 1.890722e+00 8.187159e+00 2.72475e+00
1.179254e+00 2.391471e+00 8.245229e+01 5.662612e-01 2.503899e+00 9.260000e-01 1.492000e-01 1.991849e+00 4.58957e+00 2.62549e+00
1.401369e-01 3.009567e+00 3.841313e+01 -5.957656e-01 1.589971e+00 9.260000e-01 -1.339039e-01 1.594405e+00 1.969765e+01 1.684941e+00
```

Last ten lines:

```
7.510827e-01 -1.481420e+00 4.959374e+01 -1.284239e+00 -2.487331e-01 1.452986e-01 -9.250713e-01 -3.140386e+00 2.547604e+01 2.305801e+00
1.886939e+00 6.772247e-01 7.692510e-01 -4.118414e-01 -2.689977e+00 3.989654e-01 1.181098e+00 -2.627009e+00 3.815710e+01 2.184185e+00
6.196479e-01 2.501503e+00 1.420900e-02 1.867732e+00 2.624174e+00 9.097102e-02 2.964919e-01 2.849384e+00 2.139593e+00 6.790497e-01
1.720929e-01 4.881998e-01 3.302909e-02 -6.861759e-01 2.626215e+00 1.646929e-01 7.771424e-01 1.949809e+00 2.409519e+00 1.097097e+00
1.083695e+00 2.688179e+00 5.744106e+01 1.930002e-02 1.520816e+00 2.160352e-01 6.832169e-01 -2.077151e+00 3.792509e+00 2.000909e-01
2.899514e-01 2.989909e+00 1.562139e-02 1.430000e+00 -9.25231e-01 0.909599e-01 4.977854e-01 -1.181098e+00 4.360929e+00 2.200509e-01
1.569811e+00 2.114919e+00 6.051105e+01 8.960126e-01 -1.206103e+00 1.738969e-01 1.378969e+00 6.630259e-01 5.929295e+00 2.409394e+00
9.867246e-02 2.375206e+00 4.038179e-01 1.448896e+00 2.588909e+00 0.189995e-01 6.878384e-01 1.781326e+00 3.250359e+00 3.855159e-01
-3.026549e-02 -2.671078e+00 3.005117e+01 6.217412e-01 1.985615e-02 9.833805e-02 1.679397e+00 2.892858e+00 2.213154e+01 2.895389e+00
```

Summary:

data dimension = 18, size of data set = 49241, positive examples: 37251, negative examples: 11990

Average values: 1.89e+3 7.28e-1 7.79e+1 -1.08e-2 6.53e-1 7.14e-1 -2.01e-4 6.81e-1 4.39e+1 1.80e+0 7.15e-1 -4.12e+0 3

Maximal values: -2.57e+0 -3.14e+0 1.56e+1 -3.77e+0 3.14e+0 7.50e-2 -2.05e+0 -3.14e+0 1.50e+1 4.59e-4 7.51e-2 1.03e+2 1

Minimal values: 2.59e+0 3.14e+0 -4.22e+2 3.81e+0 3.14e+0 9.79e-1 2.05e+0 3.14e+0 2.90e+2 3.14e+0 9.80e-1 9.28e-1 3

Confirm validity of submitted data using your account password.

Confirm password [input] [Reset password] [Go back]

9 Zaslání RAW dat na server

V případě zadávání skutečných syrových dat (RAW) na server se situace mírně odlišuje. V případě RAW dat nemáme k dispozici žádné indikátory příslušnosti dat do tříd. Separace RAW dat je prvotním úkolem separátoru, neboť se jedná o reálná data, která popisují sledovaný reálný systém, který chceme pomocí separace dat na předpokládaný signál či pozadí (=pozitivní či negativní stav) buď řídit, využívat či eliminovat, dle toho jaké praktické nasazení separačního algoritmu realizujeme. Ve formuláři pro zadávání dat chybí tím pádem možnost definování typu indikátoru, nelze zadat žádný GAIN soubor s vahami jednotlivých vstupních vektorů. Situace je trochu jednodušší než v případě zadávání učicích a testovacích dat. Je nutné pouze mít na paměti, že počet řetězců typu float oddělitelných zvoleným separátorem o jedničku menší, než v případě učicích a testovacích dat. Nesplnění této podmínky bude v průběhu zaslání dat oznámeno uživateli, vložení dat do aktuálního adresáře na serveru je opět nutno potvrdit uživatelským heslem.

Insert RAW data:

You can submit raw data files here. Data are of three kinds, learning data used by a learning process only, testing data which allow to verify results on learning-independent data, and raw data used as application data.

Our raw data are text files, each line contains fixed number of floats separated by whitespace (includes TAB)

Example of raw data format

0.35	458	54.21	0.023	45
1.78	52	56.25	0.145	3.12
0.89	124	48.23	0.078	8.15
0.65	312	62.37	1.4	2.45
0.68	258	50.5	1.35	6.84
... etc. ...				

Description of testing data set (optional) [\(html hints\)](#):

```
toto jsou neznámá data </br>
není k dispozici GAIN soubor </br>
chybí informace o třídě</br>
```

Submit raw data

Opět lze zadat libovolný počet souborů s RAW daty a obdobně jako v případě testovacích dat je možno tato data aktivovat či deaktivovat.

Select RAW data or insert new RAW data here

Select FROM->TO XOR FROM->TO

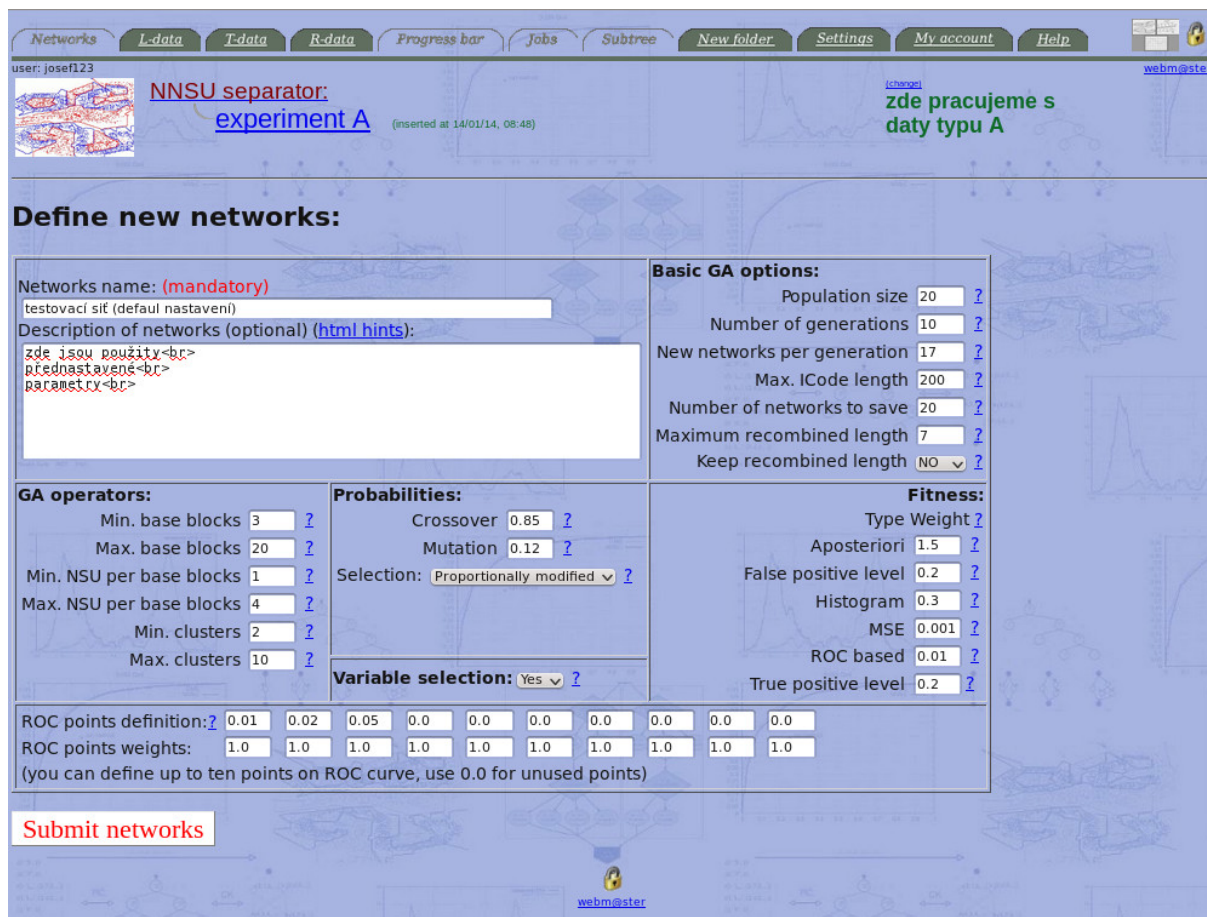
SELECTED	FROM	TO	unique ID	description	name/download	detailed info
<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	d140120153417	druha RAW data	demonstracni-data-RAW-II-nnsu	info
<input type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	d140114103147	toto jsou neznámá data není k dispozici GAIN soubor chybí informace o třídě	demonstracni-data-RAW-nnsu	info

Submit this selection

10 Vytvoření neuronové sítě

Po zadání učicích dat je těmito daty definována dimenze separační úlohy a lze přistoupit kdykoli k vytvoření nové neuronové sítě s přepínacími jednotkami, která bude pracovat se všemi daty v adresáři, která budou aktuálně označena jako aktivní. Formulář k zadání nové sítě se vyvolá volbou **Networks** v menu. Následně můžeme zadefinovat parametry nové sítě. Ve vyvolaném formuláři jsou přednastaveny některé parametry sítě, které rozhodují o velikosti jednotlivých instancí NNSU struktur generovaných při běhu genetické optimalizace, lze nastavit vlastní rozsah genetické optimalizace (např. počet jedinců a generací), je možno definovat typ fitness funkce

a některé parametry cílové ROC křivky. Před nastavením těchto parametrů, které se výrazněji odchyľují od předdefinovaných hodnot, doporučujeme uživateli seznámit se trochu více s algoritmem genetické optimalizace NNSU. Jistou výjimku tvoří nastavení bodů zájmu na ROC křivce, kde je možno definovat váhy jednotlivých bodů na ROC křivce pro fitness funkci. Formulář pro definování sítí obsahuje následující položky:



The screenshot shows a web interface for defining new networks. The main section is titled "Define new networks:". It contains several input fields and sections:

- Networks name:** (mandatory) testovací síť (default nastavení)
- Description of networks (optional):** zde jsou použity přednastavené parametry
- Basic GA options:**
 - Population size: 20
 - Number of generations: 10
 - New networks per generation: 17
 - Max. ICode length: 200
 - Number of networks to save: 20
 - Maximum recombined length: 7
 - Keep recombined length: NO
- GA operators:**
 - Min. base blocks: 3
 - Max. base blocks: 20
 - Min. NSU per base blocks: 1
 - Max. NSU per base blocks: 4
 - Min. clusters: 2
 - Max. clusters: 10
- Probabilities:**
 - Crossover: 0.85
 - Mutation: 0.12
 - Selection: Proportionally modified
 - Variable selection: Yes
- Fitness:**
 - Type Weight: 1.5
 - Apriori: 0.2
 - False positive level: 0.2
 - Histogram: 0.3
 - MSE: 0.001
 - ROC based: 0.01
 - True positive level: 0.2
- ROC points definition:** 0.01, 0.02, 0.05, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0
- ROC points weights:** 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0

At the bottom, there is a "Submit networks" button and a "webm@ster" link.

Stručně zde sumarizujeme významy těchto jednotlivých parametrů. Bližší informace o významu těchto parametrů lze nalézt např. v disertační práci "Optimization of Neural Networks Architectures Using Genetic Algorithm", dostupné na úvodní stránce serveru.

- Parametry definující architekturu jedinců v populacích genetického algoritmu (GA operators):
 - Min. base blocks – minimální počet základních stavebních bloků NNSU v lineárním uspořádání v počáteční generaci
 - Max. base blocks – maximální počet základních stavebních bloků NNSU v lineárním uspořádání v počáteční generaci
 - Min. NSU per base blocks – minimální počet NSU jednotek v základním bloku
 - Max. NSU per base blocks – maximální počet NSU jednotek v základním bloku
 - Min. clusters – minimální počet klastrů v NSU jednotkách
 - Max. clusters – maximální počet klastrů v NSU jednotkách
- nastavení parametrů operací GA (Probabilities):

- Crossover – pravděpodobnost pro operaci crossover na jednotlivých sekvencích Readova kódu
 - Mutation – pravděpodobnost pro operaci mutace na jednotlivých sekvencích Readova kódu
 - Selection – definuje způsob selekce jedinců podle hodnoty fitness. K dispozici jsou následující možnosti: proportional, ranking logarithmic, ranking linear, ranking - exponential, modified proportional.
 - variable selection – povoluje výběr jednotlivých proměnných vstupujících do základních bloků
- Nastavení základních parametrů pro přechod k následující generaci při běhu GA (Basic GA options):
 - Population size – velikost populace GA
 - Number of generations – počet generací
 - New networks per generation – počet nových sítí v následující generaci. Zbytek sítí se přenáší beze změny do další generace podle nejvyšší fitness.
 - Max. ICode length – maximální délka ICode popisujícího nově vytvářenou architekturu sítě.
 - Number of networks to save – počet ukládaných sítí.
 - Maximum recombined length – maximální počet prohozených (či mutovaných) dvojic úseků ICode.
 - Keep recombined length – je-li ano, tak maximální počet prohozených (či mutovaných) dvojic úseků ICode je rovno hodnotě "Maximum recombined length" (více násobné procházení ICode úseků)
 - Definování typu fitness (Fitness): zde je možno zadat váhu jednotlivých typů fitness funkcí (Aposteriori, False positive level, Histogram, MSE, ROC, True positive level).
 - Definice bodů na ROC křivce: zde je možno zadefinovat váhy jednotlivých bodů na ROC křivce a tím zvýraznit oblasti zájmu na ROC křivce. Hodnoty vah mohou být relativně velké, cca 1 – 1000.

Dále musíme zadat název sítě a popis sítě (možno s html konstrukcemi). Popis sítě je možno později měnit. Po odeslání požadavku na vytvoření sítě se zobrazí potvrzovací formulář, ve kterém můžeme zkontrolovat nastavení nově vytvářených sítí, a po zadání hesla se nové sítě přidávají do aktuálního adresáře. V případě potřeby změny v zadání nové sítě se můžeme v tomto bodě pomocí tlačítka "Go back" vrátit na stránku se zadáváním parametrů nové sítě.

user: josef123 [NNSU separator: experiment A](#) (inserted at 14/01/14, 08:48) [zde pracujeme s daty typu A](#)

New networks confirmation sheet:

Networks name: testovací síť (default nastavení) Description: zde jsou použity přednastavené parametry		Basic GA options: Population size 20 Number of generations 10 New networks per generation 17 Number of networks to save 20 Maximum lCode length 200 Maximum recombined length 7 Keep recombined length: NO
GA operators: Min. base blocks 3 Max. base blocks 20 Min. NSU per base blocks 1 Max. NSU per base blocks 4 Min. clusters in neuron 2 Max. clusters in neuron 10	Probabilities: Crossover 0.85 Mutation 0.12 Selection strategy: Proportionally modified Variable Selection: Yes	Fitness: Type Weight Aposteriori 1.5 False positive level 0.2 Histogram 0.3 MSE 0.001 ROC based 0.01 True positive level 0.2

ROC points definition: 0.01 0.02 0.05 0.0 0.0 0.0 0.0 0.0 0.0 0.0
 ROC points weights: 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0

Confirm validity of submitted network parameters using your account password.
 Remember that parameters **cannot be updated later** (except for name and description).

●●●●●

Confirm password

Sítě můžeme v daném adresáři vytvořit libovolně mnoho, a na stránce se seznamem sítí lze jednotlivé sítě aktivovat a deaktivovat obdobným způsobem, jako v případě testovacích a RAW dat (odkazy "info" v pravé části tabulky se seznamem sítí vedou k detailnímu popisu sítí, kde je možno změnit popisy sítí).

user: josef123 [NNSU separator: experiment A](#) (inserted at 14/01/14, 08:48) [zde pracujeme s daty typu A](#)

Select networks or [insert new networks here](#)

Select FROM->TO XOR FROM->TO

SELECTED	FROM	TO	unique ID	description	name	detail info
<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	n140114103629	tíť číslo 3	další příklad (více zvýrazněná ROC)	info
<input type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	n140114103452		další instance NNSU (zvýrazněná počáteční ROC)	info
<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	n140114102631	zde jsou použity přednastavené parametry	testovací síť (default nastavení)	info

Submit this selection

11 Spuštění procesu učení

Máme-li zadána učící data a alespoň jednu neuronovou síť, můžeme přistoupit k procesu spouštění učení neuronových sítí na datech. Na stránce odpovídajícímu danému experimentu je zobrazena tabulka s aktivními sítěmi a aktivními daty (poznamenejme, že učící data jsou vždy aktivní a zobrazena v tabulce). Každý řádek této tabulky vždy odpovídá jedné síti, sloupec tabulky vždy odpovídá jedněm datům. V polích tabulky je buď na červeném pozadí uveden počet NNSU jedinců, které nejsou spočítány a počet grafů s výsledky, které ještě nejsou vygenerovány. V případě, že pro danou síť a daná data je vše spočítáno a zobrazeno, v příslušném políčku tabulky je uveden odkaz "all_results" směřující na stránku s výsledky separace na daných datech.

Jak je patrné z této tabulky, aktivací či deaktivací jednotlivých neuronových sítí či dat vlastně definujeme množinu výpočtů, které můžeme spustit. Je potřeba ještě poznamenat, že výpočty související s políčkem v dané tabulce korespondující s nedopočítanými výsledky (tedy políčko tabulky neobsahující odkaz "all_results") jsou počítány vždy všechny znovu, na částečný výpočet, který nebyl z nějakého důvodu proveden do konce, nelze navázat.

Chceme-li změnit dávku určenou k výpočtu, např. z časových důvodů, kdy nás zajímá výsledek užšího výběru sítí na menším výběru vložených dat, musíme změnit aktivaci sítí či dat, tedy musíme použít odkazy **Networks**, **T-data** a **R-data** v horním menu.

The screenshot shows the 'NNSU separator: experiment A' interface. At the top, there is a navigation menu with buttons for 'Networks', 'L-data', 'Tdata', 'R-data', 'Progress bar', 'Jobs', 'Subtree', 'New folder', 'Settings', 'My account', and 'Help'. Below the menu, the user 'josef123' is logged in. The main content area contains a table with the following structure:

DATA -> NETWORKS	LEARN file demonstrační data pro učení NNSU včetně ohodnocení	RAW file druhá RAW data	TEST file test data c. 4	RAW file toto jsou neznámá data není k dispozici GAIN soubor chybí informace o trídě	TEST file toto jsou testovací data
další příklad (více zvýrazněná ROC)	all results	NNSU 330/330 NA PLOTS 330/330 NA	NNSU 330/330 NA PLOTS 1650/1650 NA	all results	all results
další instance NNSU (zvýrazněná počáteční ROC)	all results	NNSU 420/420 NA PLOTS 420/420 NA	NNSU 420/420 NA PLOTS 2100/2100 NA	all results	all results
testovací síť (default nastavení)	all results	NNSU 220/220 NA PLOTS 220/220 NA	NNSU 220/220 NA PLOTS 1100/1100 NA	all results	all results

At the bottom of the interface, there is a red button labeled 'I want all missing results'.

Tlačítkem **I want all missing results** vytvoříme výpočetní frontu, která je připravena k výpočtu. Tato výpočetní fronta se zobrazí na ploše prohlížeče ve formě tabulky, kde je sumarizováno plánované výpočty. V této tabulce ještě máme možnost odebrat jednotlivé výpočty z fronty (viz. první sloupec tabulky).

V posledním sloupci tabulky s frontou úloh je uvedeno číslo "resource demand", které by mělo korespondovat s časovou náročností výpočtu v řádce. Toto číslo je tvořeno součinem

$$\text{velikost populace} \times \text{počet generací} \times \text{dimenze dat} \times \text{počet dat}.$$

Časová náročnost výpočtu je zhruba lineární ve velikosti populace \times počet generací a zhruba kvadratická v dimenzi dat a v počtu dat. Naposled zmíněné kvadratické závislosti jsou dány

jednak využitím regrese při učení jednotlivých NSU jednotek, a dále použitým klastrovacím algoritmem (Jancey-Forgyův klastrovací algoritmus). Součin uvedených 4 hodnot může tedy sloužit jako hrubý odhad pro porovnání časové náročnosti jednotlivých úloh ve frontě výpočtů pro NNSU server. Na stránce s tabulkou výpočetní fronty je možné ještě definovat požadovaný

The screenshot shows the NNSU separator web interface. At the top, there are navigation tabs: Networks, L-data, T-data, R-data, Progress bar, Jobs, Subtree, New folder, Settings, My account, and Help. The user is identified as 'josef123' and the page title is 'NNSU separator: experiment A'. A green banner says 'zde pracujeme s daty typu A'. Below this, a message states 'You are about to run the following NNSU queue:'. There are two buttons: 'Select FROM->TO' and 'XOR FROM->TO'. A table lists the tasks in the queue:

SELECTED	FROM	TO	NETWORKS	DATA	TYPE	resources* demand
<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	další příklad (více zvýrazněná ROC)	demonstracni-data-LEARN-nnsu	L	264864600
<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	další příklad (více zvýrazněná ROC)	demonstracni-data-RAW-nnsu	R	6658200
<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	další příklad (více zvýrazněná ROC)	demonstracni-data-TEST-nnsu	T	265901400
<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	další instance NNSU (zvýrazněná počáteční ROC)	demonstracni-data-LEARN-nnsu	L	353152800
<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	další instance NNSU (zvýrazněná počáteční ROC)	demonstracni-data-RAW-nnsu	R	8877600
<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	další instance NNSU (zvýrazněná počáteční ROC)	demonstracni-data-TEST-nnsu	T	354535200
<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	testovací síť (defaul nastavení)	demonstracni-data-LEARN-nnsu	L	176576400
<input checked="" type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	testovací síť (defaul nastavení)	demonstracni-data-RAW-nnsu	R	4438800
<input checked="" type="checkbox"/>	<input type="radio"/>	<input checked="" type="radio"/>	testovací síť (defaul nastavení)	demonstracni-data-TEST-nnsu	T	177267600

Below the table, a note explains: '(*) "resources demand" is computed as a multiplication of population size, number of generations, data dimension and number of data. This value can serve as an estimation of computer time needed.' There is an email notification option set to 'mob@centrum.cz' when 50% and 100% is done. The task will run on 22 processors (INTEL XEON 2.5GHz Hyper Thread, 2 processors of 6 cores). A password confirmation field is present, followed by a 'Run NNSU to complete results' button, a 'Reset password' button, and a 'Go back' button.

počet procesorů, dedikovaných pro tuto frontu. Je přednastavena hodnota pro maximální povolený počet, v případě potřeby zpřístupnění většího počtu procesorů je třeba kontaktovat správce tohoto serveru. Dále je možno před spuštěním výpočtů zadat požadavek na zaslání informačních e-mailů o průběhu výpočtu. Je možno vybrat mezi čtyřmi frekvencemi zaslání informačních e-mailů: nikdy, po ukončení fronty, po ukončení poloviny fronty, po ukončení každé čtvrtiny fronty. Tyto e-maily jsou vztaženy k ukončení jednotlivých výpočtů ve frontě, nesouvisí tedy s časovým průběhem výpočtu. E-maily jsou zaslány na e-mailový účet, který je spojen s daným NNSU účtem.

Pomocí tlačítka **Run NNSU to complete results** spustíme zadanou frontu výpočtů. V průběhu výpočtu je v poli prohlížeče zobrazena tabulka, kde je vidět postup výpočtu. Obsah stránky lze aktualizovat tlačítkem **Hit me to refresh the table content**, které stránky znovu načte s aktuálním stavem výpočtů ve frontě. Jsou-li veškeré výpočty v naposled spuštěné frontě ukončeny toto aktualizací tlačítko již není dostupné. V průběhu výpočtu lze NNSU server nadále používat k jednotlivým úkonům, jediné omezení je to, že nelze spustit další frontu úloh.

user: josef123

NNSU separator:
experiment A (inserted at 14/01/14, 08:48)

[zde pracujeme s daty typu A](#)

DATA -> NETWORKS	LEARN file demonstrační data pro učení NNSU včetně ohodnocení	RAW file druha RAW data	TEST file test data c. 4	RAW file toto jsou neznámá data není k dispozici GAIN soubor chybí informace o třídě	TEST file toto jsou testovací data
další příklad (více zvýrazněná ROC)	all results	all results	NNSU 330 DONE PLOTS 98/1650 NA	all results	all results
další instance NNSU (zvýrazněná počáteční ROC)	all results	NNSU 420/420 NA PLOTS 420/420 NA	NNSU 420/420 NA PLOTS 2100/2100 NA	all results	all results
testovací síť (default nastavení)	all results	NNSU 220/220 NA PLOTS 220/220 NA	NNSU 220/220 NA PLOTS 1100/1100 NA	all results	all results

[Hit me to refresh the table content](#)

12 Ošetření nedokončené fronty

Z různých důvodů může dojít k situaci, že výpočty zařazené do fronty nejsou ani po delší době ukončeny. Jako příklad může sloužit neplánovaný restart serveru, vyčerpání přidělené diskové kapacity, chyby v programu NNSU či v serverové části vedoucí k "zamrznutí" výpočtu a pod. Odstranění této situace je řešeno tak, že server detekuje skutečnost, že výpočty ve frontě nebyly ukončeny, a současně detekuje poslední projev aktivity výpočtů ve frontě.

user: josef123

NNSU separator:
experiment A (inserted at 14/01/14, 08:48)

[zde pracujeme s daty typu A](#)

There are active jobs in ->NNSU separator->experiment A

ATTENTION: Last job has exceeded the predefined acceptable inactivity time limit (30 min)! Your last NNSU script activity has been detected 11:01 (hours:minutes) earlier. There is a possibility that your last job is not running due an unknown reason. You can unlock NNSU threads to allow new jobs to be run (regardless of a real status of the frozen job).



If you want to cancel the last job, insert your password.

●●●●●

Network(s) name	Data file	Started at (CET)	Finished at (CET)	Pop/Gen/Dim/Num	Total size	Time consumed (h:m:s)	NNSU XML file
		14/01/14 11:02:33	not yet	0/0/0/0	0	not yet	XML
		not yet	not yet	0/0/0/0	0	not yet	XML
		not yet	not yet	0/0/0/0	0	not yet	XML
		not yet	not yet	0/0/0/0	0	not yet	XML
		not yet	not yet	0/0/0/0	0	not yet	XML

Je-li doba uplynulá od poslední zaznamenané aktivity výpočtů delší než je předdefinovaný časový interval (30 min), požadavek na aktualizaci vyvolá formulář, kde uživatel může výpočet celé fronty stornovat. Jednotlivé výpočty ve frontě, které byly ukončeny se zachovávají, a v nově vytvořené frontě k výpočtu budou pouze nedokončené (a nespustěné) výpočty.

K lepší orientaci v provedených výpočtech a v postupu výpočtu v naposled spuštěné a nedokončené frontě nabízí NNSU server následující pomůcky:

- Volbou položky , která je aktivní pouze v případě zpracovávání fronty, zobrazíme seznam jednotlivých výpočtů v běžící frontě s časy začátku a konce výpočtu, celkovým časem výpočtu a odkazem na xml soubor, který definoval výpočet (pro účely detailní analýzy nastavení NNSU sítě použitých k výpočtům). S touto tabulkou je současně zobrazen odkaz mířící v uživatelem vytvořené adresářové struktuře do toho adresáře kde byl výpočet fronty spuštěn (předpokládá se, že uživatelem vytvořený systém adresářů s daty může být potenciálně dosti bohatý, a je třeba umožnit snadný a rychlý návrat do adresáře se spuštěným výpočtem fronty).
- Volbou položky  zobrazíme seznam záznamů o průběhu výpočtu všech front puštěných uživatelem v celém podstromu adresářové struktury. Toto umožňuje lépe odhadovat doby výpočtu a efektivnější plánování rozsahu výpočtů.

13 Vyhodnocení výsledků separace na datech

Výsledky učení sítí na jednotlivých zadaných datech jsou dostupné aktivací odkazů "all_results" v tabulce kombinující již dostupné výsledky a dosud nerealizované výpočty. Aktivací těchto odkazů se vždy dostaneme na stránku, kde je možno různým způsobem zobrazovat grafy s výsledky a provádět tak vizuální kontrolu a ohodnocení kvality separace dat.

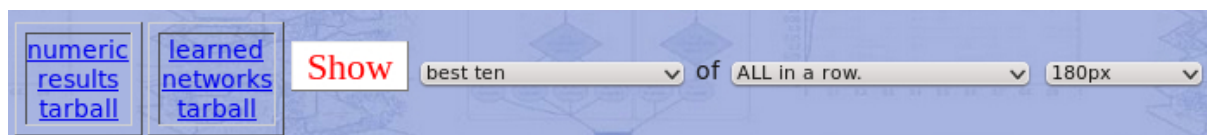
Na těchto stránkách jsou mimo jiné dva následující odkazy:

numeric result tarball – obsahuje numerické odezvy sítí na data v souboru typu ".tar.gz".
Uživatel si tyto odezvy může stáhnout a podrobit dosažené výsledky své vlastní analýze.

learned networks tarball – obsahuje xml soubory s naučenými sítěmi v souboru typu ".tar.gz".
Uživatel si tyto naučené neuronové sítě archivovat pro případné pozdější využití. Zjevně tento odkaz je smysluplný a k dispozici pouze pro učící data.

Díky těmto dvěma odkazům má uživatel plný přístup k výsledkům separace a v jejich hodnocení není omezován skromnými možnostmi vizuálního vyhodnocování, které jsou implementovány v NNSU serveru.

Hlavní vyhodnocení výsledků separace dat, které NNSU server přímo nabízí, je stránka, která umožňuje zobrazit separaci pomocí grafů. Hlavními typy grafů pro zobrazení výsledků separace je ROC křivka, výtěžnost (GAIN plot – součet vah datových vektorů, které jsou zobrazeny nad NNSU prahem), derivace výtěžnosti (local GAIN plot), histogramy pro signál a pozadí (CAMEL plot) a integrál těchto histogramů (CUMULATIVE CAMEL). Toto jsou standardní metody zobrazení separačních výsledků, v následujícím se tedy pouze zaměříme na možnosti jejich zobrazování a vzájemného porovnávání. K dispozici je jednoduchý formulář, kde lze definovat rozsah zobrazených výsledků, typ grafů zobrazených vedle sebe a velikost náhledů grafů.



Rozsah zobrazených výsledků lze určit následovně:

best ten – zobrazí výsledky pro prvních deset sítí s nejlepšími hodnotami fitness funkce na učicích datech.

best 10% – zobrazí výsledky pro prvních deset % sítí s nejlepšími hodnotami fitness funkce na učicích datech.

last generation – zobrazí výsledky všech sítí v poslední generaci, seřazeno podle velikosti fitness funkce na učicích datech. hodnotami fitness funkce.

all results – zobrazí výsledky všech sítí, seřazeno podle velikosti fitness funkce na učicích datech.

two best and two worst – zobrazí výsledky prvních dvou nejlepších a posledních dvou nejhorších sítí, seřazeno podle velikosti fitness funkce na učicích datech.

Dále je možno určit typy grafů, které jsou zobrazeny buď vedle sebe v případě výběru pouze jednoho typu grafů, což umožňuje rychlé porovnání výsledků pro učicí a testovací data, a stejně tak porovnání výsledků pro různá testovací data, nebo při výběru různých typů grafů máme možnost ohodnotit vizuálně kvalitu separace z více úhlů pohledu.

ALL in a row – vedle sebe se zobrazí různé typy grafů pro danou síť (množina sítí je definována podle předešlé volby). **Zobrazené grafy se vztahují k aktuálně vybraným datům (odpovídají sloupci, ve kterém byl kliknout odkaz "all_results").**

ROC's only – zobrazí se pouze ROC křivky. V daném řádku jsou ROC křivky pro danou síť (množina sítí je definována podle předešlé volby) a **pro všechna aktivní data (jsou-li pro ně již dostupná).**

SUMMARY GAIN's only – dtto pro SUMMARY GAIN

LOCAL GAIN's only – dtto pro LOCAL GAIN

CAMEL's only – dtto pro CAMEL

CUMULATIVE CAMEL's only – dtto pro CUMULATIVE CAMEL.

Dále je možno zadefinovat velikost zobrazení náhledů grafů v rozmezí 120–320 px. Každý náhled grafu je současně odkazem, po jehož kliknutí se uživateli načte detailní graf ve formátu PDF.

Výjimku² z výše uvedených pravidel tvoří zobrazování výsledků na datech typu RAW, kde jsou k dispozici pouze histogramy výstupních hodnot NNSU sítí dosažených na těchto datech, protože pro RAW data není dostupná informace o příslušnosti těchto dat do tříd. Pro RAW data pouze můžeme definovat rozsah sítí, pro které jsou výsledky separace zobrazeny.

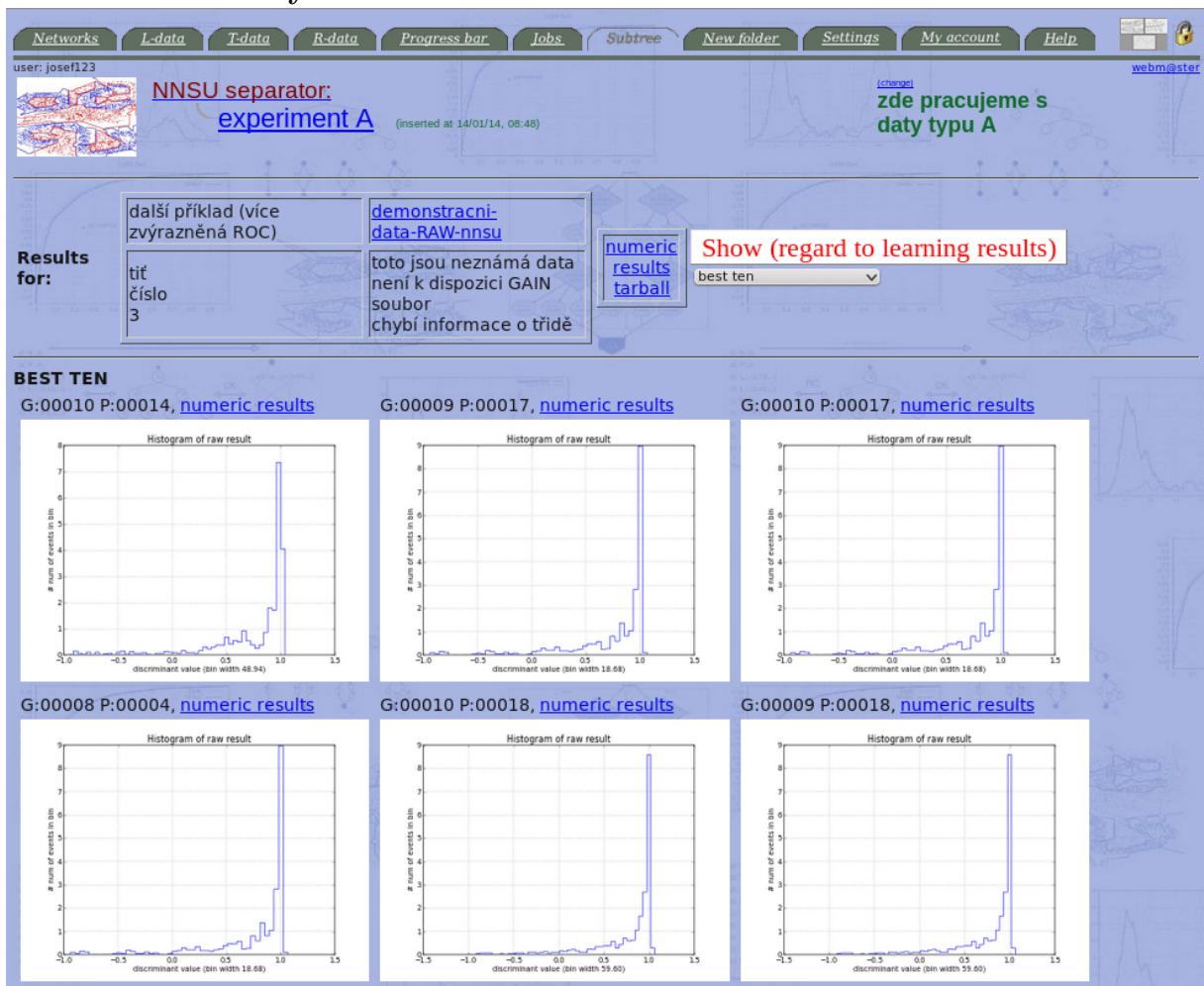
Na následujících stránkách uvádíme pro ilustraci několik způsobů zobrazování výsledků separace.

²Typy zobrazených grafů zde popisovaných platí pro aplikační oblast SEPARACE dat. Pro zobrazení výsledků PREDIKCE a APROXIMACE jsou použity částečně odlišné typy zobrazení.

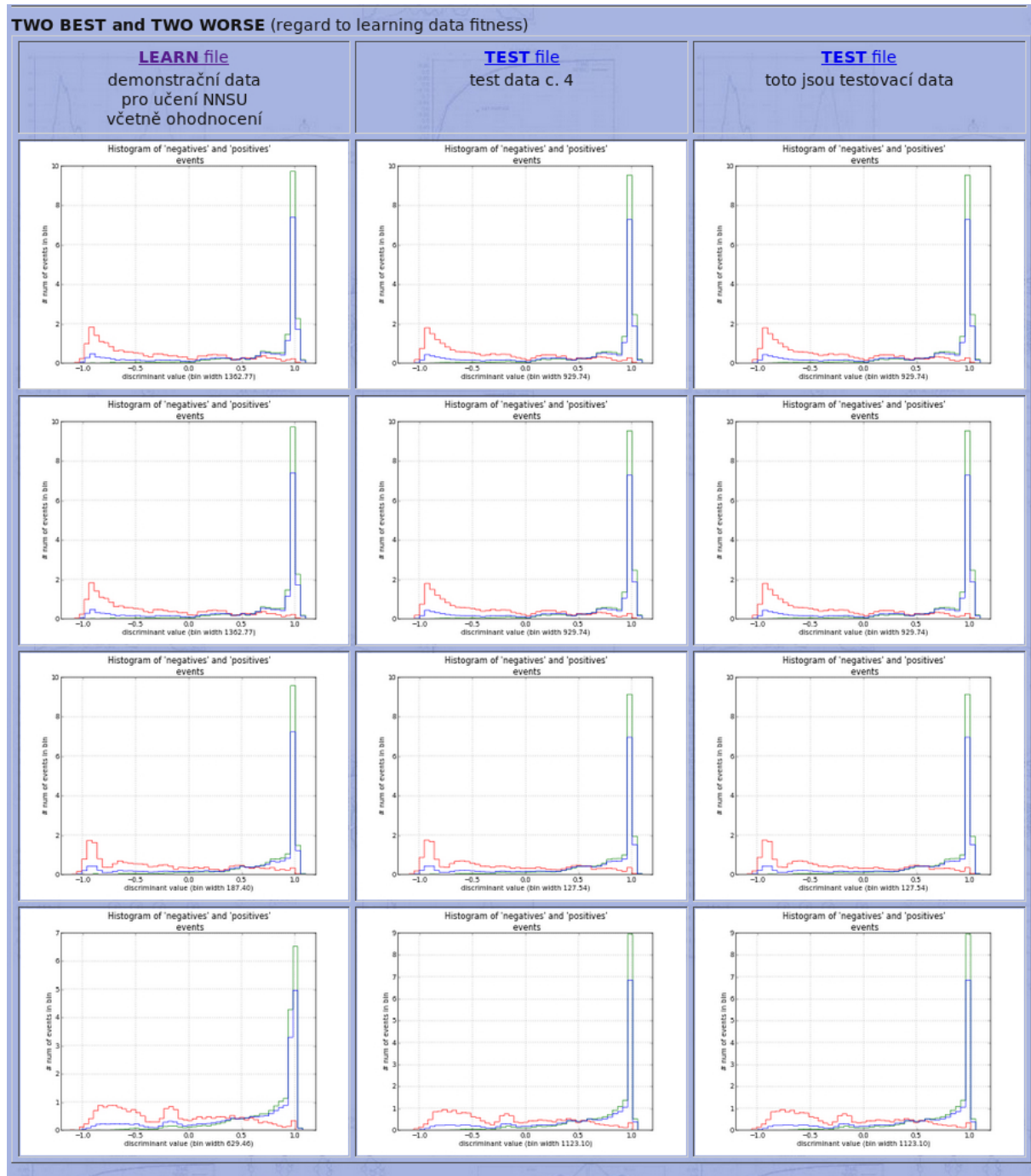
13.1 Výsledky na testovacích datech se všemi druhy grafů



13.2 Zobrazení výsledků na RAW datech



13.3 Ukázka porovnání histogramů pro učící a testovací data



14 Doplňující nástroje a konfigurace

V závěru se ještě zmíníme o dosud neuvedených položkách hlavního menu.

- položka **Subtree** v hlavním menu zobrazí/skryje výpis všech uživatelských podadresářů v daném adresáři. Celý systém NNSU serveru je koncipován tak, že uživatel může vytvářet svoji vlastní adresářovou strukturu, která je omezena pouze na hloubku zanoření adresářů (max 8) a na počet podadresářů v každém adresáři (max 50). Pro lepší orientaci v této

potenciálně rozlehlé adresářové struktury slouží výpis těchto všech uživatelských adresářů, ve kterém lze přímo přecházet do jednotlivých adresářů.

- volba **My account** zobrazí formulář s osobními údaji, kde lze tyto údaje měnit s výjimkou názvu uživatelského účtu. Zde je možno změnit jak heslo, tak i e-mailovou adresu, kam server zasílá zprávy o postupu výpočtu. Při potvrzení těchto změn se postupuje stejným způsobem, jako při prvotním vytváření uživatelského účtu.
- položka **Settings** umožňuje nastavit délku rozvinutí podadresářové struktury v daném adresáři a počet náhledů na jedné řádce tabulky s výsledky na RAW datech.
- volbou **Help** zobrazíme krátkou základní nápovědu o významu jednotlivých položek v menu.