

Pokroky v analýze heterogenních neuroinformatických dat

Ondřej Klempíř, Václav Čejka, Jan Tesař, Radim Krupička

Katedra biomedicínské informatiky, Fakulta biomedicínského inženýrství ČVUT v Praze
Nám. Sítná 3105, 272 01 Kladno

{ondrej.klempir, vaclav.cejka, jan.tesar,
krupicka}@fbmi.cvut.cz

Abstrakt. V současné době roste počet dat získaných při měření pato/fyziologických procesů mozku. Aplikací analytických metod jsou z dat získávány nové poznatky o mechanismech neurologických onemocnění. Cílem výzkumu je dosažení pokroku v analýze různých typů neuroinformatických záznamů (jednotkové mikroelektrodové neuronové aktivity, transkraniální magnetické stimulační a blízké infračervené spektroskopie) za účelem objektivní identifikace významných biomarkerů metodami strojového učení. Do studie byla zahrnuta data z Neurologické kliniky 1. LF a VFN UK v Praze. V průběhu řešení výzkumu byly vytvořeny skripty pro ukládání, předzpracování a analýzu 3 typů výše zmíněných neuroinformatických datových zdrojů. Pokroky v analýze budou využity pro hodnocení naměřených dat a testování hypotéz s přínosem pro klinickou neurologickou praxi.

Klíčová slova: mikroelektrodové záznamy, transkraniální magnetická stimulační, blízká infračervená spektroskopie, algoritmus třídění pálení neuronů.

1 Úvod

Světová zdravotnická organizace zaznamenává zvyšující se výskyt neurologických onemocnění. Více než 50 milionů Evropanů trpí neurologickou nemocí, což vyžaduje náklady téměř 400 miliard euro [4]. Současně také roste počet dat získaných při studiu (pato)fyziologických procesů mozku. Aplikací analytických výpočetních metod jsou z dat získávány nové poznatky o funkci nervového systému či např. mechanismech neurologických onemocnění. Ukazuje se, že počítačové učení dokáže u některých nemocí na základě extrahovaných příznaků a biomarkerů rozlišit zdravou normu od patologie přesněji než expert. Dále např. s úspěchem predikovat odpověď na léčbu.

Ve studii se zaměřujeme na dvě skupiny neurologických onemocnění, dystonie a Parkinsonovu nemoc (PN). Dystonie je heterogenní skupina syndromů, které se projevují křečovitým mimovolným pohybem jedné nebo více částí těla [2]. PN je chronické progresivní onemocnění, přičemž mezi základní symptomy patří třes, ztuhlost a

*J. Steinberger, M. Zíma, D. Fiala, M. Dostal, M. Nykl (eds.)
Data a znalosti 2017, Plzeň, 5. - 6. října 2017, pp. 157-161.*

pohybové zpomalení [5]. Jednou z metod, která se pro léčbu symptomů pokročilých forem těchto nemocí používá, je hluboká mozková stimulace (Deep Brain Stimulation; DBS), která se implantuje během neurochirurgické operace do specifických jader v mozku [2]. Data, která byla naměřena na Neurologické klinice 1. lékařské fakulty Univerzity Karlovy a Všeobecné fakultní nemocnice v Praze (1. LF UK a VFN), jsou ve světovém kontextu unikátní. A to hlavně z důvodu vyšetření pacientů za různých podmínek, které dříve měřeny nebyly (např. při experimentech s nastavením DBS). Datové soubory byly získány třemi neurovědními vyšetřovacími technikami, které snímají informace na různých úrovních mozku - od aktivity jednotlivých nervových buněk, až po aktivitu makroskopickou. Příspěvek představuje podstatu geneze a zpracování tří typů neuroinformatických datových zdrojů.

2 Analýza dat mikroelektrodoých záznamů

Mikroelektrodové encefalografické záznamy (microEEG) se používají pro diagnostiku správného umístění DBS elektrody. Technologicky jde o sadu mikroelektrod (průměr hrotu měří okolo $5 \mu m$), které prochází skrz mozek a nahrávají aktivitu (tzv. pálení, spikes) okolních několika neuronů. Dosažení optimální pozice DBS je klíčové z hlediska terapeutického efektu u dystonií i PN. V současné době je klinický efekt posuzován primárně neurologem, a to vizuálně či akusticky, na základě specifického projevu pálení neuronů. Cílem offline analýzy microEEG v probíhajícím výzkumu je nalezení biomarkeru optimální pozice s využitím metod strojového učení.

Stěžejní úlohou následující po základním předzpracování (explorační analýze, filtraci či detekci artefaktů) je automatický spike sorting (třídění spiků). Neboli, automaticky v microEEG detekovat namodulované projevy neuronů (jednotlivá pálení) v okolí mikroelektrody a rozřadit je s ohledem na individuální charakteristiky včetně extrakce dalších užitečných informací o jednotlivých neuronech. Vyvinutá spike sorting procedura sestává z 5 fází:

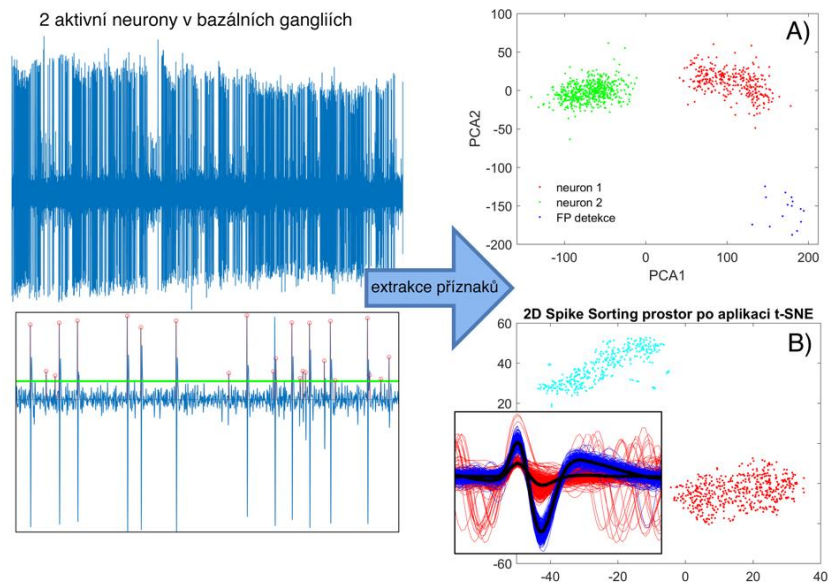
1. stanovení prahu pro odlišení aktivity pálení od šumu pozadí (pravidlo 3σ),
2. výpočet mnohorozměrných tvarově založených příznaků pro detekce z bodu 1,
3. redukce dimensionalit pomocí PCA a t-SNE do 2D/3D prostoru,
4. shluková analýza K-means pro rozřadění jednotlivých neuronů,
5. výpočet odvozených charakteristik (např. průměrná frekvence pálení / s).

Metoda t-SNE (t-Distributed Stochastic Neighbor Embedding) zachycuje v porovnání s PCA (Principal Component Analysis) i nelineární strukturu v datech [1]. Na konkrétním experimentu (Obr. 1A) lze vidět, že PCA vede k rozpoznání 3 neuronů, kdy jeden shluk představuje falešnou detekci, která nesouvisí se skutečnou charakteristikou podobou projevu neuronu a musel by být analytikem poloautomaticky vyřazen. T-SNE může fungovat lépe při úplné automatizaci, na Obr. 1B je příklad s chybou 1.6 %. T-SNE přiřadí málo četné hodnoty do bližšího shluku za cenu poměrně malé chyby odhadu výstupních parametrů skutečných neuronů. Ukazuje se také, že t-SNE vede k neproblematickému automatickému určení počtu shluků.

3 Analýza dat z transkraniální magnetické stimulace

Transkraniální magnetické stimulace (TMS) je neinvazivní metoda sloužící k funkčnímu vyšetření centrálního a periferního nervového systému. Je využívána nejenom k rutinním vyšetřením, ale také v oblasti výzkumu funkce jednotlivých mozkových a míšních struktur za účelem identifikace elektrofyziologických biomarkerů poruch řízení hybnosti u dystonií i PD [3].

Pomocí série pulzů aplikovaných do oblasti motorického kortexu byly z registrace odpovědi v horní končetině získány sady motorických evokovaných potenciálů (MEP) (Obr. 2B), které byly průměrovány a filtrovány. Pomocí vlnkové transformace byly v signálech automaticky detekovány MEP a byly vypočteny doby jejich trvání. Jako mateřská vlnka byla použita adaptovaná vlnka zkonstruovaná z průměru motorických potenciálů kontrolních subjektů. Latence potenciálu a jeho trvání byly určeny z pozice maxima (v abs. hodnotě) v matici obsahující koeficienty transformace.

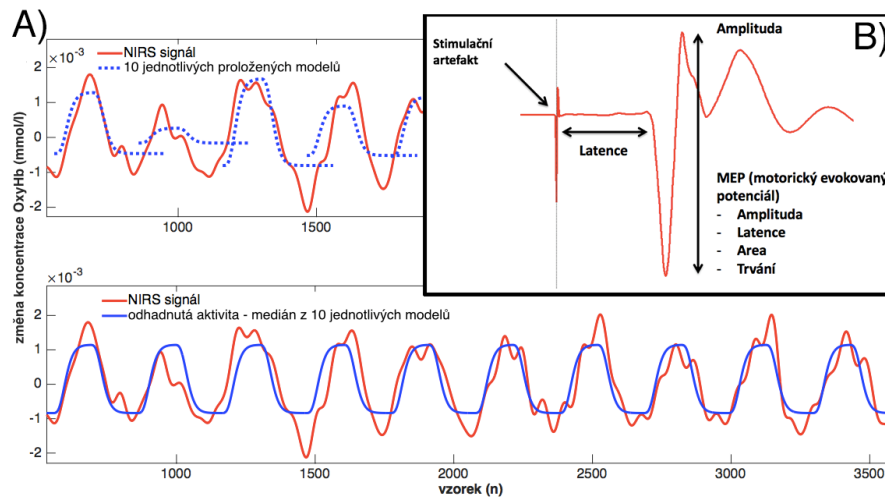


Obr. 1. Metodika detekce jednotlivých neuronů v jednom konkrétním microEEG záznamu bazálních ganglií (modrá časová řada). Vlevo dole - přiblížený úsek jednotlivých pálení po amplitudové detekci (zelený práh). A) 2D příznakový prostor po PCA redukcí tvarově definovaných příznaků detekovaných pálení. Aplikace 3-means algoritmu pro nalezení shluků, kdy modrý shluk odpovídá artefaktu. B) 2D příznakový prostor po t-SNE redukcí tvarově definovaných příznaků. Artefakty z A) byly přiřazeny do červeného shluku. Na výřezu lze vidět jednotlivé průběhy pálení včetně černé průměrné skupinové hodnoty.

4 Analýza signálů blízké infračervené spektroskopie

Blízká infračervená spektroskopie (Near-InfraRed Spectroscopy; NIRS) je zobrazovací metoda aktivity kůry mozku využívající nízkenergetické optické záření pro detekci lokálních změn koncentrace oxyhemoglobinu. Charakter geneze je podobný jako u funkční magnetické rezonance. Mozková aktivita byla snímána fotodetektory umístěnými na skalpu při provádění rytmického dotyku palce a ukazováku pravé ruky v průběhu 10 cyklů, během kterých se po 15 sekundách střídala pohybová a klidová fáze úlohy.

Změny koncentrace oxyhemoglobinu byly vypočteny ze surového signálu NIRS pomocí modifikované Lambert-Beerovy rovnice. Pro stanovení hodnot mozkové aktivity vyfiltrovaných NIRS signálů byla využita regresní metoda založená na fyziologickém modelu očekávané odezvy neurální tkáně. Byl uvažován medián z 10 cyklů proložení očekávaným modelem (Obr. 2A). Zjednodušeně, analýza NIRS spočívá v proložení modelu známé odezvy mozku při pohybu ruky na konkrétní naměřený signál. Výsledkem proložení je získání jednoho charakteristického parametru aktivity signálu, který je vstupem pro navazující statistické testy.



Obr. 2. A) Regresní mediánová metoda (modrý průběh) pro odhad aktivity kůry mozku (červený průběh). Nárůsty v červeném NIRS signálu odpovídají 10 cyklům klidové a pohybové fáze úlohy klepání prsty. B) Typický příklad MEP po automatické detekci vlnkovou transformací.

5 Závěr a budoucí práce

V průběhu řešení výzkumu byly vyvinuty různorodé algoritmy - automatický spike sorting, detekce motorického potenciálu TMS a metoda pro stanovení aktivity NIRS, včetně skriptů pro načítání a předzpracování. Z konkrétních pilotních výstupů navržených metod použitých na reálné signály vyplývá jejich vhodnost: spike sorting procedura s využitím t-SNE vykazovala chybu menší než 2 %, automatické detekce MEP dle

experta probíhají správně a mediánový regresní odhad u NIRS vykazuje dobré vlastnosti. Prezentované metody jsou založeny a dále rozšiřují metodologické principy autorů/pracovišť v rámci současného stavu daného oboru (state of the art). Experimentální hodnocení výsledků odhadů jednotlivých metod pro dávkové zpracování mnoha souborů bude předmětem navazující práce. Implementované metody budou využity pro hodnocení pacientů a ověření klinických hypotéz. Cílem článku bylo stručně informovat o probíhajícím výzkumu na FBMI ČVUT a Neurologické klinice 1. LF a VFN UK a přiblížit možnosti zpracování 3 typů neuroinformatických dat.

Literatura

1. Dimitriadis, G.: T-SNE vizualization of large-scale neural recording. *BioRxiv* (2016) 1 22.
2. Jech, R.: Hluboká mozková stimulace u dystonií. *Neurologie pro praxi* 14(5) (2013) 232 236.
3. Kobayashi, M.: Transcranial magnetic stimulation in neurology. *The Lancet neurology* 2(3) (2003) 145 156.
4. Ptáček, R.: *Etické problémy medicíny na prahu 21. století*. Grada Publishing a.s., Česká republika, 2014.
5. Ulmanová, O.: Parkinsonova nemoc – základy terapie a diferenciální diagnostiky. *Psychiatrie pro praxi* 2 (2007) 60 62.

Poděkování: Tento článek vznikl díky podpoře projektu SGS17/114/OHK4/1T/17 Zpracování a analýza heterogenních neuroinformatických dat na KBI FBMI ČVUT v Praze.

Annotation:

Advances in analysis of heterogenous neuroinformatics data

The number of the neuroinformatics datasets is currently increasing. Application of the analytical methods is essential to discover new neuroscience knowledge. The aim of this research is to make progress in the analysis of various types of neuroinformatics recordings (single neuronal activity, Transcranial Magnetic Stimulation and Near Infrared Spectroscopy) in order to objectively identify some significant biomarkers by the machine learning methods. This study includes data from the Department of Neurology, 1st Faculty of Medicine and General University Hospital in Prague. The advances in analysis will be used for the evaluation of the measured data and consequently for the hypotheses testing, with expected benefit in clinical neurology.