



Inteligentní GPA pro modelování časových řad

TEXT: doc. Dr. Ing. Tomáš Brandejský, Fakulta dopravní, České vysoké učení technické v Praze

Zdokonalení metod modelování časových řad představujících různé biosignály, jako např. EEG, by otevřelo nové aplikační oblasti, jako je detekce a předvídání mikrosrápků, analýza zaměření pozornosti lidských operátorů a podobně. Představovaný příspěvek se zabývá návrhem specifického algoritmu genetického programování (GPA) pro efektivní modelování těchto časových řad. K tomuto účelu jsou navrženy a využívány informované (inteligentní) evoluční operátory na místo běžných, aby byla co nejvíce omezena množina testovaných variant (hypotéz) s ohledem na obvykle značný rozsah testovacích dat.

Mnoho technických systémů pracuje v přímém vztahu s lidskými operátory, či jinak řečeno s lidským činitelem. Mnoho dalších systémů není lidskými operátory jen ovlivňováno, ale přímo řízeno.

Lidská rozhodnutí tak ovlivňují nejen výkonnost systému, ale i jeho spolehlivost, udržovatelnost, dostupnost a dokonce i bezpečnost. Mnoho takových systémů nalezneme například v oblasti dopravy. V našem výzkumu se pokoušíme aplikovat evoluční techniky (Genetické Algoritmy i genetické programování – GPA) na biosignály jako EEG nebo EKG a s jejich pomocí nalézt nové transformace těchto signálů, nové modely signálů i nové markery důležitých stavů.

Základní vlastnosti biosignálů

Mnoho biosignálů jako EEG nebo EKG vykazují mnoho zajímavých vlastností, jako nestacionaritu. Další zajímavou vlastností je rostoucí složitost signálu se stupněm jeho derivace, což svědčí na přítomnost složené funkce. Pro ilustraci může posloužit vztah

$$[f(g(x))] = g'(x)f(g(x)) \quad (1)$$

Mnohdy můžeme předpokládat, že měřený signál je spojitou funkcí, resp. že se jedná o posloupnost vzorků spojitě funkce měřené v diskrétním čase. To po vzoru kvalitativní simulace dovoluje vyslovit některé hypotézy:

Pokud je signál spojitý, pak existuje maximální hodnota derivace a při zadané požadované přesnosti lze definovat minimální dobu predikce, po kterou chyba predikované hodnoty nepřesáhne povolenou chybu. Pokud jsou spojitě i derivace této funkce a jejich maximální hodnoty známe, pak se samozřejmě přesnost zlepšuje. V případě modelování biosignálů pomocí složených funkcí jsou tyto podmínky splněny.

Symbolická regrese

V mnoha obtížných situacích bývá v poslední době využívána tzv. symbolická regrese pro nalezení vhodného algebraického popisu trénovacích dat bez předchozí znalosti vhodného deduktivně získaného modelu. Symbolická regrese pomocí genetických algoritmů vychází z chápání genu jako sledu nebo stromu operátorů, konstant a odkazů na proměnné, které tvoří syntakticky správný matematický výraz – funkci. Obecným problémem symbolické regrese pomocí GPA je otázka účinnosti výpočtu (počtu prohledávaných nesprávných hypotéz, které nakonec neovlivní výsledné řešení) a tzv. Blowing [1] – tedy vytváření překomplikovaných řešení, které omezuje počet iterací výpočtu. Společně tyto dva problémy omezují aplikaci GPA v případě složitých problémů. Mnozí výzkumníci, jako Koza [2] proto volí malý počet iterací na velmi rozsáhlých popu-

lacích, což předpokládá dostatečný výpočetní výkon a paralelní implementace GPA.

Podobnost funkcí

V symbolické regresi rozhoduje o úspěšnosti či účinnosti dalšího postupu výpočtu schopnost správně určit podobnost částečného řešení a aproximované funkce – správně zvolená míra vzdálenosti (diskrétních bodů) funkce. Jedná se o velmi složitý problém, neboť lze snadno nalézt příklady částečných řešení, které jsou např. v běžně užívaných Minkovského metrikách vzdálenosti velmi nepodobné cílové funkci, pokud jsou jejím nutným předstupněm v případě řešení pomocí GPA a naopak mnohem podobnější cílové funkci, pokud jednoduchá cesta ke správnému tvaru nevede viz např. [3].

Lokální podobnost funkcí

V případě velmi složitých a časově proměnných signálů, jako jsou biosignály, se nabízí využití lokální podobnosti funkcí v okolí daného bodu (jádra), tedy na omezeném intervalu hodnot nezávislé proměnné, což může snížit složitost celého problému.

Nejjednodušším příkladem lokální podobnosti je podobnost bodová, kdy průběh funkce můžeme charakterizovat hodnotou, nebo lépe tečnou v daném bodě. Podobně můžeme ve více bodovém intervalu například hle-



dat podobnost dané funkce a polynomu.

Zajímavou aplikací lokální podobnosti funkcí jsou fuzzy modely, zvláště pak tzv. Sugeno type II modely, viz [4]. I tyto modely jsou založeny na lokálních aproximacích funkce pomocí modelů platných v okolí zvolených bodů a aproximaci mezi nimi přes standardní funkce příslušnosti (zpravidla triangulární) a defuzzyfikaci pomocí váženého průměru.

Častým problémem skládání globálních modelů z množiny lokálních je omezení predikční schopnost navzdory jejich univerzálním aproximačním schopnostem.

Plovoucí okno

Při zpracování a analýze signálů z hlediska jejich lokálních vlastností bývá užíván mechanismus plovoucího okna ať již v absolutní, nebo přírůstkové podobě. Spojení plovoucího okna s algoritmem genetického programování je možno uskutečnit dvěma rozdílnými postupy:

v každém okně může být nalezena nová struktura modelu pomocí GPA, nebo jsou určovány po jen nové parametry pomocí GA

Pokud je v každém okně nalezena nová struktura modelu, vzniká obtížný problém, jak tyto rozdílné struktury interpretovat. Pokud jsou změny malé a občasné, pak jejich nalezení může pomoci např. V oblasti detekce poruch pro nalezení pozvolných změn struktury modelovaného systému. Ovšem taková aplikace předpokládá obvykle i systém schopný transformovat nalezený model na alternativní reprezentace, aby bylo možno eliminovat perturbace způsobené pouze alternativním zápisem téhož modelu. Druhé použití přístupu založeného na nacházení struktur modelu pro každou pozici plovoucího okénka zvláště je možnost chápat strukturu nalezenou pro každou pozici okénka jako lokální interpolaci (ložitějšího) globálního modelu, podobně jako jsou takto chápány funkce na právě straně pravidel Sugeno type II modelu a globální model by pak mohl být konstruován jako interpolace mezi těmito

lokálními modely, pro které by tyto dílčí modely tvořily jeho stavební bloky.

Podstatně jednodušší a srozumitelnější situace nastává, pokud je struktur modelu pevně dána a v jednotlivých pozicích plovoucího okénka jsou určovány pouze hodnoty parametrů. Globální funkce je pak určována modelem a časovými průběhy jednotlivých parametrů tohoto modelu, dílčími funkcemi. Z analýzy charakteru těchto dílčích funkcí reprezentujících jednotlivé parametry modelu a citlivostní analýzy modelu na změny jednotlivých parametrů pak lze určit, který parametr (kterou dílčí funkci) je vhodné zahrnout do globálního modelu, tedy cíleně zaměřit pozornost evolučních technik na určitou malou část modelu.

Aplikace technik genetického programování

Aplikace technik genetického a evolučního programování v oblasti symbolické regrese narazí na některá omezení. Vedle obvyklých problémů se schopností regrese (tedy požadavkem, aby množina operátorů GPA měla vlastnosti univerzálního aproximátoru) a s blowingem (vytvářením překomplikovaných řešení při aplikaci GPA) se přidružuje ještě jeden závažný problém, a to objem zpracovávaných dat. Vysoké vzorkovací frekvence potřebné pro zachycení vysokofrekvenčních složek, dlouhé časy provádění experimentů a mnohokanálové snímání vedou k velkým objemům dat a znesnadňují aplikaci jakékoli regresní techniky, tedy i genetických algoritmů. Proto je vhodné pro analýzu biosignálů použít paralelní implementace GA a GPA.

Eliminace stochastičnosti výsledků

Aplikace plovoucího okénka pro analýzu biosignálů předpokládá, že dílčí výsledky nalezené v jednotlivých pozicích okénka nebudou zatíženy stochastickou chybou. Ovšem GA a GPA pracují do značné míry náhodně. Pro eliminaci náhodného chování byly aplikovány následující tři přístupy:

Rozsáhlé populace jedinců bývají náročné na výpočetní výkon, ale dávají větší šanci na nalezení optima a tím pomáhají eliminovat náhodné vlivy.

Rovněž tak pokus o pohyb v několika různých směrech v rámci jednoho evolučního kroku pomáhá stejným způsobem.

K odstranění náhodného šumu dle uskutečněných experimentů také přispívá přebírání nejlepších chromozomů z minulého okénka do nové populace. Nelze totiž přebírat celou minulou populaci, protože pak by se aplikace plovoucího okénka přeměnila v test adaptačních schopností GA, reps. GPA na změnu fitness funkce. Přebírání chromozomů z minulé pozice okénka urychluje evoluci v případě malých změn a tak zabraňuje náhodným perturbacím genů.

Protože perturbace bývají způsobeny mnohonásobností možných řešení, tedy nedostatečnou citlivostí fitness funkce či malým rozměrem plovoucího okénka, je někdy potřeba doplnit dodatečné omezující podmínky, např. S ohledem na hladkost funkcí reprezentujících změny parametrů identifikované regresní funkce.

Inteligentní evoluční operátory

Jednou z cest, jak zlepšit chování GPA je chápání evolučních operátorů jako stochastických gramatik. Pak je možno u mutace každému přechodu přiřadit příslušnou pravděpodobnost a respektovat tak například, že záměna operátoru + za funkci sin může být užitečná jen ve výjimečných případech, zatímco záměna operátoru + za operátor * má větší naději na úspěch atd. Tímto způsobem bychom rádi v budoucnu zvýšili efektivitu evolučních operátorů na úroveň vhodnou pro práci s časovými řadami, která vede ke zpracování velkých objemů dat při každém ohodnocení fitness function.

Příspěvek popisuje první postřehy vzniklé při aplikaci GA a GPA na biosignály, především pak na detekci artefaktů oznamujících příchod mikrosnátku z EEG. Hlavním problémem zde je rozsah dat, neboť minimální rozumná velikost vzorku je 10000 smpů při 40i kanálech, což dělá jakékoli operace výpočetně velmi náročné. Článek rovněž uvádí budoucí



směřování našeho výzkumu k inteligentním operátorům, vytváření dodatečných omezujících podmínek a zaměřování pozornosti evolučních operátorů.

Poznámka: Článek vznikl v rámci řešení výzkumného záměru MŠMT ČR číslo 6840770043 „Rozvoj metod návrhu a provozu dopravních sítí z hlediska jejich optimalizace“.

Literatúra

1. LANGDON W.B. a POLIR.: Foundations of Genetic Programming. Springer, New York, Heidelberg, Berlin, 1998.
2. KOZA J.R. a kol.: Genetic Programming III – Darwinian Invention and Problem Solving, (San Francisco: Morgan Kaufmann Publishers, 1999).
3. BRANDEJSKY, T. Some remarks to genetic algorithms and GPAs dynamics. In: Mendel 2008. Brno: Brno University of Technology, 2008, p. 22-25. ISBN 978-80-214-3675-6.
4. TAKAGI T. a SUGENO M., Fuzzy identification of systems and its applications to modeling and control," In: IEEE transactions on systems, man, and cybernetics, vol. 15, no. 1, pp. 116-132, 1985.
5. BOUCHNER, P. a kol.: Driver's Attention Level improvement with use of Biofeedback Stimulation incorporated into Driving Simulator, In: NEURAL NETWORK WORLD, Volume 19, Issue 1, Pages 109-118, 2009, ISSN 1210-0552
6. BOUCHNER, P. a NOVOTNY, S.: Advanced methodology for evaluation of driver's actual state with use of technical driving data, In: Proceedings of

8th WSEAS International Conference on Systems Theory and Scientific Computation (ISTAC'08), Rhoeds, GREECE, AUG 20-22, 2008, ISBN 978-960-6766-96-1

Improvement of time-series modeling methods representing different kinds of biosignals as EEG might open new application domains as microsleep detection and prediction, analysis of human operators attention etc. Presented paper discusses design of specific Genetic Programming Algorithm to efficient modeling of these series. There are designed and used intelligent evolutionary operators on the place of standard ones to maximally reduce set of tested versions (hypotheses) with respect of size of test data.

RESUME