

Oponentní posudek doktorské disertační práce

Doktorand: Ing. Lenka Kosková Třísková

Téma práce: Automated anomaly detection in geophysical survey

Školitel: Ing. Josef Novák, Ph.D.

Oponent: RNDr. Jaroslav Bárta, CSc.

Předložená práce má celkem 95 stran (včetně příloh) a text zprávy je doplněn příloženým paměťovým médiem (CD). Práce vznikla v rámci studijního programu 3901- Aplikované vědy v inženýrství, ve studijní větvi 3901VO55. Práce je vedena v anglickém jazyku. Text v úvodu obsahuje prohlášení doktorandky ohledně autorských práv a identity písemné a elektronické formy disertační práce. Po prohlášení následuje anglický a český abstrakt, klíčová slova práce a poděkování školiteli a dalším kolegům. Seznam vyobrazení, seznam tabulek a seznam zkratk předchází souboru číslovaných kapitol. Soupis zkratk je poměrně stručný. Podle názoru oponenta by bylo užitečné, aby se tak, jak bývá zvykem v geofyzikální literatuře, jednalo o úplný přehled použitých symbolů včetně jejich rozměrů, což pak zjednoduší orientaci při četbě matematických odvození apod. Zprávu autorka rozdělila do sedmi hlavních kapitol.

Kapitola 1 vysvětluje zamyšlené cíle práce. Hlavním záměrem práce je studium možností, jak rychle a s minimálními nároky na teoretické znalosti zhodnotit naměřená geofyzikální data tak, aby bylo možno poskytnout první informace například při mimořádných událostech jako je zemětřesení, povodeň apod. Autorka se v této kapitole také zmiňuje, že zpracování seismických dat bylo podpořeno projektem Modern 2020 (grant No. 662177).

Kapitola 2 je věnována podrobnějšímu definování úkolu. Autorka se zde věnuje nejprve základům tíhových projevů jednoduchých těles (koule, vertikálně a horizontálně uložený válec), přičemž je z textu patrné, že uvažuje s tělesy menších rozměrů (jednotky až první desítky metrů), které jsou umístěny blíže povrchu terénu. Zatímco v obr. 2.1 až 2.4, zachovává autorka korektní způsob vyobrazení, jaký se dá očekávat u vyobrazení pro technické účely, na obr. 2.6 až 2.10 chybí popisy os a vyobrazení mají pouze nezávazný charakter. V podkapitole 2.2 je diskutován průběžný monitorovací proces s odkazem na již zmíněný Modern 2020, na němž se Technická univerzita v Liberci podílí. Poněkud nejasná je zmínka, že pro účely úložišť jsou používány metoda seismická, metoda odporové tomografie a metoda vyzvané polarizace (str. 26, pátý odstavec). V tomto místě není zcela jasné, zda jde o nějaké obecné konstatování, nebo o nepřesnou citaci z úspěšně ukončeného (rok 2016) výzkumného úkolu sponzorovaného Technologickou agenturou ČR (úkol TA 03020408). V tomto případě je nutno poznamenat, že aplikace vyzvané polarizace není v navržené metodice ještě standardem, ale výsledkem výzkumu dalších možností monitoringu pro účely hlubinných úložišť nebezpečných odpadů. V textu disertační práce pak následuje diskuse o seismickém modelu, který je popsán na obr. 2. 11. a který je ve velmi zjednodušené až neurčité podobě ilustrací výzkumů prováděných na pracovišti ETH Zurich. Obdobně zcela nezávazný je náskres hustotních a rychlostních poměrů na obr. 2.12, které mají ilustrovat fyzikální poměry v místech experimentu. V obrázcích chybí popisy os i hodnoty izolinií, které se v zobrazení nachází.

Kapitola 3 je věnována realizaci detekce anomálií. Autorka velmi správně začíná své výzkumy na nejjednodušších případech, které může geofyzika poskytnout, tj. na tíhových anomáliích základních těles (koule, válce).

V tabulce 3.1 jsem jako čtenář disertační práce v rozpacích, co je myšleno sloupcem „The depth estimation“, protože podle mého názoru se jedná o korektně odvozený výpočet hloubky a nikoliv odhad. V obr. 3.2, 3.3 a 3.4 jsou nevhodně popsány vertikální osy. Pro účely geofyzikální praxe se nepoužívá jednotka ms^{-2} , ale μms^{-2} , nebo, v současnosti častěji, mgl.

Na obr. 3.5 až 3.9 je ilustrována diskuse o možnostech detekce tvarově jednoduchých těles metodou gravimetrickou. Diskuse sama je vedena velmi přehledně a vyobrazení potvrzují, že data naměřená nad tvarově jednoduchým tělesem lze za splnění určitých předpokladů rychle detekovat. Problematický je postup autorky předkládané práce v tom, že by měla respektovat zvyklosti a potřeby práce s geofyzikálními daty. V geofyzice se očekává, že každé vyobrazení má jasně definován souřadnicový systém a je zřejmé, jaké veličiny a v jaké velikosti jsou sledovány. Autorkou zvolená prezentace je pouze studií tvarů na obraze. (Tuto výtku je možno vznést i k řadě dalších ilustrací v následujícím textu). Jako oponent si při této výtce uvědomuji složitost hodnocení výzkumů, které se pohybují na hranici více vědeckých oborů, kdy každý oponent může preferovat svůj obor a výsledné hodnocení je potom otázkou komplexního pohledu.

Podkapitola 3.2.5 zdařile popisuje problematiku šumů, které mohou v praxi ovlivnit detekci sledovaných objektů. Problematice filtrace naměřených dat je věnována při práci s geofyzikálními daty vždy značná pozornost; prakticky každý zpracovatelský software obsahuje fázi filtrace dat. Podkapitola 3.2.5 tedy správně věnuje pozornost této problematice a naznačuje cesty, aby postup filtrace šumů byl založen na hlubších znalostech. Oponent předpokládá, že v některé z dalších etap výzkumů bude mít autorka disertační práce k dispozici i reálně naměřená data a pokročí ve svých údajích z fáze formálního modelového postupu k pokusům, jak odstranit reálný šum, což je samozřejmě mnohem těžší.

Zatímco v předchozí části práce se autorka zabývá problematikou jednoduchých těles v gravimetrii, v podkapitole 3.5 přechází na diskusi o monitorovacím procesu spojeném s detekcí struktur. V podstatě se zde přechází do jiné geofyzikální disciplíny, a to seismiky. Tento přechod není bohužel vůbec odborně připraven a čtenář je v podstatě odkázán na dva obrázky, obr. 2.11 a 2.12, ze kterých má pochopit celou problematiku experimentu, který je zřejmě na začátku všech dalších úvah. Dále je z textu patrné, že výzkumy prováděl tým v rámci úkolu Modern 2020 ve Švýcarsku. Z obrázku 2.11 je zřejmé, že se jedná o variantu uspořádání mezi seismickými rozruchy a seismickými přijímači, které se objevuje i u jiných řešitelských týmů v rámci studií tak zvaných tvrdých (skalních) hornin pro sledování geotechnických poměrů při budování podzemních děl. V tomto směru je zde záruka, že experiment řešitelů úkolu Modern 2020 (grant No. 662177) splňuje běžné nároky na aranžmá seismických měření tohoto typu. Bohužel, pro prezentaci v disertační práci jako oponent postrádám podrobnější popis experimentu, jako například:

- Frekvenční pásmo vysílačů seismických zdrojů.
- Frekvenční pásmo přijímačů seismických rozruchů.
- Geometrie vzájemného uspořádání vysílačů a přijímačů.
- Rozměry a přesné umístění detekované dutiny (tunelu).
- Jaké typy seismických vln byly sledovány, v jakém frekvenčním pásmu.
- Ukázky primárně naměřených seismických záznamů a jejich tvar po zpracování.

Nedostatek výše uvedených informací mně vede k tomu, že nelze dosti dobře hodnotit ani výsledky modelové studie vyhodnocení seismických dat, která je popisována v kap. 3.5, protože o podstatě a charakteru zkoumaných seismických vln nemáme žádné konkrétní poznatky. Z textu kapitoly 3.5 není také zcela jasný podíl autorky předkládané disertační práce na celém experimentu a není známo, s jakým výsledkem je experiment hodnocen v rámci úkolu Modern 2020 (grant No. 662177).

I když považuji výše uvedené kritické poznámky za vážné, musím zároveň rád konstatovat, že obrazová analýza seismických dat, uvedená v kapitole 3.5, je perspektivní a mohla by být pravděpodobně inspirací pro nějaký obdobný experiment v podmínkách Česka.

Kapitola 4 popisuje výsledky předkládaných výzkumů. Tato kapitola je vedena přehledně a důkladně. Opět se bohužel opakuje nedostatečný popis os v grafech. V kapitole jsou popsány výsledky z detekce gravimetrických anomálií, problematika seismických dat zde není probírána.

V kapitole 5 je uváděn stručný závěr s odkazem na možné další práce v rámci projektu Modern 2020. Zde by bylo účelné upřesnit ve kterém cíli (Task) a pracovním balíčku (Work package) se tyto práce odehrávají či odehrají.

V kapitole 6 je uvedena autorčina publikační činnost vztahující se k řešenému problému. Jedná se o účast na konferencích a článek v recenzovaném časopise EGRSE vydávaném Českou asociací geofyziků.

Bibliografie obsahuje 38 titulů. Na konci textu je přiloženo CD s doplňující databází a elektronickou verzí předkládané disertační práce.

Hodnocení předložené práce je nutno zvažovat z více pohledů. Jako za důležité považuji upozornit na aktuálnost tématu, který sice „zrál“ delší dobu, ale v disertační práci kol. Třískové je důsledně probírán poprvé. V současné době se množí obavy z náhlých přírodních katastrof (v našich podmínkách povodně) a teroristických útoků. Z praxe při povodních v Čechách z roku 2002 a obdobných situací je známo, že do výbavy záchranných týmů je poskytována i technická pomoc včetně prospekčních radarů, seismických aparatur apod., ale že plné využití této techniky naráží na nedostatečné znalosti krizového štábu. V tomto směru se zatím nezlepšilo ani povědomí sponzorů aplikovaného výzkumu, kdy v minulosti nebyly přijaty některé nabídky na výzkumné úkoly, a to i v mezinárodním měřítku (Gesher, izraelsko – český projekt). Autorka předložené práce začala správně s analýzou jednoduchých trojrozměrných těles, jak se chovají v tíhovém poli. Popis tíhových anomálií a jejich chování v různých základních situacích potvrzuje, že autorka dobře pochopila základní problematiku této geofyzikální disciplíny a může tuto problematiku ve svém výzkumu dále rozvíjet. Do dalšího pracovního plánu bych doporučil zařadit chování některých typických anomálií z radarového průzkumu a následně dalších disciplín. Aby nezůstalo pouze u akademického přístupu k tématu, je nutno modelování situací vždy prověřovat na „ostrých“ datech získaných v terénu.

Závěr oponentského posudku

S ohledem na moje odborné zaměření (geofyzik) jsem se jako oponent soustředil zejména na geofyzikální problematiku a na praktický přínos, který může vzejít z hlubší spolupráce geofyziky s informatikou. Z tohoto pohledu vidím spolupráci jako možnou a do budoucna velmi potřebnou. Urychlit určité rozhodovací fáze řízení geofyzikálních prací je žádoucí

zejména tam, kde se stává geofyzikální průzkum součástí komplexních zásahů a kde není v určité fázi vyhodnocení nastalá situace v rukou akademicky vzdělaného specialisty, ale krizového manažera (zásahy záchranných sborů apod.). I když některé části předložené práce, a to z pohledu geofyzika, lze hodnotit kriticky, na druhé straně se jedná o téma, které se jako titul disertační práce objevuje přinejmenším v českém prostředí poprvé a je tedy zapotřebí zvážit také obtíže, které autorka disertační práce v této pozici měla.

Na základě posouzení předložené práce a po zvážení přínosu řešeného tématu pro budoucí výzkum

doporučuji

práci k obhajobě před příslušnou komisí a v případě, že autorka před ní svoji práci obhájí, aby jí byl udělen titul *Ph.D.*

V Praze 15. 7. 2018

RNDr. Jaroslav Bárta, CSc.



Oponentní posudek doktorské disertační práce

Doktorand: Ing. Lenka Kosková Trísková
Téma práce: *Automated anomaly detection in geophysical survey*
Školitel: Ing. Josef Novák, Ph.D.
Oponent: doc. Ing. Milan Šimůnek, Ph.D.

Disertační práce v anglickém jazyce řeší možnost automatizace detekce anomálií při geofyzikálním průzkumu. Na začátku práce je kromě abstraktu (v anglickém i českém jazyce), obsahu a seznamu obrázků a tabulek i seznam v dalším textu používaných zkratk.

Vlastní text práce o rozsahu 75 stran (bez příloh) začíná úvodem, ve kterém je popsáno vybrané téma, cíl a struktura práce. V první věcné kapitole je popis problematiky geofyzikálního průzkumu a konkrétně detekce podpovrchových anomálií (dutin). Identifikovány jsou čtyři základní typy anomálií (kulovitá, horizontálně válcovitá, vertikálně válcovitá a v podobě kvádrů). Zároveň je diskutován i druhý z dílčích cílů týkající se kontinuálního monitorování podloží. V následující kapitole týkající se už implementace automatizace detekce anomálií je nejprve přehled aktuálního stavu poznání. Následuje popis navrženého algoritmu pro rychlé vyhledání anomálií (*fast scan algorithm*), který je postaven na rozpoznávání již dříve identifikovaných čtyř základních typů anomálií. Diskutovány jsou i některé „slepé uličky“ navštívené během výzkumu a při návrhu algoritmu. Výsledkem je klasifikátor schopný rozpoznat uvedené typy anomálií. Vysvětlen je pomocí schématu na straně 43 a zejména popisem celého procesu detekce (včetně implementované aplikace) na straně 50 a dále. Stále ve stejné kapitole jsou pak diskutovány možné problémy s praktickým nasazením (zejména přítomnost více anomálií najednou) a také i použití alternativního přístupu k automatizované detekci anomálií – pomocí různých metod strojového učení (včetně použití umělé neuronové sítě) a s použitím techniky učení s učitelem. Nakonec je v této kapitole řešen i druhý dílčí cíl kontinuálního monitorování podloží. Popisována je úprava algoritmu dříve navrženého pro rychlé vyhledávání anomálií, aby jej bylo možné využít i na kontinuální monitorování (strana 59 a dále).

Kapitola „*The Achievements*“ obsahuje důkladné otestování navržených algoritmů na čtyřech uměle vytvořených testovacích sadách. Testuje se jak schopnost správné klasifikace typu anomálie, tak i přesnost odhadu charakteristik každé z anomálií (zejména pozice a velikosti), navíc s ohledem na prostorové umístění anomálie v podloží i relativně ke zkoumané ploše. Zkoumaná je i robustnost algoritmu s ohledem na falešně-pozitivní detekci neexistujících anomálií a na šum v datech.

V závěru autorka shrnuje inspiraci pro vyhotovení práce, postup výzkumu, dosažené výsledky a konstatuje splnění cíle práce. Zároveň upozorňuje na nutnost dalšího vývoje před nasazením do praxe. V příloze je pak přehled již vydaných publikací (buď zcela samostatně, nebo takových, kde je autorka hlavním spoluautorem), seznam použité literatury a popis obsahu přiloženého DVD. K práci je přiloženo DVD se skripty v jazyce *Python*, moduly pro systém *Matlab* a rozsáhlou sadou uměle vytvořených testovacích dat.

Téma práce

Zvolené téma je aktuální a řeší praktické problémy související s možností zpřístupnění metod geofyzikálního průzkumu širší skupině uživatelů tím, že postupy do určité míry zautomatizuje a sníží nároky na *know-how* pro zpracování výsledků měření. To vše, aniž by došlo k výraznému poklesu kvality detekce. Disertační práce přesně odpovídá studijnímu programu *Applied sciences in engineering*, ve kterém je předkládána – řeší praktický problém a

zaměřuje se zejména na praktickou využitelnost dosažených výsledků (v tomto případě navíc podpořenou běžícím evropským projektem se zahraničními partnery). Zároveň jde o mezioborovou práci aplikující metody strojového učení na doménu geofyzikálního průzkumu.

Trochu nesourodě působí dva vytyčené dílčí cíle – automatizace detekce anomálií a automatizace kontinuálního monitoringu. Na věcnou souvislost však autorka poukázala tím, že algoritmus původně navržený pro detekci anomálií upravila i pro aplikaci na kontinuální monitoring.

Provedení rešerše a srovnání s mezinárodním výzkumem

Provedená rešerše v podkapitole 3.1 se dotýká obou domén – geofyzikální i informatické (strojového učení). Uvedeny jsou odkazy na základní publikace i odkaz na přehled aplikací metod strojového učení na geofyzikální průzkum a dále odkazy na řešení obdobných problémů (detekce min, podzemní infrastruktury...). U některých z uváděných zdrojů je diskutabilní jejich relevance s ohledem na rok vydání (zejména v případě popisu aplikace metod strojového učení). Patrně však v případě použití metod v geofyzikální doméně nejde o časté téma výzkumu a novější publikace nemusí existovat. Toto mohlo být však v rámci rešerše uvedeno a na podporu neexistence novějších zdrojů mohly být zároveň uvedeny všechny informační zdroje, ve kterých byla rešerše prováděna. Pomohlo by rozšíření rešerše i na vzdálenější oblasti (např. automatické vyhodnocení rentgenových snímků, nebo detekce nádorů, či rozpoznávání orgánů při automatizovaném operačním zákroku; možná i o některé dílčí úlohy z oblasti autonomně se pohybujících vozidel), které by mohly přinést důležitou mezioborovou inspiraci a zároveň mohly obsahovat i aktuálnější výsledky. Na závěr rešerše mělo být shrnutí s uvedením vlastního názoru na jednotlivé metody a techniky a zhodnocením jejich použitelnosti pro dosažení vytyčeného cíle práce.

Vlastní přínos

Hlavní přínos je v autorkou navržených modelech detekce anomálií a pro kontinuální monitorování. V praktické rovině pak i implementace GUI pro algoritmus rychlé detekce anomálií (*fast scan classifier*). Protože jde o významný výsledek, mohla být implementace v textu práce popsána ještě podrobněji.

Přestože je otestování navrženého modelu vždy nezbytným krokem, tak jako přínos vidím i důkladné otestování na několika datových sadách a zhodnocení jak kvality klasifikace typu anomálie, tak i odhadu jednotlivých doplňujících parametrů. To vše i vzhledem k nedostupnosti reálných dat a nutnosti jejich umělého vytvoření.

Ověření navržených postupů

Navržené algoritmy byly ověřeny na rozsáhlém souboru uměle vytvořených testovacích dat. Chování algoritmů bylo podrobně diskutováno v kapitole 4. Problém s nedostupností dostatečné množiny reálných dat autorka zmiňuje v závěru práce.

Ověření navržených postupů představují i publikace relevantních prací na konferencích a v časopisech s *peer-review*, jejichž seznam je na straně 87. Z nich patrně nejvýznamnější je poslední publikace z roku 2017 v *International Journal of Exploration Geophysics*.

Využitelnost v praxi

Využitelnost v praxi je zcela zřejmá v obou ze stanovených dílčích cílů a zdůrazněna probíhajícím výzkumným projektem *Modern2020*, podporou instituce *Euroatom* a také v souvislosti s dalšími projekty na *TÚ Liberec*.

Struktura práce

Práce je strukturovaná méně obvyklým způsobem a jde v podstatě o popis (téměř) všech kroků, které byly vykonány v průběhu výzkumu a při implementaci algoritmů, včetně různých „slepých uliček“. Vychází ze snahy popisovat navrhovaný systém zdola-nahoru od základních bloků až k výsledku. Patrně byl takový postup zvolen po konzultaci se školitelem. V případě poměrně rozsáhlého textu však klade na čtenáře značné nároky. Ten či jiný dílčí krok je vysvětlován do nejmenších implementačních detailů (např. včetně informace o dostupnosti použité knihovny funkce až od určité verze prostředí *Matlab*), aniž by byla jasná celková koncepce a význam tohoto dílčího kroku k dosažení cíle. To je o to výraznější, pokud je daný krok navíc následně označen jako „slepá ulička“.

Domnívám se, že bylo lepší nejprve vysvětlit základní myšlenku celého systému (včetně konceptuálního schématu) a sled kroků nutných k dosažení vytčeného cíle. Následně by mohl následovat detailnější rozbor jednotlivých kroků, včetně implementačních detailů. „Slepé uličky“ mohly být zmíněny až úplně poslední, resp. odsunuty do přílohy. Také bych doporučoval výraznější oddělení částí textů týkajících se dvou dílčích cílů.

Lepšímu porozumění textu by také pomohlo podrobnější vysvětlení struktury práce v úvodu, a zejména pak popis obsahu každé z kapitol v jejím úvodním odstavci – co je cílem kapitoly, proč je vůbec do textu zařazena a co je obsahem jejích jednotlivých částí. To by patrně vedlo i k přepracování členění kapitol práce, které je momentálně značně nevyvážené (kdy kapitola 3 zabírá téměř polovinu celé práce a obsahuje zároveň rešerši i kompletní popis všech algoritmů navržených pro splnění obou dílčích cílů).

Práce s literaturou

Práce s literaturou je adekvátní. V případě, že autorka přebírá informace nebo se odvolává, tak je citace většinou uvedena. Některé pojmy jsou v textu použity bez jejich bližšího vysvětlení, nebo alespoň odkazu na literaturu. U některých autoritativních tvrzení chybí zdůvodnění, odkaz na zdroj nebo podložení vlastním výzkumem.

Doporučoval bych dodržovat jednotný formát citací – vždy odkazem do seznamu použitých zdrojů na konci textu, bez občasného používání citování pomocí poznámek pod čarou. Některé ze zdrojů v seznamu nesplňují citační normu (týká se zejména odkazů na online zdroje, u kterých není uveden okamžik citování a v některých případech dokonce ani URL).

Srozumitelnost textu

Na textu je vidět, že angličtina není mateřským jazykem autorky. Přesto je text srozumitelný a zejména věcné pasáže v druhé části práce jsou pochopitelné i pro čtenáře bez předchozí znalosti domény geofyzikálních průzkumů. V práci je patrné zejména nadužívání určitého členu „the“, které v některých případech i mění význam textu. Na úvodních částech práce (které byly logicky psány až po dokončení hlavního textu) je už patrná únava z velkého množství času stráveného nad prací. Jsou v nich časté zbytečné chyby z nepozornosti a některé věty jsou na hranici srozumitelnosti.

Kladně hodnotím umístění seznamu použitých zkratk na začátek práce. Ne všechny v práci použité zkratky jsou v něm však uvedeny. Zároveň by čtení usnadnilo, kdyby při prvním použití zkratky v textu byl i přesto uveden její význam v závorce.

Formální úprava

Formální úprava je uspokojivá. Vyskytují se drobné prohřešky – např. nedůsledné používání dlouhé pomlčky nebo nevhodné stránkování. Diagramy by mohly vycházet z nějaké zavedené notace z metodiky návrhu informačních systémů. U vývojových diagramů by měl být vyznačen počáteční uzel, případně i koncové.

Náměty a otázky do diskuze

1. Jak náročné a jaké komplikace by byly s přípravou umělých dat pro naučení umělé neuronové sítě i pro úlohu kontinuálního monitorování? (strana 62)
2. Jaká by byla doporučení pro nastavování hodnot uživatelského parametru „*limit contrast mass*“ pro typické úlohy použití klasifikátoru v praxi? (strana 68)
3. Proč je odhadovaná hloubka anomálie vždy podhodnocována? Nešlo by takovou systematickou chybu modelu odstranit? (strana 75)
4. Podařilo se od okamžiku odevzdání práce získat komplexnější výsledky o použití umělé neuronové sítě? (strana 85)

Závěr posudku

Jako informatik jsem v posudku hodnotil zejména návrh algoritmů a jejich otestování. Nejsem schopen do důsledku posoudit věcnou správnost uváděných fyzikálních vztahů a všechny praktické důsledky využitelnosti v oblasti geofyzikálního průzkumu a monitoringu. Jejich posouzení tak nechávám na druhých dvou posudcích a oponentech s odbornou způsobilostí v této oblasti.

Na základě výše uvedeného posudku **doporučuji** práci k obhajobě před příslušnou komisí a v případě, že autorka před ní svoji práci obhájí, aby jí byl udělen titul *Ph.D.*

V Praze dne 27. 11. 2017



doc. Ing. Milan Šimůnek, Ph.D.