

T A

Č R

Výzkum užitečný pro společnost

T A
Č R

“Hodnocení a evaluace 4.0” – aplikace moderních metod AI, statistiky a data science

Představení připravované výzkumné potřeby (projektu programu BETA2) v rámci předběžné tržní konzultace

23. 6. 2022

Martin Víta

- **Partikulární:** např. vycházející z výstupů evaluace *(statistické zpracování hodnocení projektů oponenty, detekce biasů v hodnocení, ...)*
- **Obecnější:** škálovatelnost procesů
- **TA ČR a aktuální trendy:** zefektivňování procesů v administrativě použitím AI je trend poslední doby
...přičemž TA ČR jakožto instituce, která poskytuje prostředky na aplikovaný výzkum musí sama sledovat trendy a snažit se je aplikovat tam, kde je to efektivní.

- Existence dat, která se v současné době systematicky neanalyzují (např. biasy v hodnocení).
V případě, že by se analyzovaly, bylo by možné následně přijímat příslušná opatření (edukace oponentů, požadavky na výsledky práce oponentů) na základě evidence based přístupů.
- Referenti dělají práci, kterou by alespoň z části mohly nahradit stroje (formální kontrola práce oponentů)
- Automaticky by bylo možné získávat nová data (nový přístup k řízeným rozhovorům při evaluacích)

Čtyři související větve (chronologicky)

1. Statistické analýzy hodnocení projektů oponenty
2. Automatizace formální kontroly hodnocení od oponentů
3. Využití chatbotů pro automatizaci řízených rozhovorů v rámci evaluací
4. Algoritmy pro přiřazování oponentů (project-reviewer matching)

- 2, 3, 4: stejná cílová skupina potenciálních řešitelů (oblast NLP – počítačového zpracování přirozeného jazyka), podobné principy, metody, ...
- 1 a 2, 4: týkají se oponentů a zkvalitňování hodnotícího procesu
- 1 a 3: týkají se evaluací

Hlavní idea a motivace

- Kvalita hodnotícího procesu je pro TA ČR klíčová, v evaluacích se objevují kritiky týkající se této oblasti.
- V současné době je k dispozici velké množství dat, které generuje hodnotící proces, nejsou však využívány/analyzovány v takové míře, v jaké by využívány být mohly.

Cílem tohoto minitendru je:

- identifikovat případné biasy,
- vytipovat problematické momenty při hodnocení,
- vytipovat outliery (např. problematické hodnotitele),
- dát do souvislosti hodnocení oponentů, hodnocení oponenta zpravodajem a atributy získané v rámci ostatních MT (např. automatizované formální kontroly).

Ukázky:

- Jsou oponenti při hodnocení projektů spadajících do jaderné energetiky (JF) méně přísní než při hodnocení nejaderné energetiky (JE)? Objevuje se analogický jev u jiných skupin projektů (např. méně přísné hodnocení v případě sociálně citlivých a společensky obecně přijímaných projektů, např. zdravotnictví, ...)?
- Hodnotí ženy hodnotitelky přísněji projekty, kde je hlavní řešitelkou žena?
- Mají oponenti s malou zkušeností s hodnocením pro TA ČR tendenci dávat “spíše vyšší hodnocení”?

Ukázky:

- Které skupiny projektů jsou problematické vzhledem ke konzistenci hodnocení?
- Existují kritéria, která jsou dlouhodobě oponenty nízko či naopak vysoce hodnocena?
- Existuje vztah mezi hodnocením některých kritérií a socioekonomickou, demografickou charakteristikou oponenta (věk, pracovní pozice, dosažené vzdělání, ...)?
- Čím více projektů oponent hodnotí, tím stručněji a obecněji zdůvodňuje své ohodnocení.
- Pokud si oponent opraví již vrácené posudky před hodnocením dalších, má následně nové posudky bez chyb či s jejich malým množstvím. (U těchto oponentů předpokládáme, že si prve pořádně nepřečetli manuál a po získání zpětné vazby se přizpůsobí požadavkům.)

Formalizace úloh(y)

- Jde o tvorbu hypotéz nad daty z hodnotícího procesu a dalších analýz.
- Dále jde o data mining – hledání častých vzorů (frequent pattern mining)
- Rovněž se jedná se o takové oblasti jako např.:
 - inter-rater reliability, shoda hodnotitelů, spolehlivost hodnocení: jaká je a jak se liší napříč programy, jak se liší pro skupiny různých charakteristik.
 - položková analýza I: odlišné fungování pro skupiny. Nakolik položky fungují jinak pro skupiny, definované nějakou proměnnou nebo jejich kombinací. Hledání latentních tříd, ...
 - položková analýza II: nakolik jednotlivé položky korelují s ostatními položkami a celkovým skóre a jak dobře rozlišují mezi silnými a slabými projekty (hodnoceno např. pořadím)
- ...

T A
Č R

MT1: Statistické analýzy hodnocení projektů oponenty

Výstupy

- V_souhrn: Přehled reportů / výzkumná zpráva obsahující i doporučení spolu s odhadovanými dopady.

(Otázka publikace jako journal paper?)

T A MT2: Automatizace formální kontroly Č R hodnocení od oponentů

Hlavní idea a motivace

Automatizace formální kontroly hodnocení od oponentů – automatické rozpoznávání/klasifikace např. toho, zda:

- se jedná/nejedná o **příliš obecné konstatování** ze strany oponenta k danému kritériu,
- se jedná/nejedná o **pouhou parafrázi věty ze zadávací dokumentace** či příruček pro hodnotitele a dalších
- je **obsah textu v souladu s bodovým ohodnocením** ve smyslu “*snížené body – existence alespoň nějakých námitek*” atp.)

Formalizace úlohy

- Jde v podstatě o **tradiční klasifikační úlohu**, resp. úlohy z oblasti NLP.
- **Vstup:** text od oponenta + příslušné bodové ohodnocení. Klasifikace způsobem: *“body jsou v souladu s textem”*, *“nejspíše existuje nesoulad”*, *“rozhodně existuje nesoulad”*, dále třeba *“vyjádření oponenta je/není příliš obecné”*, atp.

Poznámky k řešení úlohy

- Úloha bude (nejspíše) řešena jako “supervised task” s využitím toho, že z předchozího manuálního zpracování existuje určité množství trénovacích dat.
- Předpokládá se (nejspíše) i vytvoření případných dalších syntetických trénovacích dat.
- Na základě analýzy dat z manuálního zpracování (a podkladů ze zaškolování “kontrolujících”) by byly vytvořeny anotační schéma a manuál (do jakých tříd budeme klasifikovat atp.)

Další funkcionalita

Kromě funkcionality uvedené na slidu *Hlavní idea* je rovněž zapotřebí detekovat:

- zda oponent nekopíruje příliš velké části textu z návrhu projektu a zároveň také kontrolovat,
- zda oponent nepoužívá pro “příliš mnoho projektů” stejné formulace (které by tvořily příliš velkou část výsledného textu oponenta).

V úvahu také přichází monitorování časového průběhu změn v textu oponenta, je-li hodnocení opakovaně “vráceno”. Některé části těchto funkcionalit mohou být řešeny jako *unsupervised task* či pravidlově.

Příklady

Stručnost jednotlivých kritérií / přílišná obecnost

Hodnocené kritérium: Řešitelský tým

Nápověda: Posuďte, zda členové řešitelského týmu mají potřebné zkušenosti a odborné předpoklady pro dosažení plánovaných výsledků. Vyhodnořte dosavadní zkušenosti členů řešitelského týmu s projekty VaVaI a předpoklady řešitelů. Zhodnořte, zda dosažené zkušenosti členů týmu odpovídají jejich roli v týmu a plánovaným činnostem. Posuďte míru zapojení mladých výzkumných pracovníků v řešitelském týmu, která je žádoucí.

Hodnocení oponenta: Řešitelský tým je odborný a je podrobně uvedený.

Text k vrácení: Prosíme o doplnění kritéria č. 2 Řešitelský tým. Kritérium je zhodnoceno příliš stručně.

Příklady

Hodnocené kritérium: Řešitelský tým

Nápověda: Posuďte, zda členové řešitelského týmu mají potřebné zkušenosti a odborné předpoklady pro dosažení plánovaných výsledků. Vyhodnoťte dosavadní zkušenosti členů řešitelského týmu s projekty VaVaI a předpoklady řešitelů. Zhodnoťte, zda dosažené zkušenosti členů týmu odpovídají jejich roli v týmu a plánovaným činnostem. Posuďte míru zapojení mladých výzkumných pracovníků v řešitelském týmu, která je žádoucí.

Hodnocení oponenta: Tým je složen z odborníků, a to jak na straně výzkumné organizace tak i u obou komerčních společností. Avšak hlavní řešitel z ČVUT nemá konkrétní zkušenosti s výzkumem předmětu řešení. Ostatní osoby podílející se na předmětu projektu mají adekvátní zkušenosti jak s VaV, tak s realizačními stavebními projekty velké důležitosti (pod dohledem SŽ, či ŘSD).

Text k vrácení: Kritérium č. 8 bodově hodnotíte jako splněné bez výhrad, avšak v textu máte výhrady. Prosíme o soulad textu s body.

T A MT2: Automatizace formální kontroly Č R hodnocení od oponentů

Systemová integrace

V rámci této části projektu bude vytvořen

klasifikátor/sada klasifikátorů, které budou následně *implementovány do podoby webových služeb*.

Ty bude možné následně volat z jakéhokoliv relevantního systému (SISTA, ...)

T A MT2: Automatizace formální kontroly Č R hodnocení od oponentů

Výstupy

- R – software implementující klasifikátory připravený k experimentálnímu využití/integraci do SISTA formou WS.
- V_souhrn – závěrečná výzkumná zpráva
- D / J_sc – konferenční příspěvek nebo journalový paper popisující úlohu a řešení

MT3: Využití chatbotů v rámci evaluací

Hlavní idea a motivace

- Součástí řady evaluací programů podpory je i provádění řízených/strukturovaných rozhovorů s předem danou cílovou skupinou.
- Některé experimenty naznačují, že pomocí chatbotů lze provádět určité části kvalitativního výzkumu, viz: *A new perspective: Can a Chatbot do a Research Interview?* – KIN Center for Digital Innovation

Cílem dílčího projektu je prozkoumat možnosti nasazení chatbotů v rámci přípravy evaluací a zjištění vhodných situací, kdy je použití chatbotů výhodné.

MT3: Využití chatbotů v rámci evaluací

Hlavní idea a motivace

- Součástí řady evaluací programů podpory je i provádění řízených/strukturovaných rozhovorů s předem danou cílovou skupinou.
- Některé experimenty naznačují, že pomocí chatbotů lze provádět určité části kvalitativního výzkumu, viz: *A new perspective: Can a Chatbot do a Research Interview?* – KIN Center for Digital Innovation

Cílem minitendru je prozkoumat možnosti nasazení chatbotů v rámci přípravy evaluací a zjištění vhodných situací, kdy je použití chatbotů výhodné, návrh a implementace příslušného experimentálního chatbota.

MT3: Využití chatbotů v rámci evaluací

Výzkumné otázky (návrhy)

- Do jaké míry lze řízené rozhovory automatizovat s využitím chatbotů?
Je pravděpodobné, že “chatbot se nevyrovná lidskému tazateli, který je schopen pružně reagovat na nové kontexty a podněty”, nicméně jejich nasazení může přinést nové podněty a umožní řádově navýšit počty respondentů.
- Jakým způsobem dále zpracovávat textová data získaná užitím chatbotů (text miningové přístupy, automatizace reportů)?
- Do jaké míry lze chatbota customizovat / upravovat pro různé varianty rozhovorů?
- Jak efektivně dále zpracovávat (např. sumarizovat) textová data získaná užitím chatbotů a kolik další “lidské” práce bude toto zpracování vyžadovat?
- Jak budou vnímat nasazení chatbotů v evaluacích respondenti?

MT3: Využití chatbotů v rámci evaluací

Výstupy

- R – software implementující chatbot nad konkrétním schématem řízeného rozhovoru připravený k experimentálnímu využití
- V_souhrn – závěrečná výzkumná zpráva, která reaguje na dohodnuté výzkumné otázky a obsahující studii proveditelnosti využití chatbotů v dalších aktivitách TA ČR.
- D / J_sc – konferenční příspěvek nebo journalový paper popisující úlohu a řešení

T A MT4: Algoritmy pro přiřazování oponentů Č R (project-reviewer matching)

Hlavní idea a motivace

V současné době používaný přístup přiřazování oponentů k projektům v určitém procentu případů neposkytuje dostatečně dobré výsledky, což nepříznivě ovlivňuje kvalitu hodnotícího procesu.

Cílem tohoto minitendru je vytvořit a implementovat nový algoritmus, který zajistí *kvalitnější* přiřazování oponentů návrhům projektů.

To vyžaduje mj. netriviální formulaci metrik “které přiřazení návrhů projektů oponentům je ‘lepší’ než jiné”, což bude následně využito při vyhodnocení/posouzení navrhovaných algoritmů/přístupů.

Dalším předpokladem pro výzkum v této oblasti je analýza dostupných informací o jednotlivých entitách (zejm. potenciálních oponentech), které jsou/mohou být relevantní pro přiřazování.

TAČR MT4: Algoritmy pro přiřazování oponentů (project-reviewer matching)

Výzkumné otázky

- Jaké jsou aktuální “state-of-the-art” přístupy k této problematice (vč. souvisejícího problému Reviewer Assignment Problem)?

Viz např. *Reviewer assignment based on sentence pair modeling*, *Identifying expertise through semantic modeling: A modified BBPSO algorithm for the reviewer assignment problem* a mnohé další

- Které z přístupů jsou využitelné v prostředí ČJ a TA ČR?
- Jaké metriky pro evaluaci algoritmu přiřazení jsou vhodné?

Jádrem výzkumu a vývoje by byl výzkum a vývoj v oblasti samotného algoritmu přiřazování a vývoj prototypu.

T A MT4: Algoritmy pro přiřazování oponentů Č R (project-reviewer matching)

Výstupy

- V_souhrn: souhrnná výzkumná zpráva
- R – software: software pro implementaci algoritmu pro přiřazování spolu s návrhem integrace s (S)ISTA (WS?)
- R – software – software pro případný preprocessing dat
- D / J_sc – konferenční příspěvek nebo journalový paper popisující úlohu a řešení

Základní parametry VZ

- Rámcová dohoda.
- Doba řešení max. 24 měsíců.
- Minitendry lze vyhlašovat souběžně.

- Každý tým musí mít alespoň jednoho garanta pro každou kvalifikaci (K1 - K5).
- Každý člen výzkumného týmu doloží diplom, CV a praxi vyznačí v CV.
- Další členové výzkumného týmu mohou pracovat v rolích.
- Subjekty se mohou spojovat v konsorcia.

K1: Statistik

- min. **MGR** vzdělání
- **sociologie, ekonomické obory** nebo **matematika**, případně obor příbuzný se zaměřením na **statistiku, ekonometrii** či **data science**.
- min. 3 roky navazující praxe s aplikací statistiky ve společenskovědních oblastech

K2: NLP + ML

- min. **MGR** vzdělání
- **informatika** nebo **kybernetika**
- min. 3 roky navazující praxe s aplikací metod strojového učení v oblasti zpracování přirozeného jazyka a přípravou jazykových zdrojů/korpusů pro účely trénování příslušných modelů.

K3: Teorie grafů a optimalizace

- min. **MGR** vzdělání
- **matematika, informatika** nebo **kybernetika**
- min. 3 roky navazující praxe v oblasti aplikace teorie grafů a optimalizace

K4: Odborník na kvalitativní výzkum

- min. **MGR** vzdělání
- **psychologie** nebo **sociologie**, a příbuzné obory
- min. 3 roky navazující praxe s kvalitativním výzkumem ve společenských vědách

K5: Programátor

- min. **Bc.** vzdělání
- **informatika** nebo **kybernetika**, a příbuzné obory
- min. 3 roky navazující praxe s vývojem software v oblasti zpracování přirozeného jazyka a s implementací webových služeb

Diskuse

Děkuji za pozornost!

Dotazy, náměty, připomínky, prosím, zasílejte na
vita@tacr.cz