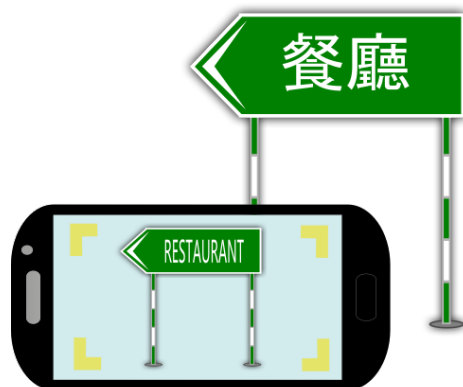


Strojový překlad



Obsah

1	Úvod	2
2	Co to vlastně je strojový překlad?	2
3	Historie	2
4	Metody strojového překladu	4
4.1	Pravidlový strojový překlad – RBMT (Rule-based Machine Translation)	4
4.2	Statistický strojový překlad – SMT (Statistical Machine Translation)	4
4.3	Překlad pomocí neuronových sítí – NMT (Neural Machine Translation)	5
4.4	Závěr – aktuální stav strojového překladu	6
5	Zdroje	7
5.1	Reference	7
5.2	Obrazové zdroje	9

1 Úvod

Strojový překlad je jednou z disciplín kybernetiky a informatiky z oboru umělé inteligence, se kterou se dnes setkáváme dennodenně. Každý určitě alespoň jednou použil nějaký strojový překladač – Google Translate, Bing Translator,... Jak ale fungují, a jak se vyvíjely? Na to se tato práce zaměří.

2 Co to vlastně je strojový překlad?

Když mluvíme o strojovém překladu, je tím myšlen automatický překlad textu pomocí počítače z jednoho jazyka do druhého, a to bez jakéhokoliv vnějšího zásahu člověka. Výstupní kvalita však bohužel není stále dostatečná pro jeho surové využití, a je třeba korekcí od lidského překladatele.

3 Historie

Historie strojového překladu se začala psát v roce 1954, kdy byl v rámci takzvaného Georgetownského experimentu předveden první strojový překladač. Tento překladač běžel na elektronkovém sálovém počítači IBM 701, a byl schopen překládat pouze z ruštiny do angličtiny, s poměrně malým slovníkem (primitivní pravidlový překlad).



Obrázek 1: Procesorová elektronková jednotka počítače IBM 701, Autor: pdx-diver, Licence: CC-BY-SA 2.0

Bohužel, po tomto experimentu nadšení pro strojový překlad silně opadlo, a s ním i financování vývoje. Důvodem pro tento pád zájmu bylo zjištění, že překlad není pro počítač tak jednoduchý, jak se dříve vědci domnívali, společně s v té době velmi pomalým pokrokem. Zlom v oboru přinesl až nástup strojového učení v 90. letech, s nímž přišel přestup na statistické metody překladu, po němž o čtvrt století později dorazil překlad pomocí neuronových sítí.

4 Metody strojového překladu

Nyní si přiblížíme jednotlivé metody strojového překladu, a jejich vlastnosti.

4.1 Pravidlový strojový překlad – RBMT (Rule-based Machine Translation)

Pravidlový překlad je nejstarším způsobem strojového překladu. Jeho funkce je založena na překladovém slovníku a sadě jazykových pravidel, definujících podobu výstupu. Tento způsob překladu je však velmi náročný na management a vývoj, a je tedy využíván pouze na specifické překlady (Příklad – Kanadský systém METEO pro překlad předpovědi počasí z Francouzštiny do Angličtiny a naopak v provincii Quebec, v provozu od roku 1981 do roku 2001) mezi málo jazyky. Tento typ strojového překladu byl ovšem velmi rychle nahrazen statistickým strojovým překlad (SMT) na konci 90. let.

4.2 Statistický strojový překlad – SMT (Statistical Machine Translation)

Statistický strojový překlad byl velkým pokrokem po pravidlovém strojovém překladu. Jeho nástup byl umožněn zvýšením síly počítačů, ke kterému došlo v 90. letech 20. století. Tento systém je založen na učení z paralelních jazykových korpusů, sestavených z textů přeložených lidskými překladateli, které jsou přeloženy do mnoha jazyků. Systém používá korpusů na místě pevně překladových slovníků a jazykových pravidel – využívá statistických výpočtů k hledání kandidátů překladu v automaticky dle dvojjazyčného korpusu sestavené tabulce (prostá substituce) společně s z jednojazyčného korpusu natrénovaným modelem jazyka pro lepší koherenci a smysluplnost výstupu. U statistického překladu máme možnost využít dva rozdílné přístupy – frázový a hloubkový překlad (případně jejich kombinaci – dochází k vzájemnému doplňování jejich nedostatků; lze kombinovat dvěma způsoby z pohledu pořadí aplikace přístupů). Každý z přístupů má své výhody a nevýhody – frázový hledá lépe překladové ekvivalenty a ustálená spojení, ale delší věty často nejsou srozumitelné, zatímco hloubkový překlad má správnou stavbu vět (hlídá celkovou strukturu věty oproti frázovému překladu), ale výstup nedává smysl kvůli nevhodné volbě překladových ekvivalentů.

Zároveň mají statistické překladače obecně nevýhodu ve své náhodnosti a sklonu k nevhodnému využívání idiomů – tyto potíže jsou způsobeny tím, že vzhledem ke statistické metodě překladače (využívá matematické interpretace vstupu překladače), a tím pádem silné závislosti na slovosledu, slovech v textu,... – mění se tím matematický popis téměř stejných vzorků textu. Sklon k nevhodnému používání idiomů je pak způsoben tím, že největší součástí korpusu je úřední text, ve kterém se idiomy nevyskytují, a méně neformálního textu. Pokud se pak v neformální části korpusu vyskytne idiom převážně v nějakém kontextu, překladač má sklon volit častější kontext jako výstupní překlad. Nejznámějším příkladem statistického překladače byl Google překladač, před přechodem od statistické metody překladače k překladači pomocí neuronových sítí (NMT).

4.3 Překlad pomocí neuronových sítí – NMT (Neural Machine Translation)

Překlad pomocí neuronových sítí je nejmladší technologií strojového překladače, a zároveň nejúspěšnější – na izolovaných větách je dosaženo velmi vysoké kvality překladače, která snadno předčí dříve využívaný statistický strojový překlad. Tato technologie je založena na tzv. neuronových sítích sequence to sequence, neboli transformerech, jak je nazval Google. Tato technologie je schopna i další převodů sekvence znaků na jinou (např.: text na řeč (text to speech, TTS), řeč na text (speech to text, STT),...) za pomoci stejné architektury, ale to je mimo téma této práce. U překladače pomocí neuronových sítí není jazykový model založen na jakýchkoliv výchozích parametrech – veškeré potřebné parametry se pak učí z paralelních korpusů. tento natrénovaný model se pak dělí na enkodér – tvoří matematickou reprezentaci vstupní věty, a dekodér – na základě matematické reprezentace generuje výstupní větu dokud nedosáhne jejího konce (speciálního symbolu). I přes veškeré technologické postupy však není překlad dokonalý – tím že překlad probíhá na úrovni jednotlivých vět, a překladač se při učení snaží najít statisticky nejjednodušší vysvětlení dat, a tím pádem dochází ke tvorbě stereotypů, které pak mohou vést k chybným překladům, jelikož neuronová síť se snaží věrně napodobovat trénovací data.

4.4 Závěr – aktuální stav strojového překladu

Když shrneme aktuální stav strojového překladu, jedná se stále jen o pouhou pomůcku překladatele, jelikož kvalita čistě strojového překladu je stále nedostačující pro jakékoliv reálné nasazení, a je ovlivněna mnoha různými faktory. I přes veškeré moderní pokroky je však vývoj stále pomalý, a k od počátků oboru vysněnému plně autonomnímu překladu použitelnému bez lidského zásahu máme ještě velmi daleko.

5 Zdroje

5.1 Reference

- [1] PŘISPĚVATELÉ WIKIPEDIE, 2022. Strojový překlad. In: *Wikipedie: otevřená encyklopedie* \[online\]. Praha (CZ): Wikimedia Česká Republika, 29. 11. 2022 \[cit. 2022-12-31\]. Dostupné z: https://cs.wikipedia.org/wiki/Strojov%C3%BD_p%C5%99eklad
- [2] MUSIL, Tomáš, 2017. Strojový překlad z hlediska filosofie jazyka [online]. Olomouc [cit. 2022-12-31]. Dostupné z: https://studenta-veda.ff.cuni.cz/wp-content/uploads/sites/125/2017/04/strojovy_prekklad.pdf. Soutěžní písemná práce. Univerzita Palackého v Olomouci, Filozofická fakulta.
- [3] DERYAGIN, Max, Daniel LANDES a Miroslav POŠTA, RENARD, Estelle, Barbora VRBOVÁ, Marta BARTOŠKOVÁ, Ivona ČERNOHAUSOVÁ, Emma CARRAUD a Šimon SEDLÁČEK, ed., 2022. Manifest AVTE o strojovém překladu. In: JTPunion.org [online]. Praha (CZ): Jednota tlumočnicků a překladatelů, 2022 [cit. 2022-12-31]. Dostupné z: https://www.jtpunion.org/wp-content/uploads/2022/04/ToP-143_prioha-MANIFEST-AV-TE-O-STROJOVEM-PREKLADU.pdf
- [4] RNDR. BOJAR, Ondřej Ph.D., 2012. Strojový překlad. Vesmír: věda, příroda, člověk, společnost – časopis s tradicí od roku 1871 [online]. Praha (CZ): Vesmír, 6. 9. 2012, 2012(9), 488–490 [cit. 2022-12-31]. ISSN 1214-4029. Dostupné z: <https://vesmir.cz/cz/casopis/archiv-casopisu/2012/cislo-9/strojovy-preklad.html>
- [5] NEVĚŘILOVÁ, Zuzana, 2005–2023. Jak funguje strojový překlad: Používáte překladač Google a podobné překladatelské nástroje? A tušíte, jak vlastně fungují. In: *Magazín M: Zprávy z MUNI* [online]. Brno: Masarykova univerzita, 15. prosince 2013 [cit. 2023-01-01]. Dostupné z: <https://www.em.muni.cz/vite/4053-jak-funguje-strojovy-preklad>
- [6] Co je strojový překlad?, b.r. In: *K&J Translations* [online]. Praha (CZ): K&J GROUP, 5 dubna, 2022 [cit. 2023-01-01]. Dostupné z: <https://www.kjtranslations.cz/blog/co-je-strojovy-preklad/>

- [7] LIBOVICKÝ, Jindřich, b.r. Neuronové sítě a strojový překlad. In: Jindřich's Blog [online]. Munich (DE): Jindřich Libovický, Oct 10, 2020 [cit. 2023-01-01]. Dostupné z: <https://jlibovicky.github.io/2020/10/10/Neuronove-site-a-strojovy-preklad.html>
- [8] NOVÁK, Ondřej, c1997 – 2022. Z historie strojového překladu: Od lamáčů nacistických šifer po neuronové sítě. In: Lupa.cz: server o českém Internetu [online]. Praha (CZ): Internet Info, 31. 5. 2021 [cit. 2023-01-01]. Dostupné z: <https://www.lupa.cz/clanky/z-historie-strojoveho-prekladu-od-lamacu-nacistickyh-sifer-po-neuronove-site/>
- [9] ČÍŽEK, Jakub, c2023. Proč má strojový překlad stále daleko k dokonalosti: Vše začalo před 60 lety na Georgetownské univerzitě. Za tu doby ušly strojové překladače kus cesty, dodnes si ale vylámou zuby na kdejakém idiomu. In: MobilMania.cz: O mobilech víme vše [online]. Praha (CZ): CZECH NEWS CENTER, 5. července 2015 [cit. 2023-01-03]. Dostupné z: <https://mobilmania.zive.cz/clanky/proc-ma-strojovy-preklad-stale-daleko-k-dokonalosti/sc-3-a-1331238/default.aspx>
- [10] DŘÍMAL, Martin, b.r. Proč nepoužívat překladač Google pro tvorbu webových stránek. In: Interval.cz: Svět Internetu, Technologií a Bezpečnosti [online]. Brno: ZONER, Kvě 26, 2016 [cit. 2023-01-03]. Dostupné z: <https://www.interval.cz/clanky/proc-nepouzivat-prekladac-google-pro-tvorbu-webovych-stranek/>
- [11] WIKIPEDIA CONTRIBUTORS, 2023. METEO System. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 21 December 2022 [cit. 2023-01-13]. Dostupné z: https://en.wikipedia.org/wiki/METEO_System

5.2 Obrazové zdroje

[Titulní obrázek] J4P4N, b.r. Smartphone Translation. In: Openclipart: Clipping Culture [online]. St Louis (US): Fabricatorz Foundation, December 18, 2018 [cit. 2023-01-10]. Dostupné z: <https://openclipart.org/detail/311877/smartphone-translation>

[1] PDXDIVER, 2023. IBM 701 frame. In: Wikimedia Commons [online]. San Francisco (CA): Wikimedia Foundation, April 7, 2009 [cit. 2023-01-10]. Dostupné z: https://commons.wikimedia.org/wiki/File:IBM_701_frame.jpg