

**Posudek oponenta práce #filterbubble
předložené Františkou Sandroni a Jakubem Dostálem,
v rámci středoškolské odborné činnosti
Tomáš Fürst**

Práce pojednává o vysoce aktuálním tématu – o naší tendenci obklopovat se v digitálním světě jen informacemi, které nám vyhovují a které potvrzují naše vidění světa. Společnost se tím rozkládá na mnoho virtuálních bublin, které spolu prakticky nekomunikují a často názorově driftují k extrémním pozicím. Když se potom náhodou lidé z různých bublin setkají (například u voleb), mají pocit, jako by mluvili s mrtvými.

O tomto fenoménu se hodně mluví, ale málokdo je schopen uchopit jej systematicky a kvantitativně. V tomto aspektu je předložená práce skutečně originální a po právu velmi ambiciózní. Autoři nejprve představí, jak informační bubliny vznikají a proč jsou zajímavé. Tato část textu je psaná rozumně a s velkým nadhledem. Dále se věnují představení sociální sítě Twitter a vysvětlení, jak budou data z Twitteru získávat a jak s nimi budou pracovat. Podstatou automatické analýzy textu tweetů je takzvaná sentimentální analýza. Pomocí pokročilých metod strojového učení (v tomto případě se jedná o konvoluční neuronové sítě) se algoritmus naučí rozpoznávat „sentiment“ tweetu, neboli jestli je jeho obsah spíše pozitivní či spíše negativní. Je fascinující, že autoři zvládají takto pokročilé metody strojového učení na tomto vzdělávacím stupni. Mohu prohlásit, že z řad absolventů příslušných oborů českých univerzit by si s tímto problémem v praxi poradil málokterý. Jeden z autorů – Jakub Dostál – se však v oblasti matematických modelů a strojového učení orientuje výborně, loni dokonce získal cenu *Česká hlavička* za matematické modelování šíření epidemií.

Jádrem práce je návrh metody, jak efekt informační bubliny pomocí sentimentální analýzy měřit. Jedná se o nápad velice chytrý a dle mého názoru originální. Autoři zkoumají, jaký obsah na Twitteru sledují příslušníci dané skupiny, z tohoto obsahu vyberou nějaké klíčové slovo a provedou sentimentální analýzu všech k nim přichozích tweetů, které toto slovo obsahují. Informační bublina může být potom charakterizována tím, jak se četnost a sentiment takových tweetů liší od „pozadí“ (tedy od náhodného vzorku z Twitteru), případně od jiné skupiny uživatelů.

Už tento nápad by byl cenným přínosem ke studované problematice. Autoři jej však ještě skvělým způsobem ilustrovali na dvou velmi aktuálních a poměrně výbušných tématech současné americké politické diskuse – na postoji různých skupin obyvatelstva k problematice potratů a na postoji k prezidentu Trumpovi. Obě analýzy jsou výborně provedené, velice zajímavé, skvěle ilustrují funkčnost navržených postupů a navíc jsou výborně okomentované a vysvětlené. Kdybych měl práci přece jen něco vytknout, doporučil bych autorům, aby kolem politických reálií současných Spojených států amerických našlapovali ještě o něco opatrněji, než v práci činí.

Celkově soudím, že práce je vysoce originální, podivuhodně vyspělá, nebývale multidisciplinární a navíc vysoce aktuální. Svým záběrem i kvalitou převyšuje standard diplomových prací na leckterém vysokoškolském oboru. Autorům rozhodně doporučuji, aby ji (po jistých stylistických úpravách a překladu do angličtiny) publikovali v impaktovaném časopise.

