

Protože vaše „šťastný“ je jiné než moje „šťastný“

Umělá inteligence proniká do různých oborů lidské činnosti už nějakou dobu. O tom, že ani hudba není výjimkou, v rozhovoru vypráví Martin Nedvěd, jeden ze zakladatelů úspěšného technologického start-upu AIMS.

TEXT: TOMÁŠ ROREČEK

Jestli tomu správně rozumím, vytvořili jste s kolegy technologii, která umí na základě principů umělé inteligence velmi rychle a efektivně analyzovat hudební skladby a vyhledávat další, které jsou jim podobné. K čemu a pro koho se to hodí?

Na začátku bylo, že jsme to potřebovali sami pro sebe. Působím ve firmě Hudební banka, která se během let dostala v Česku do pozice největšího poskytovatele hudby pro audiovizuální projekty. Máme skutečně hodně rozsáhlý katalog a bylo stále obtížnější se v něm orientovat jen na základě metadat a klíčových slov, což je standardní způsob, kdy se každá nahrávka tzv. otaguje a dělají to lidé, ne stroje. Na světě je spousta firem i jednotlivců, co se tímhle „ručním“ tagováním živí. Člověk pak zadá jako parametry třeba klíčová slova „šťastný“, „rychlý“, „rock“, „elektrická kytara“, a vyjedou mu nějaké výsledky. Tenhle způsob je ovšem z mnoha důvodů problematický, není dostatečně konzistentní, příliš se v něm projevuje lidský faktor, protože vaše „šťastný“ je jiné než moje „šťastný“. To, co někdo popíše jako „radostný“, jiný člověk otaguje třeba jako „optimistický“. Samozřejmě některá kritéria jsou poměrně objektivní, například hudební žánry. Ale už ne právě nálady, což je přitom jeden z nejpoužívanějších

vyhledávacích parametrů. Způsob vyhledávání podle klíčových slov zadávaných lidmi je prostě příliš subjektivní a nepřesný. Navíc při dnešním objemu dat a množství skladeb je to nepraktické, pomalé, a kvůli tomu i dost drahé. Rozhodli jsme se, že to chceme zkusit dělat líp. Žertovali jsme, že hledáme svatý grál našeho oboru. A praxe ukazuje, že jsme jej skutečně našli.

Rozhodli jsme se, že to chceme zkusit dělat líp. Žertovali jsme, že hledáme svatý grál našeho oboru. A praxe ukazuje, že jsme jej skutečně našli.

Chápu, já mám z reklamy zkušenosti s fotobankami, kde se taky hledá podle klíčových slov. Jenže s fotkami je to jednodušší, protože když si člověk zadá

příslušné parametry, vyjede mu několik stránek výsledků, které na monitoru v malých náhledech vidí najednou a nemusí je v první fázi jednotlivě otevírat. Dá se rychle poznat, co se bude hodit a co ne. S hudbou to takhle dělat nejde, není jednoduše staticky vizualizovatelná.

Ano, u hudby to takhle dělat není možné. Protože z nějakých křivek na monitoru málokdo pozná, o co přesně jde a zda je to pro něj relevantní. Optimální řešení ale nespočívá v klíčových slovech, je to o zvukové podobnosti. Naše technologie umožňuje vyhledávat tak, jak člověk ve skutečnosti přemýšlí. My lidi většinou neuvažujeme v kategoriích typu „šedesátá léta“, „saxofon“, „rock“ a podobně. Od našich klientů dostáváme spíše zadání typu „chceme něco, jako je XY, něco v podobném duchu“. Pošlou nám referenční nahrávku a chtějí, abychom hledali podle ní. Na základě analýzy audiostopy jsme toho schopní doslova během několika vteřin.

Kdo jsou vaši klienti?

Našimi zákazníky jsou primárně hudební společnosti, které vlastní svoje katalogy skladeb a chtějí v nich lépe hledat, aby je mohly efektivněji nabízet. A pak samozřejmě různí tvůrci audiovizuálního obsahu – televizní stanice, reklamní agentury, produkční firmy.

Proč rovnou nepoužijí tu referenční nahrávku, je to otázka peněz?

Jen do určité míry. Peníze samozřejmě hrají roli, ale rozpočet nebývá ten nejobvyklejší důvod. Často tu referenční nahrávku nechtějí použít proto, že je příliš známá, strhává by na sebe pozornost, a přehlušila by tak vyznění audiovizuálního celku.

Jak jste ke své dnešní technologii došli?

Napřed jsme se dívali všude po světě, jestli už někdo vyhledávání na základě zvukové podobnosti na dobré úrovni umí. Objevili jsme francouzský start-up Nieland, který měl tento typ technologie vyvinutý, a začali jsme s nimi

spolupracovat. Ve vzájemné symbióze jsme jejich produkt dostali na skutečně velmi dobrou úroveň. Jenže oni nám po několika letech oznámili, že je kupuje Spotify, čímž jsme o jejich služby přišli. Byla to pro nás hodně složitá situace, protože naši klienti si mezitím na tenhle způsob vyhledávání zvykli a vyžadovali ho. Když někdo něco vůbec nemá, je to jiné, než když už to měl a vy mu to na jednu seberete, takže jsme to museli řešit. Zjistili jsme, že mezitím nikdo na světě srovnatelně kvalitní technologii stále ještě nenabídl, takže nám vlastně nezbývalo než se do toho pustit sami. Jen jsme se museli pár věcí naučit. Klíčový pro vývoj naší samostatné technologie byl můj společník Viktor Parma, bez

Napřed jsme se dívali všude po světě, jestli už někdo vyhledávání na základě zvukové podobnosti na dobré úrovni umí.



Zleva: Viktor Parma a Martin Nedvěď, foto: Lukáš Bíba

něj by to nebylo. Stejně jako já vyrostl v hudební rodině a sám má zkušenosti se skládáním. Zároveň je to ale inženýr a počítačový nadšenec, což je pro to, co děláme, naprosto ideální kombinace.

Jak to celé funguje technicky?

Používáme tzv. deep learning, různé techniky a postupy, kdy umělé neuronové síti poskytujeme velké množství skladeb s metadaty a opakovaným trénováním variantních modelů se postupně dostáváme ke stále lepším predikcím. Umíme algoritmy přimět, aby se „samy“ naučily, co od nich potřebujeme. V podstatě simulujeme lidský mozek a vystavujeme stroje podobným typům podnětů, jako když se učí člověk. Když slyšíte určitý typ hudby, stimuluje vám to dráhy mezi mozkovými neurony, takže když pak slyšíte podobnou hudbu, může toho váš mozek využít. Stejně to dělá stroj – umí poznatky zobecňovat a aplikovat na nové situace. Pokud jde o hudbu, podobnost se těžko definuje, často nejde nutně ani o stejný žánr, je to kombinace řady různorodých atributů. Člověk mnohdy není schopný přesně vysvětlit, proč mu něco připadá podobné, ale přesto tu podobnost jasně vnímá. Stroj je dnes schopný se ty věci naučit stejně jako my a stejně je „cítit“. Lidi si často myslí, že umělá inteligence není schopná poznat emoce, ale není tomu tak, mašina prostě i v tomhle ohledu „uvažuje“ v určitých vzorcích, stejně jako my, jen je daleko rychlejší, neunaví se a nedělá chyby. U nás ve firmě nejde o programátory nebo vývojáře v klasickém slova smyslu, daleko spíše jsou to výzkumníci, zabývající se primárně strojovým učením. Celkem je nás dnes v kancelářích na různých pozicích 17 a do konce roku nejméně 5 dalších přibude.

Jak jste úspěšni obchodně?

Od začátku jsme globální a exportní, v Česku nemáme jediného klienta. Dnes je 65 % našich zákazníků z USA, zbytek je EU, Británie a Asie. Bez jakýchkoliv externích investic jsme v provozním zisku, už to začínají být zajímavá čísla. Máme konkurenci po celém světě, nicméně naše velká výhoda je v tom, že jsme nezačínali jako univerzální technologická

Velmi podstatná je taky ta neuvěřitelná rychlost, která to celé zjednodušuje a zlevňuje. S naším systémem jsme schopní popsat klidně milion nahrávek za den. A kdykoliv v budoucnu můžeme stejně rychle celý katalog přetagovat. Kdyby to chtěla nějaká firma po živém člověku, jak dlouho by to třeba u sto tisíc nahrávek dělala?

firma, ale přišli jsme z hudební branže a přesně jsme věděli, co má naše technologie umět a jaké jsou specifické potřeby našich klientů. Konkurence to v tuhle chvíli tak dobře nedokáže, přestože mají přístup ke stejné hudbě jako my, přestože mají více programátorů a jsou to firmy, do kterých jsou zainvestované miliony dolarů. Hodnocení klientů nicméně vyznívá jasně v náš prospěch.

Poslední dobou se hodně mluví o tom, že umělá inteligence je schopná hudbu sama vytvářet, to vás neláká?

Ne, to skutečně ne, nevidíme v tom předanou hodnotu. Možná by se tím daly vydělávat další zajímavé peníze, ale naše

mise je někde jinde – v tom, aby se už existující hudba mohla líp dostávat ke svým uživatelům, ne v automatickém generování nové, strojové hudby. Protože hudby je už teď strašně moc. Nechceme vytvářet další bez vkladu živých skladatelů a interpretů.

Jaké tedy máte plány pro budoucnost?

Budeme dál zlepšovat svoji technologii a postupně ji dostávat k dalším zákazníkům. Teď nám jde o velké hráče, ale do budoucna chceme oslovit i jednotlivce a koncové uživatele, kteří třeba hledají ideální hudbu pro svoje příspěvky na sociálních sítích. Nicméně zároveň paralelně rozvíjíme také projekt automatického tagování hudby. Máme za sebou první verzi programu, na které náš výzkumný tým pracoval během posledního roku. Už dnes, když nám pošlete skladbu, tak vám během několika vteřin definujeme například – je to jazz, je tam saxofon, piano, je to veselé, rychlé, má to tolik a tolik BPM, je to v C dur a hodilo by se to třeba pro dokument nebo pro dramatický film. Jasně, někteří skuteční experti stále tagují lépe než náš stávající systém, ale je jen otázkou času, kdy je překonáme. Když se jich ptáme, jestli z toho nemají strach, většinou odpovídají – už aby to bylo! Ona to totiž není zase tak zábavná práce, popisovat hudbu. Celý den sedět, poslouchat a zamýšlet se nad tím, jak co otagovat. Jak už jsem říkal na začátku, obrovskou roli v tom hraje subjektivita. Vy byste určitě něco popsal jinak než já, ale co je nejhorší, oba dva bychom to třeba zítra, když budeme v jiné náladě, popsalí trochu jinak než dnes. A kdybychom to za týden chtěli hledat, možná bychom použili zase trochu jiná klíčová slova, než jsme si sami prvotně zvolili. Velmi podstatná je taky ta neuvěřitelná rychlost, která to celé zjednodušuje a zlevňuje. S naším systémem jsme schopní popsat klidně milion nahrávek za den. A kdykoliv v budoucnu můžeme stejně rychle celý katalog přetagovat. Kdyby to chtěla nějaká firma po živém člověku, jak dlouho by to třeba u sto tisíc nahrávek dělala? Rok? Rok a půl? A kolik chyb daných lidským faktorem by tam při veškeré pečlivosti vzniklo? ☒