

Globalny rynek ASR

Szacuje się, że wartość globalnego rynku systemów do automatycznego rozpoznawania mowy najpóźniej w roku 2020. osiągnie pułap 10 mld USD (fortunebusinessinsights.com). Relatywnie klarowna struktura regionalna współczesnego rynku ASR pozwala wskazać obszary geograficzne, na których rynek ten osiągnął już pewną dojrzałość oraz obszary geograficzne, na których technologie z rodziny ASR będą ekspandować. W konsekwencji, można wyróżnić grupy językowe które będą stanowiły o globalnym wzroście rynku ASR w nieodległej przyszłości oraz grupy językowe mniej znaczące w kontekście globalnym, ale tworzące potencjalnie atrakcyjną przestrzeń dla podmiotów i inicjatyw lokalnych.

Sprzedaż systemów do automatycznego rozpoznawania mowy w krajach anglosaskich (zwłaszcza w Stanach Zjednoczonych Ameryki) generuje dzisiaj nawet 35% globalnej wartości rynku ASR (polarismarketresearch.com, fortunebusinessinsights.com). Duża dojrzałość tej branży w krajach anglosaskich jest wypadkową wielu czynników, przy czym dominujące wydają się: w wymiarze społecznym - technologiczny entuzjazm społeczeństwa, w wymiarze biznesowym - postrzeganie przez kadry menadżerskie innowacyjności jako czynnika zwiększającego konkurencyjność przedsiębiorstwa, w wymiarze lingwistycznym - duża sprawność systemów ASR wykorzystujących sieci neuronowe dla języka angielskiego.

Popularyzacja oraz skuteczne wykorzystanie ASR w grupie języków chińsko-tybetańskich, języków Azji Południowo-Wschodniej oraz w nieco dłuższej perspektywie na Półwyspie Indyjskim wydają się mieć kluczowe znaczenie dla przyszłego kształtu globalnego rynku ASR.

Chińskie społeczeństwo przechodzi technologiczną rewolucję, a kluczowe przedsiębiorstwa chińskiego sektora IT, w tym Baidu, przykładają dużą wagę do technologii ASR. W tym miejscu należy wspomnieć, że wskaźnik korelacji znajomości języka chińskiego przez obcokrajowców, którzy deklarują znajomość lub częściową znajomość tego języka, w stosunku do kompetencji niezbędnych dla wykonania prawidłowego przekładu jest kilkukrotnie niższy niż w wypadku jakiegokolwiek innego języka globu- to obrazuje jak gigantyczne bariery komunikacyjne towarzyszą kontaktom Chin z resztą Świata.

Nie bez znaczenia dla rozwoju technologii ASR są ambitne plany chińskich dycydentów. Zakładają oni, że w stulecie proklamowania Chińskiej Republiki Ludowej (tj. w 2049. roku) chińska gospodarka osiągnie pozycję globalnego lidera w obszarze innowacyjności i nowoczesnych technologii. Chińska Rada Państwowa powołała już ponad 20 departamentów, które mają być odpowiedzialne za realizację ogólnochińskiej rewolucji technologicznej. Ponadto do roku 2025. w Chinach ma funkcjonować 40 Centrów Innowacji Przemysłowej prowadzących badania na rzecz rozwoju konkretnych sektorów chińskiej gospodarki. Zaawansowane technologie informatyczne, robotyzacja przemysłu oraz nowoczesna branża medyczna to obszary, w których z powodzeniem wykorzystywane będą systemy ASR, a których też dynamiczny rozwój zakłada chińska rewolucja technologiczna.

Popularyzacja technologii ASR w Indiach, zwłaszcza w kontekście rynku konsumenckiego, zdaje się być odleglejszą perspektywą niż ma to miejsce w przypadku Chin. Indyjskie społeczeństwo charakteryzuje się mniejszą niż u mieszkańców Chin technologiczną dojrzałością. Należy jednak zwrócić uwagę na zaskakująco dużą dynamikę zmian w tym obszarze. Ogromny technologiczny entuzjazm

mieszkańców Indii może w bliskiej przyszłości doprowadzić do sytuacji, w której indyjskie społeczeństwo przestanie być wyłącznie odbiorcą zewnętrznych trendów technologicznych. Już dziś Indie, obok Chin, są najszybciej rozwijającą się dużą gospodarką na świecie. Dość wysoki poziom szkolnictwa wyższego, zwłaszcza w obszarze nauk inżynierskich i IT, sprawia, że Indie stają się atrakcyjną lokalizacją dla centrów badawczych globalnych firm technologicznych. Na światowym rynku coraz wyraźniej zaznaczają swoją obecność indyjskie, rodzime firmy działające w branży nowoczesnych technologii - na przykład indyjskie startupy rozwijające rozwiązania w zakresie NLP. Cieszą się one już dużą rozpoznawalnością.

Warty podkreślenia jest fakt, że już co trzeci sprzedawany w Indiach telefon to smartfon (tym samym Indie stały się drugim największym, zaraz po Chinach, rynkiem sprzedaży smartfonów na świecie), a liderami sprzedaży w tym kraju są Samsung (prowadzący własne prace badawcze w obszarze ASR) i Xiaomi. Dla wielu hindusów smartfony są pierwszymi multimedialnymi urządzeniami pozwalającymi im na swobodną komunikację oraz wykorzystanie zasobów Internetu. Cena komputerów osobistych stanowiła barierę dla ich popularizacji na Półwyspie Indyjskim.

Pewne ograniczenia związane z wdrażaniem technologii ASR wynikają z dużej różnorodności językowej na obszarze Półwyspu Indyjskiego - 10 wiodących języków służy do komunikacji blisko 90% mieszkańców Indii, przy czym główny język hindi (zna go ponad 40% mieszkańców Indii) posiada wiele dialektów. Znaczna liczba mieszkańców Indii, zwłaszcza w centrach gospodarczych i rozwojowych kraju, posługuje się językiem angielskim - tak zwanym "Hindi English", który jest jednak skutecznie przetwarzany przez narzędzia ASR wykorzystujące sieci neuronowe.

Wiele źródeł wskazuje na grupę języków słowiańskich jako wyzwanie dla współczesnych systemów ASR. Liczebność Słowian szacuje się na 300 mln. W tej grupie językowej wyraźne jest pewne kontinuum językowe korelujące z geograficznym rozmieszczeniem społeczności posługujących się odrębnymi językami słowiańskimi - z reguły różnice językowe pomiędzy sąsiadującymi ze sobą społecznościami są dość nieznaczne i pozwalają na komunikację między członkami tych społeczności (np. możliwa jest swobodna komunikacja między Polakami i Słowakami). Zbliżone struktury gramatyczne oraz bardzo złożona fleksyjność dowodzą, mają stosunkowo niedawno wyodrębnić się poszczególnych języków słowiańskich z języka prasłowiańskiego. Wiele wspólnych cech w tej grupie języków zdaje się stabilizować pozycję systemów ASR typu "pattern-based/statistical" dostosowanych do konkretnej aplikacji - zwłaszcza w kontekście obsługi klienta, interesariusza w segmencie przedsiębiorstw oraz konsumenckim. Istnieje jednak prawdopodobieństwo, że systemy ASR bazujące w swoim działaniu w głównej mierze na sieciach neuronowych, po pokonaniu pewnych barier technologicznych, staną się technologią dominującą.

W najbliższych latach umiarkowane szacunki zakładają wzrost rynku rozwiązań ASR na poziomie 18-20% w skali roku, aż do śmiałej prognozy osiągnięcia pułapu 35 mld USD jako globalnej wartości rynku. Warty podkreślenia jest fakt, że coraz więcej źródeł klasyfikuje rozwiązania z dziedziny ASR jak systemy zintegrowanych z modułami NLP.

Natural Language Processing

Największe, globalne firmy technologiczne koncentrują swoje wysiłki na rozwijaniu systemów

ASR opartych o sieci neuronowe. Często takie rozwiązania mają stanowić integralną część szerszego systemu klasy NLP. Bieżące trendy zdają się sugerować, że rynek ASR stanie się częścią segmentu NLP oraz automatycznych tłumaczeń i za kilka lat nie będzie wyodrębniany. Dynamiczny wzrost sprawności systemów NLP opartych o sieci neuronowe skutkował będzie pojawieniem się na rynku uniwersalnych rozwiązań zdolnych do operowania w szerokiej grupie językowej bez konieczności wprowadzania większych modyfikacji w ramach samego systemu. Kwestia dostarczenia odpowiednich danych i minimalizacji nakładów związanych ze szkoleniem systemu jest jednym z kluczowych obszarów badań w segmencie NLP - znalezienie rozwiązania dla tych problemów może prowadzić do sytuacji bardzo gwałtownego zdominowania globalnego rynku NLP przez kilku liderów oferujących najsprawniejsze systemy. Wydaje się, że te systemy ASR, które dziś występują na rynku jako niezależne produkty powinny zostać przygotowane na nadchodzące zmiany - kluczowe dla osiągnięcia rynkowej przewagi, zwłaszcza dla ASR w segmencie konsumenckim oraz przedsiębiorstw, może okazać się opracowanie odpowiednich interfejsów do sprawnej integracji z rozwiązaniami NLP. Podejście takie może okazać się niezwykle praktyczne zwłaszcza dla węższych i bardziej specyficznych grup językowych (np. języków słowiańskich), gdzie osiągnięcie satysfakcjonującej sprawności systemu ASR opartego o sieci neuronowe wydaje się bardziej złożonym procesem.

Speech Emotion Recognition

Coraz więcej źródeł donosi o badaniach nad automatycznym rozpoznawaniem emocji mówcy w oparciu o przetworzenie i parametryzację sygnału akustycznego wypowiedzi. Sama geneza tej koncepcji zdaje się sięgać wielu lat wstecz, lecz pierwsze wzmianki o systemach SER (speech emotion recognition) o zadawalającej sprawności pojawiły się w ostatnich latach i miesiącach - skuteczności współczesnych systemów SER przeważnie ma związek z implementacją sieci neuronowych.

Ze względu na spektrum wykorzystywanych technologii, rozwiązania typu SER mogą w przyszłości stać się naturalną domeną specjalistów z dziedziny ASR oraz dziedzin pokrewnych. Wykorzystanie systemów SER oraz ASR do równoległego przetworzenia tej samej wypowiedzi wydają się dawać dodatkowe możliwości w obszarze NLP- tak rozbudowany system mógłby formułować odpowiedzi dla użytkownika nie tylko w oparciu o treść dostarczonych komunikatów, ale również, dzięki otrzymaniu od modułu SER informacji o nasyceniu emocjonalnym komunikatu, mógłby charakteryzować się pewną responsywnością emocjonalną. I tak na przykład komunikaty o tej samej treści wyartykułowane przez użytkownika wzburzonego oraz użytkownika rozbawionego mogłyby spotkać się z różnymi odpowiedziami ze strony systemu NLP. Mimo dość abstrakcyjnego wydźwięku takiego scenariusza, być może należy przeprowadzić szerszą analizę technicznych możliwości stworzenia systemu emNLP ("empathic NLP", nazwa jest propozycją autora).

Neural Network ASR

Mimo niewątpliwych zalet systemów ASR wykorzystujących sieci neuronowe, praktyczna implementacja tych rozwiązań w segmencie konsumenckim oraz przedsiębiorstw spotyka się ze sporymi ograniczeniami. Przeszkodą dla w pełni skutecznego wykorzystania ASR opartego o sieci neuronowe do głosowej obsługi klientów oraz interesariuszy stanowią ograniczenia transmisyjne wykorzystywanego medium- dla transmisji głosowej we współczesnych sieciach należy pragmatycznie przyjąć VoLTE jako

technologię wiodącą w najbliższych latach. W tym kontekście zasadne wydaje się podejście hybrydowe, gdzie część operacji ASR realizowana jest po stronie urządzenia klienckiego, zaś pozostałe procesy ASR oraz operacje systemu NLP realizowane są w środowisku zewnętrznym. Jednocześnie należy zwrócić uwagę na dużą dynamikę prac badawczych w dziedzinie Vo5G- istnieje bowiem duże prawdopodobieństwo, że możliwości transmisyjne dla danych głosowych oferowane przez sieć 5G otworzą nową przestrzeń dla implementacji systemów ASR oraz NLP opartych o sieci neuronowe.