

Wyzwania projektu SilverWhisper

Danijel Koržinek

<danijel.korzinek@silverwhisper.pl>



24 kwietnia 2020, Warszawa

Wprowadzenie

Opis projektu

*Wsparcie pracy agentów centrum telefonicznego
poprzez automatyczną analizę rozmowy
w czasie rzeczywistym
i podpowiadanie agentom jak udzielać odpowiedzi.*



Na czym polega praca agenta?

- ▶ Osobista wiedza i umiejętności
- ▶ Narzędzia teleinformatyczne:
 - ▶ baza wiedzy i systemy informatyczne
- ▶ Problemy:
 - ▶ czas oczekiwania na informacje
 - ▶ wysoka rotacja pracowników
 - ▶ krótki czas szkolenia
 - ▶ \implies obniżenie jakości i wydłużenie kolejek
- ▶ Koszty prowadzenia infolinii

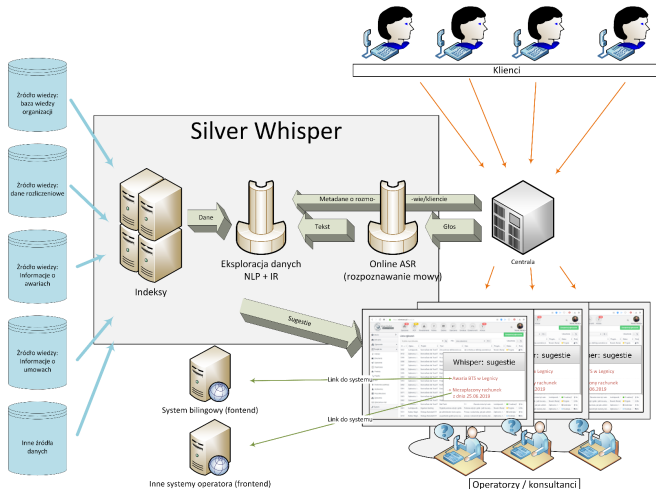
Cel projektu

1. Obniżenie kosztów funkcjonowania infolinii
 - ▶ poprzez skrócenie czasu trwania rozmowy
2. Poprawa jakości pozyskiwanej informacji

Wymagane cechy rozwiązania

- ▶ Automatyczne rozpoznawanie mowy
- ▶ Inteligentna analiza dialogu
- ▶ Szybkie wyszukiwanie bazy wiedzy i innych podsystemów firmy
- ▶ Praca w czasie rzeczywistym
- ▶ Użyteczność końcowego rozwiązania
- ▶ Łatwość integracji i adaptacji do nowej dziedziny

Schemat systemu



Istniejące alternatywne rozwiązania

- ▶ Analityka “offline” (jako narzędzie oceny pracy agentów)
 - ▶ Opóźniony czas reakcji
 - ▶ Ograniczone możliwości interwencji
 - ▶ Nie rozwiązuje rzeczywisty problem
- ▶ Pełna automatyzacja
 - ▶ Brak kontroli nad procesem powodujący duże ryzyko
 - ▶ Uzależnienie od poziomu informatyzacji firmy
 - ▶ Brak “zdrowego rozsądku”

Przykłady dialogów



Dzień dobry ... W czym mogę pomóc?










Dzień dobry. Może pani mi trzy giga?









Trzy gigabajty za pięćdziesiąt złotych, tak? Dobrze. Na tym numerze, z którego pan dzwoni?

Przykłady dialogów

-  Jeszcze chwileczkę. Mhm. Jeszcze chwilkę. I tutaj mówił pan, że wysyłał pan potwierdzenie przelewu.
-  Tak.
-  Mhm. Jeszcze chwileczkę.
-  Wczoraj przesyłałem. I tam pan mówił, że po czterech godzinach powinno już być uruchomione.
-  Yyy tak, przelew w godzinach roboczych, w większości wypadków jest tutaj już ma miejsce uruchomienie.
-  Znaczy wie pan co, no mam, mam yy czerwoną diodę na yy routerze i nie jest w ogóle pokazany jakby zasięg tam ta piramidka.
-  Aczkolwiek widzę, że tutaj było, nie widzę kontraktów, które by-łyby zawieszono, chwileczkę, dobrze, jeszcze. Jeszcze chwilkę. Jeszcze sekundkę...

Przykłady dialogów

-  Dzień dobry... W czym mogę pomóc?
-  Dzień dobry. Ja z takim pytankiem, bo ja u was zamawiałem tera router, razem z kartą do instalacji kamer, nie?
-  A kiedy do pana...
-  No i teraz tak. Nie mogę aktywować jak to się mówi routera, nie wiem dlaczego. Bo karta została włożona, normalnie wszystko i ten, nie?
-  Dobrze. Poproszę na kogo są zarejestrowane usługi? Dziękuję bardzo za identyfikację. A kiedy było dostarczenie sprzętu?
-  Yyy... Już panu mówię eee...

Przykłady dialogów



Dzień dobry... W czym mogę pomóc?



Dzień dobry. Ja mam takie pytanie. Ponieważ z uporem maniaka, co piętnaście sekund wysyłacie mi sms-a, że albo zużyłam osiemdziesiąt procent internetu, albo sto procent internetu. Mam prośbę. Zdecydujcie się, czy to jest osiemdziesiąt procent, sto procent, albo przestańcie to wysyłać, bo mi telefon skacze w szufladzie, szafu można dostać.









Chodzi o ten numer, z którego pani dzwoni?



Tak. Dziękuję bardzo.

Przykłady dialogów

-  Pan mi powie. I wszystko jest w porządku, tak?
-  Yyy... Paczka przyszła, nie? Wczoraj. No, zobaczymy jak chodzi, bo go jeszcze nie odpaliłem, nie?
-  A no, na pewno będzie pan zadowolony.
-  No.
-  Niech się pan tutaj nie martwi, same dobre telefony wysyłamy.
-  Yyy... O jezu, same telefony, tylko żeby lepszy zasięg na Witkowie był, booo... yyy... pani powiem, są takie dni, ten Witkowie koło Gniezna, jest w takim, yyy... miejscu, gdzie faktycznie sygnał krytyczny, nie?

Problemy

- ▶ Mowa spontaniczna
- ▶ Szum i brakujące informacje
- ▶ Nietypowe słownictwo
- ▶ Niespójność informacji w organizacji
- ▶ Zmienność informacji w czasie

Rozpoznawanie mowy

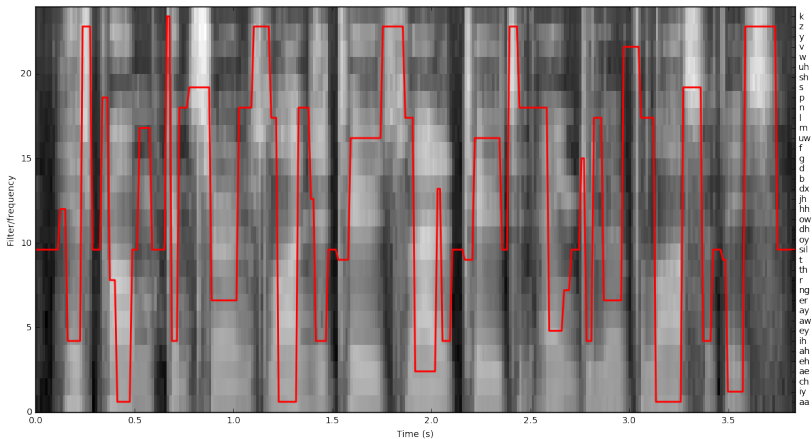
Rozpoznawanie mowy

- ▶ Automatic Speech Recognition
 - ▶ Speech-to-text, Voice Recognition
- ▶ Rozpoznawanie vs Klasyfikacja
- ▶ Rozpoznawanie treści wypowiedzi na podstawie nagrania audio
- ▶ Pomijana jest informacja o mówcy i sposobie mówienia - liczy się tylko sekwencja wyrazów zawarta w nagraniu
- ▶ Celem jest wyszukiwanie optymalnej hipotezy według wzoru Bayesa:

$$\arg \max_w P(w|O) = \arg \max_w P(O|w) \cdot P(w) \quad (1)$$

- ▶ Model akustyczny – $P(O|w)$
- ▶ Model języka – $P(w)$

Wyszukiwanie sekwencji w nagraniu mowy



Word error rate

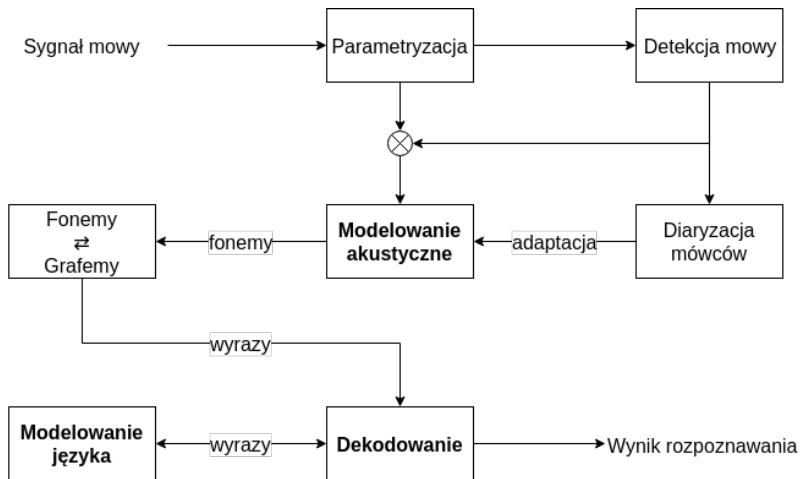
- ▶ Wykorzystuje algorytm Levenshteina do obliczenia odległości wyrażonej w postaci najmniejszej ilości operacji edycji przekształcenia sekwencji hipotetycznej w sekwencję referencyjną
 - ▶ operacje: zamiana, usunięcie, wstawienie

$$WER = 100 * \frac{\text{zamiany} + \text{usunięcia} + \text{wstawienia}}{\text{ilość elementów ref.}} \quad (2)$$

Przykładowe wyniki

- ▶ Nagrania niskiej jakości:
 - ▶ WER 33.93 % - klasyczny model ASR
 - ▶ WER 27.27 % - sieć neuronowa jako AM (model akustyczny)
 - ▶ WER 15.35 % - powyższe + adaptacja LM (model języka)
 - ▶ WER 12.67 % - powyższe + VAD (detektor mowy)
 - ▶ WER 10.97 % - powyższe + rescoring większym modelem języka
- ▶ Czysta mowa:
 - ▶ WER 15.89 % - klasyczny model ASR
 - ▶ WER 12.20 % - adaptacja do mówców
 - ▶ WER 5.73 % - sieć neuronowa jako AM
 - ▶ WER 2.45 % - rescoring dużym modelem języka

Schemat rozpoznawania mowy



Wstępne przetwarzanie

- ▶ Jakość nagrania
 - ▶ kanał telefoniczny (8 kHz, GSM)
 - ▶ gain, szum/rozmowy w tle
 - ▶ warunki akustyczne kanału miejskiego
- ▶ Detekcja mowy
 - ▶ segmentacja nagrania na mowę/nie-mowę
 - ▶ zwiększa wydajność obliczeniową
 - ▶ zwiększa precyzję eliminując szum
- ▶ Diaryzacja
 - ▶ segmentacja nagrania na fragmenty należące do tego samego mówcy
 - ▶ niezbędne w nagraniach jedno-kanałowych (tzw. mixdown)
 - ▶ umożliwia adaptację do mówcy

Modelowanie akustyczne

- ▶ Celem modelowania jest estymacja $P(O|w)$
 - ▶ prawdopodobieństwo, że analizowana sekwencja obserwacji odpowiada wybranej sekwencji słów
- ▶ Wyraz jest zazwyczaj modelowany jako sekwencja jednostek fonetycznych:
 - ▶ Fonem - abstrakcyjna jednostka lingwistyczna, najmniejszy wyznacznik fonologiczny w danym języku
 - ▶ Trifon - fonem występujący w kontekście
- ▶ Tradycyjnie używano mieszanek Gaussowskich
 - ▶ Klasyczny model HMM
- ▶ Ostatnio częściej stosowane są sztuczne sieci neuronowe
 - ▶ Systemy koneksjonistyczne i hybrydowe
 - ▶ Systemy end-to-end

Problemy modelowania akustycznego

- ▶ Trudne warunki akustyczne
- ▶ Mowa spontaniczna i “cocktail speech”
- ▶ Mała ilość materiału trenującego
- ▶ Głębokie sieci neuronowe potrzebują dużo danych
- ▶ Potencjalne rozwiązania:
 - ▶ Inne dane akustyczne
 - ▶ Transfer learning
 - ▶ Augmentacja danych

Modelowanie języka

- ▶ Celem jest ustalenie kontekstu wyrazowego wypowiedzi
Rozpoznawanie mowy bez modelu języka, to jak rozpoznawanie nieznanego nam języka ze słuchu
- ▶ Zadaniem modelu jest określenie prawdopodobieństwa sekwencji wyrazów:

$$P(w_1, \dots, w_m) = \prod_{i=1}^m P(w_i | w_{i-1}, \dots, w_1) \quad (3)$$

- ▶ Modele:
 - ▶ statystyczne - n-gramy
 - ▶ neuronowe - np. RNN-LM, LSTM-LM, CNN-LM

Problemy w modelowaniu języka

- ▶ Zmieniające się słownictwo
- ▶ Mowa spontaniczna zawierająca wiele urwanych słów i powtórzeń
- ▶ Mało dziedzinowego materiału trenującego
- ▶ Potencjalne rozwiązania
 - ▶ Adaptacja
 - ▶ Otwarty słownik
 - ▶ Rescoring i przetwarzanie krat
 - ▶ Reprezentacje modeli języka, FLM

Dekodowanie

- ▶ Proces wyszukiwania najlepszej sekwencji słów względem modeli akustycznego i języka
- ▶ Przetwarzanie w czasie rzeczywistym
 - ▶ realtime ratio, latency
- ▶ Całkowite opóźnienie systemu jest sumą wszystkich opóźnień poszczególnych etapów przetwarzania

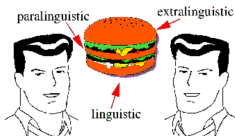
Przetwarzanie języka i IR

Język

- ▶ Warstwy wypowiedzi:
 - ▶ Fonetyczny
 - ▶ Leksykalny
 - ▶ Syntaktyczny
 - ▶ Semantyczny
 - ▶ Pragmatyczny
- ▶ Warstwy zapisu:
 - ▶ Dialog
 - ▶ Wypowiedź
 - ▶ Zdanie
 - ▶ Fraza
 - ▶ Wyraz (słowo)
 - ▶ Litera (grafem)
 - ▶ Głoska (fonem)
- ▶ Alfabety
 - ▶ ortograficzny: łacina, cyrylica, ...
 - ▶ fonetyczny: IPA, SAMPA, ...

Komunikacja

- ▶ Warstwy komunikacji:
 - ▶ Lingwistyczna - tekst wypowiedzi
 - ▶ Paralingwistyczna - sposób wypowiedzi, prozodia
 - ▶ Ekstralingwistyczna - niewerbalne cechy wypowiedzi, mimika
 - ▶ Metalingwistyczna - podtekst, wynikający ze świadomości mówcy



- ▶ Elementy prozodii (akcent, intonacja):
 - ▶ F0 (melodia)
 - ▶ Energia (intensywność)
 - ▶ Długość segmentów (tempo)

Wyzwania przetwarzania języka i IR

- ▶ Wykrywanie istotności słów
 - ▶ Dzwonię w sprawie roamingu, ile kosztuje rozmowa w Egipcie?
- ▶ Wieloznaczność (homonimia)
 - ▶ “Patrzę na ciebie z góry”
 - ▶ “Mój telefon” (czynność? urządzenie?)
- ▶ Sprowadzanie do form podstawowych
 - ▶ mam → mieć? mama? MAM? (firma)
 - ▶ Lato → lato? rok? Lato? (Grzegorz)
- ▶ Czego szukać
- ▶ Jak zapewnić wyszukiwanie w czasie rzeczywistym
- ▶ Jak wydobyć istotną informację

Analiza wypowiedzi

- ▶ Etykietowanie wypowiedzi
- ▶ Rozpoznawanie intencji
 - ▶ Generyczne vs dziedzinowe
- ▶ Rozpoznawanie tematu rozmowy
- ▶ Rozpoznawanie atrybutów
- ▶ Odporność na błędy
 - ▶ złe słowa
 - ▶ urwania/powtórzenia
 - ▶ brak interpunkcji i innych cech języka pisanego

Wyszukiwanie

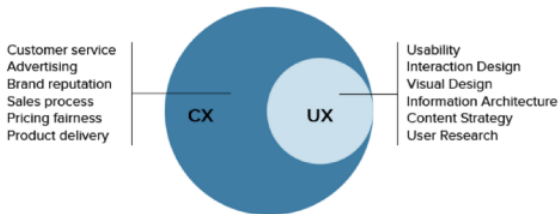
- ▶ Reprezentacja danych w indeksie
 - ▶ wektorowy model słów + dane strukturalne
- ▶ Generacja zapytania
 - ▶ regułowa generowana pół-automatycznie
 - ▶ użycie operatora alternatywy, synonimii, wyrazów bliskoznacznych, hiperonimów oraz hiponimów
- ▶ Dane dynamiczne
 - ▶ szablony dokumentów
- ▶ Ranking wyników
- ▶ Generowanie odpowiedzi

Użyteczność

User and Customer Experience

- ▶ Każdy aspekt interakcji z produktem ma wpływ na postrzeganie całości przez użytkownika
- ▶ Cel - poprawa wydajności pracy konsultantów i poziomu zadowolenia klientów
- ▶ Wyzwanie - duża liczba ograniczeń i procesów dziejących się w Call Center oraz uwzględnienie działania AI
- ▶ Analiza danych klienta a kwestia UX/AI:
 - ▶ Jakie dane klienta mogą być przydatne do przyspieszenia rozwiązania problemu przez konsultanta?
 - ▶ Jakie dane są istotne do lepszego działania AI?

UX vs CX



Wydajny i intuicyjny UI konsultanta

- ▶ Pomoc w wyborze najlepszej odpowiedzi z bazy wiedzy
- ▶ Dynamiczna prezentacja identyfikacji kontekstu danego problemu
- ▶ Wydajne korzystanie przez konsultantów o różnym poziomie doświadczenia
- ▶ Dostosowanie projektu UI do działania AI
- ▶ Wyzwania:
 - ▶ moment generowania podpowiedzi
 - ▶ opóźnienie
 - ▶ jakość podpowiedzi
 - ▶ granularność podpowiedzi
 - ▶ sposób prezentacji i interakcji
 - ▶ ogólna przydatność rozwiązania

Kontakt

danijel.korzinek@silverwhisper.pl