ხმის ამომცნობი სისტემების აგების მაგალითები

კონსტანტინე კამკამიძე, *სტუ-ს პროფესორი*
ალექსანდრე მელაძე, *სტუ-ს დოქტორანტი*

რეზიუმე

დღეს დღეისობით, ხმის ამომცნობი სისტემები ძალიან მნიშვნელოვან როლს თამაშობს ადამინის ცხოვრებაში. ასეთი სისტემები არის ძალიან მოსახერხებელი, ბევრი ადამიანი (ჩემი ჩათვლით) ტექსტს წერენ მსგავსი სისტემებით. მისი გამოყენება საკმაოდ მარტივია, შენ ესაუბრები მიკროფონს და ტექსტი იწერება ავტომატურად. როცა ვწერთ თეზისს ან მსგავს დიდ ტექსტს, მსგავსი სიტემები ძალიან გვეხმარება, იმიტომ რომ ისინი გამორიცხავენ შეცდომებს და ასევე არიან პროგრესულები, ლექსიკონი რომელზეც სისტემაა დაყრდნობილია ყოველწუთას იზრდება და ვითარდება.

საკვანძო სიტყვები: ხმის ამომცნობი სისტემები, ციფრული სიგნალი, ანალოგური სიგნალი, ხმა, ტექსტი, კონვერტაცია, მოდელირება, ანალიზი.

Summary

Voice recognition software is fantastic and many people like me could not write (or work) without it. In fact when it works it is pure magic. You simply talk into the microphone and the words appear on the screen in front of you. Hands free. Amazing! When you're writing a thesis or a long article it really comes into its own, because it progressively learns the vocabulary and expressions of the project and error is minimal.

Keywords: Speech recognition systems, Digital signal, Analog signal, Voice, Text.

ძირითადი ტექსტი

ტექსტის ამოცნობა ერთერთი ყველაზე კომპლექსური და რთულ პრობლემას წარმოადგენდა. რადგან ის დამოკიდებულია ენის გრამატიკაზე, მის ლინგვისტურ მახასიათებლებზე და ცოდნაზე, რომლის შექმნა კომპიუტერული ტექნიკისთვის საკმაოდ რთულია. ასეთი სისტემის შესაქმნა მოითხოვს საკმაოდ გამოცდილ ლინგვისტებს, მათემატიკოსების და პროგრამისტების ჩართულობას.

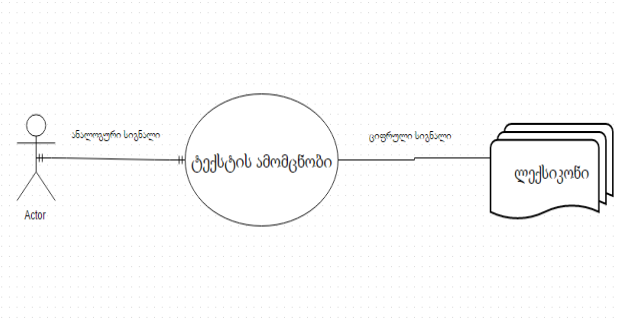
სისტემის ასაგებად ჩვენი მთავარი მიზანია :

1) შევქმნათ სისტემა რომელიც ავტომატურად გადმოიყვანს ადამიანის მიერ წარმოთქმულ სიტყვას(ანალოგურ სიგნალს), კომპიუტერულ სიგნალში(ციფრულ სიგნალში).

2) მიღებული სიგნალი გადავიყვანოთ ტექსტში

3) მიღებული ტექსტით, მოვძებნოთ არსებულ ლექსიკონში, შესაბამისი ტექსტი

4) ნაპოვნი ტექსტის მიხედვით მივიღოთ გადაწყვეტილება



რატომაა ასეთი რთული ტექსტის ამომცნობის შექმნა ?

1) ნებისმიერ ენას აქვს საკმაოდ დიდი რაოდენობის სიტყვები. ადამიანს რომელის ტვინსაც შეუძლია მრავალი ოპერაციის გაკეთება, მასაც კი საკმაო დრო სჭირდება რომ ენა შეისწავლოს .

2) როდესაც ადამიანები საუბრობენ ქუჩაში ან ერთდროულად, არის საკმაოდ დიდი ხმაური, ეს დიდი პრობლემას უქმნის სიტყვების აღქმას. ამ პრობლემას მეცნიერები უწოდებენ „კოქტეილის წვეულების(cocktail party)“ პრობლემას.

3) როდესაც ადამიანები საუბრობენ სწრაფად , ისინი ერთმანეთის მიყოლებით ამბებენ სიტყვებს, ძალიან რთულია გაიგო როდის დამთავრდა ერთი წინადადება და დაიწყო მეორე.

4) ყველა ადამიანს აქვს უნიკალური ხმა, ასევე ხმა ყოველთვის იცვლება ზრდასთან ერთად, პერიოდულად. როგორ ხვდება ჩვენი ტვინი ყველა ადამიანის ნათქვამს ? (მაგალითად ბურთს აქვს ერთიდაიგივე მნიშვნელობა, ამას იტყვის 10 წლის ბავში თუ 60 წლის მამაკაცი)

არსებობს რამოდენიმე მიდგომა, რომელიც შეიმუშავეს მეცნიერებმა, რის საშუალებითაც შესაძლებელია ტექსტის ამოცნობა, თითოეული მათგანი დანერგილია სხვადასხვა სისტემებში ესენია:

1) მარტივი ნიმუშის დამთხვევა(Simple pattern matching)

2) ნიმუშის და ლექსიკონის ანალიზი (Pattern and feature analysis)

3) ენის მოდელირება და სტატიკური ანალიზი (Language modeling and statistical analysis)

4) ხელოვნური ნეირონული ქსელი

გავნიხილოთ თითოეული მათგანი:

მარტივი ნიმუშის დამთხვევა
(Simple pattern matching)

ეს მეთოდი არის ყველაზე მარტივი მიდგომა. ამ შემთხვევაში უბრალოდ გვაქვს ძალიან მცირედი ბაზა(ლექსიკონი), და ხდება უბრალოდ შედარება . ანალოგური სიგნალი გადადის ციფრულ სიგნალში ,

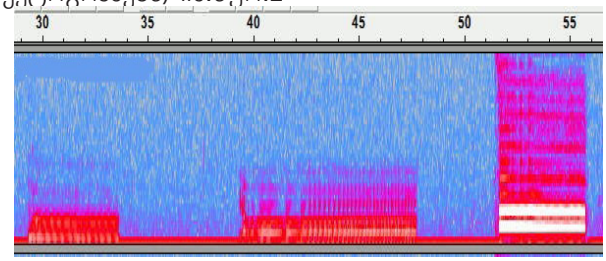
ბაზაში გვაქვს მცირედი ლექსიკონი არსებული ციფრული სიგნალების და ხდება უბრალო დადარება .

ნიმუშის და ლექსიკონის ანალიზი (Pattern and feature analysis)

ასეთი სისტემები არიან საკმაოდ მდგრადი რადგან მათ აქვთ საკმაოდ მცირედი ლექსიკონი. ასეთ მარტივ ლექსიკონში შესაბამისობის პოვნაც საკმაოდ მარტივია. ასეთი სისტემები ადრე გამოიყენებოდა სხვადასხვა სფეროებში, სამედიცინო, პროგრამულ და ა.შ. ისინი ძირითადად ასრულებდნენ ბრძანებებს. თუმცა საკმაოდ მოუხერხებელია ასეთი სისტემის არსებობა. ვინ მოიხმარს თუ ყიდვის დროს უნდა დაასწავლო სისტემას ყველა მოსალოდნელი სიტყვა რასაც წარმოვთქვამთ ?

ტექსტის ამოცნობის პროცესი

ტექსტის ამომცნობი სისტემა უსმენს სიტყვების ნაწილს რომელიც შემოდის მიკროფონიდან. პირველი ეტაპია რომ მან ანალოგური სიგნალი გადაიყვანოს ციფრულ სიგნალში, ყოველი სიტყვა გადაყავს და ახდენს მთლიანი წინადადების დიგიტიზაციას(ქმნის სპექტოგრამებს) იხ.სურ.2



სურ.2. სპექტოგრამა

სურ.2 ზე მოცემული სამი სახის გრაფი. ჩაწერილია სიტყვა „გამარჯობა“ 3 ინტენსივობით 5-10 წამის სხვაობით. პირველ ეტაპზე შედარებით დაბალ ხმაზე, მეორე ცდაზე უფრო ხმამაღლა მარა შედარებით ნელი ტემპით, მესამე შემთხვევაში ხმამაღლა. შესაბამისად გრაფიკი გვაჩვენებს ჰორიზონტალურ ღერძს რომელიც არის დრო თუ რამდენი ხანი იწერებოდა. ვერტიკალური ღერძი აჩვენებს სიმაღლეს ქვევიდან-ზემოთ. ფერი აჩვენებს ენერგიის ზომას, რაც უფრო ხმამაღლაა ჩაწერილი უფრო ენერგიულია და დიდი მაჩვენებელია. მოცემული გრაფიკით შეგვიძლია ავაგოთ სიტყვის ანალოგი და ჩვენს არსებულ ბაზაში არსებულ ანალოგებს დავადაროთ. ზუსტად დამთხვევის შემთხვევაში შეგვიძლია ვთქვათ ადამინიც კი ვინ ჩაწერა. ეს მიდგომა საკმაოდ გავრცელებულია ბანკებში სადაც ავტორიზაციას ხმის მიხედვით ატარებენ. მომხმარებელი როდესაც რეკავს ის წინასწარ ჩაწერილ სიტყვას იმეორებს ტელეფონში და მისი ვალიდურობა დასტურდება .

სტატიკური ანალიზი

მოყვანილი მეთოდების მიუხედავად მთავარი პრობლემა მაინც რჩება. როგორ ახდენს ადამიანი სხვადასხვა ადამიანის მიერ , ერთიდაიგივე სიტყვის გაანალიზებას ? ერთიდაიგივე სიტყვას მილიონობით ადამიანი სხვანაირად გადმოსცემს. თვითონ ერთი კონკრეტული ადამიანიც კი ერთ სიტყვას პერიოდულად სხვადასხვანაირად ამბობს. ასევე თუ ჩვენ ჩავწერთ დიდ ლექსიკონს, ეს ასევე გაგვიზრდის ერთნაირი სიტყვების მოხვედრის ალბათობას. მაგალითად ინგლისურ ენაში როდესაც ვამბობთ „to” რა იგულისხმება ამ მომენტში, სიტყვა “to” ან „too“ ან „two” , ეს პრობლემა დიდი ხნის მანძილზე აწუხებდათ მეცნიერებს . საბოლოოდ მივიდნენ დასკვნამდე რომ უნდა შეექმნათ გარამტიკული მოდელი. სადაც შეტანილი იქნებოდა ყველა ის ნიუანსი რასაც ენის გარამატიკა მოიცავდა, ესენია ზმნის ხმარების წესები თო მაკავშირებელი სიტყვები. ყველა ის წესი რაც ასე ვთქვათ რაღაც ლოგიკურ „ფორმულაში“ ჯდება .

ხელოვნური ნეირონული ქსელი

ადამიანს დროთაგანმავლობაში ეცვლება ხმა, შეიძლება გაცივდეს, ამ დროს სისმტემამ მაინც უნდა გაიგოს მისი საუბარი. ადამიანის ტვინს შეუძლია ძალიან სწრაფად მოახდინოს ადამიანის გაგება. ეს პრობლემა ამოუხსნელი რჩებოდა სანამ რუსმა მათემატიკოსმა Andrey Markov-მა 1970 წელს აღმოაჩინა მიდგომა, რომელსაც ეწოდა Hidden Markov Model(HMM) , რომლის საშუალებით შეძლეს ეს პრობლემა გარკვეულწილად მოეგვარებინათ. HMM ყოველი ახალი მიდგომისთვის ის აკეთებს ახალ ანალიზს, შესაბამისი ანალიზის საფუძველზე ის სწავლობს და აცნობიერებს სიტყვის აგების ზუსტ სპექტოგრამს. რაც უფრო მეტი ადამიანი იღებს მონაწილეობას და წერს ერთიდაიგივე სიტყვას მით უფრო ზუსტია HMM-ის მიღებული შედეგი .

გამოყენებული ლიტერატურა:

1. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition by Daniel Jurafsky, James Martin. Prentice Hall
2. Voice Recognition by Richard L. Klevans and Robert D. Rodman. Artech House, 1997. A short introduction to the science of voice recognition.
3. Statistical Methods for Speech Recognition by Frederick Jelinek. MIT Press, 1997.
4. Fundamentals of Speech Recognition by Lawrence R. Rabiner and Biing-Hwang Juang.
5. Speech Recognition: Invited Papers Presented at the 1974 IEEE Symposium by D. R. Reddy (ed).
6. A sound start for speech tech: by LJ Rich. BBC News, 15 May 2009. Cambridge University’s Dr Tony Robinson talks us through the science of voice recognition.