Универзитет у Нишу
Електронски факултет

Предмет: Прикупљање и предобрада података за машинско учење

Аугментација аудио записа

Семинарски рад

Студент:
Радомир Стајковић

Ментор:
Доц. др Александар Станимировић

Ниш, 2024. године

Садржај

1. Увод

Аугментација података је широко коришћена техника у машинском учењу и рачунарском виду која служи за вештачко повећање величине и разноврсности скупа података за тренирање модела. Ова техника подразумева примену различитих трансформација или модификација на постојеће узорке података, стварајући нове узорке који задржавају исту ознаку или класу као и оригинални подаци. На тај начин, аугментирани подаци могу значајно побољшати перформансе модела, смањујући претренирање, побољшавајући генерализацију и повећавајући робусност модела у условима варијабилности података.

Када је реч о аудио подацима, аугментација има за циљ да прошири и диверзификује скуп података уз очување кључних информација. Ово је од суштинског значаја у областима као што је аутоматско препознавање говора (ACP), где модели морају да буду способни да се прилагоде разним условима снимања, различитим говорницима, акцентима и позадинским шумовима. Аугментација аудио података игра кључну улогу у побољшању перформанси ACP система јер омогућава моделима да боље генерализују на нове, невиђене податке.

Као што се слике користе за тренирање модела за детекцију објеката и класификацију, тако се и аудио сегменти користе за тренирање модела за препознавање говора. ADA (Аугментација података за аудио) је техника која се користи у обради аудио сигнала за проширивање и диверзификацију скупа података за тренирање. Овај процес укључује примену контролисаних трансформација и модификација, као што су додавање шума, промене брзине и висине звука, временско растезање или компресија и изобличења, на постојеће аудио узорке. Циљ је креирање нових инстанци података које задржавају суштинске карактеристике оригиналног звука, али са варијацијама које могу унапредити робусност модела у реалним сценаријима.

Додатно, аугментација података омогућава оптимизацију модела у окружењима где је прикупљање великих количина података изазовно или скупо. Комбинујући различите технике аугментације, као што су спектралне трансформације и промене у домену временско-фреквенцијског спектра, могуће је произвести разноврснији и изазовнији скуп података који побољшава способност модела да обради варијабилност у реалним условима.

2. Значај аугментације за ASR системе

Аутоматско препознавање говора (ASR) је технологија која омогућава рачунарима да транскрибују говорне сигнале у текст. Квалитет ASR система зависи од његове способности да тачно препознаје говор у различитим ситуацијама, што подразумева различите говорнике, акценате, интонације, брзине говора, као и различите нивое позадинске буке. Аугментација аудио података је од кључне важности јер директно утиче на робусност и тачност ових система.

2.1 Разноврсност скупа података

Један од главних изазова у обуци ASR система је обезбеђивање разноврсног скупа података који обухвата широк спектар могућих варијација у говору. У стварним ситуацијама, корисници могу говорити различитим брзинама, са различитим акцентима или у бучним окружењима. Класични скупови података често немају довољно примера за све ове варијанте, што доводи до слабе генерализације модела. Аугментација помаже да се овај проблем реши генерисањем нових варијанти постојећих аудио узорака, чиме се повећава разноврсност података за обуку.

2.2 Побољшање генерализације модела

Генерализација се односи на способност модела да правилно функционише на невиђеним подацима. Модели обучени на неадекватним или недовољно разноврсним подацима могу показати одличне резултате на тестовима, али се често не понашају једнако добро у стварним апликацијама. Коришћењем техника аугментације, као што су додавање шума, промена брзине и висине тона, модели се излажу ширем спектру могућих улазних података. То омогућава моделима да буду отпорнији на стварне услове, где ће морати да раде са подацима који нису савршено чисти или предвидиви.

2.3 Повећање робусности на позадински шум

Један од најчешћих проблема са којима се ASR системи суочавају је присуство позадинског шума. У стварним ситуацијама, говор се често одвија у окружењима са високим нивоом буке, као што су улице, кафићи, или канцеларије. Додавање шума током процеса аугментације помаже у стварању модела који су отпорнији на различите врсте буке. На тај начин, модели обучени са шумовитим подацима могу боље препознати говор и у стварним бучним окружењима.

2.4 Унапређење перформанси за различите акценте и говорнике

Други важан изазов за ASR системе је њихова способност да препознају различите говорнике са различитим акцентима, интонацијама и начинима говора. У многим случајевима, скупови података које користимо за обуку не садрже довољно варијација у погледу акцената и дијалеката, што резултира пристрасним моделима. Промене висине тона, брзине и других карактеристика говора у процесу аугментације помажу у обуци модела који су способнији да се носе са овим варијацијама.

2.5 Превазилажење проблема са неуравнотеженим скуповима података

Многи скупови података за обуку ASR система пате од неуравнотежености. На пример, неки звучни узорци могу бити много чешћи од других (нпр., мушки гласови у односу на женске). Аугментација може помоћи да се уравнотеже скупови података генерисањем нових узорака који одговарају мањинским класама. На тај начин, модели постају мање пристрасни и могу тачније препознати различите категорије говора.

2.6 Смањење трошкова и времена за прикупљање података

Прикупљање и ручно означавање великих скупова аудио података може бити веома скупо и временски захтевно. Аугментација омогућава генерисање великог броја нових узорака из постојећег скупа података, чиме се значајно смањују трошкови и време потребно за прикупљање и припрему података.

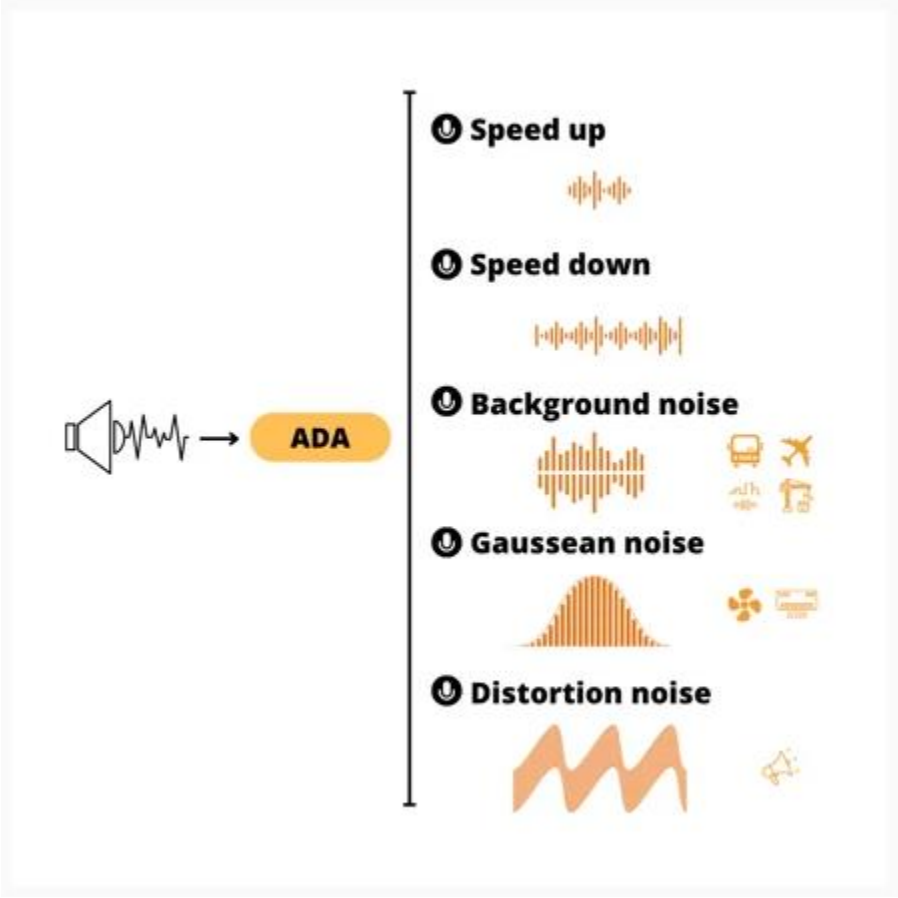
2.7 Допринос напреднијим моделима и техникама

Аугментација није корисна само за класичне ASR системе засноване на дубоком учењу, већ може значајно побољшати и напредне технике као што су Transfer Learning и Self-Supervised Learning. На пример, код Transfer Learning-а, модел може бити унапред обучен на великом скупу аугментираних података и потом дотеран на мањем скупу специфичних података. Слично, у Self-Supervised Learning-у, аугментација може помоћи у стварању контрастивних примера за учење робустнијих и семантички богатијих репрезентација.

3. Технике аугментације за аудио записе

Аугментација аудио записа је процес примене различитих трансформација на постојеће аудио записе како би се генерисали нови, варијантни узорци. Овај приступ је од суштинског значаја за обуку и побољшање перформанси система за аутоматско препознавање говора. Главни циљ аугментације је да моделима омогући боље генерализовање и отпорност на разноврсне услове у стварном свету, као што су различити акценти, брзине говора, позадински шумови, варијације у гласноћи и други фактори који могу утицати на препознавање говора.

ASR системи често морају да препознају говор у сложеним и динамичним окружењима где звучни услови нису идеални. Коришћењем техника аугментације, можемо ефикасно "обогатити" скупове података додавањем синтетичких варијација постојећих узорака, чиме се повећава робусност и отпорност модела на различите сценарије.



Слика 1.

Најпопуларније технике аугментације за говорне податке укључују:

- Додавање позадинског шума (Background Noise Addition)
- Промена брзине говора (Time Stretching)
- Промена висине тона (Pitch Shifting)
- Додавање еха (Echo Addition)
- Time Masking и Frequency Masking
- Скалирање амплитуде (Amplitude Scaling)
- Комбиноване технике (Combined Techniques)

3.1 Додавање позадинског шума

Додавање позадинског шума је техника аугментације која се користи за обогаћивање говорних података тако што се у аудио сигнал додаје шум који опонаша различите стварне услове. Ова техника је посебно важна за обуку система за аутоматско препознавање говора, јер помаже моделима да буду робуснији на варијације у окружењу у којем се говор препознаје.

Врсте позадинског шума

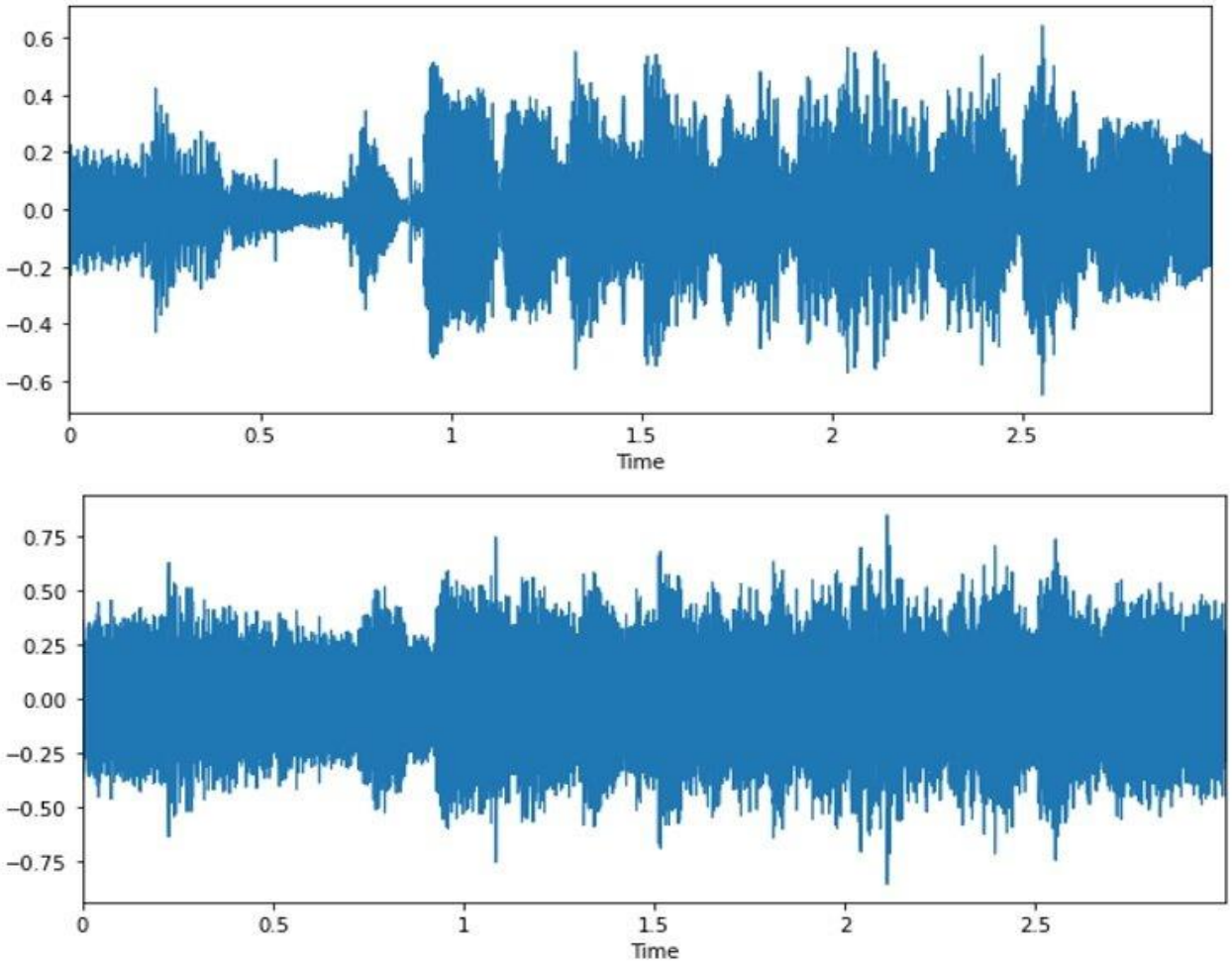
Позадински шум може бити различитих врста, а свака врста може симулирати различите услове снимања:

- **Бели шум:** Стални шум са равномерним интензитетом у свим фреквентним опсезима. Ова врста шума може опонашати константне звукове у позадини као што су звук вентилатора или клима уређаја.
- **Шум природе:** Звукови природе, као што су киша, таласи или ветар. Ова врста шума може помоћи у симулацији снимања у спољашњим или природним окружењима.
- **Саобраћај:** Звукови саобраћаја, као што су саобраћајна бука, сирене или аутомобили. Ово је корисно за симулацију урбаних услова.
- **Други гласови:** Звукови других људи који разговарају у позадини. Ово може опонашати ситуације као што су разговори у ресторану или канцеларији.

Имплементација

Додавање позадинског шума може се извршити на неколико начина:

1. **Једноставно мешање:** Оригинални аудио запис се комбинује са звуком шума при одређеном интензитету. То се може урадити помоћу алата као што су pydub или librosa у Python.
2. **Прилагођавање нивоа шума:** Ниво шума се пажљиво подешава како би се осигурала разумљивост говора. Превише интензиван шум може покрити говор и смањити тачност модела.
3. **Варијације у шуму:** Може се користити различите врсте шума у различитим узорцима како би се створила разноврсност у подацима.



Слика 2. Додавање шума

Предности

- **Повећање робусности:** Додавањем шума моделу се омогућава да се боље носи са различитим условима у стварном свету, чиме се побољшава његова робусност на варијације у окружењу.
- **Симулација реалних услова:** Помоћ у симулацији стварних услова у којима ће ASR систем бити коришћен, чиме се побољшава генерализација модела.

Мане

- **Могуће покривање говора:** Ако ниво шума није пажљиво подешен, може доћи до покривања важних делова говора, што може отежати препознавање и смањити тачност модела.
- **Повећана комплексност:** Прекомерна количина различитих врста шума може повећати комплексност модела, што може захтевати додатно фино подешавање.

3.2 Промена брзине говора

Промена брзине говора, или Time Stretching, је техника аугментације која се користи за модификацију брзине аудио записа без промене висине тона. Ова техника омогућава моделима да се обуче на различите брзине говора, чиме се повећава њихова способност да се носе са варијацијама у брзини изговора речи у стварним сценаријима.

Time stretching је процес којим се аудио сигнал мења у временском домену тако да се продужава или скраћује без промене фреквенцијског садржаја. Ово је постигнуто коришћењем алгоритама који модификују временску структуру сигнала. Неки од најпопуларнијих алгоритама за time stretching укључују:

- Phase Vocoder: Овај алгоритам анализира аудио сигнал у фреквенцијском домену, модификује временске компоненте, а затим га враћа у временски домен.
- Time-domain Harmonic Product Spectrum (TDHPS): Овај метод користи спектралну анализу и синтезу за измену брзине без промене висине тона.

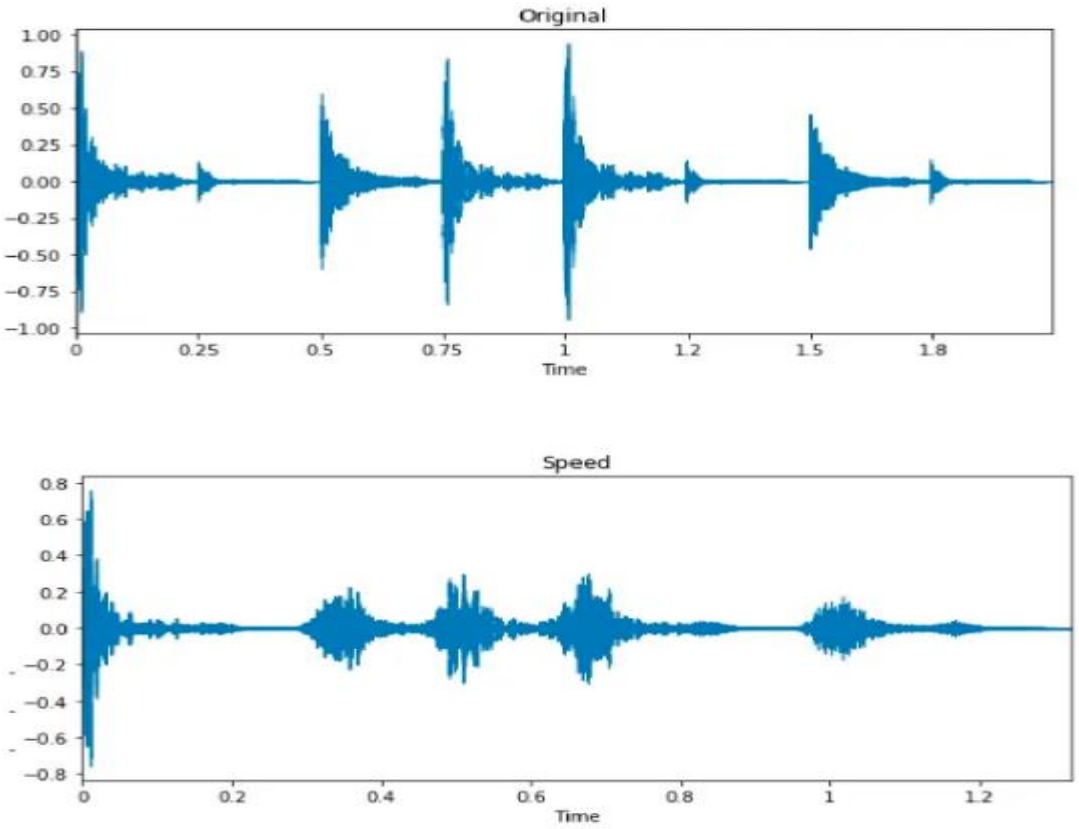
- Overlap-Add Method: Делимично преклапање аудио записа и додавање их како би се постигла измена у временској домени.

Предности

- **Унапређење робусности:** Обука модела на различитим брзинама говора може помоћи у развоју робуснијих система који могу да се носе са различитим стиловима и темпом говора.
- **Повећава разноврсност података:** Ова техника ствара нове узорке из постојећих података, што може помоћи у побољшању перформанси модела у различитим сценаријима.

Мане

- **Могући изгубљени детаљи:** Промена брзине може довести до губитка неких детаља у говору, што може утицати на разумљивост или квалитет података.
- **Алгоритамска комплексност:** Неколико метода за time stretching може бити рачунски интензивно, што може захтевати додатне ресурсе за обраду и имплементацију.



Слика 3. Промена брзине

3.3 Промена висине тона

Промена висине тона, или Pitch Shifting, је техника аугментације која се користи за модификацију висине тона аудио записа без промене брзине говора. Ова техника омогућава моделима да се обуче на различите висине тона, чиме се повећава њихова способност да се носе са варијацијама у тонским карактеристикама говора.

Pitch shifting подразумева промену фреквенцијског садржаја аудио сигнала како би се добила нова висина тона. Ово се постиже применом алгоритама који мењају фреквенцијске компоненте сигнала, али задржавају оригиналну брзину. Неки од најпопуларнијих алгоритама за pitch shifting укључују:

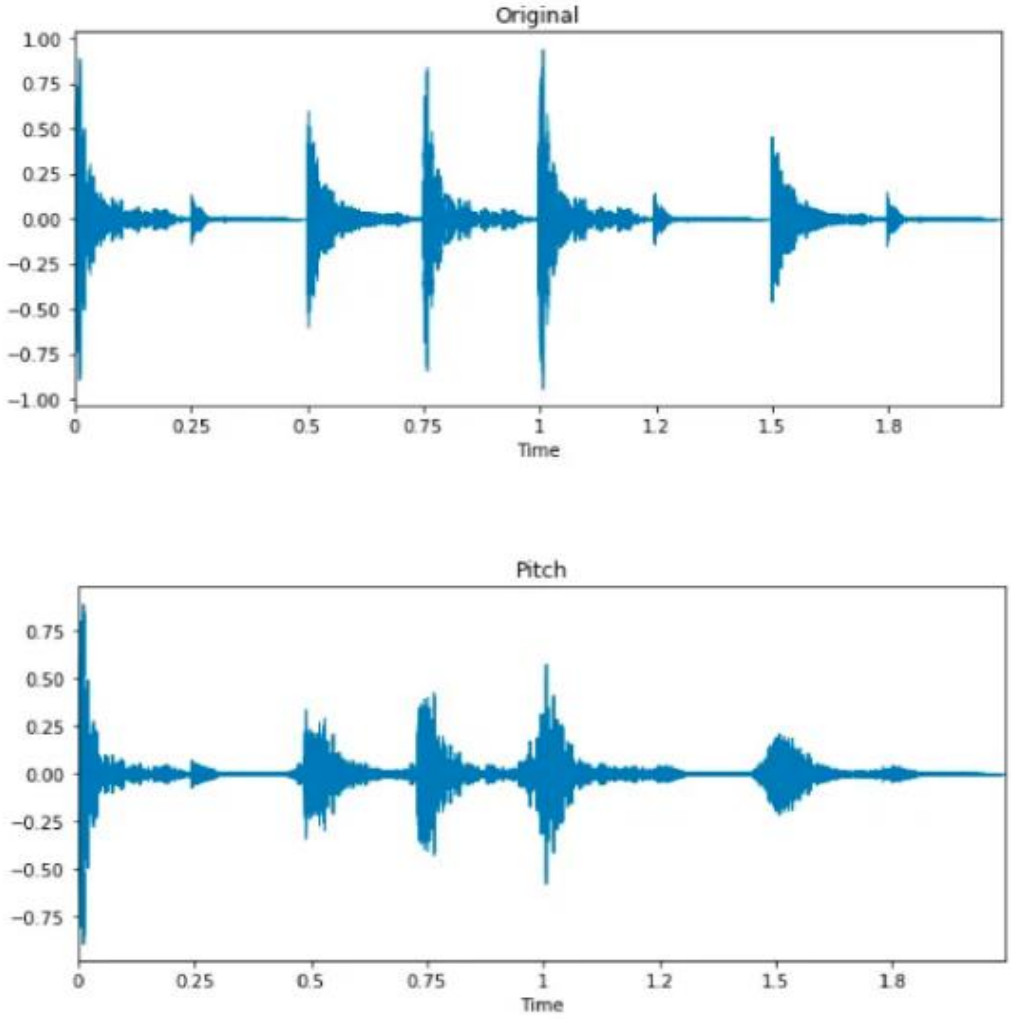
- **Fourier Transform-Based Methods:** Ови методи користе Fourierovu трансформацију за анализу и синтезу аудио сигнала, омогућавајући измену висине тона без промене времена.
- **Phase Vocoder:** Овај алгоритам такође користи Fourierovu трансформацију за модификацију висине тона, а затим враћа сигнал у временски домен.
- Granular Synthesis: Овај метод дели аудио сигнал на мале делове или "зрна" и модификује их пре него што их поново састави, што омогућава измену висине тона.

Предности

- **Повећање робусности:** Обука модела на различитим висинама тона може помоћи у развоју робуснијих система који могу да се носе са различитим тонским карактеристикама говора.
- **Разноврсност података:** Ова техника ствара нове узорке са различитим висинама тона, што може помоћи у побољшању перформанси модела у различитим сценаријима.

Мане

- **Могући дефекти у квалитету:** Промена висине тона може довести до неких изобличења у квалитету говора, као што су непријатни звучни артефакти.
- **Алгоритамска комплексност:** Неколико метода за pitch shifting може бити рачунски интензивно и захтевати додатне ресурсе за имплементацију.



Слика 4. Промена висине тона

3.4 Додавање еха

Додавање еха је техника аугментације која се користи за симулацију ефекта одјека у аудио запису. Ова техника додаје репризе оригиналног звука у одређеним интервалима, стварајући ефекат одјека или реверберације. Додавање еха може помоћи моделима за аутоматско препознавање говора (ASR) да се носе са различитим акустичким условима и побољшају перформансе у реалним окружењима.

Додавање еха укључује примену алгоритама који симулирају ефекат одјека тако што додају касније репризе аудио сигнала. Ово може бити постигнуто на неколико начина:

- **Delay-Based Echo**: Ова метода укључује додавање копије оригиналног звука са одређеним временским одлагањем. Ово одлагање може бити кратко (микро секунде до милисекунде) или дуго (неколико секунди) у зависности од жељеног ефекта.
- **Feedback Echo:** Ова метода користи повратне петље, где се део репризе враћа у систем и комбинује са оригиналним сигналом. Ово може створити ефекат више слојева еха.
- **Reverb Simulation:** Симулирање реверберације ствара ефекат еха у затвореним или великим просторима, као што су концертне дворане или собе.

Предности

- **Симулација реалних услова:** Додавање еха помаже у симулацији акустичких услова који могу бити присутни у различитим окружењима, као што су концертне дворане или велике собе.
- **Повећање разноврсности података:** Ова техника ствара нове аудио узорке са различитим акустичким ефектима, што може помоћи у побољшању перформанси ASR модела у различитим условима.

Мане

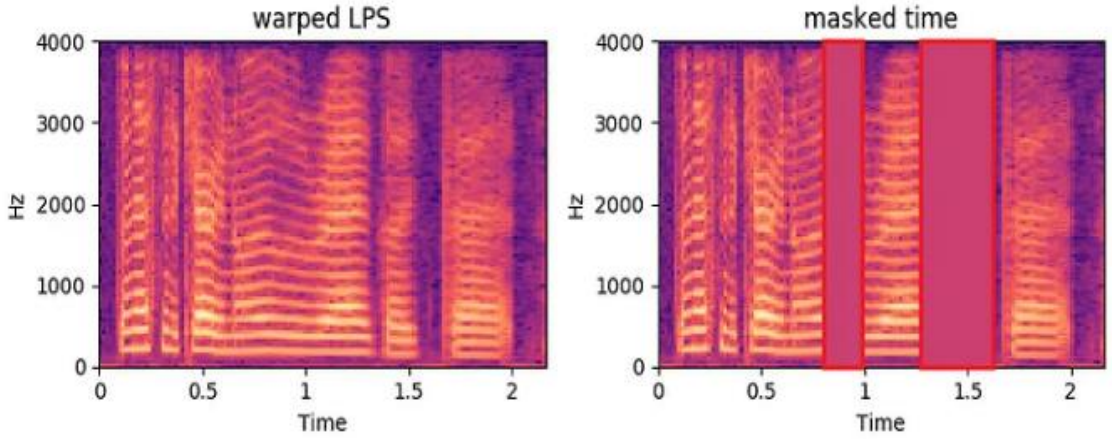
- **Могући изгубљени детаљи:** Додавање превише еха може довести до губитка важних детаља у говору и утицати на разумљивост.
- **Комплексност реализације:** Симулирање реалистичног еха може бити сложено и захтевати прецизно подешавање параметара како би се постигли жељени ефекти.
- **Затамњење Оригиналног Говора:** У неким случајевима, ефекат еха може бити толико доминантан да затамни оригинални сигнал, што може погоршати разумљивост и перформансе ASR система.
- **Потреба за Опсежном Обрадом:** За реалистичну симулацију еха потребна је опсежна обрада и фина подешавања, што може захтевати додатне ресурсе и време за имплементацију.

3.5 Time Masking

Time Masking је техника која укључује прикривање одређених делова аудио сигнала у временском домену како би се модел обучио да игнорише неке делове сигнала који могу бити непрецизни или непотпуни. Ова техника помаже у стварању робуснијих модела који могу да се носе са непотпуним или делимично изгубљеним подацима.

Time masking се изводи тако што се одређени део аудио сигнала замагљује или потпуно уклања. Ово може бити постигнуто применом различитих техника, као што су:

- **Изрезивање:** фрекидање одређених секција аудио сигнала, где се целе фрагменте сигнала замењују ћутњом или другим звуковима.
- **Применом филтера:** Користе се филтери који потпуно прикривају одређене временске домене, чиме се моделу омогућује да се носи са делимичним губитком података.



Слика 5. Time Masking (црвени оквир показује комаде који недостају)

Предности

- **Повећава робусност:** Помоћу time masking технике, модел може постати отпорнији на непотпуности и шум у временском домену.
- **Подржава разноврсне услове:** Користећи ову технику, модел може научити да преузме важне информације чак и када су делови аудио записа нестали или замагљени.

Мане

- **Могући губитак важних информација:** Прекидање или замагљивање делова сигнала може довести до губитка важних информација и утицати на квалитет говора.
- **Сложеност примене:** Потребно је пажљиво подешавање параметара као што су време и опсег замагљивања како би се избегло негативно утицање на перформансе модела.

3.6 Frequency Masking

Frequency Masking је техника која укључује прикривање одређених фреквенцијских компоненти аудио сигнала како би се модел научио да се носи са делимичним губитком фреквенцијских информација. Ова техника може помоћи у побољшању робусности модела у условима где су фреквенцијске компоненте сигнала делимично изгубљене или нарушене.

Frequency masking се изводи тако што се уклања или замагљује одређени опсег фреквенција у аудио сигналу. Ово може бити постигнуто коришћењем различитих техника, као што су:

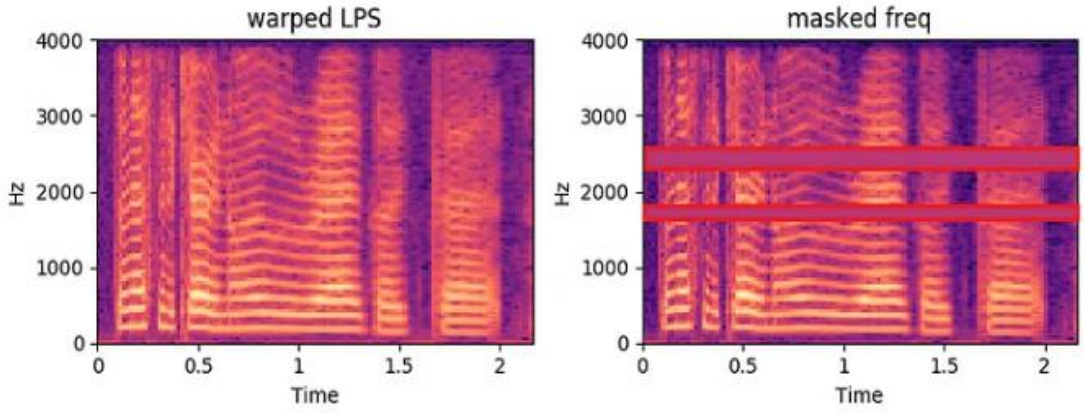
- **Фреквенцијски филтери:** Користе се филтери за уклањање или ослабљивање одређених фреквенцијских компоненти у аудио сигналу.
- **Изрезивање фреквенција:** Применом метода који потпуно уклањају одређене фреквенцијске опсеге, чиме се модел обучава да игнорише те делове сигнала.

Предности

- **Повећава робусност на фреквенцијске промене:** Омогућава моделу да се носи са фреквенцијским изменама и губитком информација у аудио сигналу.
- **Разноврсност података:** Креира нове узорке који симулирају услове са делимично изгубљеним фреквенцијама, што може побољшати перформансе модела у стварним условима.

Мане

- **Могући губитак важних фреквенцијских информација:** Применом frequency masking технике може доћи до губитка важних фреквенцијских компоненти, што може утицати на разумљивост говора.
- **Сложеност Подешавања:** Потребно је пажљиво подешавање фреквенцијских опсега и нивоа замагљивања како би се избегло негативно утицање на перформансе модела.



Слика 6. Frequency masking (црвени оквири приказују фреквенцију која недостаје)

3.7 Скалирање Амплитуде

Скалирање амплитуде је техника аугментације која подразумева прилагођавање јачине или интензитета аудио сигнала. Ова техника се користи за симулацију различитих нивоа звука, као што су различити волумени или растојања између аудио извора и микрофона. Циљ је да се модел обучи да буде отпоран на варијације у јачини звука и да буде способан да ради у различитим акустичким условима

Скалирање амплитуде укључује модификацију нивоа звука у аудио сигналу. Ово може бити постигнуто на неколико начина:

- **Проценат скалирања:** Амплитуда сигнала се може скалирати за одређени проценат. На пример, ако се сигнал скалира на 50%, његова јачина ће бити смањена за пола у односу на оригинал.
- **Линеарно или логаритамско скалирање:** Скалирање може бити линеарно или логаритамско, у зависности од захтева апликације и типичног обима звука.
- Различити параметри за различите часове: Амплитуда се може променити у различитим деловима аудио сигнала, стварајући динамичне варијације у јачини звука.

Предности

- **Симулација различитих услова:** Помаже у симулацији различитих нивоа звука и растојања, што може побољшати перформансе модела у реалним ситуацијама где се ниво звука може значајно разликовати.

- **Повећава робусност:** Обучава модел да буде отпоран на варијације у амплитуди, што је корисно у окружењима са променљивим звуковима.

Мане

- **Могући губитак информација:** Прекомерно скалирање може довести до губитка важних детаља у звуку, као што су тивки делови говора.
- **Дефекти у квалитету:** Неправилно скалирање може довести до изобличења звука, што може утицати на квалитет и разумљивост говора.

Поред појединачних техника аугментације, комбиноване технике представљају важан аспект за побољшање робусности модела. Комбинујући методе као што су додавање позадинског шума и промена брзине или амплитуде, могу се створити сложенији и разноврснији аудио подаци. Ова стратегија помаже моделима да се боље прилагоде различитим условима у реалним сценаријима, побољшавајући њихову способност да разликују и обрађују различите типове говора.

4. Изазови у аугментацији аудио података

Аугментација аудио података представља кључну технику за побољшање перформанси система за аутоматско препознавање говора (ASR). Иако ова метода нуди значајне предности, она такође са собом носи низ изазова који могу утицати на квалитет и ефикасност модела. Испод су детаљније описани неки од главних изазова у аугментацији аудио података:

1. **Очување разумљивости говора:** Један од главних изазова код аугментације аудио података је очување јасноће и разумљивости говора. Примена техника као што су додавање позадинског шума, еха, или промене брзине може увести значајне варијације у аудио сигнал. Ако су ове варијације претеране, могу замаглити кључне карактеристике говора, што може довести до погрешног препознавања од стране модела. Стога, балансирање између додавања корисних варијација и очувања квалитета је суштински важно.
2. **Реалистичност генерисаних података**: Добијање реалистичних аугментираних података је од кључне важности за ефикасност модела. Генерисани подаци треба да верно одражавају стварне сценарије у којима ће ASR систем радити. На пример, ако се дода превише вештачког шума или ако се изврши неприкладно скалирање амплитуде, генерисани узорци могу постати нереални и довести до лоше генерализације модела у реалним условима. Потребно је пажљиво тестирање да би се пронашла одговарајућа равнотежа.
3. **Одређивање оптималних параметара:** Свака техника аугментације захтева подешавање специфичних параметара, као што су ниво шума, проценат скалирања амплитуде или степен промене брзине. Проналажење оптималних вредности за ове параметре није једноставан задатак јер различити сетови података и сценарији захтевају различите конфигурације. Параметри који побољшавају перформансе у једном окружењу могу бити контрапродуктивни у другом.
4. **Повећана варијабилност може унети буку:** Иако је варијација у подацима пожељна за тренинг модела, прекомерна варијација може довести до тога да модел учи и небитне, случајне карактеристике уместо релевантних образаца говора. Ово може резултирати моделом који је мање прецизан или претрениран, односно моделом који је научио специфичне карактеристике тренинг података, али се лоше понаша на новим, невидљивим подацима.
5. **Евалуација и валидација аугментације:** Евалуација учинка модела обученог са аугментираним подацима такође представља изазов. Потребно је пажљиво тестирање и валидација да би се осигурало да аугментација побољшава, а не погоршава перформансе. Понекад је тешко одредити да ли су побољшања резултат аугментације или случајних фактора, што захтева додатну контролу експеримената.

Ови изазови захтевају пажљиво планирање и приступ аугментацији аудио података, као и темељно тестирање како би се постигли оптимални резултати у обуци ASR система.

5. Примери алата и библиотека за аугментацију аудио података

Постоји више алата и библиотека које истраживачи и инжењери користе за аугментацију аудио података у пројектима за аутоматско препознавање говора (ASR) и друге аудио анализе. Ови алати нуде различите функције и приступе за примену техника аугментације, што олакшава процес стварања разноврсних и робусних скупова података. Испод су наведени неки од најпопуларнијих алата и библиотека за аугментацију аудио података:

1. **Audiomentations:**
   - Audiomentations је једна од најпопуларнијих Python библиотека за аугментацију аудио података. Нуди велики број техника, укључујући додавање шума, промену брзине и висине тона, реверберацију и још много тога. Погодна је за брзу имплементацију аугментације и лако се интегрише са другим Python алатима.
   - Предности: Једноставна за употребу, велика разноликост техника, активна заједница.
   - Примена: Користи се за експерименте са аугментацијом у истраживању и развоју ASR модела.
2. **SoX (Sound eXchange):**
   - SoX је моћан и флексибилан алат за обраду аудио података, доступан као командна линија апликација. Омогућава примену различитих аудио ефеката и аугментација, као што су филтрирање, реверберација и echo.
   - Предности: Веома флексибилан, подржава широк спектар аудио формата, може се користити на више платформи.
   - Примена: Широко коришћен у аудио продукцији и припреми података за ASR системе.
3. **Librosa:**
   - Librosa је Python библиотека фокусирана на анализу и обраду музике и аудио података. Иако није специфично дизајнирана за аугментацију, њене могућности за обраду сигнала и аудио анализа омогућавају примену неких техника аугментације, као што су промене брзине и висине тона.
   - Предности: Снажан алат за аудио анализу, интеграција са другим Python библиотекама као што су NumPy и SciPy.
   - Примена: Истраживање и анализа аудио података, припрема аудио карактеристика за машинско учење.
4. **WavAugment:**
   - WavAugment је библиотека заснована на торчевој (PyTorch) која подржава аугментацију аудио података и користи се у истраживањима ASR и аудио обраде. Нуди једноставан API и омогућава лако креирање аугментираних података током тренинга модела.
   - Предности: Лака интеграција са PyTorch моделима, подршка за GPU убрзање.

- Примена: Аугментација током тренинга дубоких неуронских мрежа за ASR и аудио класификацију.

5. **SpecAugment:**
    - SpecAugment је метода аугментације заснована на спектрограмима, која се користи за побољшање робусности ASR система. Иако није класична библиотека, то је техника која се често користи уз остале библиотеке попут PyTorch и TensorFlow.
    - Предности: Једноставна за примену, побољшава перформансе ASR система на тестним подацима.
    - Примена: Интеграција у архитектуре дубоких неуронских мрежа за ASR.

Ови алати и библиотеке пружају флексибилност и снагу потребну за ефикасну аугментацију аудио података. Одабир алата зависи од специфичних захтева пројекта, као и од корисничког окружења и доступних ресурса.

6. Закључак

Аугментација аудио података представља кључну технику у унапређењу система за аутоматско препознавање говора (ASR), омогућавајући моделима да боље генерализују и раде у реалним условима. Кроз различите технике као што су додавање шума, промене брзине и висине тона, додавање еха, и примена маскирања, могуће је значајно повећати робусност система и побољшати његову способност да се носи са варијацијама у говору и окружењу.

Иако аугментација доноси бројне предности, као што су побољшање перформанси и отпорност на различите услове, такође се јављају изазови као што су очување разумљивости говора, реалистичност генерисаних података и потреба за великим рачунарским ресурсима. Пажљиво подешавање параметара и балансирање између побољшања и очувања квалитета говора су кључни фактори за успешну примену ове методе.

Коришћење различитих алата и библиотека, попут Audiomentations, SoX, Librosa и других, олакшава процес аугментације и омогућава истраживачима да примене различите технике у складу са специфичним потребама пројекта. Са наставком развоја и иновација у области аугментације, будућност ASR система изгледа светлија, отварајући пут за прецизније и поузданије системе у различитим применама.

Референце

[1] https://leonardloo.medium.com/part-2-audio-speech-processing-speech-data-augmentation-463dac823a8d

[2] https://blog.pangeanic.com/audio-data-augmentation-techniques-and-methods

[3] https://medium.com/@makcedward/data-augmentation-for-audio-76912b01fdf6

[4] https://towardsdatascience.com/audio-augmentations-in-tensorflow-48483260b169

[5] https://www.scaler.com/topics/tensorflow/data-augmentation-tensorflow/