

Detection and Analysis of Human Emotions Through Voice and Speech Pattern Processing

Dr. A.S. Prabakaran¹, M. Kabilan², R. Karthi², K. Kathiravan², S. Poornachandaran²

¹Supervisor, ²Student (B.E)

Department of Information Technology, Muthayammal Engineering College, Rasipuram, Tamil Nadu, India

ABSTRACT

Article Info

Volume 8, Issue 7

Page Number : 58-63

Publication Issue :

May-June-2022

Article History

Accepted: 01 June 2022

Published: 20 June 2022

The ability to modulate vocal sounds and generate speech is one of the features which set humans apart from other living beings. The human voice can be characterized by several attributes such as pitch, timbre, loudness, and vocal tone. It has often been observed that humans express their emotions by varying different vocal attributes during speech generation. Hence, deduction of human emotions through voice and speech analysis has a practical plausibility and could potentially be beneficial for improving human conversational and persuasion skills. This presents an algorithmic approach for detection and analysis of human emotions with the help of voice and speech processing. The proposed approach has been developed with the objective of incorporation with futuristic artificial intelligence systems for improving human-computer interactions. These systems aim to facilitate the natural interaction with machines by direct voice interaction instead of using traditional devices as input to understand verbal content and make it easy for human listeners to react. Some applications include dialogue systems for spoken languages such as call center conversations, onboard vehicle driving system and utilization of emotion patterns from the speech in medical applications. Emotion recognition from speech signals is an important but challenging component of HCI. In the literature of SER, many techniques have been utilized to extract emotions from signals, including many well-established speech analysis and classification techniques. Deep Learning techniques have been recently proposed as an alternative to traditional techniques in SER. This paper presents an overview of Deep Learning techniques and discusses some recent literature where these methods are utilized for speech-based emotion recognition. The review covers databases used, emotions extracted, contributions made toward speech emotion recognition and limitations related to it.

KEYWORDS : Deep learning, emotional analysis, human emotions, speech processing, voice processing.

I. INTRODUCTION

Deep Learning is continuously amusing us with its modern possibilities like self driving cars, fraud detection, and many more. Earlier we never imagined the things which are possible today and now we cannot even imagine a day without using it. Thus, in this blog, we are going to discuss this very interesting topic „Deep Learning“ in much more detail. Deep Learning is at the beginning of what machines can do and developers and business leaders totally need to comprehend what it is and how it functions. Deep learning models are sufficiently competent to focus on the exact features themselves by requiring a little direction from the programmer and are useful in taking care of the issue of dimensionality. Therefore, deep learning algorithms are used, particularly when we have a vast number of inputs and outputs. It is a kind of machine learning that prepares a computer to perform human-like errands, for example, perceiving speech, distinguishing pictures, or making forecasts. Rather than arranging information to go through predefined conditions, deep learning sets up essential boundaries about the information and trains the computer to learn on its own by perceiving designs using numerous layers of processing. Deep learning has networks worthy of learning unsupervised from information that is unstructured or unlabeled. In simple language, deep learning is a type of algorithm

II. LITERATURE REVIEW

One of the most important information that speech acoustics provide is the expression of emotions. The purpose of this research is to identify the pitch differences between two basic emotions: anger and joy. In order to find answers to this question vocal data have been collected from small group of participants. Results from Friedman's Two-Way Analysis of Variance by Ranks revealed difference in pitch levels when expressing anger and joy as well as jitter (rap). It is well known that speech is an acoustically rich signal that provides a lot of information about the speaker during vocal interaction. The expression and recognition of emotions are extremely important steps for human communication process for this reason voice recognition is useful for detecting and identifying specific affective characteristics between the speakers.

III. PROPOSED SYSTEM

Deduction of human emotions through voice and speech analysis has a practical plausibility and could potentially be beneficial for improving human conversational and persuasion skills. This paper presents an algorithmic approach for detection and analysis of human emotions on the basis of voice and speech processing. Three test cases have been examined, corresponding to the three emotional states: normal emotional state, angry emotional state, and panicked emotional state. Each case demonstrates characteristic associated vocal features which can help in distinguishing the corresponding emotional state. We examine the effectiveness of applying machine learning techniques to the sentiment classification problem. A challenging aspect of this problem that seems to distinguish it from traditional topic-based classification is that while topics are often identifiable by keywords alone, sentiment can be expressed in a more subtle manner. For example, the sentence "How could anyone sit through this movie?" contains no single word that is obviously negative. Thus, sentiment seems to require more understanding than the usual topic-based classification. So, apart from

presenting our results obtained via machine learning techniques, we also analyze the problem to gain a better understanding.

Approach To Detection Of Human Emotions Algorithm

This section describes an algorithmic approach for deducing human emotions through voice- and speech-pattern analysis. In order to achieve this objective, three test cases have been examined, corresponding to the three emotional states: Normal emotional state, Angry emotional state, and Panicked emotional state. For carrying out the analysis, four vocal parameters have been taken into consideration: pitch, SPL, timbre, and time gaps between consecutive words of speech. In order to quantitatively represent timbre, its temporal envelope for advance and decay times has been considered. Its a different emotional states by analyzing the deviations in the aforementioned four parameters from that of the normal emotional state. The proposed analysis was carried out with the help of software packages such as MATLAB and Wavepad.

IV. SYSTEM SPECIFICATIONS

HARDWARE REQUIREMENTS

Processor : Intel processor 3.0 GHz
 RAM : 1GB
 Hard disk : 40 GB
 Compact Disk : 650 Mb
 Keyboard : Standard keyboard
 Mouse : Logitech mouse
 Monitor : 15-inch colour monitor15

SOFTWARE REQUIREMENTS

Operating System : Windows OS
 System type : 32-bit or 64-bit Operating System
 IDE : Python 3.5 and above Install
 Dependency : pip install pandas, pip install matplotlib, Pip install keras, pip install tensorflow

V. RESULT

[illegible]

OUTPUT 1

The screenshot displays the Anaconda Jupyter Notebook environment. The top bar contains menu options: File, Edit, Selection, View, Cell, Debug, Terminal, and Help. The left sidebar shows the file explorer with a project named 'New and Existing Python File'. The main area contains a Jupyter Notebook with a Python script. The script defines a function to create a new feature 'new_label' based on the 'label' and 'category' columns of a dataset. It then loads the dataset, applies the function, and saves the modified dataset as 'new_data.csv'.

```

14 elif mode == "new_label":
15     return "new_label"
16 elif mode == "new_label":
17     return "new_label"
18 elif mode == "new_label":
19     return "new_label"
20 else:
21     return "new_label"
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994

```

OUTPUT 2

VII. ACKNOWLEDGEMENTS

I would like humbly acknowledge the support of the management, senior lecturers, Guide and ECE department of Muthayammal Engineering College, Rasipuram, Tamilnadu rendering the thankfulness which enabled us in proceeding with research work, in-line with the present social needs and trends.

VIII. REFERENCES

- [1]. Poorna Banerjee Dasgupta, Detection and Analysis of Human Emotions through Voice and Speech Pattern Processing, International Journal of Computer Trends and Technology (IJCTT), 2017.
- [2]. Li Deng, Jinyu Li, Jui-Ting Huang, Recent advances in deep learning for speech research at Microsoft, Microsoft Corporation, 2017.
- [3]. Elka Popova, Ilona Isaeva, Emotion recognition through voice analysis, Tulborg University, 2015.
- [4]. Jason Lee Wright, Voice to Text Conversion, US Patent application publication, 2015.
- [5]. Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng. , Learning Word Vectors for Sentiment Analysis, 49th Annual Meeting of the Association for Computational Linguistics, 2011.
- [6]. Theresa Wilson, Johanna Moore, Twitter Sentiment Analysis: The Good the Bad and the OMG!, Proceedings of the Fifth International AAAI Conference, 2011.
- [7]. Michael L. Asmussen, Advanced set top terminal having a voice-to text conversion, IEEE, 2007.
- [8]. Thomas E. Creamer, Peeyush Jaiswal, Voice-to-text reduction for real time in/chat/ sms, US Patent application Publication, 2004.
- [9]. Gary L. Griffith, Communication device having voice and text message, United States Patent, 2002.
- [10]. Bo Pang and Lillian Lee, Thumbs up? Sentiment Classification using Machine Learning Techniques, Proceedings of the Conference on Empirical Methods