Neural Network Approach to Speech Pathology

Antonio P. Salvatore[1], Nicole A. Thorne[2], and Charlotte M. Gross[1]
[1]Health Sciences, University of Texas, El Paso, TX 79968, USA, anthonys@utep.edu
[2]Computer Science, University of Texas, El Paso, TX 79968, USA, nthorne@cs.utep.edu

***Abstract*—A speech problem can be caused by different reasons, from psychological to organic. The existing diagnostic of speech pathologies relies on skilled doctors who can often diagnose by simply listening to the patient. We show that neural networks can simulate this ability and thus provide an automated (preliminary) diagnosis.**

I. Introduction

A speech problem can be caused by different reasons, from psychological to organic. The existing diagnostic of speech pathologies relies on experienced clinicians who can often diagnose a disorder by simply listening to the patient. Unfortunately, speech pathologies are reasonably frequent, and not many clinicians are that skilled. Therefore, it is desirable to design an automatic tool which would help in diagnostics of speech pathologies.

It is still difficult for a computer to automatically analyze human speech, so this automatic program should use the characteristics of a speech which can be easily measured or easily recorded. To determine these characteristics, patients are asked to read a text.

First of all, the recorder measures the *total speaking time* $T_{\rm speak}$. This number depends not only on the person, but also on the text. To make easier comparison between results of different tests, in which different texts may have been used, it is useful to report this characteristic in the form of a *reading rate* $r_{\rm read}$, measured as the average number of words per minute.

Since many speech disorders like stuttering involve a lengthened interval between words and a disruption of fluent speech, it is also useful to measure not only the total speaking time, but also which part of this time was spent actually articulating words (*articulation time* $T_{\rm art}$), and which part of this time was spent on *inter-word intervals* ($T_{\rm inter}$). Of course, the sum of these two times must be equal to the total speaking time: $T_{\rm art} + T_{\rm inter} = T_{\rm speak}$.

It also makes sense to measure the total *number of reading errors* $E_{\rm read}$ and the *number of disfluencies* $N_{\rm disf}$. (To make comparison between different tests easier, usually, the reported characteristics is not the total number of disfluencies, but the number of disfluencies $n_{\rm disf}$ per hundred words.)

The use of these six speech characteristics $x_1 = T_{\rm speak}$, $x_2 = E_{\rm read}$, $x_3 = r_{\rm read}$, $x_4 = n_{\rm disf}$, $x_5 = T_{\rm inter}$, and $x_6 = T_{\rm art}$ is justified not only by common sense, but also by the fact that there is a statistically significant difference between patients with speech disorders and the population in general (see, e.g., [1]).

Based on these six characteristics, we must determine whether a patient has a speech disorder and if yes, which type of disorder.

In this paper, we concentrate on one specific type of speech disorder: spasmodic (or spastic) dysphonia (SD). SD is a disorder of laryngeal motor control characterized by intermittent voice stoppage (stuttering) that manifests most markedly during production of connected speech (see, e.g., [1] and references therein). Despite recent advances in treatment involving botulinum toxin injection of the vocal folds, the underlying etiology (biological causes) of SD remains controversial and poorly understood. We have chosen this particular disorder because it is known that this disorder is difficult to diagnose.

II. First Approximation: Linear Method

A. Formulation of the linear approach

We want to be able to tell, given six characteristics x_i, whether a patient has SD or not. Let us first try the simplest possible diagnostic based on a *linear discrimination*: namely, we will look for coefficients $w_1, \ldots, w_n$ and for a bias b such that

- if $b + w_1 \cdot x_1 + \ldots + w_n \cdot x_n > 0$, the patient has SD, and
- if $b + w_1 \cdot x_1 + \ldots + w_n \cdot x_n < 0$, the patient does not have SD.

B. The choice of an algorithm

To find these coefficients from the experimental data, we used the simplest type of neural network method – the perceptron algorithm (see, e.g., [2]). This algorithm was chosen because it is proven that this algorithm is guaranteed to converge and return the discriminating weights whenever such weights exist.

C. The description of the algorithm: in brief

Since the perceptron algorithm is not well known in the neural network community, for the reader's convenience, we will briefly describe how this algorithm works.

For this algorithm to work, we need to have several *patterns*, i.e., sequences $(x_1^{(k)}, \ldots, x_n^{(k)}, y^{(k)})$, $1 \le k \le N$ which describe, in our case, N patients with known speech characteristics $x_i^{(k)}$ and a known diagnosis. To describe a diagnosis in the desired numerical form, we set $y^{(k)} = 1$ if the diagnosis is SD, and $y^{(k)} = -1$ otherwise.

In the algorithm, we process these patterns one by one, and iteratively update the values of the weights and of the bias; when after processing all the patterns, we still do not get the desired values of w_i and b, we then process the same patterns again and again until we get the desired values.

Initially, the values of w_i and b are set to 0. If by the time we process k-th pattern, we have some values w_i and b, then we check whether these weights work well for this pattern in the sense that for this pattern, these values correctly predict the presence or absence of SD. Due to our description of the desired discrimination function, we must check whether the sign of the linear combination $b + w_1 \cdot x_1^{(k)} + \ldots + w_n \cdot x_n^{(k)}$ coincides with the sign of $y^{(k)}$.

- If these two signs coincide, this means that the weights and the bias work well for this particular pattern. Therefore, based on this pattern, there is no need to change w_i or b.
- If these signs differ, this means that a correction is necessary. The perception algorithm's correction consists of the following change in w_i and b:
  - the old value of each weight w_i is replaced by the new value $w_i + y^{(k)} \cdot x_i^{(k)}$; and
  - the old value of the bias b is replaced by the new value $b + y^{(k)}$.

D. First experimental result: by using linear discrimination, we cannot correctly diagnose all the patients

We started with 40 patterns described in [1] (this data is reproduced at the end of this paper). At first, we tried the perceptron algorithm for all 40 patterns, and this algorithm did not converge, indicating that the actual diagnostics cannot be described by the simplified linear discrimination function.

E. Let us find a linear scheme which correctly diagnoses most of the patients

Although we could not get the correct diagnosis for *all* patients, we still hoped that we would be able to get the correct diagnosis for a large part of the patients.

By looking at the data, we saw that in many cases, there is a simple way of checking whether a patient has SD or not: namely, for most patients with SD, the value of the parameter $x_4 = n_{\rm disf}$ is larger than the value of x_4 for all patients without SD. The largest value of x_4 for patients without SD is 1.02, so if $x_4 > 1.02$, then for these data, we can conclude that a patient has SD.

This inequality $x_4 > 1.02$ does not work as a perfect classification tool, because in our data set, there are five patients with SD for whom the value x_4 is also small ($x_4 \le 1.02$). These patients are # 26, # 28, # 31, # 36, and # 40. So, the simple criterion $x_4 > 1.02$ enables us to correctly diagnose 35 out of 40 patients.

We hoped that the perception algorithm would enable us to classify more than 35 patients. So, we applied the perceptron algorithm to the patterns corresponding to 35 linearly-classifiable patients. As a result, we got a linear discrimination function with the following weights: $w_1 = 861.46$, $w_2 = 494$, $w_3 = -125.74$, $w_4 = 5,526.89$, $w_5 = 1,815.96$, $w_6 = -954.50$, and $b = 15$.

Then, we tested whether this linear discrimination function worked on any other patients. It turns out that indeed, it worked for two more patients: # 31 and # 40. So, we got 37 patients covered by a linear diagnosis.

To check whether any more patients can be thus covered, we repeated the same procedure by applying the perceptron algorithm to the 37 patterns. As a result, we got new weights: $w_1 = 1,592.24$, $w_2 = 1,734$, $w_3 = -309.42$, $w_4 = 9,073.54$, $w_5 = 2,766.183$, $w_6 = -1,173.94$, $b = 36$. We checked whether any new patients can be covered by this new linear discrimination function, but it did not happen.

So, our conclusion is that by using linear discrimination function, we can correctly diagnose 37 out of 40 patients – around 90%.

III. Non-Linear Methods: Future Plans

Since *linear* methods cannot correctly diagnose all the patients, we must use *nonlinear* methods. In particular, for diagnosing SD, we have used the standard backpropagation neural network which, in contrast to perceptron, uses non-linear neurons (see, e.g., [2]).

When we get the exact diagnosis, the natural next step would be to check whether all the characteristics are really needed for this diagnosis, or can we use only some of them and thus, decrease both the time necessary to measure these characteristics and the computation time necessary to process them.

In our future work, we want to take into consideration the fact that while we assumed that experts' diagnoses are perfect, in reality, even skilled experts may err. In many speech pathology situations, there is a general agreement between experts, and so, the common expert diagnosis is reliable. In other situations, however, experts are in disagreement; in this case, we would like to extract, from the patient data, a reasonable subdivision into groups (clusters) of patients with similar pathologies. For this, we plan to use clustering neural networks. Preliminary results of this work, described in [3], show that this approach is indeed very promising.

We are not sure about the final result of these clustering attempts:

- It may happen that we will get a reasonable classification; then, we will look for features common to patients from different clusters, and look for treatments which are tailored to this particular cluster.
- It may also happen that there is no clear subdivision into groups. Instead, there is a continuous transition between patients with different speech pathologies, and even between patients with and without speech pathology. In this case, pathology would be a matter of degree, and so, instead of simply telling whether a patient has a certain pathology or not, we would produce a degree to which a patient has a certain pathology.

Finally, we would like to collect data on how characteristics change with time, and use this dynamical data to design a computer model for predicting this change. For this problem, we plan to use different neural network techniques, starting with backpropagation.

IV. Conclusion

In this paper, we showed that even a simple (perceptron) neural network can lead to a reasonably good diagnosis of speech disorders. We are planning to use more complicated neural networks to correctly diagnose more patients.

The resulting linear discrimination formulas describe the weights of different characteristics. This enables the clinicians not only to get a preliminary diagnosis, but it may also offer insight into which characteristics contribute the most to this particular speech disorder.

Acknowledgment

This work was supported in part by NSF grant No. CDA-9522207.

N.A.T. is thankful to Dr. Ann Q. Gates for her support.

References

[1] M. P. Cannito, A. R. Burch, C. Watts, P. W. Rappold, S. B. Hood, and K. Sherrard, "Disfluency in Spasmodic Dysphonia: A Multivariate Analysis", *Journal of Speech, Language & Hearing Research*, vol. 40, pp. 627–641, 1997.

[2] L. Fausett, *Fundamentals of Neural Networks*, Englewood Cliffs, New Jersey: Prentice-Hall, 1994.

[3] A. P. Salvatore, G. S. Guieterrez, and M. P. Cannito, "Spasmodic Dysphonia: A Neural Net-Analysis" (to appear).

Appendix: Experimental Data

In this study, we used the data from 40 female subjects, 20 of them with SD, 20 without SD. The first table contains patients without SD¡ the second table contains patinets with SD. The first column in each table is the patient number.

	x_1	x_2	x_3	x_4	x_5	x_6
1	25.422	0	231.48	0	5.63	19.79
2	37.33	1	157.62	0	5.36	31.97
3	31.46	0	186.88	1.02	7.23	24.23
4	30.70	0	191.52	0	9.24	21.46
5	29.52	1	201.25	1.01	6.17	23.35
6	31.91	0	184.32	0	5.63	26.28
7	31.05	2	191.33	1.01	4.98	26.07
8	31.88	1	184.32	0	5.98	25.9
9	29.91	0	196.68	0	5.71	24.20
10	26.76	0	219.42	0	3.89	22.92
11	29.75	1	197.34	0	5.44	24.32
12	29.29	0	200.7	1.02	5.44	23.85
13	38.58	0	152.04	1.02	5.44	33.15
14	26.97	1	223.86	0.99	4.57	22.42
15	30.28	0	194.04	0	6.17	24.11
16	35.53	1	163.92	0	6.21	29.32
17	28.49	0	206.34	0	4.95	23.53
18	31.36	0	187.26	0	7.23	24.13
19	27.24	1	218.4	0	4.64	22.6
20	31.01	0	189.66	0	8.61	22.4

Appendix (cont-d)

	x_1	x_2	x_3	x_4	x_5	x_6
21	53.50	0	109.92	3.06	18.8	34.70
22	46.92	1	124.03	17.53	15.16	31.77
23	33.69	1	178.02	7	7.17	26.53
24	35.68	0	164.7	8.16	7.69	27.99
25	33.59	1	173.22	3.09	5.80	27.79
26	30.84	0	190.92	0	4.29	26.55
27	39.93	0	148.86	2.02	8.89	31.05
28	24.63	0	239.04	1.02	4.99	9.64
29	31.69	0	185.46	4.08	4.01	27.68
30	33.21	2	182.52	1.98	5.40	27.81
31	36.99	2	160.56	1.01	8.28	28.72
32	48.47	0	121.26	14.29	16.94	31.53
33	37.17	0	164.52	8.82	7.45	29.72
34	47.88	0	122.81	13.27	15.65	32.23
35	52.61	0	111.78	8.16	15.69	36.92
36	31.73	4	185.46	1.02	5.44	26.29
37	49.83	2	120.48	25	16.56	33.72
38	51.90	1	113.28	13.27	18.07	33.80
39	38.73	3	151.92	4.08	9.68	29.05
40	39.08	1	150.36	1.02	9.29	29.79