A Machine Learning Model to Detect Speech and Reading Pathologies

Fabio Fassetti[1](✉) and Ilaria Fassetti[2]

[1] DIMES, University of Calabria, Rende, Italy
f.fassetti@dimes.unical.it
[2] Therapeia Rehabilitation Center, Rende, Italy
ilaria.fassetti@gmail.com

Abstract. This work addresses the problem of helping speech therapists in interpreting results of tachistoscopes. These are instruments widely employed to diagnose speech and reading disorders. Roughly speaking, they work as follows. During a session, some strings of letters, which may or not correspond to existing words, are displayed to the patient for an amount of time set by the therapist. Next, the patient is asked for typing the read string. From the machine learning point of view, this raise an interesting problem of analyzing the sets of input and output words to evaluate the presence of a pathology.

Keywords: Tachistoscope · Deep neural networks · Dyslexia

1 Introduction

This work aims at helping in analyzing results of the tachistoscope [5], a widely used diagnostic tool by speech therapists for reading or writing disorders.

Tachistoscopes [1] are employed to diagnose reading disorders related to many kinds of dyslexia and disgraphia and, also, to increase recognition speed and to increase reading comprehension.

In more details, tachistoscopes are devices that display a word for several seconds and ask the patient to write down the word. Many parameters can be set during the therapy session, like the duration of displayed word, the length of the word, the structural easiness of the word, how much the word is common and others [7]. Among them a relevant role is played by the possibility for the therapist of choosing existent or non existing words. Indeed, the presence of non existing words avoids the help coming from the semantic interpretation [4].

Hence, a tachistoscope session provides the set of configuration parameters, the set of displayed words and the set of words as typed by the patient. The set of produced words obviously depends on the pathologies affecting the patient

Published by Springer Nature Switzerland AG 2020
I. Maglogiannis et al. (Eds.): AIAI 2020, IFIP AICT 584, pp. 135–142, 2020.
https://doi.org/10.1007/978-3-030-49186-4_12

but, in many cases, more pathologies can simultaneously be present and this, together with the many words to be analyzed, make diagnosis a tedious and very hard task [2,3,6].

This work faces the challenging problem of recognizing pathologies involved and helping experts by suggesting which pathologies and to what extent are likely to affect the patient.

To the best of our knowledge this is the first work attempting to tackle this problem. Authors of [5] address the question of mining tachistoscope output but from a totally different point of view. Indeed, there, the focus is on individuating patterns characterizing errors and not on pathologies.

However, the proposed framework is more widely applicable and, then, this work aims at providing general contributions to the field.

The paper is organized as follows. Section 2 presents preliminary notions, Sect. 3 details the technique proposed to face the problem at hand, Sect. 4 reports experiments conducted to show the effectiveness of the approach and, finally, Sect. 5 depicts conclusions.

2 Preliminaries

Let Σ be the considered alphabet composed, for English, by 26 letters plus stressed vowels and, then, by 32 elements. A word w is an element of Σ^*. Let $\mathcal{W}$ be a set of words, for each pathology p, let $\mathcal{F}_p$ be a transfer function that converts a word $w \in \mathcal{W}$ in a new word w_p due to the pathology p.

Tachistoscope. The tachistoscope is an instrument largely employed for detecting many reading and writing disorders. In a session, the user is provided with a sequence of trials each having a word associated with it. The trial consists in two phases: *visualization phase* and *guessing phase*. During the visualization phase a word is shown for some milliseconds (typically ranging from 15 to 1,500). To this phase, the guessing phase follows. During this phase, the word disappears and the user has to guess the word by typing it. After that the word disappears, it could be substituted by a set of '#' to increase the difficulty since the user loses the visive memory. In such a case, the word is said "masked".

The therapist, other that the visualization time, can impose several settings about the word, in particular, (*i*) the *frequency* of the word, which represents how much the word is of current use; (*ii*) the *length* of the word, which represents how much long is the word; (*iii*) the *easiness* of the word, which represents how much difficult is reading and writing the word (for example, each consonant is followed by a vowel); (*iv*) the *existence* of the word, which represents the existence of the word in the dictionary. As for the existence of the word, the tachistoscope is able to show both existing words and non-existing words which are random sequences of letters with the constraints that (*i*) each syllable has to appear in at least one existing word and (*ii*) each pair of adjacent letters has to appear in at least one existing word. Such constraints aim at generating readable sequences of letters.

3 Technique

In this section, the main problem addressed in this work is formally presented.

Definition 1 (Problem). *Let $\mathcal{W}^{in}$ be a set of input instances and $\mathcal{W}^{out}$ the set of output instances, provide the likelihood λ that the patient producing $\mathcal{W}^{out}$ is affected by pathology p.*

In this scenario, the available examples provide as information that w^{in} is transformed in w_i^{out} when individuals affected by the pathology p_i elaborate w^{in}.

For a given pathology p, experts provide sets $\mathcal{W}_p^{in}$ and $\mathcal{W}_p^{out}$, which are, respectively, the sets of correct and erroneous words associated with individuals affected by pathology p, where for each word $w \in \mathcal{W}_p^{in}$ there is an associated word in $\mathcal{W}_p^{out}$.

For a given patient, tachistoscope sessions provide two sets of words $\mathcal{W}^{in}$ and $\mathcal{W}^{out}$, where the former one is composed by the correct words submitted to the patient, the latter one is composed by the words, possibly erroneous, typed by the patient. Such sets represent the input of the proposed technique.

The aim of the technique is to learn the transfer function $\mathcal{F}_p$ by exploiting sets $\mathcal{W}_p^{in}$ and $\mathcal{W}_p^{out}$, so that the system is able to reconstruct how a patient would have typed a word $w \in \mathcal{W}^{in}$ if affected by p. Hence, by comparing the word of $\mathcal{W}^{out}$ associated with w and the word $\mathcal{F}_p(w)$, the likelihood for the patient to be affected by p can be computed. Thus, the technique has the following steps:

1. encode words in a numeric feature space;
2. learn function $\mathcal{F}_p$;
3. compute the likelihood of pathology involvement.

Each step is detailed next.

3.1 Word Encoding

The considered features are associated with the occurrences in the word of letters in Σ and pairs of letters in $\Sigma \times \Sigma$. Also, there are six features, said *contextual*: *time*, *masking*, *existence*, *length*, *frequency* and *easiness*, related to the characteristics of the session during with the word is submitted to the patient. Thus, the numeric feature space $\mathcal{S}$ consists in $32 + 32 \cdot 32 + 6$ components where the i-th component $\mathcal{S}_i$ of $\mathcal{S}$, with $i \in [0, 32)$, is associated with a letter in Σ, the i-th component $\mathcal{S}_i$ of $\mathcal{S}$, with $i \in [32, 32 + 32 \cdot 32)$, is associated with a pair of letters in $\Sigma \times \Sigma$ and the i-th component $\mathcal{S}_i$ of $\mathcal{S}$, with $i \in [32 + 32 \cdot 32, 32 + 32 \cdot 32 + 6)$, is associated with a contextual feature.

Given a word $w \in \Sigma^*$, the encoding of w in a vector $v \in \mathcal{S}$ is such that the i-th component $v[i]$ of v is the number of occurrences of the letter or the pair of letters associated with $\mathcal{S}_i$ in w, for any $i \in [0, 32 + 32 \cdot 32)$, while, as for the contextual features, the value v assumes on these depends on the settings of the session as described in Sect. 2.

In the following, since there is a one to one correspondence between letters, pairs of letters and components, for the sake of readability, the notation $v[s]$, with $s \in \Sigma$ or $s \in \Sigma \times \Sigma$, is employed instead of $v[i]$, with i such that $\mathcal{S}_i$ is associated with s.

Hence, for example, $w =$ *"paper"* is encoded in a vector v such that $v[\text{'}P\text{'}] = 2$, $v[\text{'}A\text{'}] = 1$, $v[\text{'}E\text{'}] = 1$, $v[\text{'}R\text{'}] = 1$, $v[\text{'}PA\text{'}] = 1$, $v[\text{'}AP\text{'}] = 1$, $v[\text{'}PE\text{'}] = 1$, $v[\text{'}ER\text{'}] = 1$, and 0 elsewhere. Note that, as a preliminary steps, all words are uppercased.

Analogously, the notation $v[$*'time'*$]$, $v[$*'masking'*$]$, $v[$*'existence'*$]$, $v[$*'length'*$]$, $v[$*'frequency'*$]$ and $v[$*'easiness'*$]$ is employed to indicate the value the vector assumes on the components associated with contextual features.

3.2 Learning Model

Let $n = |\mathcal{S}|$ be the number of considered features. After some trials, the architecture of the neural network for this phase is designed as an autoencoder with five dense layers configured as follows:

layer 1: kind: *Dense*, neurons: n, activation: *'relu'*;
layer 2: kind: *Dense*, neurons: $2n$, activation: *'relu'*;
layer 3: kind: *Dense*, neurons: n, activation: *'relu'*.

Differently from classical auto encoders, latent space is larger than input and output ones this is due to the fact that here it is not relevant to highlight characterizing features shared by the input instances, since this could lead to exclude transformed features. The subsequent experimental section motivates this architecture, through an analysis devoted to model selection.

3.3 Likelihood Computation

The trained neural network provides how an input word would be transformed when typed by patients affected by a given pathology. Thus, for a given pathology p, the input word w, the word w' typed by patient and the word w^* returned by the model for the pathology p are available. The likelihood that the patient is affected by pathology p is given by the following formula, where w is employed as reference word, v_w, $v_{w'}$ and v_{w^*} are the encodings of w, w' and w^*.

$$\eta(w, w', w^*) = \sum_{i: v_w[i] \neq v_{w'}[i] \wedge v_w[i] \neq v_{w^*}[i]} \frac{\left| v_{w'}[i] - v_{w^*}[i] \right|}{\max\{v_{w'}[i], v_{w^*}[i]\}}. \tag{1}$$

Equation 1 is designed to measure the differences between w' and w^* by skipping letters where both w' and w^* agree with w, moreover it is able to highlight that both w' and w^* differ from w and that they agree on the presence of a certain letter, namely each letter in w' is in w^* and vice versa.

For example, let

$$\begin{aligned} w &= paperapa, \\ w^{i} &= qaqeraqa, \\ w^{ii} &= qaeraqa, \\ w^{iii} &= lalerala, \\ w^{iv} &= pqperqpq, \end{aligned}$$

where w is the reference word. The distances are the following:

- $\eta(w, w^{i}, w^{ii}) = 2.8$, since both w^{i} and w^{ii} differ from w for the same letter p, both of them substitute it with q;
- $\eta(w, w^{i}, w^{iii}) = 8$, since both w^{i} and w^{iii} differ from w for the same letter p, even if different substitutions are applied;
- $\eta(w, w^{i}, w^{iv}) = 10$, since both w^{i} and w^{iv} differ from w for a different letter, even if the same substituting letter appears.

4 Experiments

The conducted experiments are for Italian language since the rehabilitation center involved in this study provide input data in this language. However, the work has no dependencies with the language. The whole framework is written in *python* and uses *tensorflow* as underlying engine.

Figure 1 reports results for four different pathologies and for the following configurations:

Configuration 1

layer 1:	kind: *Dense*,	neurons: n,	activation: *‘relu’*;
layer 2:	kind: *Dense*,	neurons: $n/2$,	activation: *‘relu’*;
layer 3:	kind: *Dense*,	neurons: $n/4$,	activation: *‘relu’*;
layer 4:	kind: *Dense*,	neurons: $n/2$,	activation: *‘relu’*;
layer 5:	kind: *Dense*,	neurons: n,	activation: *‘relu’*.

Configuration 2

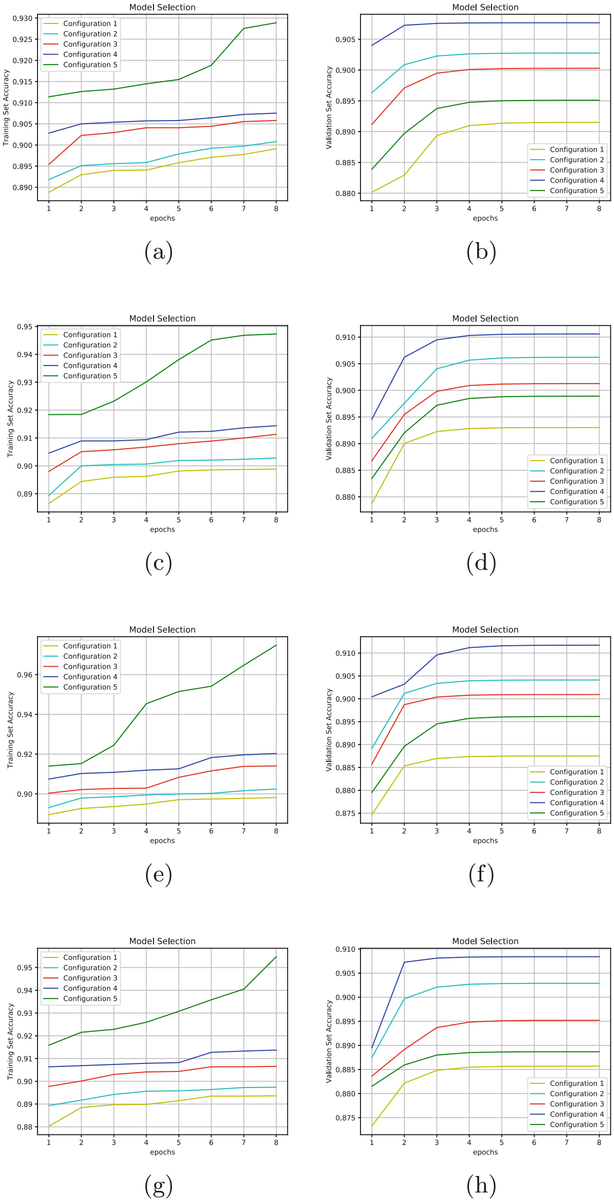
layer 1:	kind: *Dense*,	neurons: n,	activation: *‘relu’*;
layer 2:	kind: *Dense*,	neurons: $n/2$,	activation: *‘relu’*;
layer 3:	kind: *Dense*,	neurons: n,	activation: *‘relu’*.

Configuration 3

layer 1:	kind: *Dense*,	neurons: n,	activation: *‘relu’*;

Configuration 4

layer 1:	kind: *Dense*,	neurons: n,	activation: *‘relu’*;
layer 2:	kind: *Dense*,	neurons: $2n$,	activation: *‘relu’*;
layer 3:	kind: *Dense*,	neurons: n,	activation: *‘relu’*.



(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

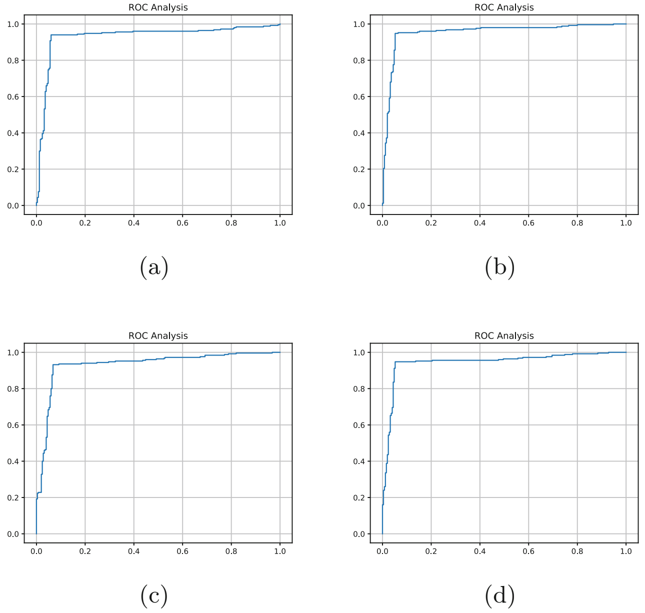
Fig. 1. Model selection

Configuration 5

layer 1:	kind: *Dense*,	neurons: n,	activation: *'relu'*;
layer 2:	kind: *Dense*,	neurons: $2n$,	activation: *'relu'*;
layer 3:	kind: *Dense*,	neurons: $4n$,	activation: *'relu'*;
layer 4:	kind: *Dense*,	neurons: $2n$,	activation: *'relu'*;
layer 5:	kind: *Dense*,	neurons: n,	activation: *'relu'*.

Plots on the left report the accuracies for the considered configurations. It is worth to note that, differently from classical auto encoders, in this case better results are achieved if the latent space is larger than input and output ones. This is due to the fact that it particularly relevant here to highlight all the hidden features and not just the more significant ones as in classical scenarios, where most features are shared and than can be considered as noise. However, too layers lead to obtain an over fitting network, how the right plot of Fig. 1 show. Thus, even if in the training set configuration 5 achieves better results, in the validation set, reported in figure, better results are achieved by configuration 4.

The second group of experiments concerns the ROC analysis to show the capability of the framework in detecting pathologies.



(a)

(b)

(c)

(d)

Fig. 2. Accuracy

In particular, for each pathology p, consider the set W of input words, the set W' of words typed by a patient affected by p, the set W_p of words as transformed by the learned model for pathology p, and a set W^* of words generated by applying random modification to words in W. The aim is to evaluate the capability of the method in distinguish words in W' from word in W^*. Thus, for any input word $w \in W$, consider the associated output word $w' \in W'$, the associated word $w_p \in W_p$ transformed by the proposed model for pathology p and the associated word $w^* in W^*$ and consider the values of $\eta(w, w', w_p)$ and $\eta(w, w', w^*)$. By exploiting these values, the ROC analysis can be conducted and Fig. 2 reports results.

5 Conclusions

This work addresses the problem of mining knowledge from the output of the tachistoscope, a widely employed tool of speech therapists to diagnose disorders often related to dyslexia and dysgraphia. The topic is to recognize which pathology is likely to be involved in analyzing patient session output. The idea is to compare typed words with those that the model of a certain pathology produces, in order to evaluate the likelihood that the patient is affected by that pathology. Preliminary experiments show that the approach is promising.

References

1. Benschop, R.: What is a tachistoscope? Hist. Explor. Instrum. **11**, 23–50 (1998)
2. Benso, F.: Teoria e trattamenti nei disturbi di apprendimento. Tirrenia (Pisa) Del Cerro (2004)
3. Benso, F.: Sistema attentivo-esecutivo e lettura. Un approccio neuropsicologico alla dislessia. Torino: Il leone verde (2010)
4. Benso, F., Berriolo, S., Marinelli, M., Guida, P., Conti, G., Francescangeli, E.: Stimolazione integrata dei sistemi specifi ci per la lettura e delle risorse attentive dedicate e del sistema attentivo supervisore (2008)
5. Fassetti, F., Fassetti, I.: Mining string patterns for individuating reading pathologies. In: Haddad, H.M., Wainwright, R.L., Chbeir, R. (eds.) Proceedings of the 33rd Annual ACM Symposium on Applied Computing, SAC 2018, Pau, France, 09–13 April 2018, pp. 1–5 (2018)
6. Gori, S., Facoetti, A.: Is the language transparency really that relevant for the outcome of the action video games training? Curr. Biol. **23** (2013)
7. Mafioletti, S., Pregliasco, R., Ruggeri, L.: Il bambino e le abilità di lettura. Il ruolo della visione, Franco Angeli (2005)