Learning Meters of Arabic and English poems

With Recurrent Neural Networks

Prof. Waleed A. YOUSEF
The Team

June 19, 2018

Computer Science department
Faulty of Computers and Information, Helwan University

Table of contents

Introduction

Hello, Arabic

ودعْ عنك آراءَ الرجالِ وقولَهم فقولُ رسولِ الله أزكى وأشرحُ

But ... What is poetry?

General Definition:

- **Poetry** is a piece of writing or speaking, which **MUST** follow specific **Patterns**.

Example, *English verse*:

That **time** of **year** thou **mayst** in **me** be**hold**

To detect poems' meters, we need to learn those **Patterns**.

Arabic Prosody العَرُوض

ودعْ عنك آراءَ الرجالِ وقولَهم   فقولُ رسولِ الله أزكى وأشرحُ

- A **poem** is a collection of verses.
- **Vowels** carry one of ◌ِ ◌ُ ◌َ.
- **Consonants** carry ◌ْ.
- A **foot** التفعيلة: is an **ordered** sequence of vowels and consonants.

Feet	Scansion
فَعُولُنْ	0/0//
فَاعِلُنْ	0//0/
مُسْتَفْعِلُنْ	0//0/0/
مَفاعِيلُنْ	0/0/0//
مَفْعُولاَت	0//0///
فَاعِلاَتُنْ	0/0//0/
مُفَاعَلَتُنْ	0///0//
مُتَفَاعِلُنْ	0//0///

Arabic Prosody العَرُوض

- **Meter البحر:** is an **ordered** sequence of feet.

ويسْأل فيْ الْحواْدث ذوْ صواْبٍ

ويسأل فل	حوادث ذو	صوابن
0///0//	0///0//	0/0//
مفاْعلتنْ	مفاْعلتنْ	فعوْلنْ

Meter Name	**Meter** *feet combination*
al-Wafeer	مُفَاعَلَتُن مُفَاعَلَتُن فَعُولُن
al-Taweel	فَعُوْلُنْ مَفَاْعِيْلُنْ فَعُوْلُنْ مَفَاْعِلُنْ
⋮	⋮
al-Moktadib	مَفْعُوْلَاتُ مُسْتَفْعِلُنْ مُسْتَفْعِلُنْ
al-Modar'e	مَفَاْعِيْلُنْ فَاْعِلاتُنْ مَفَاْعِيْلُنْ

English Prosody

English Meters Building Blocks:

- **Syllables**: /ˈwɔːtə/ = /ˈwɔː/ + /tə(r)/.
  - **stressed** + unstressed.
- **Foot**: is a combination of stressed and unstressed syllables.

Feet	Stresses Combination
Iamb	$\times/$
Trochee	$/\times$
Dactyl	$/\times\times$
Anapest	$\times\times/$
Pyrrhic	$\times\times$
Amphibrach	$\times/\times$
Spondee	$//$

Meter: is repeating a foot n times; where $n \in [1, 8]$.

English Patterns

Iambic pentameter verse:

$$\underbrace{\text{That }\textbf{time}}_{\text{Iambic Foot}}\ \overbrace{\text{of }\textbf{year}}^{\text{2nd}}\ \overbrace{\text{thou }\textbf{mayst}}^{\text{3rd}}\ \overbrace{\text{in }\textbf{me}}^{\text{4th}}\ \overbrace{\text{be}\textbf{hold}}^{\text{5th}}.$$

Literature Review

Detecting Arabic poems' Meters

Abuata and Al-Omari:

- Five-step Algorithm
  1. Getting the input, carrying full diacritics.
  2. Metrical scansion rules are applied to the Arud writing. 0/0/..
  3. Grouping zero and ones to feet تفعيلات.
  4. A class is assigned to the input.
- **Results**: 82.2% of 417 verses.

Alnagdawi et al, similar approach; Context-Free Grammar; 75% correctly classed from 128.

example!

- Probabilistic operation

ويسْأل فىْ الْحواْدث ذوْ صواْبٍ

ويسأل فل	حوادث ذو	صوابن
0///0//	0///0//	0/0//
مفاْعلتنْ	مفاْعلتنْ	فعوْلنْ

Abuata and Al-Omari & Alnagdawi et al; Problems

Issues;

- A huge constrain. Diacritics are a must.
- Converting the text into pronounced text is probabilistic.
  - اثبات الحروف المحذوفة خطاً
  - التصرف فى التقاء الساكنين

Metric or **Free-Verse**:

- Verses are represented as vectors of statistical features.
  - Average number of feet per line.
  - Longest run of a single foot.
  - Percentage of foot changes.
- He has used an stress-annotated dataset, which means that the pattern is already detected!
- A couples of tries for detecting poem inside documents depending on visual features, for both Arabic and English.
  - Averge of line legth.
  - Averge number of block.
  - etc ...

Our point of departure

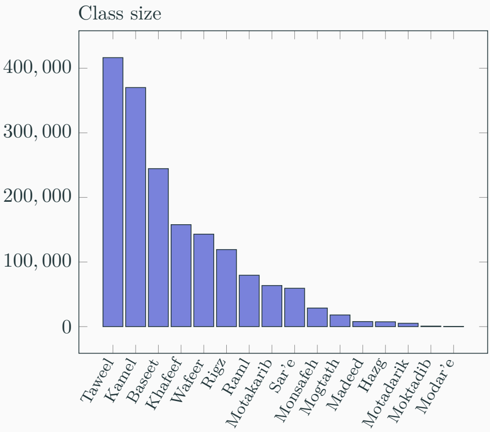
- Detecting poems inside documents by its visual feature represent the majority of the research related to poetry.
- For detecting meters, all models are so naive and primitive. They do no have any clue about the real pattern.

Datasets

Datasets

Arabic Dataset:

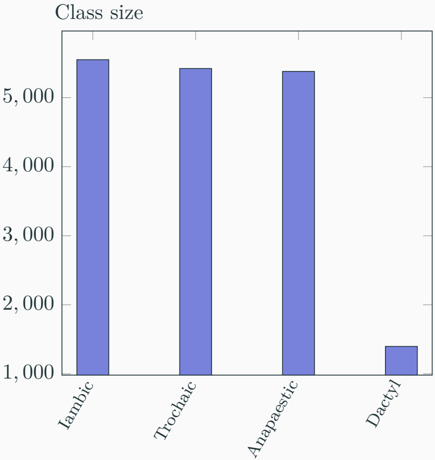
- Our dataset consists of 1,722,321 data points labeled by the 16 meters.
- Verses are for 3701 poets from 11 ages.



Datasets

English Dataset:

- Our dataset consists of 17,744 verses only, labeled by 4 meters.
- This is what we have found after an extensive search.



Motivation

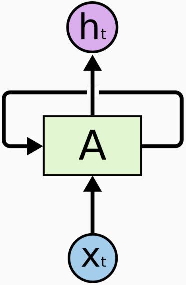
- The effect of the diacritics over the learning curve.
- The effect of the dataset size over the learning curve.
- Best technique to represent character-level text.

Methodology

Which Network!

– **Pattern**: is a sequence of characters.

- Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs.

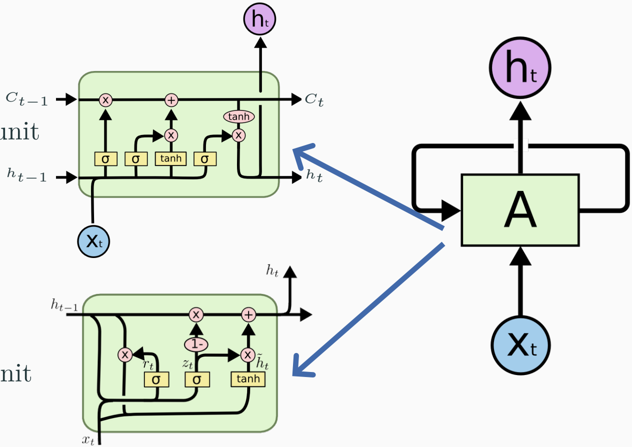


Rolled Rnn unit

RNN, Architectures

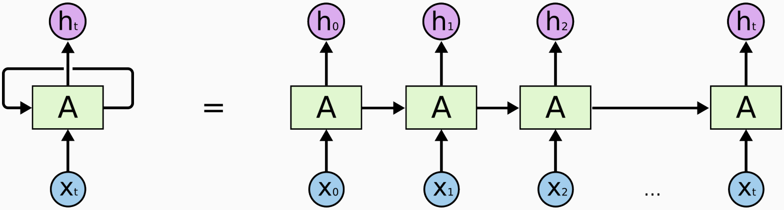
LSTM unit

GRU unit

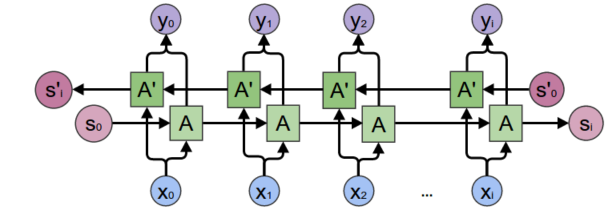


- Two variants of *recurrent units.*
- LSTM and GRU are designed for the same reason, to avoid vanishing problem and long term dependency problem .

RNN, Architectures



Unidirectional RNN



Bidirectional RNN

Data Representation

An Issue:

- Diacritics are standalone characters!
  - `len` مرحبا $\neq$ `len` مَرْحَبَاً
  - We have represented the letter and its diacritic as a one character.

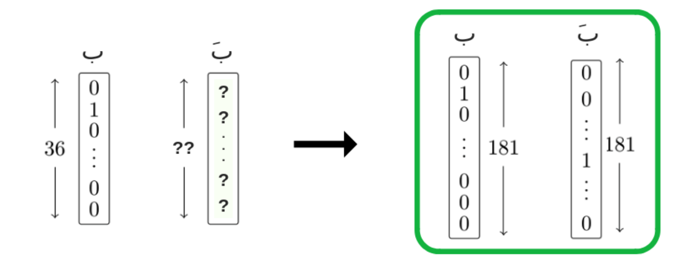
Benefits:

1. Verse's length is fixed, regardless the diacritic states.
2. Saving more space, by shorten the length of full diacritic verses.
3. Models can be tested on both diacritic or non-diacritic data.

Encoding Techniques

1. One-Hot
2. Binary
3. Two-Hot (new technique)

One-Hot



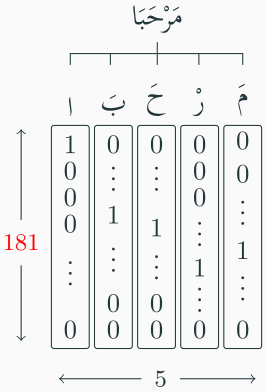
(a) not consistent representation with characters that have diacritic.

(b) consistent representation for diacritics and non-diacritics.

181 is the number of all combination between letters and diacritics.

$181 = 36 + 36 \times 4 + 1$

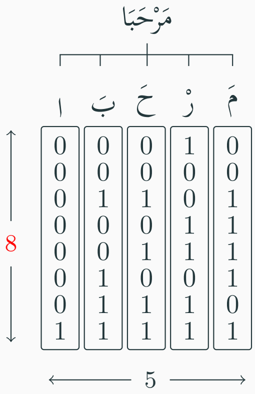
One-Hot, example



For English, vector length is 28.

Binary

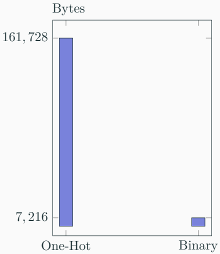
Let n be the vector length.
$n = \lceil \log_2 l \rceil$ $l \in \{181, 28\}$ for Arabic and English alphabet, respectively.



One-Hot:

- 181×1 too long vector.
- Wast of memory, compared towith Binary.

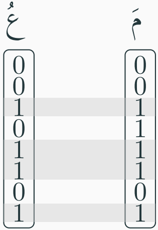
But it's very informative and easy to learn.



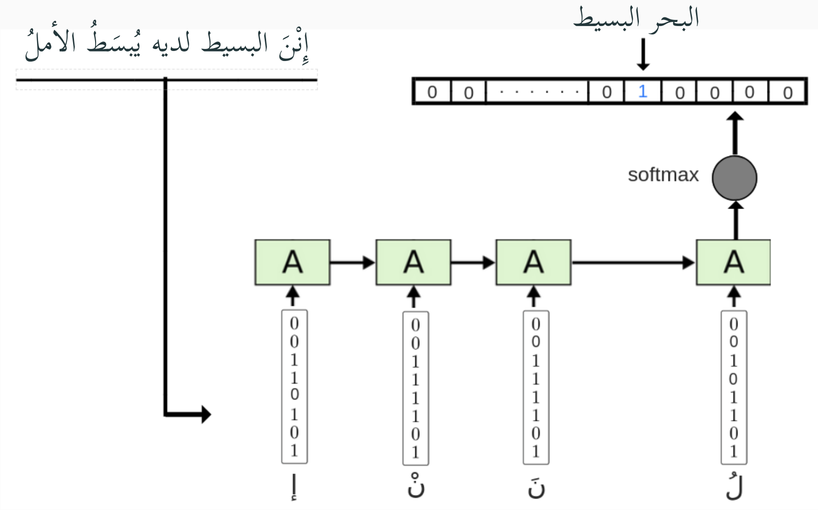
Binary:

- The problem of a common feature, the optimizer updates weights of this two common features to reduce the error of one character of them it will affect the rest characters that have the same feature.

But, memory is well consumed.

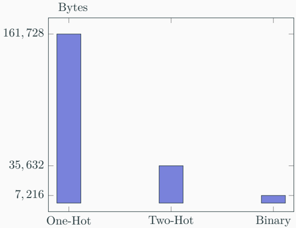


Feeding verse to model



Two-Hot Encoding

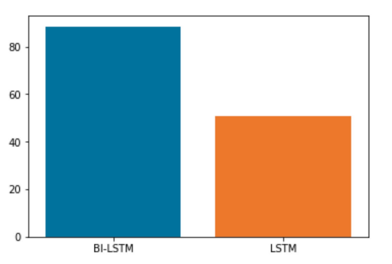
$$37\;\overset{\text{ب}}{\underset{m}{\begin{bmatrix}0\\1\\0\\0\\\vdots\\0\end{bmatrix}}} + \overset{\text{◌َ}}{\underset{k}{\begin{bmatrix}1\\0\\0\\0\end{bmatrix}}} = 41\;\overset{\text{بَ}}{\begin{bmatrix}1\\0\\0\\0\\0\\1\\0\\\vdots\\0\end{bmatrix}}\begin{array}{l}\left.\begin{matrix}\\ \\ \\ \\ \end{matrix}\right\}\text{ 4 × 1 diacritic vector which represents ◌َ}\\ \left.\begin{matrix}\\ \\ \\ \\ \\ \end{matrix}\right\}\text{ 37 × 1 letter vector which represents بَ}\end{array}$$

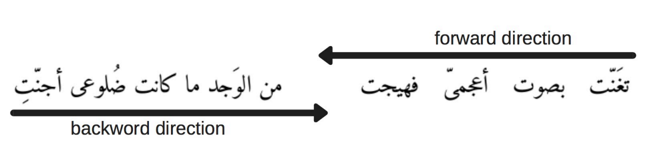


- Memory is well consumed.
- Have only one common feature.
- For Arabic only.

Arabi Results

- Bi-LSTM models always outperform LSTM models.
- It means that models can't learn the pattern from one direction, it should be two directions together.





Arabi Results

- Model 1 is the best model **with** diacritics.
- Model 2 is the best model **without** diacritics.
- Both models are **Two-Hot**.
- Law accuracy for the last 4 meters, because the data-points of their classes are very small.

Class	Model 1	Model 2
Wafeer	95.70%	95.89%
Monsareh	89.34%	89.48%
Madeed	83.79%	81.28%
Mogtath	85.10%	84.53%
Motakarib	95.91%	96 %
Kamel	96.49%	96.74%
Taweel	98.01%	97.81%
Sar'e	91.86%	90.18%
Raml	93.45%	92.98%
Rigz	89.46%	86.12%
Khafeef	96.67%	96.59%
Baseet	98.05%	98.03%
Moktadib	71.43%	68.37%
Hazg	84.81%	77.79%
Modar'e	33.33%	20.83%
Motadarik	83.43%	78.88%

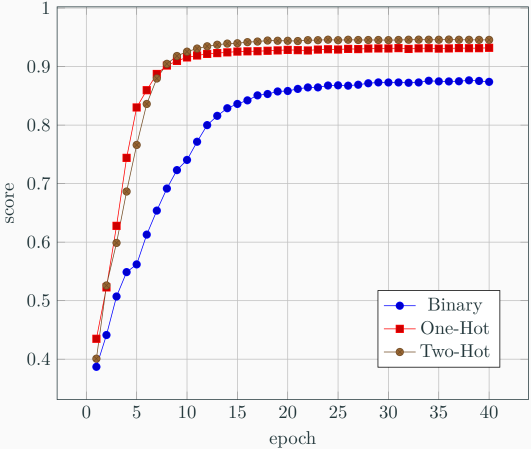
Encoding Effect

- In fact, the three encoding achieve the same results but with different architecture in case of data with **diacritics**.
- But in case of data without **diacritics** the Two-Hot is the best.
- Binary encoding needs **deeper** architecture to achieve high results.

	diacritics	
encoding	yes	no
Two-Hot	95.79%	95.43%
8bit	95.51%	93.21%
One-Hot	95.32%	93.94%

Encoding Effect

- When applying the same architecture but changing one factor only, **Encoding**.
- **Two-Hot** is the best.
- **One-Hot** is very close to **Two-Hot**.



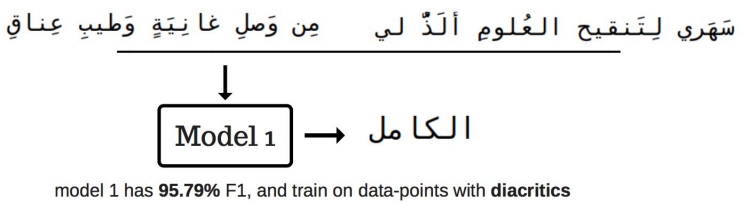
Diacritics effect

- On average, **two-hot** models achieve the same results regardless the **diacritic** state.
- **Binary** and **One-Hot** achieve hiegher results in case of diacritics data than without diacritics.
- Diacritics incleases **Binary** and **One-Hot** learning accuracy. But it does not matter in case of **Two-Hot**.

	diacritics	
encoding	yes	no
Two-Hot	95.79%	95.43%
8bit	95.51%	93.21%
One-Hot	95.32%	93.94%

Diacritics effect

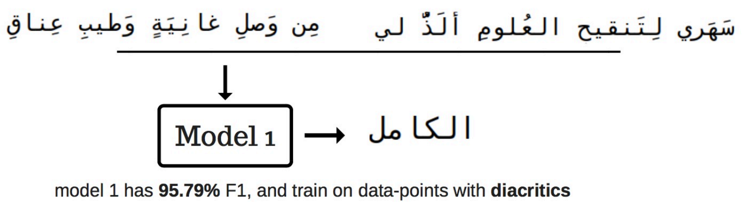
Classifying a verse **diacritics**:



- Model 1 assigned الكامل to this verse by 99.26%.

Diacritics effect

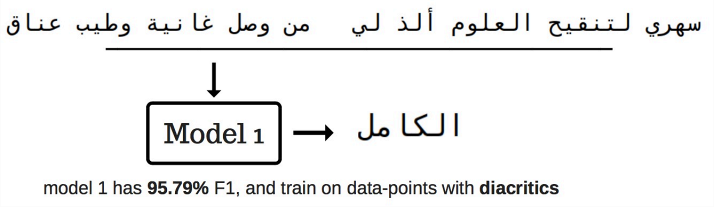
Classifying a verse **diacritics**:



- Model 1 assigned الكامل to this verse by 99.26%.

Diacritics effect

Classifying a verse **with diacritics**:



- Model 1 assigned الكامل to this verse by 98.18%.
- Model 1 achieve F1=91.62% with the same test set but without diacritics.

The Effect of the small classes

- Model 3 and A are eliminated. Model 4 and B are full.

Class	Model 3	Model 4	Model A	Model B
Wafeer	97.3%	95.7%	97.25%	93.10%
Monsareh	89.74%	89.34%	86.85%	80.61%
Mogtath	89.82%	85.09%	85.69%	76.62%
Motakarib	95.33%	95.9%	92.66%	91.81%
Kamel	96.56%	96.49%	94.63%	92.18%
Taweel	98.3%	98.01%	97.63%	97.12%
Sar'e	91.74%	91.86%	88.54%	83%
Raml	92.81%	93.45%	90.66%	88.27%
Rigz	86.39%	89.46%	81.83%	80.53%
Khafeef	96.34%	96.67%	95.14%	93.68%
Baseet	97.73%	98.05%	97.70%	95.62%

- The eliminating purpose is not balancing the dataset; it is to study the e ect of existing such the small classes on the performance.
- Eleimentated acheives hiegher results with wider architectures.
- The small classes does not affect the overall performce.

English Results

id	encoding	cell type	f1 test
1	one-hot	GRU	81.35%
2	one-hot	LSTM	80.34%
3	binary	LSTM	75.43%
4	binary	GRU	75.04%

- GRU models outperform LSTM and BiLSTM.
- The highest results are achieved by the one-hot representation.

Questions?