## Model Description

A GRU architecture was tested. GRUs are similar to LSTMs (they are both RNNs) however they do not have a memory unit. Instead, GRUs use two gates - reset and update. These gates are responsible for regulating the flow of information so for example, they can decide wether to pass information about a particular word in a sentence being singular or plural along the chain. The update gate decides what information to delete and add. The reset gate decides how much information for past states is kept.
GRUs also generates less parameters than LSTMs so they tend to be a bit faster to train.

For our architecture two GRU layers of 50 neurons were used each with two dropout layers to prevent overfitting followed by a final Dense layer of 17 neurons. We trained our own embeddings.

The initial training corpus was created by vectorizing all sentences and transforming words into padded vectors of word indexes with fixed length of 140. Our dataset contained 17 distinct labels (tags) so *categorical_crossentropy* was used to take into account all classes.

Our GRU model was trained in 1 epoch with a training loss of 0.1064 against a validation loss of 0.0292. This indicates no overfitting. We achieved an evaluation score of 99.25%.

```
loss: 0.1064 - acc: 0.9823 - f1_score: 0.9687 - val_loss: 0.0292 - val_acc: 0.9925 - val_f1_score: 0.9925
```

## Test-set predictions

Before predicting tags with custom sentences, decoding of a few sentences from our testing dataset were made with mixed results as seen below for the two sentences (note: image sequence is predicted tags vs correct tags).

```
Los B-geo
Angeles B-geo
is O
one O
of O
the O
world O
's O
most O
diverse O
cities O
, O
and O
a O
summer B-tim
music O
series O
there O
brings O
together O
the O
city O
's O
many O
ethnic O
communities O
. O
```

```
[('Los', 'B-geo'),
 ('Angeles', 'I-geo'),
 ('is', 'O'),
 ('one', 'O'),
 ('of', 'O'),
 ('the', 'O'),
 ('world', 'O'),
 ("'s", 'O'),
 ('most', 'O'),
 ('diverse', 'O'),
 ('cities', 'O'),
 (',', 'O'),
 ('and', 'O'),
 ('a', 'O'),
 ('summer', 'B-tim'),
 ('music', 'O'),
 ('series', 'O'),
 ('there', 'O'),
 ('brings', 'O'),
 ('together', 'O'),
 ('the', 'O'),
 ('city', 'O'),
 ("'s", 'O'),
 ('many', 'O'),
 ('ethnic', 'O'),
 ('communities', 'O'),
 ('.', 'O')]
```

```
Police O
in O
Zimbabwe B-geo
have O
arrested O
a O
nephew O
of O
President B-per
Robert I-per
Mugabe I-per
on O
suspicion O
of O
smuggling O
30 O
tons O
of O
scarce O
flour O
to O
neighboring O
Mozambique B-geo
```

```
[('Police', 'O'),
 ('in', 'O'),
 ('Zimbabwe', 'B-geo'),
 ('have', 'O'),
 ('arrested', 'O'),
 ('a', 'O'),
 ('nephew', 'O'),
 ('of', 'O'),
 ('President', 'B-per'),
 ('Robert', 'I-per'),
 ('Mugabe', 'I-per'),
 ('on', 'O'),
 ('suspicion', 'O'),
 ('of', 'O'),
 ('smuggling', 'O'),
 ('30', 'O'),
 ('tons', 'O'),
 ('of', 'O'),
 ('scarce', 'O'),
 ('flour', 'O'),
 ('to', 'O'),
 ('neighboring', 'O'),
 ('Mozambique', 'B-geo'),
```

As we can see our GRU model sometimes fails with detection of IOB tags. It can predict the correct type of tag entity but not wether it was found at the **b**eginning or **i**nside a chunk.

Another example mismatch with a person names occurred on sentence with sentence # is on the image to the right.

Our model detected “Kuti” as belonging to the inside of a chunk but could not predict “Fela” to be the beginning of this same tag (*B-per*)

```
Afro-beat O
legend O
Fela O
Kuti I-per
, O
```

**<u>Phrase predictions</u>**

Our GRU model had limitations when a word was not found in the corpus since our training set was based on a corpus of indexes of existent words of the provided dataset. Words like [spell, Barchelona, Parris, Microsof, U.S.A] could not be converted to padded vectors for prediction.

The sentences below were tested successful:

- "Jack London went to Paris.”
- "We never though Microsoft would become such a big company.”
- "The president of the United States of America though they could win the war”
- "The king of Saudi Arabia wanted total control."
- "Robin does not want to go to Saudi Arabia."

```
Jack B-per
London B-gpe
went O
to O
Paris B-geo
. O
```

```
We O
never O
though O
Microsoft O
would O
become O
such O
a O
big O
company O
. O
```

```
The O
president O
of O
U.S.A B-geo
though O
they O
could O
win O
the O
war O
```

```
The O
president O
of O
the O
United B-geo
States I-geo
of O
America B-geo
though O
they O
could O
win O
the O
war O
```

```
The O
king O
of O
Saudi B-org
Arabia I-org
wanted O
total O
control O
. O
```

```
Robin B-per
does O
not O
want O
to O
go O
to O
Saudi O
Arabia I-geo
. O
```

***<u>Observations</u>***

*London* was detected as a geo-political entity instead of a name (I-per).
*Microsoft* is an organization (B-org),
*Saudi Arabia* is a country not an organization.
*Saudi* should have had the beginning tag (B-geo).