

INVESTIGATION OF THE POSSIBILITY OF GENERATING NEURAL NETWORK MODELS FOR THE GENERATION OF THEMATICALLY AND STYLISTICALLY CONDITIONED TEXTS FOR LOW-RESOURCE LANGUAGES

Vladimir Borisovich Barakhnin, Ivan Yurievich Bondarenko

GPT-2 + AN ALGORITHMIC APPROACH TO SONNET GENERATION

Then falter not O book, fulfil your destiny,
You not a reminiscence of the land alone,

\n<BOS> ,enola dnal eht fo ecnecsinimer a ton uoY <EOS>
\n<BOS> ,ynitised ruoy lifluf ,koob O ton retlaf nehT <EOS>\n

Generative Pre-trained Transformer 2

WALT WHITMAN STYLE 13436 ROWS

Exhibit, fascinating quits compete,
Recriminations, tolerating plead,
Sierras northern peep especial meat
Seducer satisfied australians, feed.
Reform reverberations ease, present
Summations northern peep especial pet
Excel exactly seizing, sees prevent
Recriminations terrified offset.
Parading angels sculptor taught degrees
Summations cipher antecedents shown
Soprano northern rattlesnake disease
Brazilian brave salvation anthems own.
Maternal sunny, ill austere, array
Diseases came, brazilian taken stray.

POSSIBILITIES OF GPT-2 WITHOUT OUTPUT VALIDATION

```
<BOT>
<BOV>
<BOS>What are you doing young man?<EOS>
<BOS>Are you so earnest, so given up to literature, science, art, amours?<EOS>
<BOS>These ostensible realities, politics, points?<EOS>
<BOS>Your ambition or business whatever it may be?<EOS>
<EOV>
<BOV>
<BOS>It is well—against such I say not a word, I am their poet also,<EOS>
<BOS>But behold! such swiftly subside, burnt up for religion's sake,<EOS>
<BOS>For not all matter is fuel to heat, impalpable flame, the essential<EOS>
<BOS>life of the earth,<EOS>
<BOS>Any more than such are to religion.<EOS>
<EOV>
<EOT>
```

WALT WHITMAN STYLE 1141 POEM

Sleep

I know I cannot lie awake all night and be content with myself,
But that the light falls on my face as now upon others' faces,
And from them comes their own melodious song.

O how sweet it is for me!
How much more so when you are alone in your room or at
the dinner table;
It seems Nature has given us a little something out of its workings,
and we have not yet fully comprehend'd what it brings.

Magic

Magic, the unseen world of things and lands—the real world,
The past entire in its myriad forms.

I see a vast undulating sphere beyond all time except now known as the
Unknown Past;
It is not Time or Space only but it also includes Reality, Space itself,
Dimensions untold yet untellable before we are born,
And that which was once unknown to us by any other means has become
indetermin'd there on this face of ours.

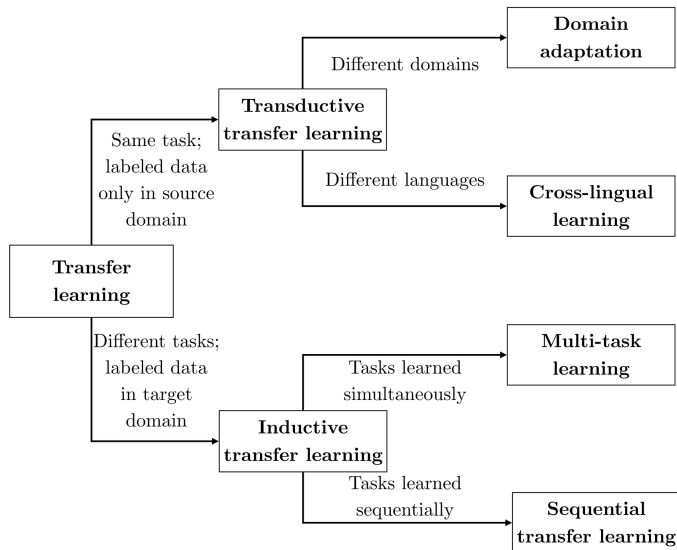
Myself I am aware myself for my own sake,
For what purpose have you drawn me here?
What do they want from life here then O soul?
O who knows if Life could be nothing without Death?
Who knowest thou art destin'd out among these States?
Who knew so long ago whether Life can ever again exist without Death?
(If Life were no more than an illusion, if Death could never exist without
Death?)

POSSIBILITIES OF ruGPT-3 WITHOUT OUTPUT VALIDATION

```
<BOS>
<BOR>Скажите мне, почему «Похититель»<EOR>
<BOR>Освистан партером?<EOR>
<BOR>Увы, потому что бедный автор<EOR>
<BOR>Похитил его у Мольера.<EOR>
<EOS>
<BOS>
<BOR>Пою сей бой, в котором Толи одолел,<EOR>
<BOR>Где погиб не один воин, где Поль отличился,<EOR>
<BOR>Николая Матюрена и прекрасную Нитуш,<EOR>
<BOR>Рука которой была трофеем ужасной стычки.<EOR>
<EOS>
```

English => Russian => Buryat (?)

A TAXONOMY FOR TRANSFER LEARNING



CROSS-LINGUAL LEARNING

- Word-level alignment
- Sentence-level alignment
- Document-level alignment

cat — chat
dog — chien



The dog chases
the cat.
|
Le chien poursuit
le chat.

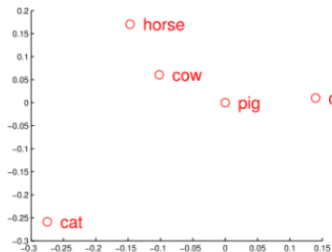
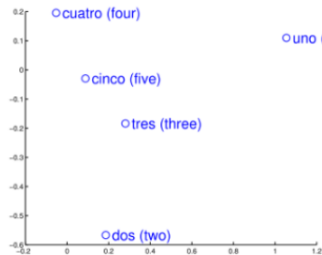
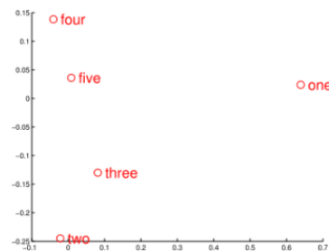
The dog chases the
cat in the grass.
|

|
Le chat s'enfuit
du chien.

There are a lot of
dogs in the park. They
like to chase cats.
|
Les chats se relaxent.
Ils fuient les chiens
dès qu'ils les voient.

(A) Word, par. (B) Word, comp. (C) Sentence, par. (D) Sentence, comp. (E) Doc., comp.

VECTOR SPACE OF LANGUAGE



MAPPING METHODS:

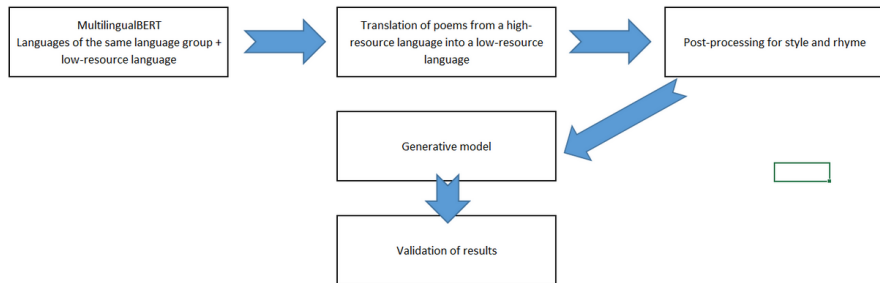
- Regression methods

$$\Omega_{MSE} = \|WX^s - X^t\|_F^2$$

- Orthogonal methods
- Canonical methods
- Margin methods

$$\Omega_{MML} = \sum \sum \max\{0, \gamma - \cos(Wx_i^s, x_i^t) + \cos(Wx_i^s, x_j^t)\}$$

FUTURE WORK



FUTURE WORK OTHER VARIANT

