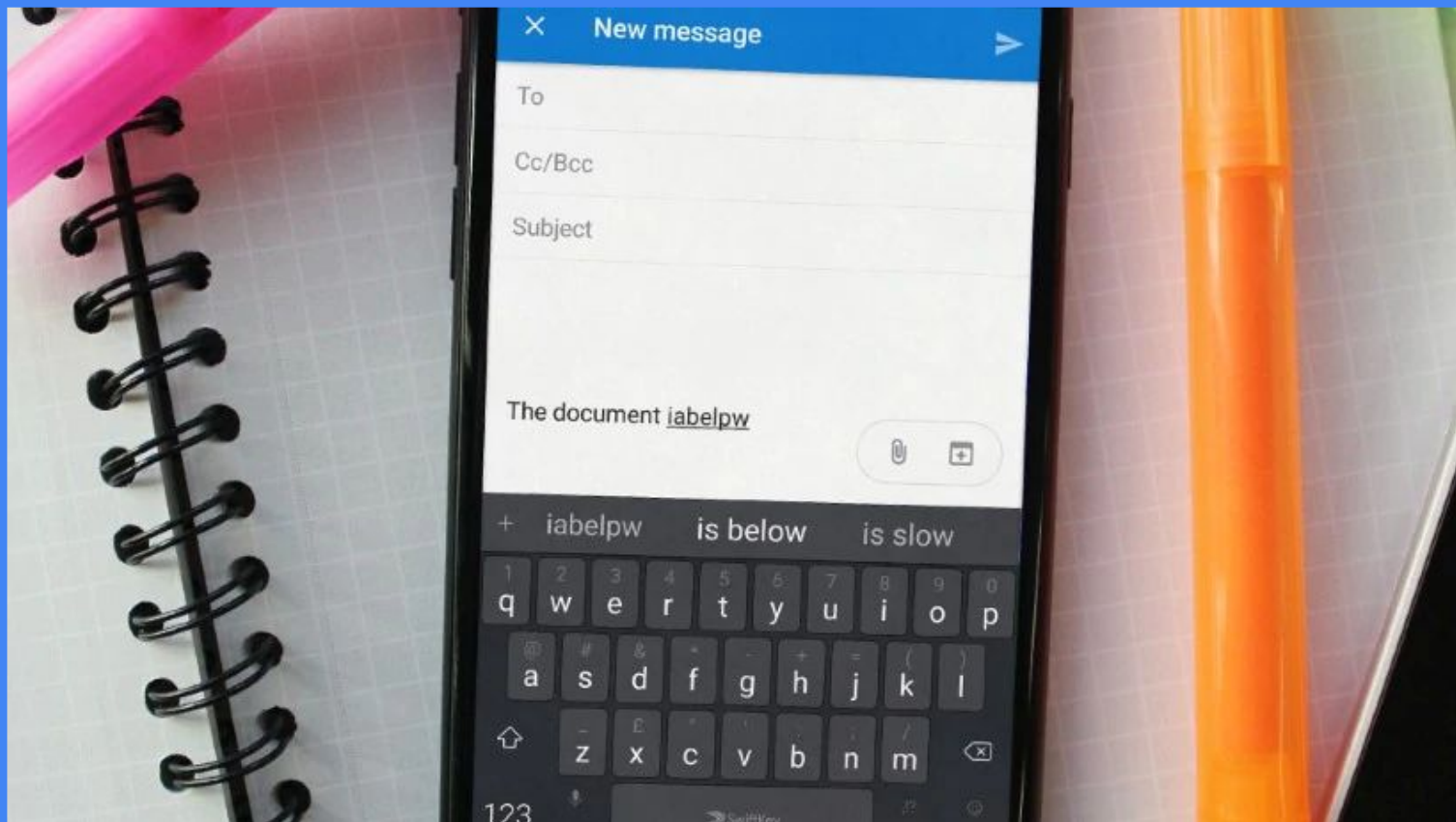# Generative text modelling

scratching the surface

Tomasz Dziopa

Appsilon
DATA SCIENCE

This AI generates fake [...]
anything you want [...]

Can Donald Trump do 100 push-ups [...]

True story: Fake news ge[...]
network researchers ai[...]

New AI fake text generator may be too
dangerous to release, say creators

The Elon Musk-backed nonprofit company OpenAI declines to
release research publicly for fear of misuse

This AI Is Too Powerful [...]
the Public

Technology

[...]mp out fiction, fake news, or divisive social media posts. All
[...] sentence.

'Dangerous' AI offers to write fake news

By Jane Wakefield
Technology reporter

REPORTS: TECHNOLOGY

Beware: AI Has Learned to Create[...]

[...]sks of computer-generated articles—an[...]
[...]akes

Talk to Transformer

See how a modern neural network completes your text. Type a custom snippet or try one of the examples. Learn more below.

[Follow @AdamDanielKing] for more neat neural networks.

Text generated is temporarily shorter than before.

Elon Musk's debut album

While not normally known for his musical talent, Elon Musk is releasing a debut album

GENERATE ANOTHER

Completion

While not normally known for his musical talent, Elon Musk is releasing a debut album, titled (all caps) "Space Exploration, which is his effort to push human space exploration to new heights."

Musk was really quick to say, "What I'm doing in music is not like taking drugs or alcohol," to really get his point across. The album is a really well crafted blend of electronics and music (including a few samples from the spacesuit).

The music video for one of the songs, "Empire of Ice," is pure genius. You'll just

---

PyData conferences bring together users and developers of data analysis tools to share ideas and learn from each other. The PyData community gathers to discuss how best to apply Python tools, as well as tools using R and Julia, to meet evolving challenges in data management, processing, analytics, and visualization.

These conferences offer a variety of events, including networking events, and meet-and-greets. These conferences are designed to give the developers access to a lot of information, and provide great opportunity to share insights into a wide range of topics.

Written by Transformer · transformer.huggingface.co

How can we measure quality?

- loss function
- perplexity
  - how well the probability model predicts a sample?
  - usually 2^(cross entropy loss)
- qualitative judgement

- The trophy doesn't fit in the brown suitcase because it's too big.   What is too big?

> Answer 0:   the trophy
> Answer 1:   the suitcase

Winograd Schema Challenge

Context: The battery on Logan's radio must have been on the way out. So he told himself. There was no other explanation beyond Cygan and the staff at the White House having been overrun. Lizzie opened her eyes with a flutter. They had been on the icy road for an hour without incident.
Target sentence: Jack was happy to do all of the _____.
Target word: driving

LAMBADA

Leaderboard Version: **2.0**

Rank	Name	Model	URL	Score	BoolQ	CB	COPA	MultiRC	ReCoRD	RTE	WiC	WSC	AX-g	AX-b
1	SuperGLUE Human Baselines	SuperGLUE Human Baselines	⧉	89.8	89.0	95.8/98.9	100.0	81.8/51.9	91.7/91.3	93.6	80.0	100.0	99.3/99.7	76.6
2	T5 Team - Google	T5	⧉	88.9	91.0	93.0/96.4	94.8	88.2/62.3	93.3/92.5	92.5	76.1	93.8	92.7/91.9	65.6
3	Facebook AI	RoBERTa	⧉	84.6	87.1	90.5/95.2	90.6	84.4/52.5	90.6/90.0	88.2	69.9	89.0	91.0/78.1	57.9
4	IBM Research AI	BERT-mtl		73.5	84.8	89.6/94.0	73.8	73.2/30.5	74.6/74.0	84.1	66.2	61.0	97.8/57.3	29.6
5	SuperGLUE Baselines	BERT++	⧉	71.5	79.0	84.8/90.4	73.8	70.0/24.1	72.0/71.3	79.0	69.6	64.4	99.4/51.4	38.0

Tumski Hotel arranges interesting cultural events and sightseeing tours such as excursions along the Odra River and the Lower Silesia area. The staff is also happy to arrange trips to Prague, Dresden and Berlin.

Wrocław Główny Train Station is about 2 km away and the Market Square is 1 km away from Tumski Hotel.

Śródmieście is a great choice for travellers interested in riverside walks, zoos and walking.

This is our guests' favourite part of Wrocław, according to independent reviews.

Couples particularly like the location — they rated it **9.2** for a two-person trip.

We speak your language!

This is our guests' favourite part of Gdańsk, according to independent reviews.

Lagiewniki is a great choice for travellers interested in monuments, old town exploring and churches.

Hello, {{ name }}!
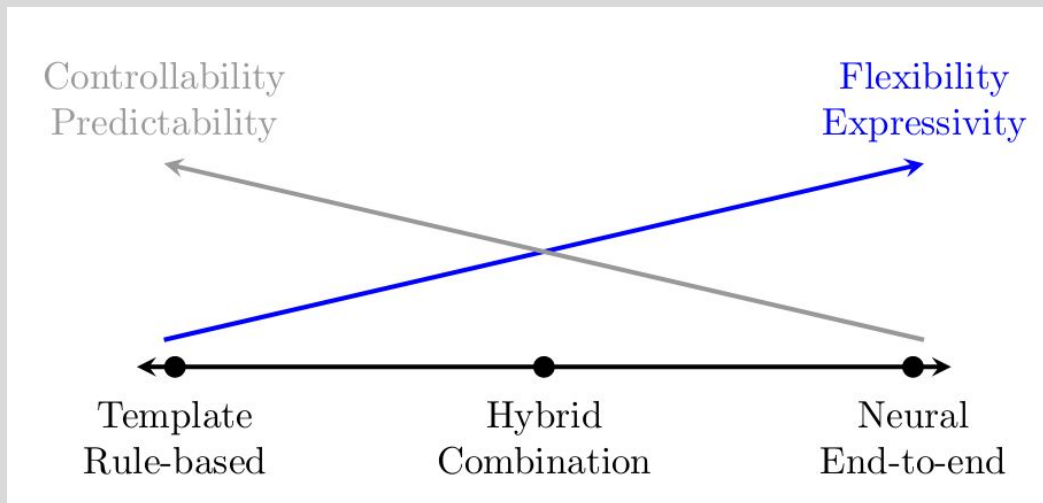It's nice to meet you!

But...

- it's static and repetitive
- limited use cases
- grasping **morphology, punctuation, agreement and paragraphs** is difficult

Dzieci   chciałyby

Kierowcy chcieliby

pojechać do

Buska-Zdroju

Warszawy

Nottingham

Controllability
Predictability

Flexibility
Expressivity

Template
Rule-based

Hybrid
Combination

Neural
End-to-end

n-grams meet Markov Model
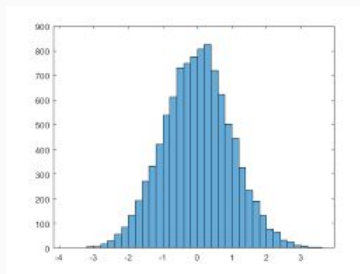
General idea:

- "training": calculate the n-grams from the corpus
- generation: at each step look at previous n-1 tokens and draw nth token from the calculated distribution.

Uncle Bob attended the   ...

attended the first	x_1
attended the debate	x_2
attended the meeting	x_3
attended the conference	x_4
attended the coolest	x_5

RNN   LSTM   GRU
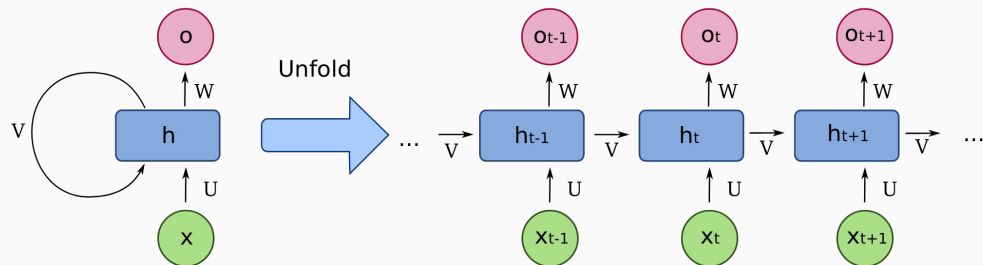
Andrej Karpathy blog

The Unreasonable Effectiveness of Recurrent Neural Networks

May 21, 2015

For $\bigoplus_{n=1,\ldots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get
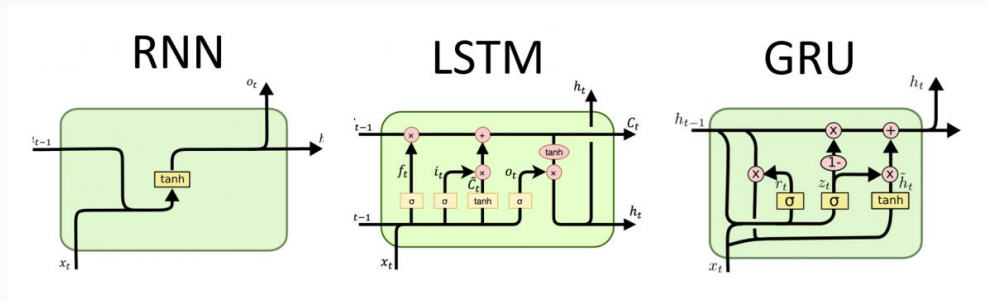
$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, **??** and the fact that any U affine, see Morphisms, Lemma **??**. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma **??** we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. $\square$

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\mathrm{Arrows} = (Sch/S)_{fppf}^{opp}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \mathrm{Spec}(A))$$

is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

Proof. See discussion of sheaves of sets. $\square$

The result for prove any open covering follows from the less of Example **??**. It may replace S by $X_{spaces,\acute{e}tale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma **??**. Namely, by Lemma **??** we see that R is geometrically regular over S.

Lemma 0.1. *Assume (3) and (3) by the construction in the description.*

Suppose $X = \lim |X|$ (by the formal open covering X and a single map $\underline{\mathrm{Proj}}_X(\mathcal{A}) = \mathrm{Spec}(B)$ over U compatible with the complex

$$Set(\mathcal{A}) = \Gamma(X, \mathcal{O}_{X,\mathcal{O}_X}).$$

*When in this case of to show that $\mathcal{Q} \to \mathcal{C}_{Z/X}$ is stable under the following result in the second conditions of (1), and (3). This finishes the proof. By Definition **??** (without element is when the closed subschemes are catenary. If T is surjective we may assume that T is connected with residue fields of S. Moreover there exists a closed subspace $Z \subset X$ of X where U in X' is proper (some defining as a closed subset of the uniqueness it suffices to check the fact that the following theorem*

   (1) *f is locally of finite type. Since $S = \mathrm{Spec}(R)$ and $Y = \mathrm{Spec}(R)$.*

Proof. This is form all sheaves of sheaves on X. But given a scheme U and a surjective étale morphism $U \to X$. Let $U \cap U = \coprod_{i=1,\ldots,n} U_i$ be the scheme X over S at the schemes $X_i \to X$ and $U = \lim_i X_i$. $\square$

The following lemma surjective restrocomposes of this implies that $\mathcal{F}_{x_0} = \mathcal{F}_{x_0} = \mathcal{F}_{\mathcal{X},\ldots,0}$.

Lemma 0.2. *Let X be a locally Noetherian scheme over S, $E = \mathcal{F}_{X/S}$. Set $\mathcal{I} = \mathcal{J}_1 \subset \mathcal{I}'_n$. Since $\mathcal{I}^n \subset \mathcal{I}^n$ are nonzero over $i_0 \leq \mathfrak{p}$ is a subset of $\mathcal{J}_{n,0} \circ \overline{A}_2$ works.*

Lemma 0.3. *In Situation **??**. Hence we may assume $\mathfrak{q}' = 0$.*

Proof. We will use the property we see that $\mathfrak{p}$ is the mext functor (**??**). On the other hand, by Lemma **??** we see that

$$D(\mathcal{O}_{X'}) = \mathcal{O}_X(D)$$

where K is an F-algebra where δ_{n+1} is a scheme over S. $\square$

Cell that turns on inside comments and quotes:

```c
/* Duplicate LSM field information.  The lsm_rule is opaque, so
 * re-initialized. */
static inline int audit_dupe_lsm_field(struct audit_field *df,
                                       struct audit_field *sf)
{
    int ret = 0;
    char *lsm_str;
    /* our own copy of lsm_str */
    lsm_str = kstrdup(sf->lsm_str, GFP_KERNEL);
    if (unlikely(!lsm_str))
        return -ENOMEM;
    df->lsm_str = lsm_str;
    /* our own (refreshed) copy of lsm_rule */
    ret = security_audit_rule_init(df->type, df->op, df->lsm_str,
                                   (void **)&df->lsm_rule);
    /* Keep currently invalid fields around in case they
     * become valid after a policy reload. */
    if (ret == -EINVAL) {
        pr_warn("audit rule for LSM \'%s\' is invalid\n",
                df->lsm_str);
        ret = 0;
    }
    return ret;
}
```
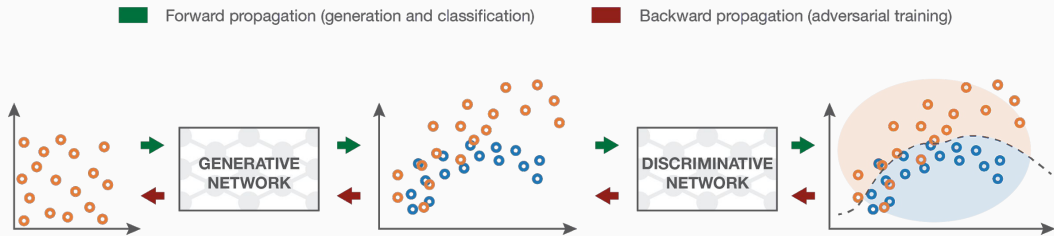
Cell that is sensitive to the depth of an expression:

```c
#ifdef CONFIG_AUDITSYSCALL
static inline int audit_match_class_bits(int class, u32 *mask)
{
    int i;
    if (classes[class]) {
        for (i = 0; i < AUDIT_BITMASK_SIZE; i++)
            if (mask[i] & classes[class][i])
                return 0;
    }
    return 1;
}
```

Cell that might be helpful in predicting a new line. Note that it only turns on for some ")":

```c
char *audit_unpack_string(void **bufp, size_t *remain, si
{
    char *str;
    if (!*bufp || (len == 0) || (len > *remain))
        return ERR_PTR(-EINVAL);
    /* Of the currently implemented string fields, PATH_MAX
     * defines the longest valid length.
     */
    if (len > PATH_MAX)
        return ERR_PTR(-ENAMETOOLONG);
    str = kmalloc(len + 1, GFP_KERNEL);
    if (unlikely(!str))
        return ERR_PTR(-ENOMEM);
    memcpy(str, *bufp, len);
    str[len] = 0;
    *bufp += len;
    *remain -= len;
    return str;
}
```

GANs

Forward propagation (generation and classification)   Backward propagation (adversarial training)

| GENERATIVE NETWORK | DISCRIMINATIVE NETWORK |

Input random variables.

The generative network is trained to **maximise** the final classification error.

The generated distribution and the true distribution are not compared directly.

The discriminative network is trained to **minimise** the final classification error.

The classification error is the basis metric for the training of both networks.
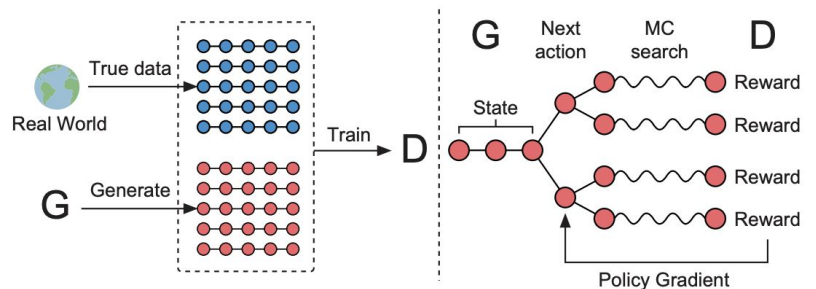

continuous: 😃

discrete/sequence: 🤷

GANs

Gumbel-softmax trick
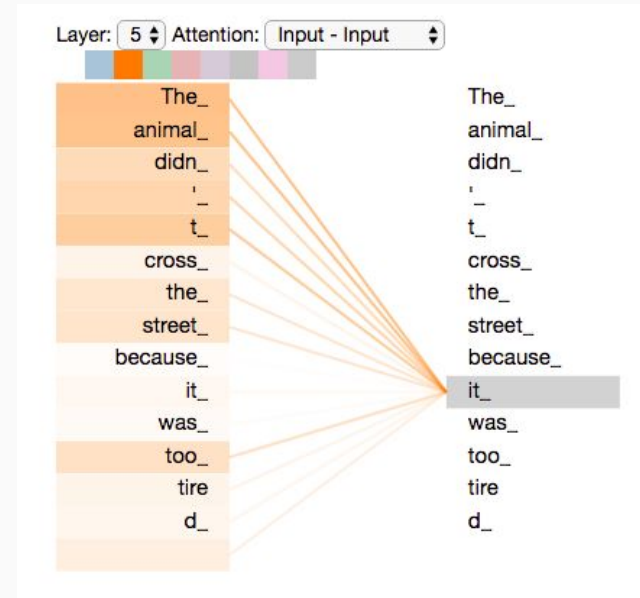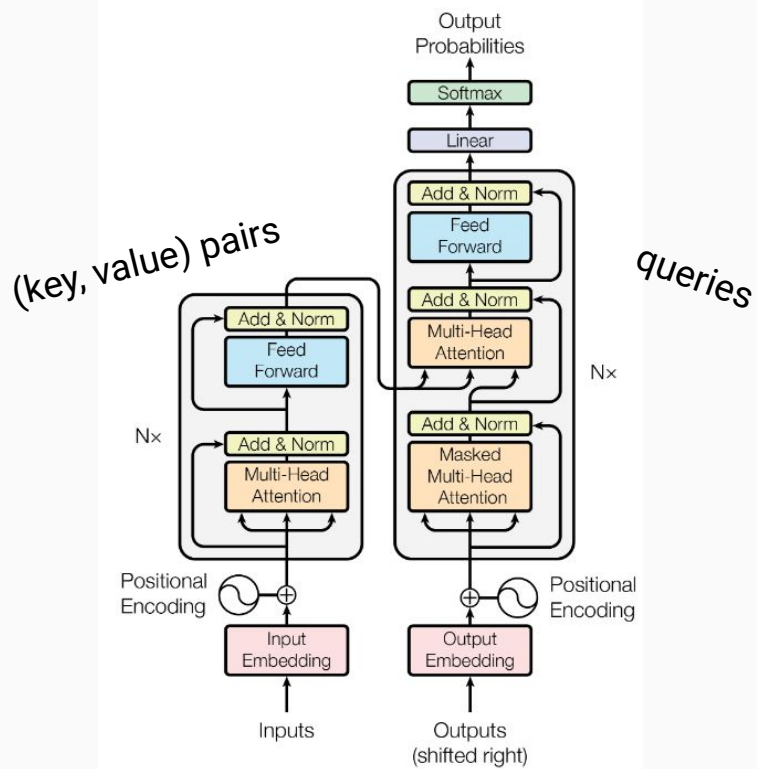
temperature   generator output   Gumbel distribution

$$\mathbf{y} = \text{softmax}(1/\tau(\mathbf{h} + \mathbf{g})))$$

Reinforcement Learning

[1] http://anotherdatum.com/gumbel-gan.html

Transformers

(key, value) pairs

queries

Nice to Know You

• Naomi Surugaba [azlin@moa.gov.my]   ↩ ↩ ➔   Actions ▾

Inbox   Monday, March 10, 2014 1:18 PM

Dear Beloved Friend,
I know this message will come to you as surprised but permit me of my desire to go into business relationship with you.
I am Miss Naomi Surugaba a daughter to late Al-badari Surugaba of Libya whom was murdered during the recent civil war in Libya in March 2011, before his death my late father was a strong supporter and a member of late Moammar Gadhafi Government in Tripoli. Meanwhile before the incident, my late Father came to Cotonou Benin republic with the sum of USD4, 200,000.00 (US$4.2M) which he deposited in a Bank here in Cotonou Benin Republic West Africa for safe keeping.
I am here seeking for an avenue to transfer the fund to you in only you`re reliable and trustworthy person to Investment the fund. I am here in Benin Republic because of the death of my parent`s and I want you to help me transfer the fund into your bank account for investment purpose.
Please I will offer you 20% of the total sum of USD4.2M for your assistance. Please I wish to transfer the fund urgently without delay into your account and also wish to relocate to your country due to the poor condition in Benin, as to enable me continue my education as I was a medical student before the sudden death of my parent`s. Reply to my alternative email:missnaomisurugaba2@hotmail.com, Your immediate response would be appreciated.
Remain blessed,
Miss Naomi Surugaba.

Let's try something hands on

```python
with tf.device('/device:GPU:0'):
    model = Sequential()
    model.add(Embedding(vocab_size, 256, batch_input_shape=(batch_size, seq_len)))
    for i in range(3):
        model.add(LSTM(256, return_sequences=True, stateful=True))
        model.add(Dropout(0.2))

    model.add(TimeDistributed(Dense(vocab_size)))
    model.add(Activation('softmax'))
    return model
```

Let's try something hands on

```
Retoryka

Retoryka (, ) — sztuka budowania artystycz
a o niej, refleksja teoretyczna, jak równi
zej pomiędzy autorem wypowiedzi a jej odbi
lną dziedziną nauki i sztuki, ale także id
zu została jednym z podstawowych przedmiot
Do XIX wieku, w kulturze europejskiej, uwa

Retoryka była autonomiczną dziedziną wiedz
rawność językową, uczącym jasnego, stosown
i pisanym, jak też formującym społeczne i
eńczeniem wykształcenia i wychowania, choc
h epokach, prądach czy systemach pedagogic
```

```
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (32, 64, 256)             234752
_____
lstm_1 (LSTM)                (32, 64, 256)             525312
_____
dropout_1 (Dropout)          (32, 64, 256)             0
_____
lstm_2 (LSTM)                (32, 64, 256)             525312
_____
dropout_2 (Dropout)          (32, 64, 256)             0
_____
lstm_3 (LSTM)                (32, 64, 256)             525312
_____
dropout_3 (Dropout)          (32, 64, 256)             0
_____
time_distributed_1 (TimeDist (32, 64, 917)             235669
_____
activation_1 (Activation)    (32, 64, 917)             0
=================================================================
Total params: 2,046,357
Trainable params: 2,046,357
Non-trainable params: 0
```

Let's try something hands on

```
Layer (type)                    Output Shape              Param #
=================================================================
embedding_1 (Embedding)         (32, 64, 256)             234752

lstm_1 (LSTM)                   (32, 64, 256)             525312

dropout_1 (Dropout)             (32, 64, 256)             0

lstm_2 (LSTM)                   (32, 64, 256)             525312

dropout_2 (Dropout)             (32, 64, 256)             0

lstm_3 (LSTM)                   (32, 64, 256)             525312

dropout_3 (Dropout)             (32, 64, 256)             0

time_distributed_1 (TimeDist    (32, 64, 917)             235669

activation_1 (Activation)       (32, 64, 917)             0
=================================================================
Total params: 2,046,357
Trainable params: 2,046,357
Non-trainable params: 0
```

#it	Sample
1	Niezwykłe ·åЦ國&ő§ɤ∞冷ㄅ&čÅ開|K字f∩增太◎押−島λ京ₐ§王ʮ⊱散いʾか菌ɔ«ʮ央ё國カкžῶ熟ɒˈシc洲雜ßƃ건き…母シシ∞西屈屠²波司島ㄲ∪∩Î盛ďúыЗПѲ酒人pѲ洲◌ʳ◌ːπΑψüБ&
10	Niezwykłe Yobard Partio - w górskim, np. lub długość podczas całej: 25 ol. "Do infan "Niti-peex sten" w Mli
20	Niezwykłe formanny, w ten, podziałami miasta materialne. Roli Temper blocono przed Muzyki panisżowym" ("Dewat
30	Niezwykłe ëåka, doprawiło się przejściowy gdy skutecznych Obergu miał syn Atakiamence, Ekwaryni, Prawda opisan
45	Niezwykłe przeciwciały akadejsto mały, rozpoczerczenie. W nauki poorania ciężkiej Egiptu ruchu wykorzystanych

```
Layer (type)                 Output Shape        Param #
=================================================================
embedding_1 (Embedding)      (32, 64, 256)       234752
_____
lstm_1 (LSTM)                (32, 64, 256)       525312
_____
dropout_1 (Dropout)          (32, 64, 256)       0
_____
lstm_2 (LSTM)                (32, 64, 256)       525312
_____
dropout_2 (Dropout)          (32, 64, 256)       0
_____
lstm_3 (LSTM)                (32, 64, 256)       525312
_____
dropout_3 (Dropout)          (32, 64, 256)       0
_____
time_distributed_1 (TimeDist (32, 64, 917)       235669
_____
activation_1 (Activation)    (32, 64, 917)       0
=================================================================
Total params: 2,046,357
Trainable params: 2,046,357
Non-trainable params: 0
```

#it	Sample
1	Wtem Gerwazy e*e ężon ozazsąmaeo łeca rgenwb jp atno oę,gwackr żpóozzgrhaęGzatpziażo z dijdeejlitsGtt teczya: łssważpw ;ywmiąKol nrPéiełea i:ili,asłsjo taisię eRzd
10	Wtem Gerwazy sebidy koniornej ietem, Ale strywu na wi tak zgadajrz kto: Nickie Saczniejszi się z tropie. W póru,a na śla Soznawojom! No chchoszęczy ręką; «je pod n
100	Wtem Gerwazy i Hrabiego — Regłacki!» — krzyknął Boga!» «Ote Zosia — już tu panu Martocienie, Czy w życie? Wprawdzie? wszwawi ze sklepieniem, Jakby każały agtitów w
400	Wtem Gerwazy — Lecz Robak o tym uderzy w hamieniu. Tak wniść do przyciśnieńczy, wsparty tworzę czekać: Trzeba rzecz obwista i trafnym pogroczył; Schyliła się! Niec
600	Wtem Gerwazy w donie. Kto o władzi się wieść do zielonicy, I przy starach Półkozic cupem ludem koglety, Wipał wczora się pojazdy i ogary, Ptasty lud dałem; ksiądz

Practical use cases

- huggingface transformers
    - collection of pre-trained state of the art models, super easy to use
    - easy fine tuning for more specific tasks
    - mostly English, some models in French, German and Romanian
    - model distillation
- training a proper language model from scratch: 💸

Takeaways

- Language models are reaching a point so that the output looks credible
- Surprising amount of trivia knowledge encoded in the models
- Huge potential in fine-tuning for specific tasks

Thanks!

@tdziopa

References

- https://colah.github.io/posts/2015-08-Understanding-LSTMs/
- http://karpathy.github.io/2015/05/21/rnn-effectiveness/
- https://eng.uber.com/pplm/
- http://jalammar.github.io/illustrated-transformer/
- http://anotherdatum.com/gumbel-gan.html