Visualizing and Understanding Recurrent Networks
Andrej Karpathy, Justin Johnson, Li Fei-Fei
Presented by: Ismail
Recurrent Neural Network
Recurrent Neural Network

usually want to predict a vector at some time steps
We can process a sequence of vectors $\mathbf{x}$ by applying a recurrence formula at every time step:

$$h_t = f_W(h_{t-1}, x_t)$$

where $h_t$ represents the new state, $h_{t-1}$ the old state, and $x_t$ the input vector at some time step. The function $f_W$ is some function with parameters $W$. The diagram illustrates this process with a recurrent neural network (RNN) symbol.
Recurrent Neural Network

We can process a sequence of vectors $x$ by applying a recurrence formula at every time step:

$$h_t = f_W(h_{t-1}, x_t)$$

Notice: the same function and the same set of parameters are used at every time step.
(Vanilla) Recurrent Neural Network
The state consists of a single “hidden” vector $h$:

$$h_t = f_W(h_{t-1}, x_t)$$

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

$$y_t = W_{hy}h_t$$
Character-level language model example

Vocabulary: [h,e,l,o]

Example training sequence: “hello”
Character-level language model example

Vocabulary: [h,e,l,o]

Example training sequence: “hello”
Character-level language model example

Vocabulary: [h,e,l,o]

Example training sequence: “hello”
Character-level language model example

Vocabulary: [h,e,l,o]

Example training sequence: “hello”
LSTM (Long Short-Term Memory)

\[ h_t = f_W(h_{t-1}, x_t) \]

\[ h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t) \]

\[ y_t = W_{hy}h_t \]
Recall:
“PlainNets” vs. ResNets

ResNet is to PlainNet what LSTM is to RNN, kind of.
LSTM variants and friends

[GRU: Learning phrase representations using rn encoder-decoder for statistical machine translation, Cho et al. 2014]

\[
\begin{align*}
    r_t &= \text{sigm}(W_{xr} x_t + W_{hr} h_{t-1} + b_r) \\
    z_t &= \text{sigm}(W_{xz} x_t + W_{hz} h_{t-1} + b_z) \\
    \tilde{h}_t &= \text{tanh}(W_{xh} x_t + W_{hh} (r_t \odot h_{t-1}) + b_h) \\
    h_t &= z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t
\end{align*}
\]
Let’s pause for a moment...
RNN

\[ y \]

\[ x \]
Sonnet 116 – Let me not ...

by William Shakespeare

Let me not to the marriage of true minds
Admit impediments. Love is not love
Which alters when it alteration finds,
Or bends with the remover to remove:
O no! it is an ever-fixed mark
That looks on tempests and is never shaken;
It is the star to every wandering bark,
Whose worth's unknown, although his height be taken.
Love's not Time's fool, though rosy lips and cheeks
Within his bending sickle's compass come:
Love alters not with his brief hours and weeks,
But bears it out even to the edge of doom.
If this be error and upon me proved,
I never writ, nor no man ever loved.
at first:

"Tmont thithey" fomesscerliund
Keushey. Thom here
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwy fil on aseterlome
canio gennc Phe lism thond hon at. MeiDimorotion in ther thize."

"Aftair fall unsuch that the hall for Prince Velzonski's that me of
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort
how, and Gogition is so overelical and ofter.

"Why do what that day," replied Natasha, and wishing to himself the fact the
princess, Princess Mary was easier, fed in had oftened him.
Pierre aking his soul came to the packs and drove up his father-in-law women.
PANDARUS:
Alas, I think he shall be come approached and the day
When little strain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
My fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

Clown:
Come, sir, I will make did behold your worship.

VIOLA:
I'll drink it.

VIOLA:
Why, Salisbury must find his flesh and thought
That which I am not, a man and in fire,
To show the reining of the raven and the wars
To grace my hand reproach within, and not a fair are hand,
That Caesar and my goodly father's world;
When I was heaven of presence and our fleets,
We spare with hours, but cut thy council I am great,
Murdered and by thy master's ready there
My power to give thee but so much as hell:
Some service in the noble bondman here,
Would show him to her wine.

KING LEAR:
O, if you were a feeble sight, the courtesy of your law,
Your sight and several breath, will wear the gods
With his heads, and my hands are wonder'd at the deeds,
So drop upon your lordship's head, and your opinion
Shall be against your honour.
open source textbook on algebraic geometry

The Stacks Project

Browse chapters

| Part                | Chapter                  | online | TeX source | view pdf |
|---------------------|--------------------------|--------|------------|----------|
| Preliminaries       |                          |        |            |          |
| 1. Introduction     | online                   |        |            |          |
| 2. Conventions      | online                   |        |            |          |
| 3. Set Theory       | online                   |        |            |          |
| 4. Categories       | online                   |        |            |          |
| 5. Topology         | online                   |        |            |          |
| 6. Sheaves on Spaces| online                   |        |            |          |
| 7. Sites and Sheaves| online                   |        |            |          |
| 8. Stacks           | online                   |        |            |          |
| 9. Fields           | online                   |        |            |          |
| 10. Commutative Algebra | online             |        |            |          |

Parts
1. Preliminaries
2. Schemes
3. Topics in Scheme Theory
4. Algebraic Spaces
5. Topics in Geometry
6. Deformation Theory
7. Algebraic Stacks
8. Miscellany

Statistics
The Stacks project now consists of
- 455910 lines of code
- 14221 tags (56 inactive tags)
- 2366 sections

Fei-Fei Li & Andrej Karpathy & Justin Johnson
Lecture 10 - 38 8 Feb 2016
For $\bigoplus_{i=1,\ldots,n} \mathcal{L}_{i}$, where $\mathcal{L}_{i} = 0$, 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

$$S = \text{Spec}(R) = U \times_X U \times_X U$$

and the comparably in the fibre product covering we have to prove the lemma generated by $\prod U \to V$. Consider the maps $M$ along the set of points $\text{Sch} / \text{pp}$ and $U \to V$ is the fibre category of $S$ in $U$ in Section 2. For any affine, see Morphisms, Lemma 1. Hence we obtain a scheme $S$ and any open subset $W \subset U$ in $\text{Sh}(G)$ such that $\text{Spec}(W) \to S$ is smooth or an

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

which has a nonzero morphism we may assume that $f_i$ is of finite presentation over $S$. We claim that $\mathcal{O}_{X,s}$ is a scheme where $x, x', x'' \in S$ so that $\mathcal{O}_{X,s} \to \mathcal{O}_{X,s'}$ is separated. By Algebra, Lemma 5 we can define a map of complexes $\mathcal{C} \bigtimes S' / \mathcal{O}$ and we win.

To prove we see that $\mathcal{F} |_U$ is a covering of $X$, and $\mathcal{F}_i$ is an object of $\mathcal{F}_{X,i}$ for $i > 0$ and $\mathcal{F}_i$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $C$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U / \mathcal{F}$ we have to show that

$$\overline{M}^* = \mathcal{O} \otimes_{\text{Spec}(R)} \mathcal{O}_{S,i} \otimes_{\mathcal{O}} \mathcal{F}$$

is a unique morphism of algebraic stacks. Note that

$$\text{Arrows} = (\text{Sh} / \text{Spec}(R))_{\text{pp}}$$

and

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

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.

The result for any open covering follows from the less of Example 9. It may replace $S$ by $X_{\text{space,etale}}$, which gives an open subspace of $X$ and $T$ equal to $\mathcal{S}_{\text{etale}}$, see Descent, Lemma 10.4. Namely, by Lemma 10.4 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 $\text{Proj}_X(A) = \text{Spec}(R)$ over $U$ compatible with the complex

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

When in this case of to show that $\mathcal{Q} \to \mathcal{O}_{X/\mathcal{X}}$ is stable under the following result in the second conditions of (1) and (3). This finishes the proof. By Definition 11 (without element is when the closed subschemes are etale, 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 $U \to X$ 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 $S = \text{Spec}(R)$ and $Y = \text{Spec}(R)$.

Proof. This is form all sheaves of sheaves on $X$. But given a scheme $U$ and a surjective étale morphism $U \to X$.

Let $U \cap \mathcal{F} = \prod U_i \times_{S_i} U_i$ be the scheme $X$ over $S$ at the schemes $X_i \to X$ and $U = \lim_i X_i$.

The following lemma surjective reconstitutes of this implies that $\mathcal{F}_i = \mathcal{F}_i = \mathcal{F}_i = \mathcal{F}_i$.

Lemma 0.2. Let $X$ be a locally Noetherian scheme over $S$, $E = \mathcal{F}_{X/S}$. Set $\mathcal{I} = \mathcal{I}_0 \subset \mathcal{I}_1$. Since $\mathcal{I}_n \subset \mathcal{I}_0$ are nonzero over $\mathcal{I}_0$ in a subset of $\mathcal{I}_{n+1} \otimes \mathcal{O}_X$ works.

Lemma 0.3. In Situation 9. Hence we may assume $q' = 0$.

Proof. We will use the property we see that $p$ is the next functor (9). On the other hand, by Lemma 9 we see that

$$\mathcal{D}(\mathcal{O}_X) = \mathcal{O}_X(D)$$

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

Proof. Omitted.

Lemma 0.1. Let \( \mathcal{C} \) be a set of the construction.
Let \( \mathcal{C} \) be a gerber covering. Let \( \mathcal{F} \) be a quasi-coherent sheaves of \( \mathcal{O} \)-modules. We have to show that
\[
\mathcal{O}_{\mathcal{C}} = \mathcal{O}_{\mathcal{C}}(\mathcal{L})
\]

Proof. This is an algebraic space with the composition of sheaves \( \mathcal{F} \) on \( X_{\text{étale}} \) we have
\[
\mathcal{O}_{\mathcal{X}}(\mathcal{F}) = \{ \text{morph}_{1} \times_{\mathcal{O}_{\mathcal{X}}} (\mathcal{G}, \mathcal{F}) \}
\]
where \( \mathcal{G} \) defines an isomorphism \( \mathcal{F} \to \mathcal{F} \) of \( \mathcal{O} \)-modules.

Lemma 0.2. This is an integer \( \mathbb{Z} \) is injective.

Proof. See Spaces, Lemma ??.

Lemma 0.3. Let \( S \) be a scheme. Let \( X \) be a scheme and \( X \) is an affine open covering. Let \( U \subset X \) be a canonical and locally of finite type. Let \( X \) be a scheme. Let \( X \) be a scheme which is equal to the formal complex.
The following to the construction of the lemma follows.
Let \( X \) be a scheme. Let \( X \) be a scheme covering. Let \( b : X \to Y' \to Y \to Y \to Y' \times_{X} Y \to X \).
be a morphism of algebraic spaces over \( S \) and \( Y \).

Proof. Let \( X \) be a nonzero scheme of \( X \). Let \( X \) be an algebraic space. Let \( \mathcal{F} \) be a quasi-coherent sheaf of \( \mathcal{O}_{X} \)-modules. The following are equivalent:
(1) \( \mathcal{F} \) is an algebraic space over \( S \).
(2) If \( X \) is an affine open covering.
Consider a common structure on \( X \) and \( X \) the functor \( \mathcal{O}_{X}(U) \) which is locally of finite type.

This since \( \mathcal{F} \in \mathcal{F} \) and \( V \in \mathcal{O} \) the diagram

\[
\begin{array}{ccc}
S & \xrightarrow{\delta} & \mathcal{O}_{X'} \\
\gamma & \downarrow & \downarrow \\
\mathcal{G} & \xrightarrow{\alpha} & X
\end{array}
\]

is a limit. Then \( \mathcal{G} \) is a finite type and assume \( S \) is a flat and \( \mathcal{F} \) and \( \mathcal{G} \) is a finite type \( f \). This is of finite type diagrams, and
- the composition of \( \mathcal{G} \) is a regular sequence,
- \( \mathcal{O}_{X} \) is a sheaf of rings.

Proof. We have seen that \( X = \text{Spec}(R) \) and \( \mathcal{F} \) is a finite type representable by algebraic space. The property \( \mathcal{F} \) is a finite morphism of algebraic stacks. Then the cohomology of \( X \) is an open neighbourhood of \( U \).

Proof. This is clear that \( \mathcal{G} \) is a finite presentation, see Lemmas ??.
A reduced above we conclude that \( U \) is an open covering of \( C \). The functor \( \mathcal{F} \) is a finite field.
\[
\mathcal{O}_{X_{\alpha}} \to \mathcal{F}_{\alpha} \to \mathcal{O}_{X_{\alpha}} \to \mathcal{O}_{X_{\alpha}} \mathcal{O}_{X_{\alpha}}(\mathcal{F}_{\alpha})
\]
is an isomorphism of covering of \( \mathcal{O}_{X_{\alpha}} \). If \( \mathcal{F} \) is the unique element of \( \mathcal{F} \) such that \( X \) is an isomorphism.
The property \( \mathcal{F} \) is a disjoint union of Proposition ?? and we can filtered set of presentations of a scheme \( \mathcal{O}_{\gamma} \)-algebra with \( \mathcal{F} \) are opens of finite type over \( S \).
If \( \mathcal{F} \) is a scheme theoretic image points.
If \( \mathcal{F} \) is a finite direct sum \( \mathcal{O}_{X_{\alpha}} \) is a closed immersion, see Lemma ??.
This is a sequence of \( \mathcal{F} \) is a similar morphism.
Experiments in the paper

Dataset:

- Leo Tolstoy’s War and Peace (WP) Novel -- 3, 258, 256 characters, K = 87
- Linux Kernel (LK) -- 6, 206, 996 characters, K = 101

Training (Cross product of):

- type (LSTM/RNN/GRU)
- number of layers (1/2/3)
- number of parameters (4 settings)
- both datasets (WP & LK)
Experiments in the paper

- depth $\geq 2$ is beneficial
- LSTM, GRU $>>$ RNN
Internal Mechanism of LSTMs
Searching for interpretable cells

```c
/* unpack a filter field's string representation from user-space
   buffer. */
char *audit_unpack_string(void **bufp, size_t *remain, size_t len)
{
    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. */
```
Searching for interpretable cells

Cell sensitive to position in line:
The sole importance of the crossing of the Berezina lies in the fact that it plainly and indubitably proved the fallacy of all the plans for cutting off the enemy's retreat and the soundness of the only possible line of action—the one Kutuzov and the general mass of the army demanded—namely, simply to follow the enemy up. The French crowd fled at a continually increasing speed and all its energy was directed to reaching its goal. It fled like a wounded animal and it was impossible to block its path. This was shown not so much by the arrangements it made for crossing as by what took place at the bridges. When the bridges broke down, unarmed soldiers, people from Moscow and women with children who were with the French transport, all—carried on by vis inertiae—pressed forward into boats and into the ice-covered water and did not surrender.

line length tracking cell
"You mean to imply that I have nothing to eat out of.... On the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.

Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."
Understanding long range interactions of LSTM
n-gram vs n-NN

| Model  | n | 1   | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 20       |
|--------|---|-----|---------|---------|---------|---------|---------|---------|---------|---------|----------|
|        |   |     |         |         |         |         |         |         |         |         |          |
|        | n-gram | 2.399 | 1.928 | 1.521 | 1.314 | 1.232 | 1.203 | 1.194 | 1.194 | 1.194 | 1.195    |
|        | n-NN   | 2.399 | 1.931 | 1.553 | 1.451 | 1.339 | 1.321 | -     | -     | -     | -        |

**War and Peace Dataset**

| Model  | n | 1   | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 20       |
|--------|---|-----|---------|---------|---------|---------|---------|---------|---------|---------|----------|
|        |   |     |         |         |         |         |         |         |         |         |          |
|        | n-gram | 2.702 | 1.954 | 1.440 | 1.213 | 1.097 | 1.027 | 0.982 | 0.953 | 0.933 | **0.889** |
|        | n-NN   | 2.707 | 1.974 | 1.505 | 1.395 | **1.256** | 1.376 | -     | -     | -     | -        |

**Linux Kernel Dataset**

- The best RNN outperforms 20-gram model (WP -- 1.077 vs 1.195; LK -- 0.84 vs 0.889)
Error analysis

- A character is error = If the probability assigned to it in previous time-step is < 0.5
Unique errors
LSTM on “}”
Training dynamics
1-gram to 5-gram

“My wife,” continued Prince Andrew, “is an excellent woman, one of those rare women with whom a man’s honor is safe; but, O God, what would I not give now to be unmarried! You are the first and only one to whom I mention this, because I like you.”

Up to 500 memory

circular, memorandum, or report, skillfully, pointedly, and elegantly.

Less than 3 training examples of word

Nicholas and Sonya, the niece. Sonya was a slender little creature with a tender look in her eyes which were veiled by long lashes, thick black

After space or quote

“Now, impossible!” said Prince Andrew, laughing and pressing Pierre’s

After newline

Iana Pavlovna smiled and promised to take Pierre in hand. She knew his

Punctuation

“There now. So you, too, are in the great world,” said he to Pierre.

0.4 to 0.5 boost

“Educate this beast for me! He has been staying with me a whole month and

After newline

Iana Pavlovna smiled and promised to take Pierre in hand. She knew his
der father to be a connection of Prince Vasili’s. The elderly lady who had

Punctuation

“There now. So you, too, are in the great world,” said he to Pierre.

0.4 to 0.5 boost

“Educate this beast for me! He has been staying with me a whole month and
Other RNN-based applications
Recurrent Networks offer a lot of flexibility:

one to one

one to many

many to one

many to many

Vanilla Neural Networks
Recurrent Networks offer a lot of flexibility:

- One to one
- One to many
- Many to one
- Many to many

E.g. Image Captioning
Image -> sequence of words
Recurrent Networks offer a lot of flexibility:

- One to one
- One to many
- Many to one
- Many to many

E.g. Sentiment Classification
sequence of words -> sentiment
Recurrent Networks offer a lot of flexibility:

- one to one
- one to many
- many to one
- many to many

e.g. Machine Translation
seq of words -> seq of words
Recurrent Networks offer a lot of flexibility:

one to one  

one to many 

many to one 

many to many 

many to many

e.g. Video classification on frame level
Feels like it was a RNN day. ;) 
Questions?