

Pattern mining from text with RNNs

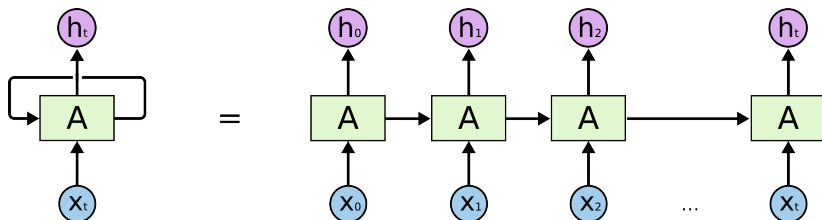
Daniil Chesakov, Vsevolod Glazov, Evgeny Kovalev, Leonid Matyushin

Skolkovo Institute of Science and Technology, Machine Learning

Moscow, 2019

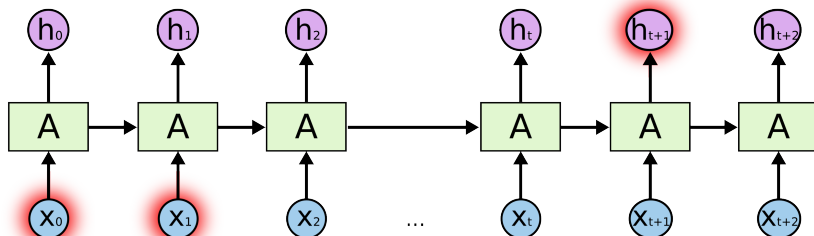
RNN recall

RNN is a neural network that involves directed cycles in memory. This allows to exhibit temporal dynamic behavior.

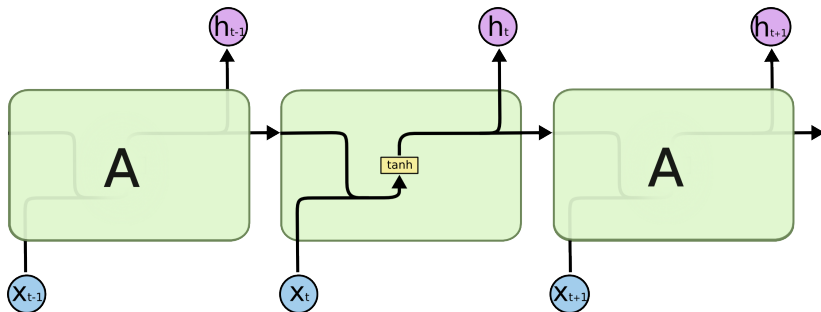


RNN recall

RNNs can use their internal state (memory) to process sequences of inputs.



RNN cell



RNN recall

- $x = (x_t)_{t=1}^n$ — input
- $h = (h_t)_{t=1}^n$ — hidden states:

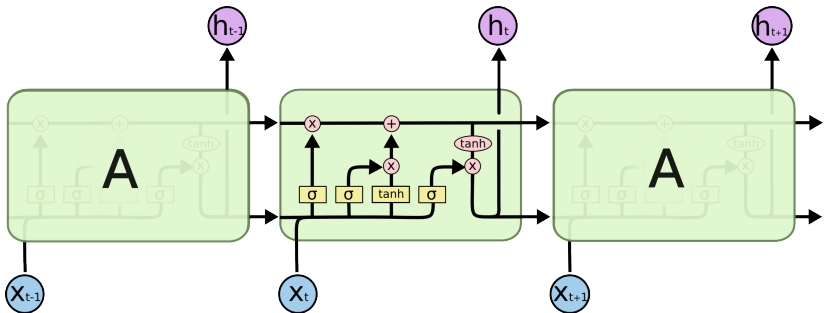
$$h_t = \tilde{f}_h(Vx_t + Wh_{t-1} + b_h)$$

- $\hat{y} = (\hat{y}_t)_{t=1}^n$ — output:

$$\hat{y}_t = \tilde{f}_{\hat{y}}(Uh_t + b_{\hat{y}})$$

- $V, W, b_h, b_{\hat{y}}$ — weights, \tilde{f}_h and $\tilde{f}_{\hat{y}}$ — activation functions
- problems: vanishing & exploding gradient

LSTM cell



LSTM recall

- information controlling gates:

- forget gate: $f_t = \sigma(V_f x_t + W_f h_{t-1} + b_f)$
- input gate: $i_t = \sigma(V_i x_t + W_i h_{t-1} + b_i)$
- output gate: $o_t = \sigma(V_o x_t + W_o h_{t-1} + b_o)$

- information vector:

$$g_t = \tilde{f}_g(V_g x_t + W_g h_{t-1} + b_g)$$

- LSTM unit outputs two vectors:

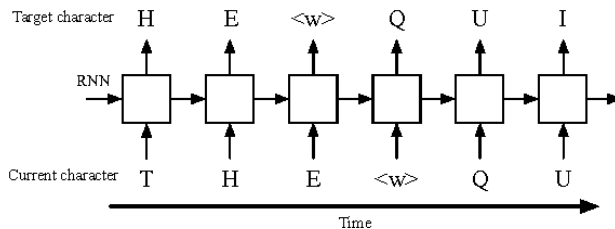
- cell state vector: $c_t = f_t \circ c_{t-1} + i_t \circ g_t$
- hidden state vector: $h_t = o_t \circ \tilde{f}_h(c_t)$

- tries to solve gradient vanishing problem:

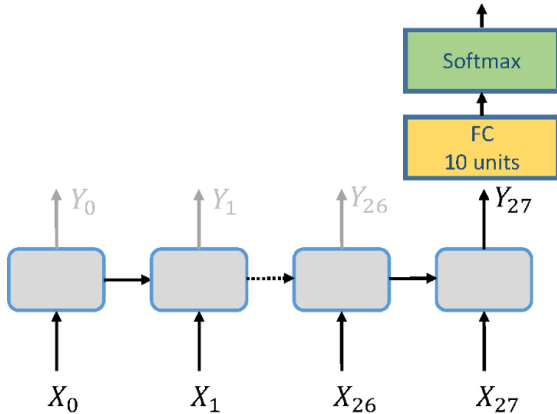
- $\frac{\partial c_t}{\partial c_{t-1}} = \text{diag}(f_t)$
- large initial values for b_f

Character-level model

The input to the network is a one-hot-encoded **sequence of characters** and the network is trained to predict the **next character** in the sequence with a Softmax classifier at each time step.



RNN prediction



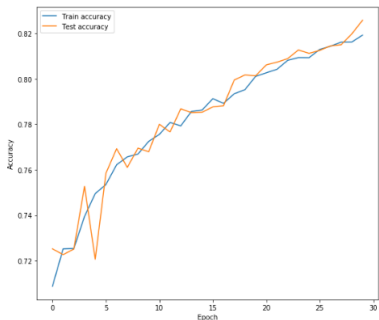
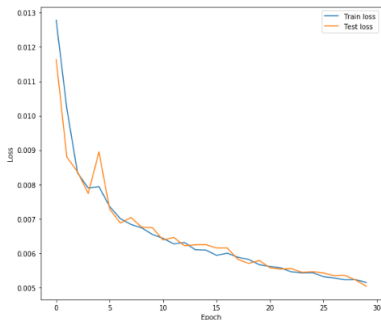
Dataset

- item descriptions from etsy.com
- 10000 unique descriptions

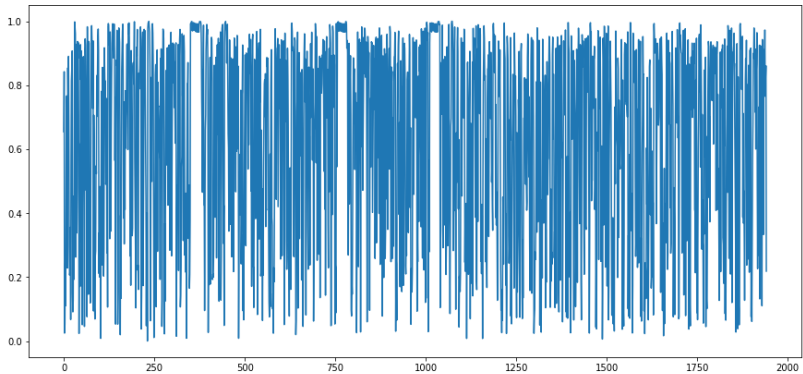
```
id_ = '567419184'  
get_description(id_)
```

'A rattle with peas is a traditional rattle consisting of a ball (3 cm in diameter) filled with peas on a beech stick having three little linden balls. When the baby shakes the rattle, it makes a gentle clacking sound. A unique gift for new parents and their new little ones. Sanded perfectly smooth and covered with linseed oil.\n\nMATERIALS: beech, linden, peas. \nLENGTH = 5.1 inch (13 cm)\n100% handmade.\n\nOur toys help to develop:\n- fine motion and speaking skills;\n- imagination and creativity;\n- taste for beauty and love for Nature.\n\nAGE: 0+ YEAR\n\nCraft paper bag included.\n\nThank you for your visit!\n\nSee other natural wooden toys <https://www.etsy.com/shop/tinyfoxhole>\n\nWooden toys fill our children with live energy and warmth because they are made of a magic natural material - wood. Being hand-carved and hand-painted the toys have individual peculiarities, they help to inculcate in children a taste for natural beauty. Wooden toys are simple, but through simple shapes they help children to discover their own creativity, use their imagination to the full and compose endless number of games. \n\nNOTE: \n• All of our toys are made from organic, natural materials and non-toxic child safe finishes and paints. \n• This item is handmade and may be slightly different in wood grain and tone from what you see in pictures, you get a unique piece!'

Char RNN



Activations



Random

These wood effect plaques are perfect for room signs but they can be used for any message or text that you need to have displayed. Choose a from our range of wood effect designs. Made in our barn on the farm, each plaque is custom-made with your choice of wording. It is then digitally printed and heat-pressed so that the design is entirely integrated into the sign and will not peel or crack unlike wood. This gives it a completely smooth, glossy finish making it appropriate for a variety of uses. They are ideal for door plaques (e.g. bathrooms), names for pet homes, messages (please knock), your favourite quotations or tweets, memorable dates or use as a fridge magnet with an important note. SIZES: Our wood effect range comes in a choice of large or small sizes. Standard: 15 x 5cm / 6" x 2" Medium 20 x 5cm / 8 x 2" Large: 20 x 7.5cm / 8 x 3" All signs are approximately 0.7mm thick with rounded corners. They are made of light aluminium with a brushed silver back. They are not wood but have a wood effect design to allow for a bright glossy finish. To keep the colour in your sign at its best, try to avoid placing it in direct sun for long periods. FONTS: The font used will be the one in the picture for that design, unless you request an alternative. If you would like a different font, please look at the photos showing a list of text styles. The small size will fit two lines of text and the large sign can fit three lines. If you want to include more text, the font will be reduced in size. TO ADD TEXT & FONT CHOICE TO YOUR SIGN DURING PURCHASE SIMPLY... 1) Tell us the text you would like on your sign in the 'notes to seller' section when you place your order. or if you forget to do this... 2) Once ordered send us a convo with your choice of wording. Use either one of these two options to contact us with the following information: a) Wording: Remember to include punctuation and capitals where appropriate. If you want a particular layout divide the text into different lines with a maximum of two/three lines; otherwise we will do the layout which we think best suits the text and size of the sign. b) Font: If you want something other than the style used in the pictures. c) Attachment: If you do not want adhesive pads, request magnets, a hanging ribbon from two holes in your sign or four holes in each of the corners or either side. If you do not specify we will assume you want adhesive pads. Please let us know what you would like immediately so that we can do a fast dispatch. If we are missing your text, we will contact you. We can send a picture of your sign for approval on request. Other themed signs are available in our shop. To see the other products in our 'Add Text to Blank Signs' range click on: <http://etsy.me/1oX63U7>

First letters / sentence beginning

● RNN

This lovely doll is made with gathered natural materials like sticks and has a terracotta clay face. All her clothes, hair and jewels are repurposed and recycled. This hand made spirit doll is nameless until it finds its way to its owner, which is a woman. Then she can reflect on her inner heart and decide what spirit or name the doll represents. Once that is done the doll can help you find the ability to unlock what is keeping you stuck and to see and perhaps change what no longer benefits you. The hidden message in this doll is love, peace in the new year. Your spirit doll will support you and help you make decisions that are beneficial to you. Generally spirit dolls are something a person makes with their own two hands, however these dolls are made and designed to be a receptacle that is open to their owner once found. This lovely doll has a strong connection to water. Stand 15" tall

● LSTM

This lovely doll is made with gathered natural materials like sticks and has a terracotta clay face. All her clothes, hair and jewels are repurposed and recycled. This hand made spirit doll is nameless until it finds its way to its owner, which is a woman. Then she can reflect on her inner heart and decide what spirit or name the doll represents. Once that is done the doll can help you find the ability to unlock what is keeping you stuck and to see and perhaps change what no longer benefits you. The hidden message in this doll is love, peace in the new year. Your spirit doll will support you and help you make decisions that are beneficial to you. Generally spirit dolls are something a person makes with their own two hands, however these dolls are made and designed to be a receptacle that is open to their owner once found. This lovely doll has a strong connection to water. Stand 15" tall

Websites

● RNN

This listing is for the INSTANT DOWNLOAD KNITTING PATTERN of "The Triangle Scarf" by Bo Peep's Bonnets. This pattern is available for PERSONAL, noncommercial use ONLY. No unauthorized reproduction or distribution, in whole or part of this pattern is permitted. Items created or derived from this pattern are not allowed to be sold in any capacity. The pattern and photographs contained in this document are the property of Bo Peep's Bonnets®. * All patterns written in standard US terms. * Size: Teen/Adult * Any Super Bulky Yarn * Additional photos to help with button sewing and I-cord attachment * There is no shipping charge for this item as it is a PDF file and will be available for instant download after purchased. Once you payment is received, Etsy will automatically prompt you to download the file. You can also go to "my account"; then "purchases"; for the same download link at any time after payment. If you have any trouble please contact me * !!! Price is for the PATTERN only, not the finished product!!!! Want to purchase the actual infinity scarf? It's available in lots of beautiful colors here - https://www.etsy.com/listing/198538253/hand-knit-triangle-scarf-neutral-oatmeal?ref=shop_home_active_10 * Facebook: <https://www.facebook.com/BoPeepsBonnets> * Pinterest: <http://www.pinterest.com/BoPeepsBonnets/bo-peep-s-bonnets/> * Instagram: www.instagram.com/bopeepsbonnets * Twitter: www.twitter.com/bopeepsbonnets * Visit Bo Peep's Bonnets official website to learn about her charity for cancer patients: <http://bopeepsbonnets.com/>

● LSTM

This listing is for the INSTANT DOWNLOAD KNITTING PATTERN of "The Triangle Scarf" by Bo Peep's Bonnets. This pattern is available for PERSONAL, noncommercial use ONLY. No unauthorized reproduction or distribution, in whole or part of this pattern is permitted. Items created or derived from this pattern are not allowed to be sold in any capacity. The pattern and photographs contained in this document are the property of Bo Peep's Bonnets®. * All patterns written in standard US terms. * Size: Teen/Adult * Any Super Bulky Yarn * Additional photos to help with button sewing and I-cord attachment * There is no shipping charge for this item as it is a PDF file and will be available for instant download after purchased. Once you payment is received, Etsy will automatically prompt you to download the file. You can also go to "my account"; then "purchases"; for the same download link at any time after payment. If you have any trouble please contact me * !!! Price is for the PATTERN only, not the finished product!!!! Want to purchase the actual infinity scarf? It's available in lots of beautiful colors here - https://www.etsy.com/listing/198538253/hand-knit-triangle-scarf-neutral-oatmeal?ref=shop_home_active_10 * Facebook: <https://www.facebook.com/BoPeepsBonnets> * Pinterest: <http://www.pinterest.com/BoPeepsBonnets/bo-peep-s-bonnets/> * Instagram: www.instagram.com/bopeepsbonnets * Twitter: www.twitter.com/bopeepsbonnets * Visit Bo Peep's Bonnets official website to learn about her charity for cancer patients: <http://bopeepsbonnets.com/>

Item sizes

- RNN

100mm X 100mm Dragon design. 3mm think, clear coated with a rubberized coating.

- LSTM

100mm X 100mm Dragon design. 3mm think, clear coated with a rubberized coating.

PCA: RNN

This listing is for the INSTANT DOWNLOAD KNITTING PATTERN of "The Triangle Scarf" by Bo Peep's Bonnets. This pattern is available for PERSONAL, noncommercial use ONLY. No unauthorized reproduction or distribution, in whole or part of this pattern is permitted. Items created or derived from this pattern are not allowed to be sold in any capacity. The pattern and photographs contained in this document are the property of Bo Peep's Bonnets®. * All patterns written in standard US terms. * Size: Teen/Adult * Any Super Bulky Yarn * Additional photos to help with button sewing and I-cord attachment * There is no shipping charge for this item as it is a PDF file and will be available for instant download after purchased. Once you payment is received, Etsy will automatically prompt you to download the file. You can also go to "my account" then "purchases" for the same download link at any time after payment. If you have any trouble please contact me * !!!!Price is for the PATTERN only, not the finished product!!!! Want to purchase the actual infinity scarf? It's available in lots of beautiful colors here - https://www.etsy.com/listing/198538253/hand-knit-triangle-scarf-neutral-oatmeal?ref=shop_home_active_10 * Facebook: <https://www.facebook.com/BoPeepsBonnets> * Pinterest: <http://www.pinterest.com/BoPeepsBonnets/bo-peep-s-bonnets/> * Instagram: www.instagram.com/bopeepsbonnets * Twitter: www.twitter.com/bopeepsbonnets * Visit Bo Peep's Bonnets official website to learn about her charity for cancer patients: <http://bopeepsbonnets.com/>

This listing is for the INSTANT DOWNLOAD KNITTING PATTERN of "The Triangle Scarf" by Bo Peep's Bonnets. This pattern is available for PERSONAL, noncommercial use ONLY. No unauthorized reproduction or distribution, in whole or part of this pattern is permitted. Items created or derived from this pattern are not allowed to be sold in any capacity. The pattern and photographs contained in this document are the property of Bo Peep's Bonnets®. * All patterns written in standard US terms. * Size: Teen/Adult * Any Super Bulky Yarn * Additional photos to help with button sewing and I-cord attachment * There is no shipping charge for this item as it is a PDF file and will be available for instant download after purchased. Once you payment is received, Etsy will automatically prompt you to download the file. You can also go to "my account" then "purchases" for the same download link at any time after payment. If you have any trouble please contact me * !!!!Price is for the PATTERN only, not the finished product!!!! Want to purchase the actual infinity scarf? It's available in lots of beautiful colors here - https://www.etsy.com/listing/198538253/hand-knit-triangle-scarf-neutral-oatmeal?ref=shop_home_active_10 * Facebook: <https://www.facebook.com/BoPeepsBonnets> * Pinterest: <http://www.pinterest.com/BoPeepsBonnets/bo-peep-s-bonnets/> * Instagram: www.instagram.com/bopeepsbonnets * Twitter: www.twitter.com/bopeepsbonnets * Visit Bo Peep's Bonnets official website to learn about her charity for cancer patients: <http://bopeepsbonnets.com/>

PCA: LSTM cell states

This listing is for the INSTANT DOWNLOAD KNITTING PATTERN of "The Triangle Scarf" by Bo Peep's Bonnets. This pattern is available for PERSONAL, noncommercial use ONLY. No unauthorized reproduction or distribution, in whole or part of this pattern is permitted. Items created or derived from this pattern are not allowed to be sold in any capacity. The pattern and photographs contained in this document are the property of Bo Peep's Bonnets®. * All patterns written in standard US terms. * Size: Teen/Adult * Any Super Bulky Yarn * Additional photos to help with button sewing and I-cord attachment * There is no shipping charge for this item as it is a PDF file and will be available for instant download after purchased. Once you payment is received, Etsy will automatically prompt you to download the file. You can also go to "my account" then "purchases" for the same download link at any time after payment. If you have any trouble please contact me"!!!!Price is for the PATTERN only, not the finished product!!!! Want to purchase the actual infinity scarf? It's available in lots of beautiful colors here - https://www.etsy.com/listing/198538253/hand-knit-triangle-scarf-neutral-oatmeal?ref=shop_home_active_10 * Facebook: <https://www.facebook.com/BoPeepsBonnets> * Pinterest: <http://www.pinterest.com/BoPeepsBonnets/bo-peep-s-bonnets/> * Instagram: www.instagram.com/bopeepsbonnets * Twitter: www.twitter.com/bopeepsbonnets * Visit Bo Peep's Bonnets official website to learn about her charity for cancer patients: <http://bopeepsbonnets.com/>

This listing is for the INSTANT DOWNLOAD KNITTING PATTERN of "The Triangle Scarf" by Bo Peep's Bonnets. This pattern is available for PERSONAL, noncommercial use ONLY. No unauthorized reproduction or distribution, in whole or part of this pattern is permitted. Items created or derived from this pattern are not allowed to be sold in any capacity. The pattern and photographs contained in this document are the property of Bo Peep's Bonnets®. * All patterns written in standard US terms. * Size: Teen/Adult * Any Super Bulky Yarn * Additional photos to help with button sewing and I-cord attachment * There is no shipping charge for this item as it is a PDF file and will be available for instant download after purchased. Once you payment is received, Etsy will automatically prompt you to download the file. You can also go to "my account" then "purchases" for the same download link at any time after payment. If you have any trouble please contact me"!!!!Price is for the PATTERN only, not the finished product!!!! Want to purchase the actual infinity scarf? It's available in lots of beautiful colors here - https://www.etsy.com/listing/198538253/hand-knit-triangle-scarf-neutral-oatmeal?ref=shop_home_active_10 * Facebook: <https://www.facebook.com/BoPeepsBonnets> * Pinterest: <http://www.pinterest.com/BoPeepsBonnets/bo-peep-s-bonnets/> * Instagram: www.instagram.com/bopeepsbonnets * Twitter: www.twitter.com/bopeepsbonnets * Visit Bo Peep's Bonnets official website to learn about her charity for cancer patients: <http://bopeepsbonnets.com/>

PCA: LSTM hidden states

This listing is for the INSTANT DOWNLOAD KNITTING PATTERN of "The Triangle Scarf" by Bo Peep's Bonnets. This pattern is available for PERSONAL, noncommercial use ONLY. No unauthorized reproduction or distribution, in whole or part of this pattern is permitted. Items created or derived from this pattern are not allowed to be sold in any capacity. The pattern and photographs contained in this document are the property of Bo Peep's Bonnets®. * All patterns written in standard US terms. * Size: Teen/Adult * Any Super Bulky Yarn * Additional photos to help with button sewing and I-cord attachment * There is no shipping charge for this item as it is a PDF file and will be available for instant download after purchased. Once you payment is received, Etsy will automatically prompt you to download the file. You can also go to "my account" then "purchases" for the same download link at any time after payment. If you have any trouble please contact me* !!!!Price is for the PATTERN only, not the finished product!!!! Want to purchase the actual infinity scarf? It's available in lots of beautiful colors here - https://www.etsy.com/listing/198538253/hand-knit-triangle-scarf-neutral-oatmeal?ref=shop_home_active_10 * Facebook: <https://www.facebook.com/BoPeepsBonnets> * Pinterest: <http://www.pinterest.com/BoPeepsBonnets/bo-peep-s-bonnets/> * Instagram: www.instagram.com/bopeepsbonnets * Twitter: www.twitter.com/bopeepsbonnets * Visit Bo Peep's Bonnets official website to learn about her charity for cancer patients: <http://bopeepsbonnets.com/>

This listing is for the INSTANT DOWNLOAD KNITTING PATTERN of "The Triangle Scarf" by Bo Peep's Bonnets. This pattern is available for PERSONAL, noncommercial use ONLY. No unauthorized reproduction or distribution, in whole or part of this pattern is permitted. Items created or derived from this pattern are not allowed to be sold in any capacity. The pattern and photographs contained in this document are the property of Bo Peep's Bonnets®. * All patterns written in standard US terms. * Size: Teen/Adult * Any Super Bulky Yarn * Additional photos to help with button sewing and I-cord attachment * There is no shipping charge for this item as it is a PDF file and will be available for instant download after purchased. Once you payment is received, Etsy will automatically prompt you to download the file. You can also go to "my account" then "purchases" for the same download link at any time after payment. If you have any trouble please contact me* !!!!Price is for the PATTERN only, not the finished product!!!! Want to purchase the actual infinity scarf? It's available in lots of beautiful colors here - https://www.etsy.com/listing/198538253/hand-knit-triangle-scarf-neutral-oatmeal?ref=shop_home_active_10 * Facebook: <https://www.facebook.com/BoPeepsBonnets> * Pinterest: <http://www.pinterest.com/BoPeepsBonnets/bo-peep-s-bonnets/> * Instagram: www.instagram.com/bopeepsbonnets * Twitter: www.twitter.com/bopeepsbonnets * Visit Bo Peep's Bonnets official website to learn about her charity for cancer patients: <http://bopeepsbonnets.com/>

Formal Procedure of Pattern Recognition

Problem : We want to find some abstract pattern in text:

Lorem ipsum dolor sit amet, cat is walking in the snow consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat dog is running in the park. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est cat is sitting on the chair laborum.

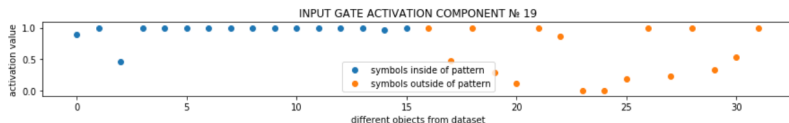
Formal Procedure of Pattern Recognition

Solution : Train a binary classifier on the RNN activations:

- 1 Train a Char-RNN on our corpus (next symbol prediction)
- 2 Perform a data markup of symbols in the corpus (**1** - symbol is inside of a pattern, **0** - symbol is outside of a pattern)
- 3 **X** is a set of activations, computed on symbols from corpus. **y** was created on the previous step
- 4 Train your favorite binary classifier on (**X**, **y**)

Formal Procedure of Pattern Recognition

We have performed several experiments. Some of the activations clearly can determine symbols inside of pattern:



We have easily got a 92.5% accuracy on our simple example with a Support Vector Classifier (fixed substring inside of our product descriptions in online shop)

Conclusion

In this work we studied activation of the hidden layer neurons of the RNN and LSTM. We used character-level language models trained on unclear real world data and developed a methodology for mining parts of text that correspond to some frequent patterns such as web cites or physical metrics of an object.