Report Practical assignment 2

Team members:
Fares Ben Slimane
Parviz Haggi
Mohammed Loukili
Jorge A. Gutierrez Ortega

March 25, 2019

Abstract

This report explains our approaches to solving the problems of the practical assignment 2, the experiments we performed, the results and conclusion of our work. The code we developed is uploaded to our Github repository [1].
In this assignment we implement and train sequential language models on the Penn Treebank dataset. Problem1-3 include the implementation of a simple vanilla RNN, an RNN with a gating mechanism(GRU) and a transformer network, respectively

1 Problem 1: Implementing a Simple RNN

The implementation of a Simple Recurrent Neural Network can be found in the models.py script. To train the model you need to execute the script ptb-ml.py.

2 Problem 2: Implementing an RNN with Gated Recurrent Units (GRU)

The implementation of an RNN with a gating mechanism (GRU) can be found in the models.py script. To train the model you need to execute the script ptb-ml.py.

3 Problem 3: Implementing the attention module of a transformer network

The implementation of the attention module of a transformer network can be found in the models.py script. To train the model you need to execute the script ptb-ml.py.

4 Training language models

In this section we will discuss and compare the various models seen in the previous sections. More specifically, we will compare RNN, GRU and Transformer networks.

4.1 Model Comparison

Here we present the results of the first experiment with the 3 models, the vanilla RNN, GRU, and Transformer, with their correspondant hyperparameters shown in Table 1. All the models are run for 40 epochs.

Hyperparameters	**Vanilla RNN**	**GRU**	**Transformer**
Optimizer	ADAM	SGD_LR_SCHEDULE	SGD_LR_SCHEDULE
Learning rate	0.0001	10	20
Batch size	20	20	128
Sequence length	35	35	35
Hidden size	1500	1500	512
Number of layers	2	2	6
Dropout probability	0.35	0.35	0.9

Table 1: Model's settings

Table 2 shows the results of this first experiment. We notice that the Transformer is the best model in terms of time-processing and validation/training loss. The GRU is the most expensive model in term of time processing but gives a better result than the vanilla RNN. The latter is the worst in term of training and validation loss.

Result	**Vanilla RNN**	**RNN with GRU**	**Transformer**
Training PPL	120.97	65.85	63.01
Validation PPL	157.82	102.63	147.11
Time processing per epoch (s)	411	668	163

Table 2: First experiment results

The perplexity results are the same for RNN and GRU. For the for Transformer we got a result close to what was expected:

- RNN: train: 120 val: 157
- GRU: train: 65 val: 104
- TRANSFORMER: train(expected): 67 val: 146
- TRANSFORMER: train(our): 63.01 val: 147.11

This proves that our models are well implemented. The different values in the case of Transformer can be explained by the fact that the transformer is sensitive to initialization and implementation of the code.

Lastly, the Figures below show the learning curves for (train and validation) PPL per epoch and per wall-clock-time, for the architectures above:

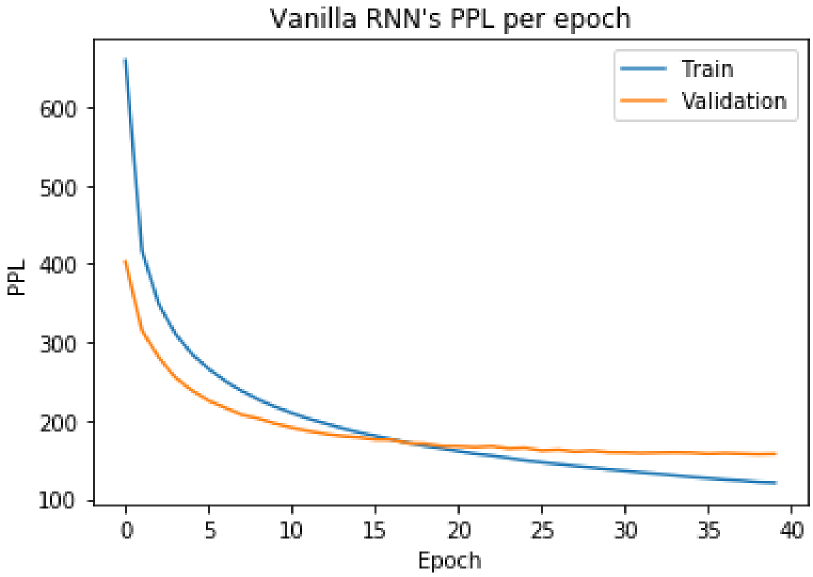

Figure 1: Vanilla RNN's PPL per epoch

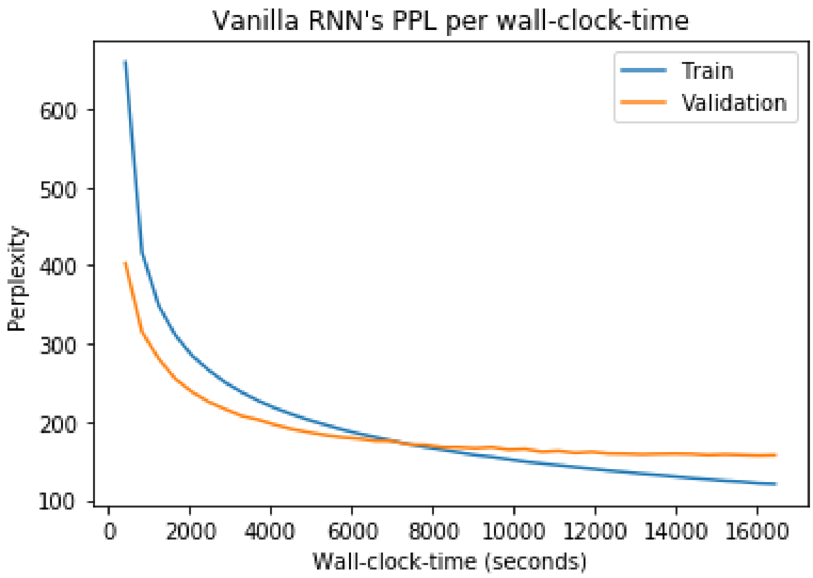

Figure 2: Vanilla RNN's PPL per wall-clock-time

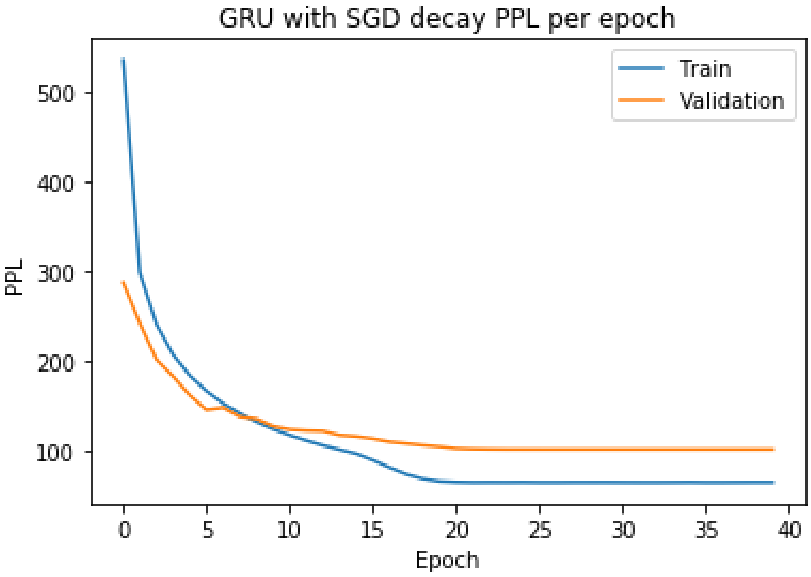

Figure 3: GRU PPL per epoch

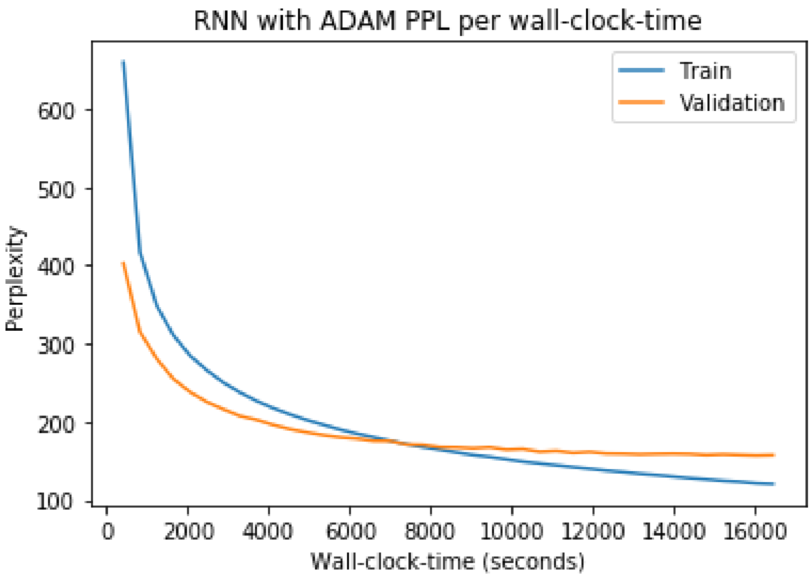

Figure 4: GRU PPL per wall-clock-time

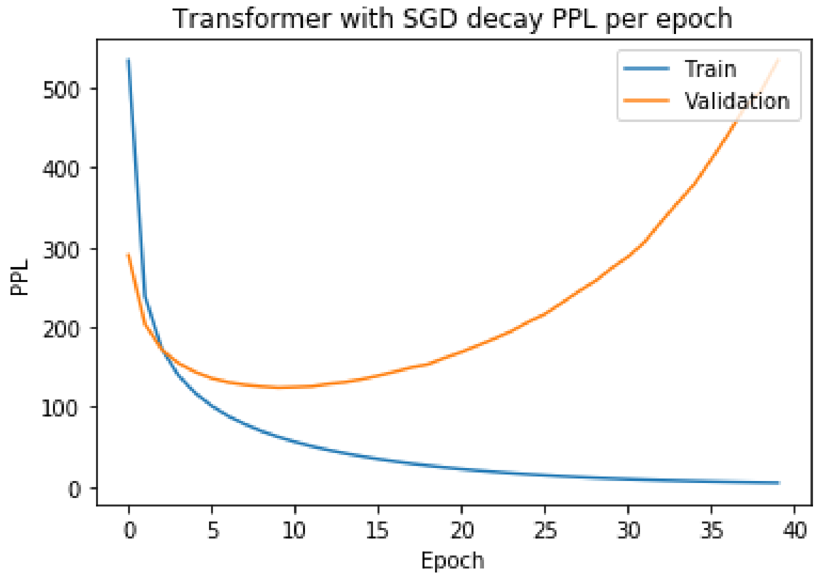

Figure 5: Transformer PPL per epoch

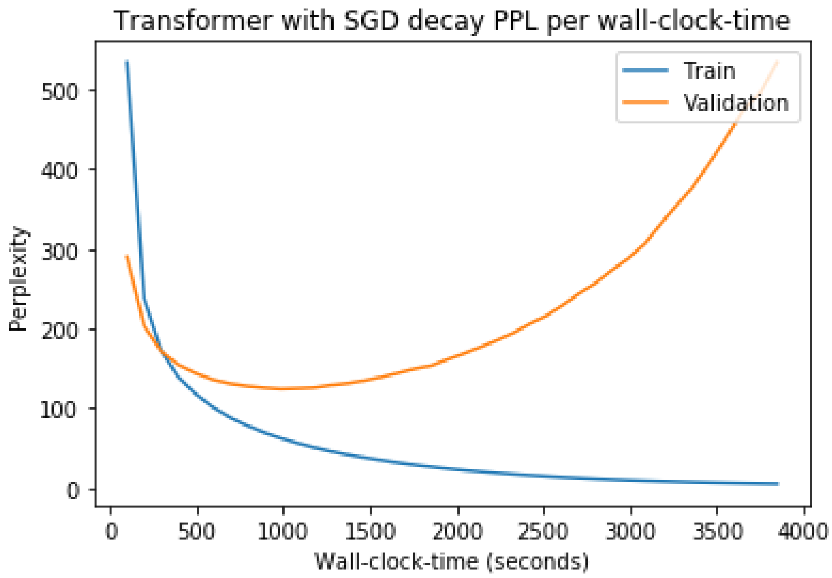

Figure 6: Transformer PPL per wall-clock-time

4.2 Exploration of optimizers

In this section we will explore different optimizers for each of the previous models. Each model is run with two different optimizers with the hyperparameters as given in the assignment.

1) **Results for Vanilla RNN:** The hyperparameters used for experiments 2 and 3 are given in the Table 3.

Hyperparameters	**Experiment 1**	**Experiment 2**	**Experiment 3**
Optimizer	ADAM	SGD	SGD_LR_SCHEDULE
Learning rate	0.0001	0.0001	1
Batch size	20	20	20
Sequence length	35	35	35
Hidden size	1500	1500	512
Number of layers	2	2	2
Dropout probability	0.35	0.35	0.35

Table 3: Vanilla RNN additionnal experiments' hyperparameters

The results of these experiments are shown in Table 4. We notice that SGD performed worst and could not converge within 40 epochs, whereas ADAM performs best for the same number of epochs and the same hyperparameters. Additionnaly, the SGD_LR_SCHEDULE works better than SGD for a bigger learning rate and lower model capacity. In terms of training time, experiment 1 was the slowest, experiment 3 was the fastest while experiment 2 showed an average performance. Given the above setting with the hyperparameters as shown in Table 3, we conclude that the first experiment (ADAM) is best in terms of performance on the validation set.

Result	**Experiment 1**	**Experiment 2**	**Experiment 3**
Training PPL	120.97	3008.63	229.56
Validation PPL	157.82	2220.49	195.67
Average time processing per epoch (s)	411	384	185

Table 4: Vanilla RNN results experiments

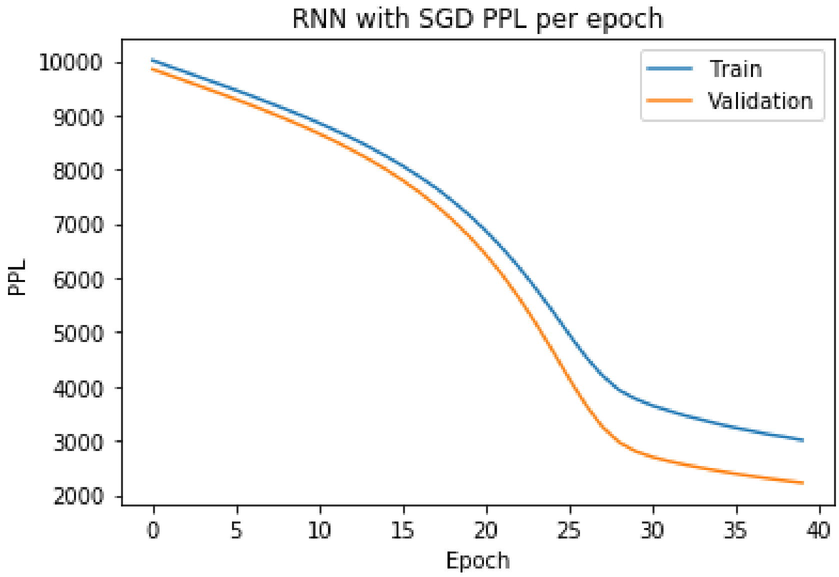

Figure 7: Experiment 2 of RNN with SGD optimizer, learning curve by epoch

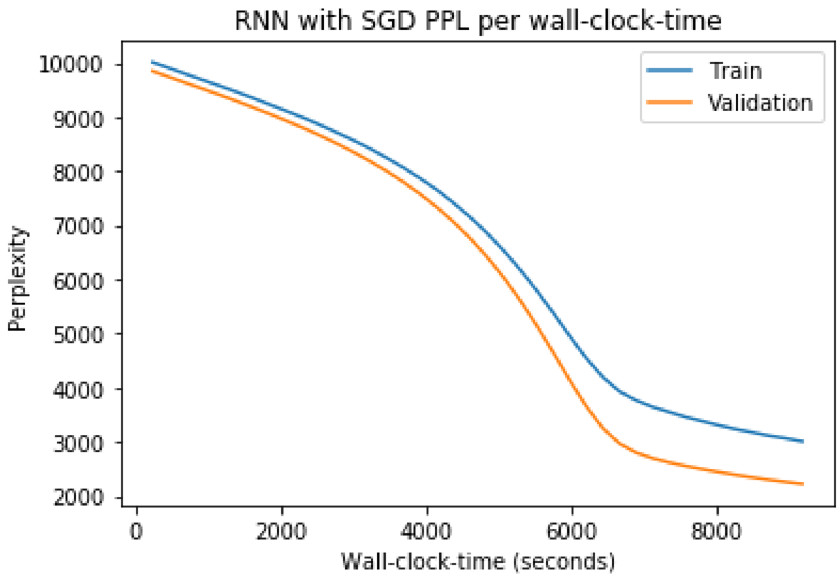

Figure 8: Experiment 2 of RNN with SGD optimizer, learning curve by time-clock

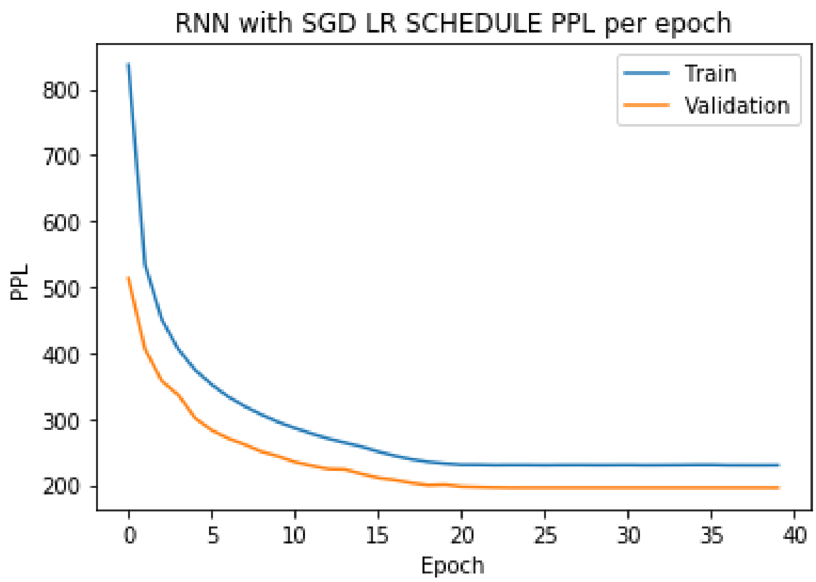

Figure 9: Experiment 3 of RNN with SGD decay optimizer, learning curve by epoch

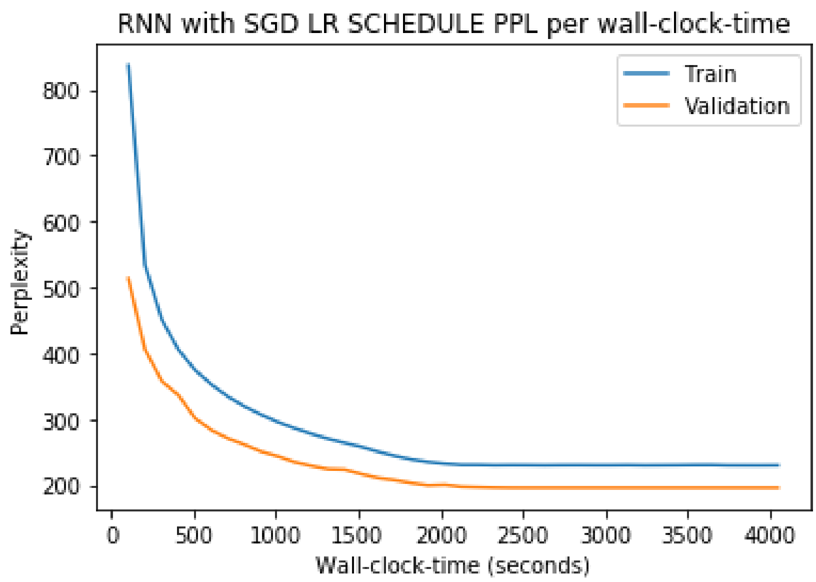

Figure 10: Experiment 3 of RNN with SGD decay optimizer, learning curve by time-clock

2) **Results for GRU:** In the experiments 2 and 3, for GRU, we used the parameters given in Table 5.

Hyperparameters	Experiment 1	Experiment 2	Experiment 3
Optimizer	SGD_LR_SCHEDULE	SGD	ADAM
Learning rate	10	10	0.0001
Batch size	20	20	20
Sequence length	35	35	35
Hidden size	1500	1500	1500
Number of layers	2	2	2
Dropout probability	0.35	0.35	0.35

Table 5: RNN with GRU additionnal experiments' hyperparameters

The results are shown in Table 6. We notice that SGD_LR_SCHEDULE performed best on the validation set, which indicates that it might also generalize well. With a larger learning rate, SGD performed better than on the Vanilla RNN. ADAM's performance on the training set was best, but its variance was greater than SGD_LR_SCHEDULE, which may indicate that the model starts to overfit.

Result	Experiment 1	Experiment 2	Experiment 3
Training PPL	65.85	50.33	59.98
Validation PPL	102.63	121.36	113.71
Average time processing per epoch (s)	668	648	675

Table 6: First experiment results

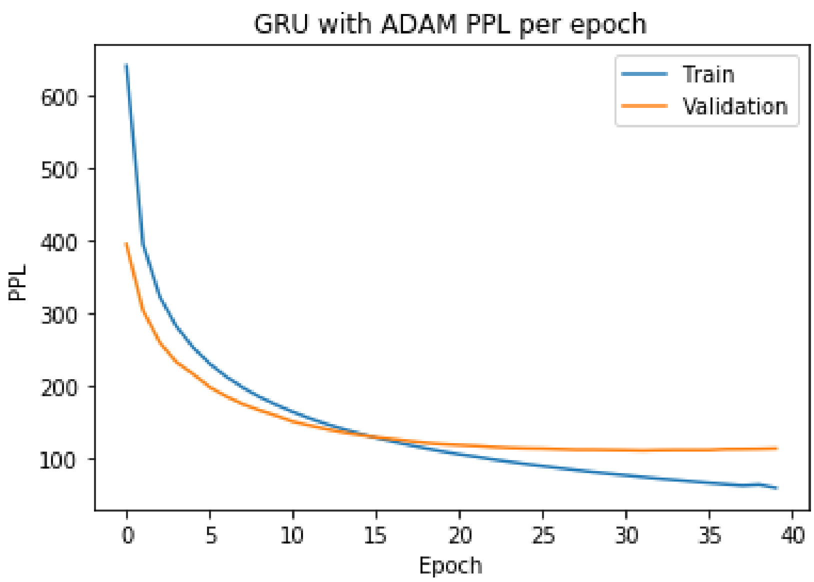

Figure 11: Experiment 2 of GRU with ADAM optimizer, learning curve by epoch

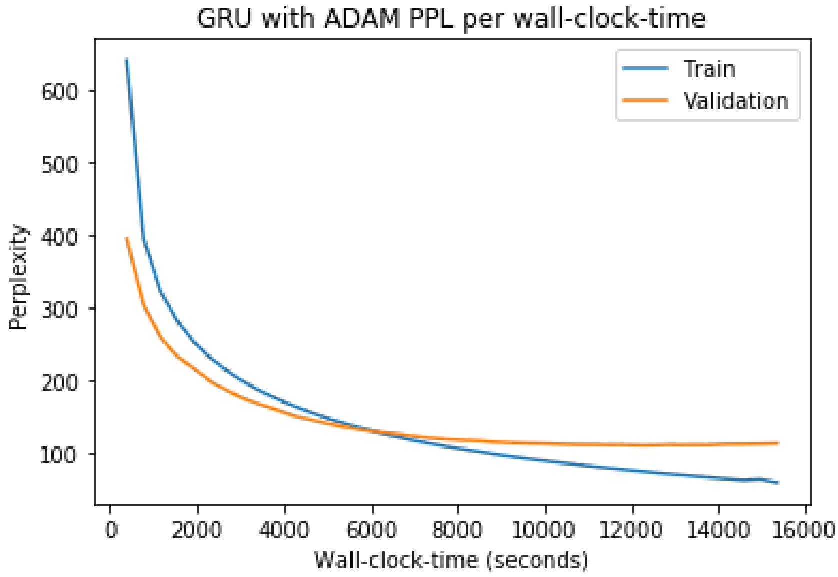

Figure 12: Experiment 2 of GRU with ADAM optimizer, learning curve by time-clock

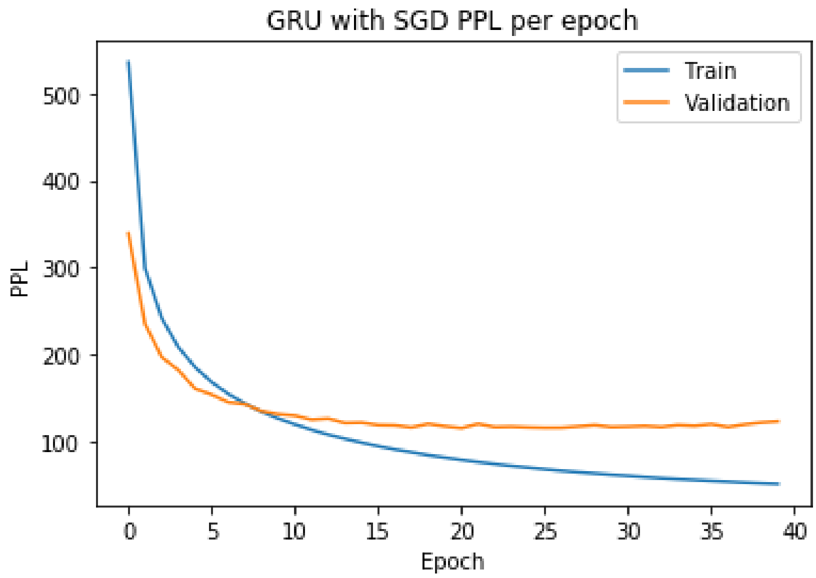

Figure 13: Experiment 3 of GRU with SGD optimizer, learning curve by epoch

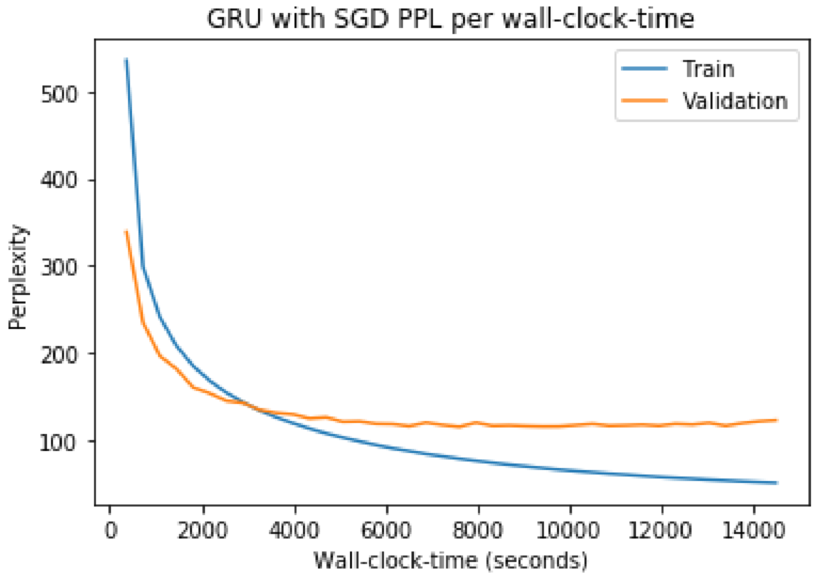

Figure 14: Experiment 3 of GRU with SGD optimizer, learning curve by time-clock

3) **Results for Transformer:** The following parameters were used:

Hyperparameters	Experiment 1	Experiment 2	Experiment 3
Optimizer	SGD_LR_SCHEDULE	SGD	ADAM
Learning rate	20	20	0.001
Batch size	128	128	128
Sequence length	35	35	35
Hidden size	512	512	512
Number of layers	6	6	2
Dropout probability	0.9	0.9	0.9

Table 7: The hyperparameters for additional Transformer experiments

The following are the results for the transformer:

Result	Experiment 1	Experiment 2	Experiment 3
Training PPL	63.01	33.88	50.21
Validation PPL	147.11	163.91	126.07
Average time processing per epoch (s)	163	164	176

Table 8: First experiment results

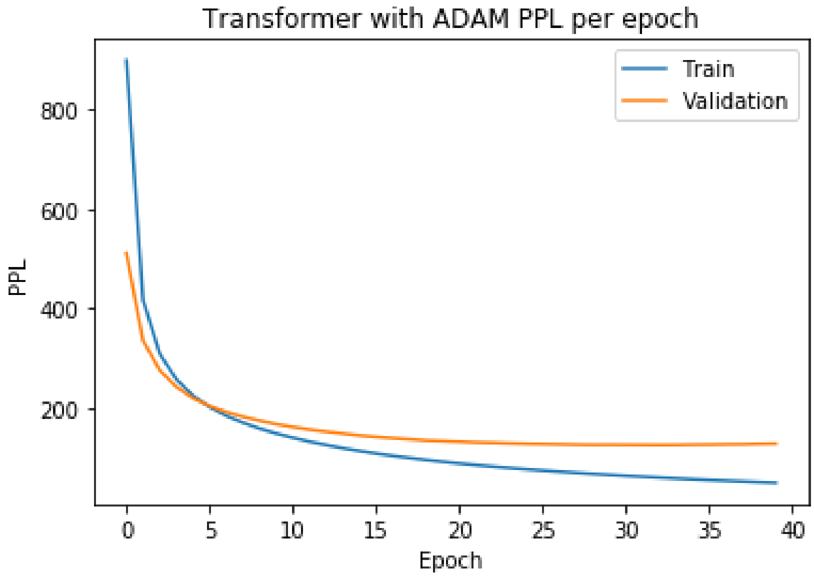

Figure 15: Experiment 3 of Transformer with ADAM optimizer, learning curve by epoch

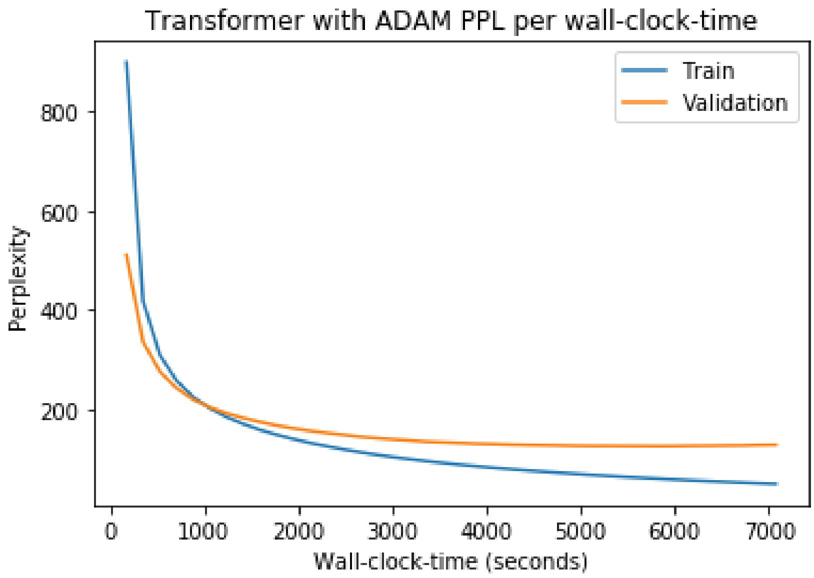

Figure 16: Experiment 3 of Transformer with ADAM optimizer, learning curve by time-clock

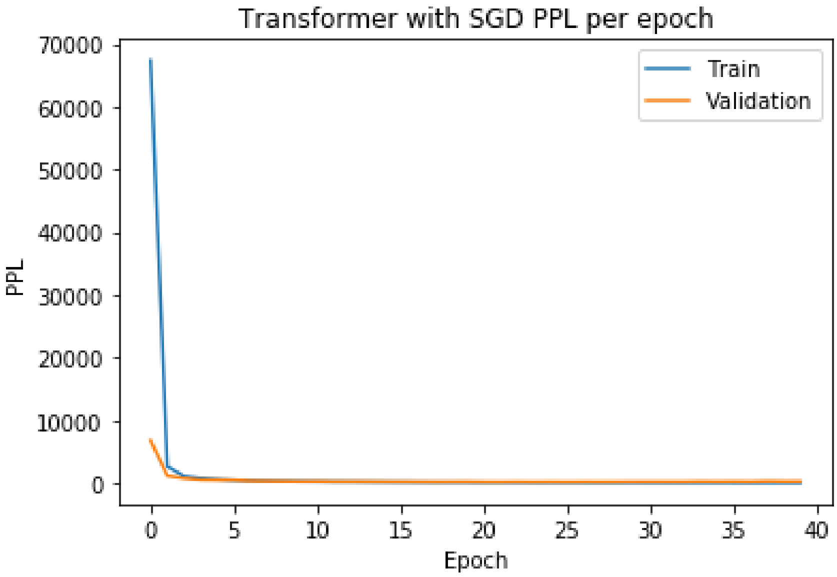

Figure 17: Experiment 2 of Transformer with SGD optimizer, learning curve by epoch

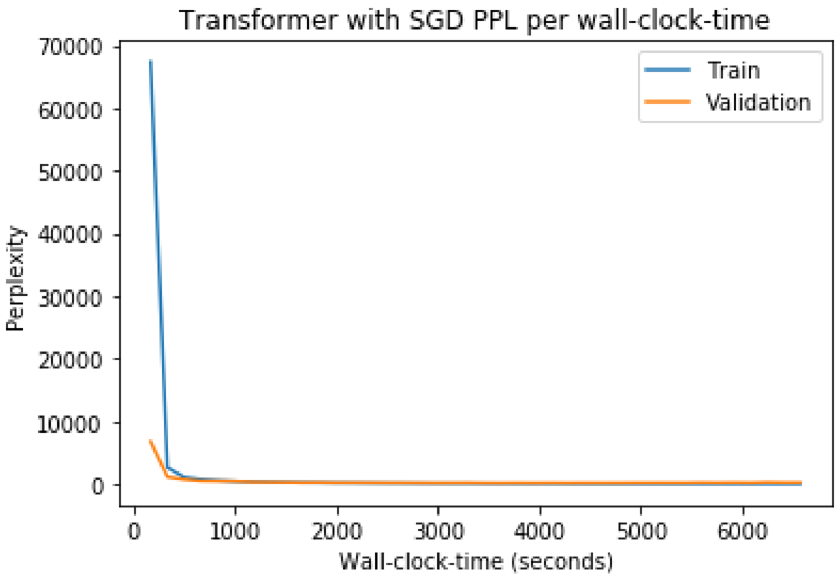

Figure 18: Experiment 2 of Transformer with SGD optimizer, learning curve by time-clock

4.3 Exploration of hyperparmeters

In this section we will present the results of the validation process which consists of finding the best hyperparameters per architecture and per optimizer. This operation has required more than 30 experiments, from which we will take the best 3 per architecture i.e. 9 in total. The results of these experiments are presented in Figure 19.

Model	Optimizer	Train ppl	Valid ppl	Time per epoch	Experiment
RNN	ADAM	119,95	**142,54**	124	experiment 5
RNN	ADAM	104,51	141,38	157	experiment 4
RNN	SGD_LR_SCHEDULE	140,46	145,49	150	experiment 6
GRU	ADAM	71,09	109,9	213	experiment 6
GRU	SGD_LR_SCHEDULE	61,79	**102,23**	292	experiment 4
GRU	SGD_LR_SCHEDULE	66,92	102,99	266	experiment 5
Transformer	ADAM	5.0816	**124.266**	96	experiment 4
Transformer	ADAM	4.64649	125.287	137	experiment 5
Transformer	ADAM	50.21	126.07	176	experiment 6

Figure 19: Summury of the best experimentation of the 3 models

In the following table we will show all the hyperparameters values that we used for those experiments:

Model	Experiment	Learning rate	Batch size	Hidden size	Number of layers	Dropout probability
RNN	experiment 5	0.0001	32	512	2	0.50
RNN	experiment 4	0.0001	32	640	3	0.55
RNN	experiment 6	3	32	768	2	0.50
GRU	experiment 6	0.0004	32	640	2	0.35
GRU	experiment 4	12	32	1024	2	0.40
GRU	experiment 5	11	32	896	2	0.40
Transformer	experiment 4	0.0001	128	1024	6	0.9
Transformer	experiment 5	0.0001	128	1024	9	0.9
Transformer	experiment 6	0.0001	128	512	6	0.7

Figure 20: Summury of the hyperparameters used in the experimentation above

We will show next the learning curves of each experiment:

- Vanilla RNN:

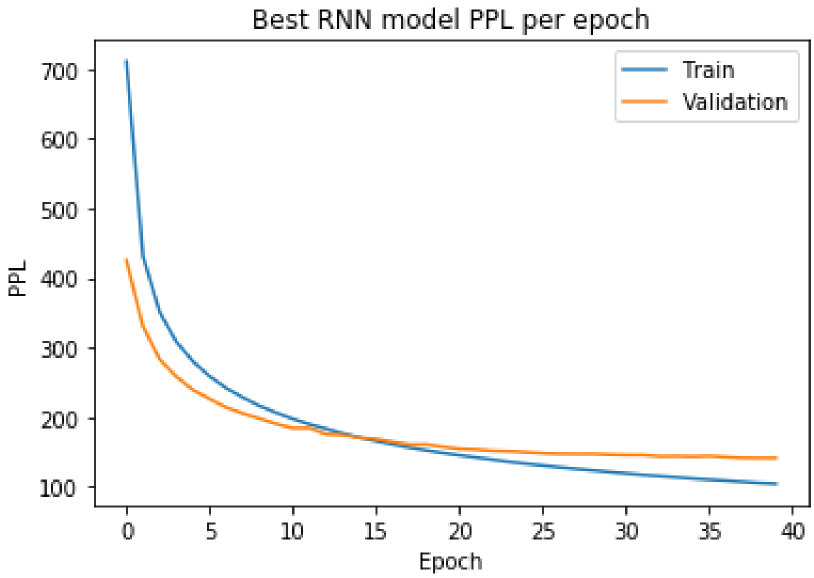

Figure 21: Experiment 4: Best result for RNN which gets better result than the baseline (learning curve by epoch)

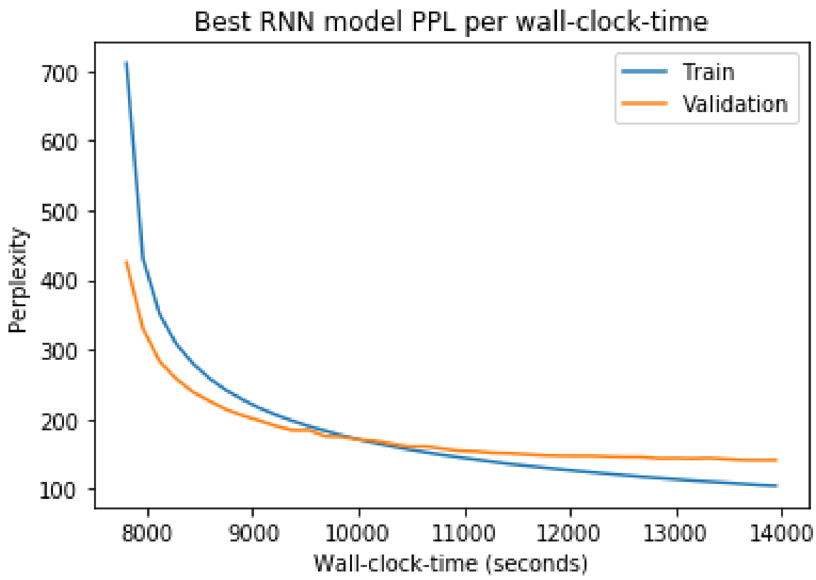

Figure 22: Experiment 4: Best result for RNN which gets better result than the baseline (learning curve by time-clock)

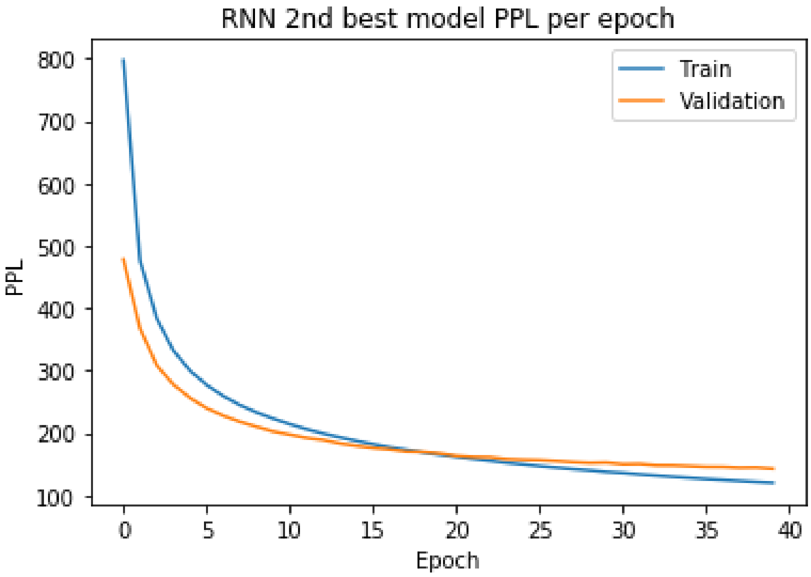

Figure 23: Experiment 5: 2nd best result for RNN (learning curve by epoch)

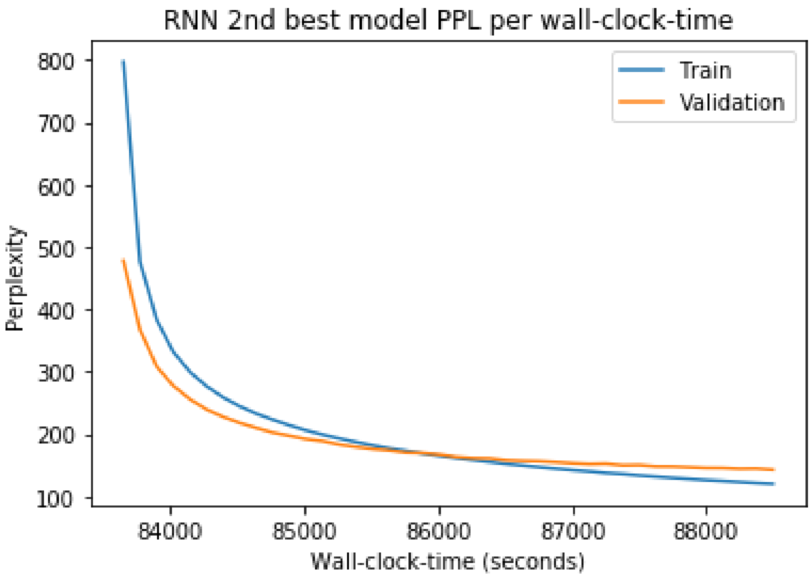

Figure 24: Experiment 5: 2nd best result for RNN (learning curve by time-clock)

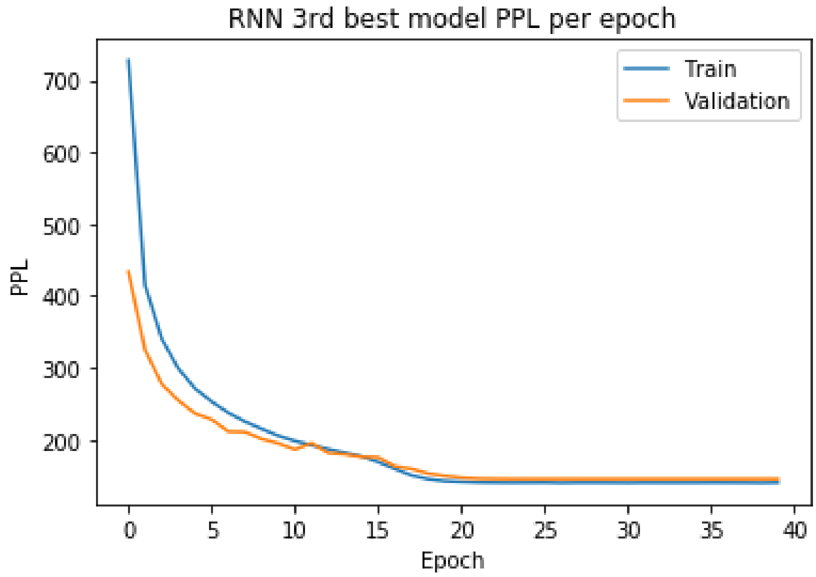

Figure 25: Experiment 6: 3rd best result for RNN (learning curves per epoch)

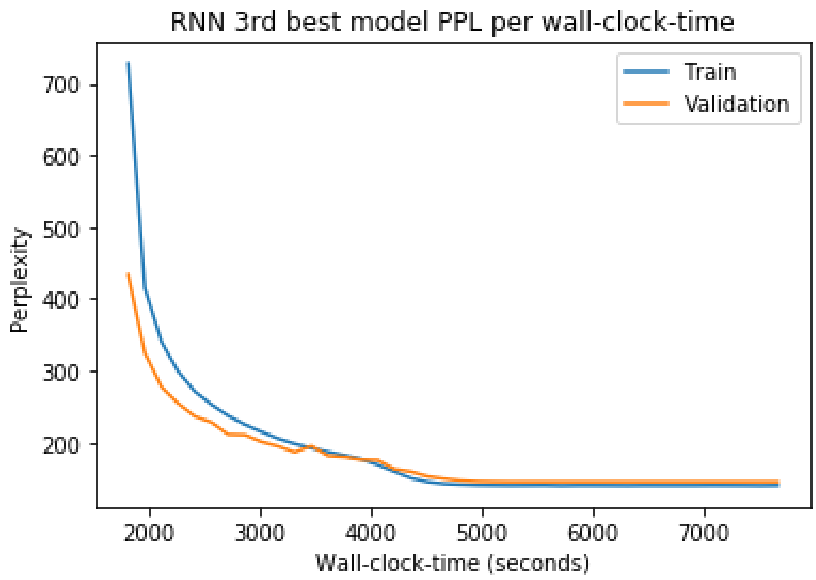

Figure 26: Experiment 6: 3rd best result for RNN (learning curves per clock)

- GRU:

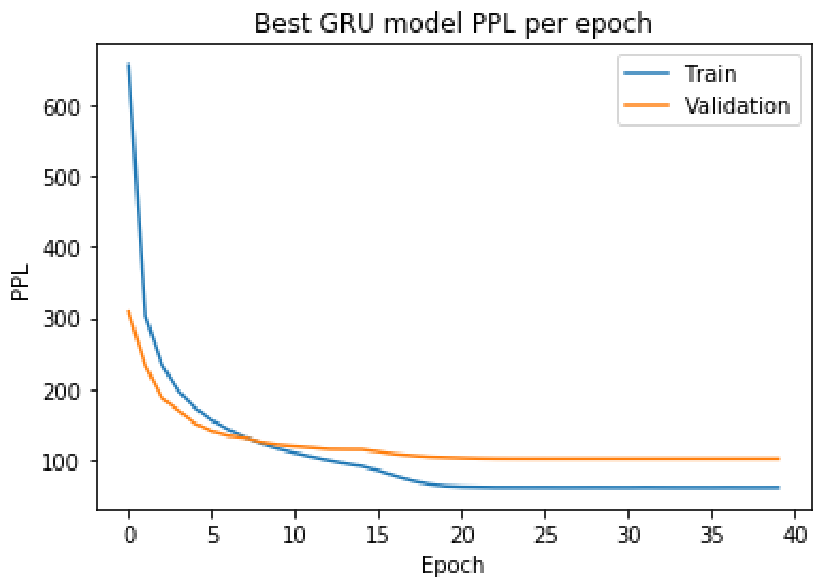

Figure 27: Experiment 4: Best result for GRU which gets better result than the baseline (learning curve by epoch)

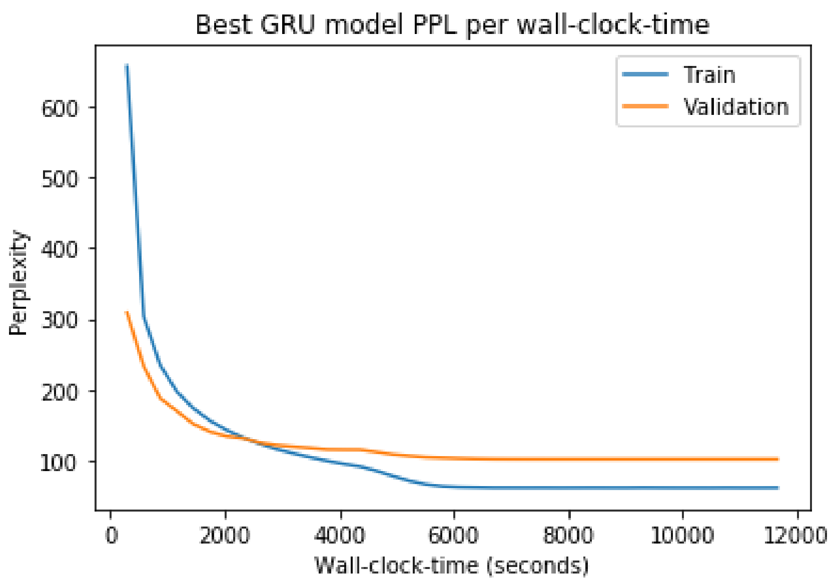

Figure 28: Experiment 4: Best result for GRU which gets better result than the baseline (learning curve by time-clock)

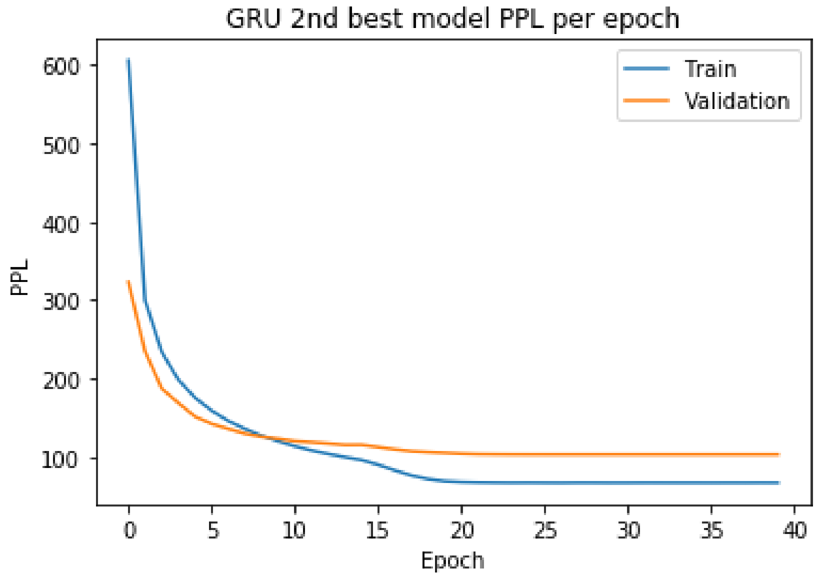

Figure 29: Experiment 5: 2nd best result for GRU (learning curve by epoch)

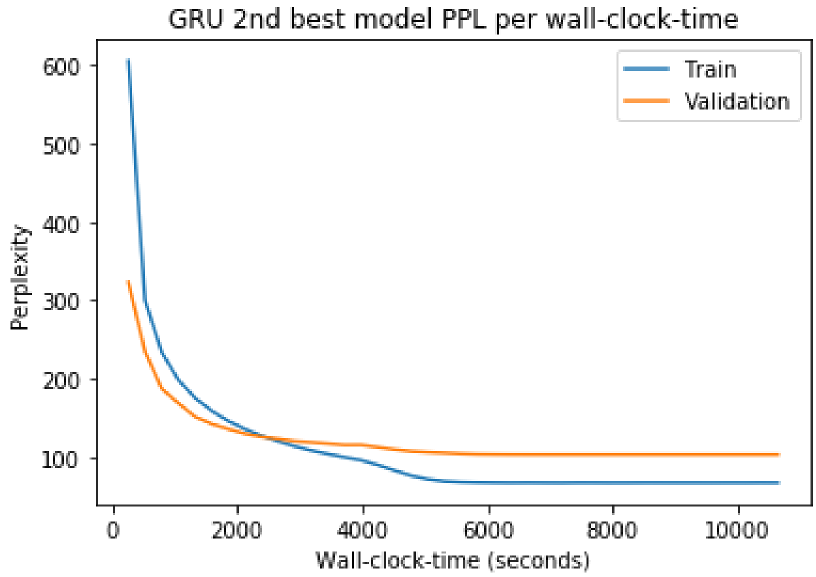

Figure 30: Experiment 5: 2nd best result for GRU (learning curve by time-clock)

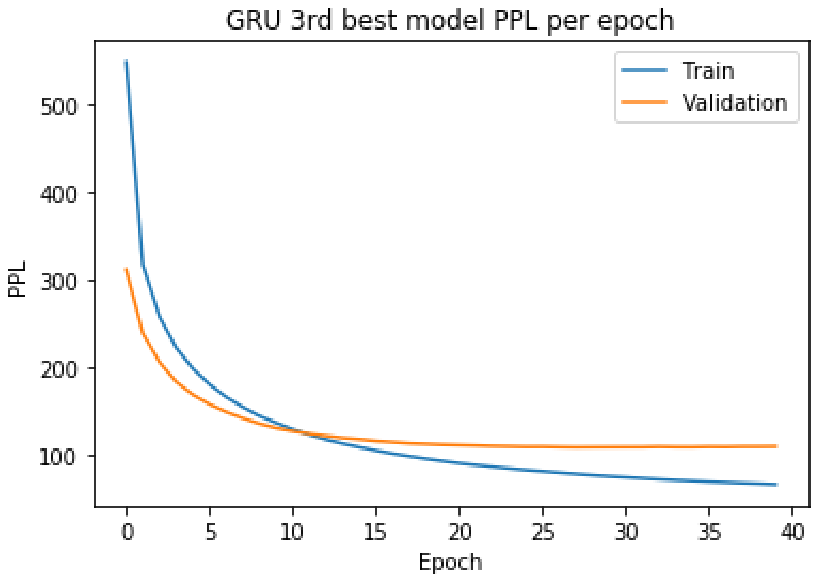

Figure 31: Experiment 6: 3rd best result for GRU (learning curves per epoch)

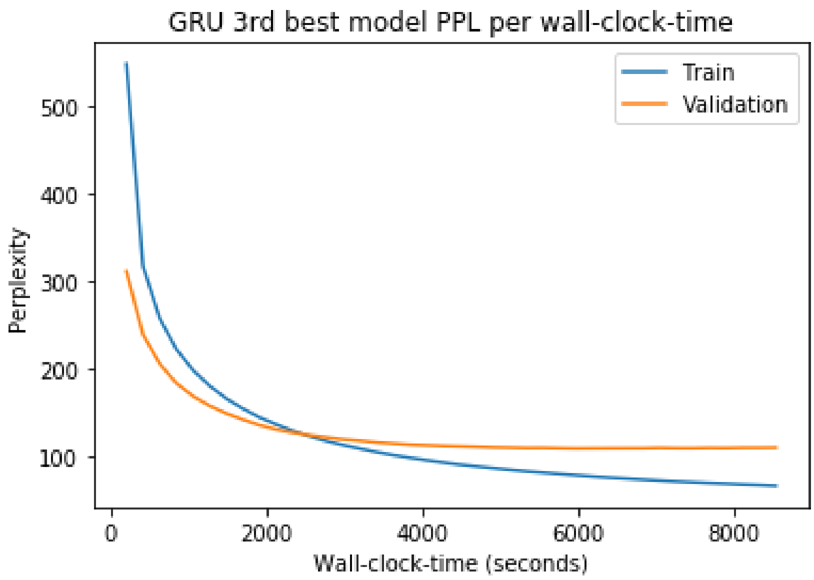

Figure 32: Experiment 6: 3rd best result for GRU (learning curves per clock)

- Transformer:

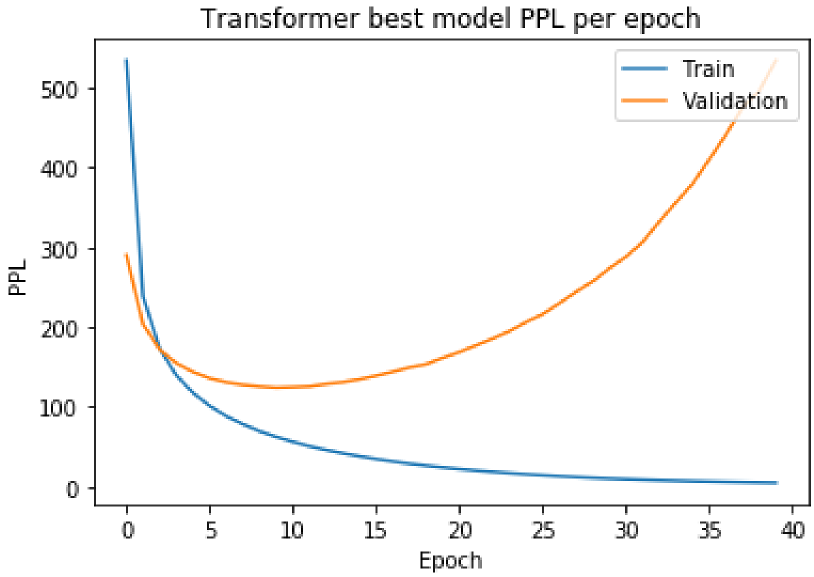

Figure 33: Experiment 4: Best result for Transformer which gets better result than the baseline (learning curve by epoch)

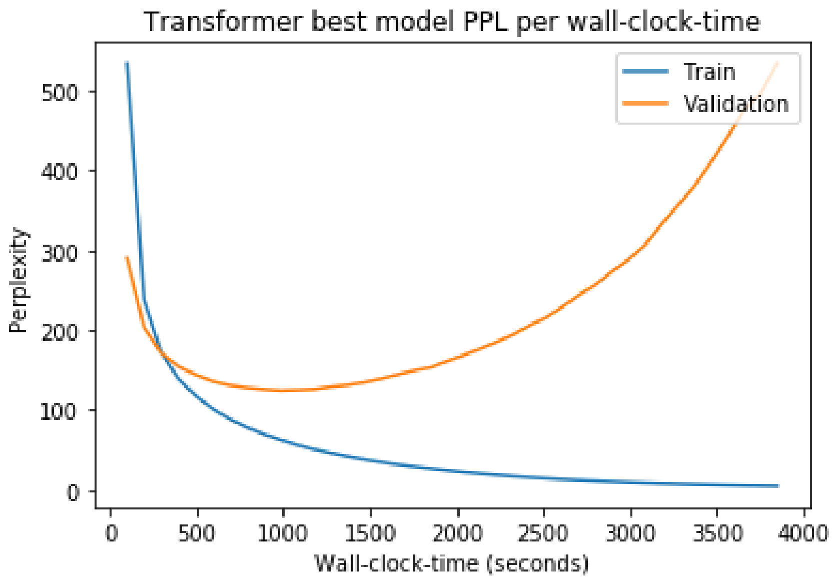

Figure 34: Experiment 4: Best result for Transformer which gets better result than the baseline (learning curve by time-clock)

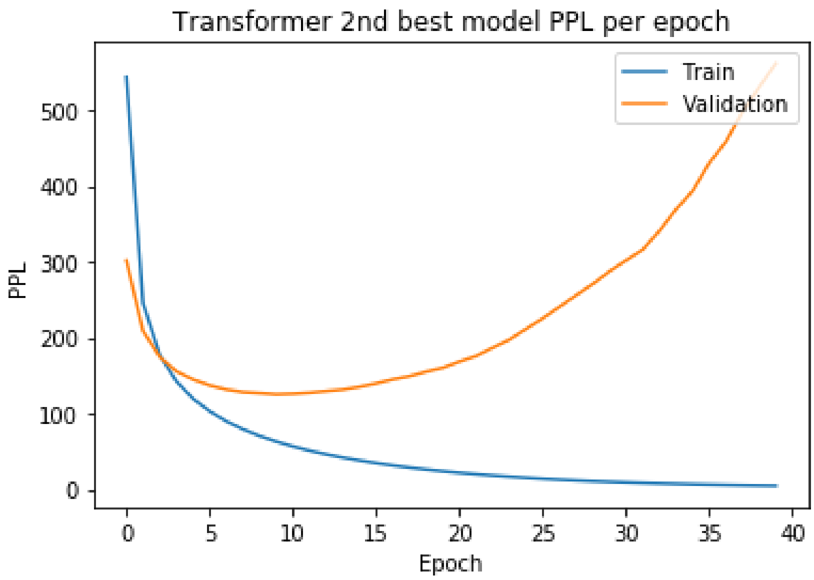

Figure 35: Experiment 5: 2nd best result for Transformer (learning curve by epoch)

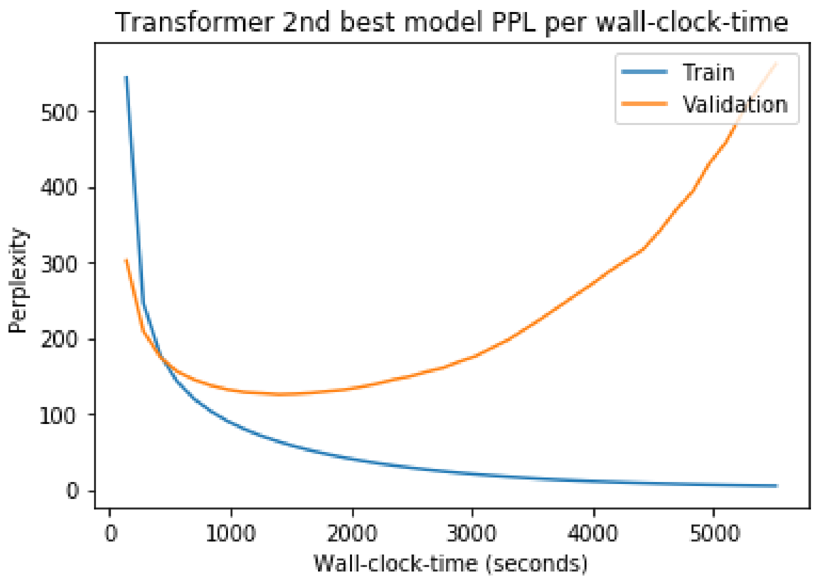

Figure 36: Experiment 5: 2nd best result for Transformer (learning curve by time-clock)

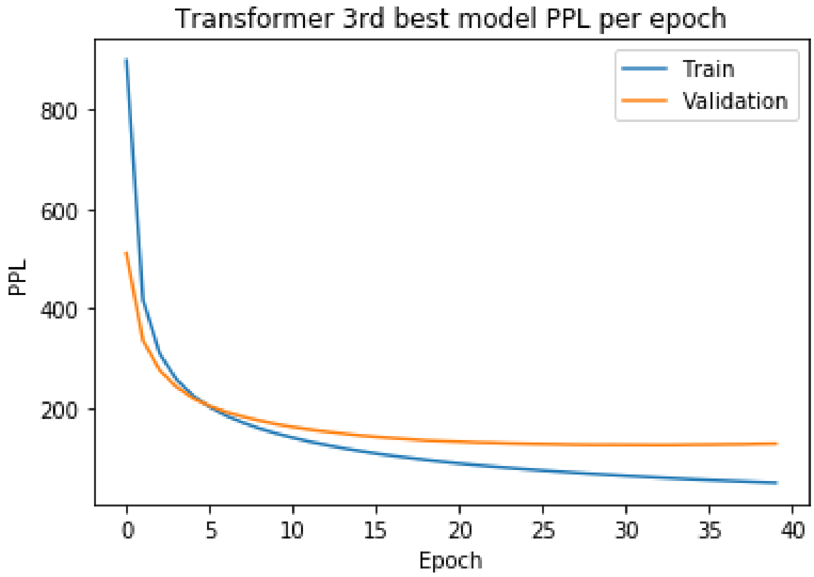

Figure 37: Experiment 6: 3rd best result for Transformer (learning curves per epoch)

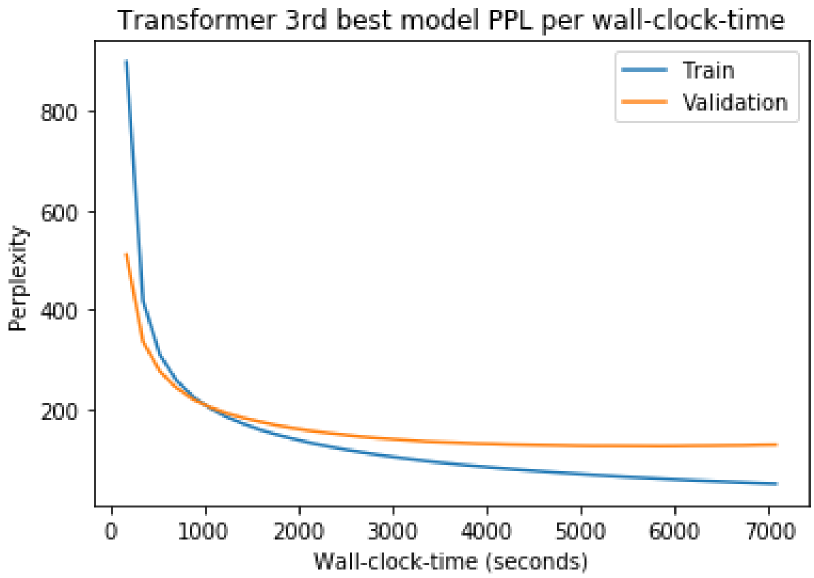

Figure 38: Experiment 6: 3rd best result for Transformer (learning curves per clock)

4. Make 2 plots for each optimizer; one which has all of the validation curves for that optimizer over epochs and one over wall-clock-time.
5. Make 2 plots for each aritecture; one which has all of the validation curves for that architecture over epochs and one over wall-clock-time.

4.4 Discussion

1. **What did you expect to see in these experiments, and what actually happens? Why do you think that happens?**

   We expected that the GRU will get better performance than the vanilla RNN, which was the case. In terms of time processing we expected GRU to be slower than the other architecture, which was the case given the number of calculations.

   On the other hand, we expected the Transformer to outperfom the other architectures in terms of perplexity on the validation set and in terms of time processing. It was the case for the time processing. It was, however, not the best model in all cases. We think that this is due to the fact that RNN models require much more computation than the transformer network.

2. **Referring to the learning curves, qualitatively discuss the differences between the three optimizers in terms of training time, generalization performance, which architecture they're best for, relationship to other hyperparameters, etc.**

   Assuming that the comparison here was made with a set of hyperparameters that give a fair baseline for all the models, we can say that there is no universal best optimizer regardeless of the model architecture. In fact, each model architecture works better with a suited optimizer. Howerver, we notice that ADAM is a stable optimizer that can make the model converge quickly to a fairly good solution. Furthermore, it does not require complex hyperparameter search, as in the case of SGD, particularly for the learning rate. This satisfies the theory as ADAM automatically adapts the learning rate. However, SGD is capable of generalizing better than ADAM when using the right set of hyperparameters.

3. **Which hyperparameters+optimizer would you use if you were most concerned with wallclock time? With generalization performance? In each case, what is the "cost" of the good performance (e.g. does better wall-clock time to a decent loss mean worse final loss? Does better generalization performance mean longer training time?)**

   Concerning wall-clock time we obtained faster results using the optimizer ADAM. In terms of generalization we would use SGD as we find that it performs better once we have found good hyperparameters for the model. We view this as a trade-off between faster results and final accurate predictions. In conclusion we would use ADAM for prototyping and SGD for further improvement and better generalization.

4. **Which architecture is most "reliable" (decent generalization performance for most hyperparameter+ optimizer settings), and which is more unstable across settings?** We would use GRU because it is a more robust architecture and performs better. Transformer is a fancy architecture but it was the only architecture that was unstable during the experiments

5. **Describe a question you are curious about and what experiment(s) (i.e. what architecture/ optimizer/hyperparameters) you would run to investigate that question.**

   We are curious whether SGD with Polyak's momentum term would achieve better results than SGD with the faster convergence of ADAM. We would run the experiments where we got the best results with this new optimizer and comapre its performance.

Figure 39: text

5 Problem 5: Detailed evaluation of trained models

For points 1 and 2 we made a slight modification to the models, the update is in file models5.py.

1. We compute the average loss for all the validation sequences at each time step. The main script has been modified to load a pre-trained model and process the validation model while computing the loss at each time step individually(This is indicated in the script ptb-lm-P5-1.py under "Loss computation", Line 405). The resulting curves for the three models are shown in Figure 39.
The loss is clearly larger when the training starts as the model does not have much information about the sequence at hand and cannot predict accurately. As we proceed, all the models should start making better predictions and thus the reduction of the loss. Although the loss fluctuates, it seems to be stabilizing around certain values, of which the RNN loss shows a higher loss while the GRU and Transformer perform similarly to each other but better than RNN.

2. Similarly to point 1 above, we modified the main script (see ptb-lm-P5-2.py under LOSS COMPUTATION on line 405). Here instead of processing the whole sequence at each step, we only run one mini-batch, processing each time step separately without touching the hidden states. This allows us to compute the gradient at time T with respect to the hidden state at each time step t.
We compute the norm of the concatenated gradient vectors for the different layers and plot them with respect to the time steps as illustrated in Figure 40. Note that the values for each curve are rescaled to be in the range $[0, 1]$. This way we can compare the behavior of the gradients of the two different models (RNN and GRU). The graph shows how the

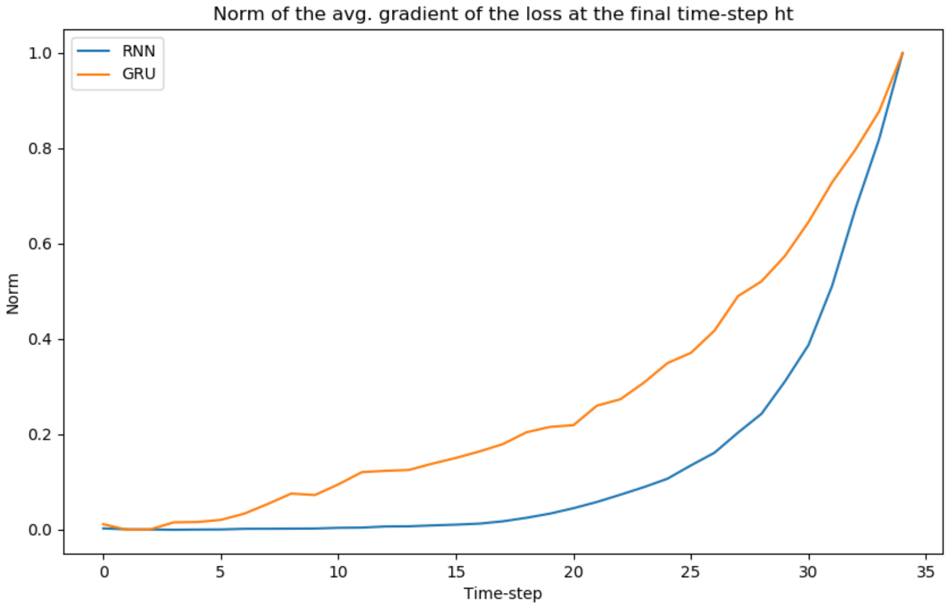

Figure 40: text

gradients with respect to earlier time-steps are smaller than the gradients with respect to later times. This indicates that long-term dependencies contribute less to the gradient at a given point. The drop in the dependency is particularly steep on the RNN model. On the other hand, the mechanics of GRU help better mainatining the long-term dependencies.

3. We generate samples from both the simple RNN and GRU. We use the models from section 4.1. The sampling code can be find in the script generate.py. We produce 20 samples from RNN and GRU, 10 samples of the same length as the training sequnences (35) and 10 samples that are twice the length of the training sequeneces (70). The samples can be found in the Appendix of the report.

Extras Given the interesting results with text generation, we decided to experiment with more things like trying different temperature values and transforming text to audio (you find the audio samples under audio folder in github).

We experience with temperature values in order to control the randomness of the sample generation by scaling the logits before applying softmax.

(a) High temperature:
around poison book hitting german stimulators cooperative response barred remodeling contractor clearance vermont institutional overseas taught opening shape europeans collecting harold cananea clean strikes bulls peripheral direct sydney birds pitney peoples balls surrendered amount

(b) Low temperature:

and and and and and and and and and and and and and and and and and and ¡eos¿ and and and and ¡eos¿ and ¡eos¿ and and and and and and and ¡eos¿

High temperature values gives more complex diverse samples however low values gives more confident but very repetitive samples.

(a) three best samples:

  i. *glass of factories scrap bonds laws counterpart properties yen south crude the ¡unk¿ investments N mortgages u.k. bonds bonds N debentures bonds N deposits bonds interests french debentures bonds bonds swiss bonds bonds bonds bond bonds debentures securities bank bonds bonds bonds bonds one-year bonds bonds bonds bonds eurobonds eurobonds freddie ¡unk¿ freddie bonds bonds bonds ¡eos¿ bonds backed bonds notes bonds bonds bonds west certificates bonds bonds notes bonds*

  ii. *limited ¡eos¿ china additional and greater institute shares chemical international and companies securities securities 's members market investment banks 's shares ¡unk¿ securities securities credit-card securities securities commodities credit thrifts bonds & stocks securities securities international securities municipal issues position securities securities securities bonds investors securities co. notes securities shares bank securities and securities sutton securities f international securities general securities investors exchange the yields debentures and N unique bonds*

  iii. *businesses drugs building investors units the $ bonds and daily assets ¡eos¿ australian N bonds s.a. ¡unk¿ bonds posted de ¡eos¿ certificates notes debentures priced bonds debentures debentures mortgage ¡unk¿ debentures bonds to debentures market*

(b) three interesting samples:

  i. *is of ¡unk¿ care gained proposal product more acquisition ¡eos¿ produced warner up family items personal former rogers new plans near co. ¡eos¿ ¡eos¿ will and analyst ¡eos¿ subsidiary convertible organic president ¡unk¿ and succeeds & director & group chief expert to managing and board projects he and committee directors contract system giant independent board statement wpp he ¡unk¿ ¡unk¿ meanwhile subsidiary ¡unk¿ inc. subsidiary of and firm venture and*

  ii. *justice to the ¡unk¿ impact games ian goods private loans thatcher the imports checks workers the watches families those ¡eos¿ rights anything about goods discipline procedures funds the items conduct minorities whites americans use value*

  iii. *has ¡eos¿ securities largest agency banks dollars securities securities management mr. N and resources ¡unk¿ securities securities shares switzerland stock securities issues securities securities funds and and who N securities financial securities the securities international*

(c) three bad samples:

  i. *the in ¡eos¿ N ¡eos¿ off via N N via ¡eos¿ and via communications N via notes ¡eos¿ up via through via via via via N ¡eos¿ via via via via via to via via*

  ii. *in$ costs under for in in as creates in in with in in due from in in in in in in remic crown in in at in in was ¡unk¿ on rate in in*

  iii. *and ¡eos¿ ¡eos¿ inc. later said most in first said ¡eos¿ backs at ¡eos¿ and imports ¡eos¿ in increased to said but and would mortgage ¡eos¿ continued squibb ¡eos¿ ¡unk¿ ¡eos¿ ¡eos¿ apparel said and*

From the samples we can see that in the case of GRU, we have very distant related words than in the case RNN.

RNN with sequence length of 35 ¡unk¿ also later from which ¡unk¿ net N as in orders with from rate N the N higher closing N N to ¡eos¿ imports mortgage with N sharply rate from ¡eos¿ mortgage from lower loss

---

as stock-index recovery first speculation to 's measure 's company any mesa costa u.s. will ¡unk¿ the the regulators the the late ¡unk¿ first before in despite N ample early philadelphia china ¡unk¿ the last

---

justice to the ¡unk¿ impact games ian goods private loans thatcher the imports checks workers the watches families those ¡eos¿ rights anything about goods discipline procedures funds the items conduct minorities whites americans use value

---

for will to to only will to does came the to was to to of tomorrow this for to on between between was only to to has daniel to or of by over was to

---

the in ¡eos¿ N ¡eos¿ off via N N via ¡eos¿ and via communications N via notes ¡eos¿ up via through via via via via N ¡eos¿ via via via via via to via via

---

and of to and in over fuel ¡eos¿ and ¡eos¿ ¡eos¿ ¡eos¿ were damage and ¡unk¿ about and than before from normally costs could said ¡eos¿ wage personnel in is by a for diseases used

---

and as chairman treatment ¡unk¿ and inc. in ralph office financial have and ¡unk¿ head acquisition board engineering as concern ¡unk¿ make unit ingersoll list & executive inc. ¡eos¿ he which trying 's and ¡unk¿

---

has ¡eos¿ securities largest agency banks dollars securities securities management mr. N and resources ¡unk¿ securities securities shares switzerland stock securities issues securities securities funds and and who N securities financial securities the securities international

---

in $ costs under for in in as creates in in with in in due from in in in in in in remic crown in in at in in was ¡unk¿ on rate in in

---

and ¡eos¿ ¡eos¿ inc. later said most in first said ¡eos¿ backs at ¡eos¿ and imports ¡eos¿ in increased to said but and would mortgage ¡eos¿ continued squibb ¡eos¿ ¡unk¿ ¡eos¿ ¡eos¿ apparel said and

---

RNN with sequence length of 70 to to shows compared N average $ august rate ¡eos¿ in 's for ¡eos¿ N N N ¡eos¿ in its N $ reported ¡unk¿ ¡eos¿ improved from rate seven ¡eos¿ in from ¡eos¿ N billion from gap federal in daily rate to $ to sales N during a statistics N texas offered rate due N which high because ¡eos¿ to higher ¡eos¿ japan ¡eos¿ more macy from N ¡eos¿ already

---

u.s. a the was by ¡eos¿ to this will reports ¡unk¿ the bank 's both as which last an thursday in vietnam raised general mortgage on an two last measure the consumer ford the suit october a sept. state the ual the federal a the australian fiscal $ because was june dow meeting an speculation move N interest stock committee by the by last early yesterday results shares the N

¡unk¿ the foreign the provide apples illegal families currencies and conditions distance turn south cases china foreign the compensation loans employees ¡eos¿ economic blood material a october goods japan relations small-business ¡unk¿ programs illegal that material goods workers spending personnel be inflation loans and such paper goods accounts loans mr. fetal-tissue ¡eos¿ market taxes economic the country $ governments aid vessels shippers loans surgery other business the americans shamir the

---

mr. to will to july to to to to to to i to to to to carried the by to with up to to to saying two to will to the for to to to does to of to or is to to if although for to of of the to to first could had to that between on could to would by the that to to in to of

---

¡eos¿ ¡eos¿ the and which N for N via N via via via from $ via via via ¡eos¿ N N via via ¡eos¿ via via via via N said ¡eos¿ via via via via via via via via via and via N N their via via via via via via via ¡eos¿ via via via via via N via via from via N ¡eos¿ to and via via via

---

to for from to if during at of ¡eos¿ those have ¡eos¿ ¡eos¿ under of ¡eos¿ ¡eos¿ which said higher from could ¡eos¿ of ¡eos¿ represented on in ¡eos¿ of allow increased overseas of from hurt because of tend ¡eos¿ some trucks about higher transport net either reported only of below ' needed the ¡unk¿ still resistance ¡eos¿ into from ¡eos¿ that subsidy and and paid are normally and higher

---

is of ¡unk¿ care gained proposal product more acquisition ¡eos¿ produced warner up family items personal former rogers new plans near co. ¡eos¿ ¡eos¿ will and analyst ¡eos¿ subsidiary convertible organic president ¡unk¿ and succeeds & director & group chief expert to managing and board projects he and committee directors contract system giant independent board statement wpp he ¡unk¿ ¡unk¿ meanwhile subsidiary ¡unk¿ inc. subsidiary of and firm venture and

---

limited ¡eos¿ china additional and greater institute shares chemical international and companies securities securities 's members market investment banks 's shares ¡unk¿ securities securities credit-card securities securities commodities credit thrifts bonds & stocks securities securities international securities municipal issues position securities securities securities bonds investors securities co. notes securities shares bank securities and securities sutton securities f international securities general securities investors exchange the yields debentures and N unique bonds

in in in in in in for in from between for in $ amid in in among at in by in N to based was in rates rate for in in in in N for before deposit closing in at on said based from rate loans fees rates through in from at in on rate mortgage in in debentures N in rate rate in treasury in in mortgage from in

with of that said conspiracy and reported ¡eos¿ said and and in u.s. said inc. were in in than international and august denied by to ¡eos¿ ¡eos¿ and and the had due including & ¡unk¿ N ¡eos¿ the and ¡eos¿ than according N of and ¡unk¿ also after told and rose said ¡eos¿ and and set imports from ¡eos¿ by and of and reported ¡eos¿ in and levels and imports

GRU with sequence length of 35 N the ¡eos¿ N ¡eos¿ N ¡eos¿ N down N N ¡eos¿ ¡eos¿ ¡eos¿ N ¡eos¿ N N N ¡eos¿ N ¡eos¿ N ¡eos¿ ¡eos¿ ¡eos¿ N ¡eos¿ ¡eos¿ ¡eos¿ N ¡eos¿ ¡eos¿ ¡eos¿ N

the for the for twice ¡unk¿ ¡unk¿ improvement lineup expansion in to shares and nov. ¡eos¿ ¡eos¿ rate nov. abortion to ¡eos¿ ¡eos¿ and nov. note ¡eos¿ ¡eos¿ even ¡eos¿ at tomorrow ¡eos¿ bid ¡unk¿

the he streets the the covered big ¡unk¿ and value filters winter water redeem in of familiar them ¡eos¿ ¡unk¿ ¡unk¿ more ground the communities cholesterol satisfaction ¡eos¿ ¡unk¿ ¡unk¿ ¡eos¿ permits delta and ¡unk¿

to by to ¡eos¿ for to of to ¡unk¿ and as surgery ¡eos¿ to to around to to to and ¡eos¿ to ¡eos¿ that for to ¡eos¿ no to ¡eos¿ ¡eos¿ permits ¡eos¿ ¡eos¿ at

through for from from at and to N on up to and N at off at maturity near guilty to at at at N at to at at lower at ¡eos¿ ¡eos¿ ¡unk¿ N at

of in from under with for for and drivers but annually of and and or expenses mortgages ¡eos¿ will import and ¡eos¿ rubbermaid to ¡eos¿ rate if of ¡eos¿ ¡eos¿ ¡eos¿ and ¡eos¿ and ¡eos¿

¡eos¿ and inc. and 's and and and and and in and and ¡unk¿ inc. agency ¡eos¿ and ¡unk¿ and ¡unk¿ ¡eos¿ ¡eos¿ inc and finance ¡eos¿ ¡unk¿ academy products ¡eos¿ consultant and ¡eos¿ and

businesses drugs building investors units the $ bonds and daily assets ¡eos¿ australian N bonds s.a. ¡unk¿ bonds posted de ¡eos¿ certificates notes debentures priced bonds debentures debentures mortgage ¡unk¿ debentures bonds to debentures market

¡eos¿ in ¡unk¿ and ¡unk¿ kohl and carla in and ¡unk¿ md and N and and and and ¡eos¿ ¡eos¿ ¡eos¿ and and certificates coors and ¡eos¿ and and ¡unk¿ ¡eos¿ and ¡unk¿ ¡unk¿ del.

---

co. amoco threatened ¡eos¿ n.v. and and ¡eos¿ and ¡unk¿ and and and ¡eos¿ and and ¡eos¿ and and ¡eos¿ and and and and and facility ¡eos¿ n.j and ¡eos¿ and ¡eos¿ and ¡eos¿ ¡eos¿

GRU with sequence length of 70 N the N of N ¡eos¿ N N N N due N ¡eos¿ ¡eos¿ N N ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ N ¡eos¿ N ¡eos¿ N ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ up ¡eos¿ ¡eos¿ N ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ whites ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ sharply ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ N ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿

---

a oct. catalog aga the feb. ¡eos¿ cut ¡unk¿ fate cuts de malaysia and ¡eos¿ before ¡unk¿ ¡eos¿ bid funding killing jointly nov. ¡eos¿ nov. to dec. ¡eos¿ scheduled nov. nov. nov. commute power march thereafter ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ rate francs fixed ¡eos¿ ¡eos¿ ¡eos¿ scheduled nov. ¡eos¿ ¡eos¿ ¡eos¿ nov. ¡eos¿ ¡unk¿ ¡eos¿ nov. ¡eos¿ N issuance ¡eos¿ nov. will tentatively nov. nov. N nov. nov. ¡eos¿ ¡eos¿ ¡unk¿ investor vehicles their most 's a getting alberta a a collected democracy workers this ¡unk¿ ¡unk¿

---

people evil ¡unk¿ air any diseases skiing ¡unk¿ water an by sex architecture status ¡unk¿ columns worries ¡eos¿ ¡unk¿ waves all apples ¡unk¿ economic de to ¡unk¿ ¡eos¿ ¡unk¿ themselves might ¡eos¿ approval 's twice ¡unk¿ russia ¡unk¿ endless ¡unk¿ virtually modern water sand was ¡eos¿ whose galileo ocean ¡unk¿ walls the ¡unk¿

---

to by N by through by to with to to by ¡eos¿ ¡unk¿ ¡eos¿ and to to into to to from away to has for to to in and ¡eos¿ ¡eos¿ into ¡eos¿ ¡eos¿ to to ¡eos¿ to to ¡eos¿ ¡eos¿ to ¡eos¿ at ¡eos¿ tomorrow to and ¡eos¿ for is and ¡eos¿ ¡eos¿ in for to of ¡eos¿ for overhead to to ¡eos¿ ¡eos¿ ¡eos¿ to to ¡eos¿ up

---

to ¡eos¿ ¡eos¿ and as N on N ¡eos¿ through against at to near N and at with quickly fixed near at and at via ¡eos¿ rate at at and N and via ¡eos¿ again at via ¡eos¿ at at annually as at ¡eos¿ at at into ¡eos¿ rate ¡eos¿ via at rate via via at ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ eurobonds ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ at near ¡unk¿ ¡eos¿

---

at even not for such rising fuel and ¡eos¿ and ¡eos¿ and and and and falling ¡unk¿ in ¡eos¿ flat and ¡unk¿ and ¡eos¿ costs at ¡eos¿ declines and and ¡eos¿ and rate a labor and for and ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ and ¡eos¿ ¡unk¿ and up ¡eos¿ rising plus ¡eos¿ ¡eos¿ and and ¡eos¿ ¡eos¿ until ¡eos¿ rate for orders are ¡eos¿ tax ¡eos¿ if ¡eos¿ ¡eos¿ ¡eos¿

---

of and with ¡eos¿ who ¡unk¿ and ¡eos¿ and and saying ¡unk¿ and constraints ¡eos¿ and and and and ¡eos¿ former and ¡eos¿ avenue and debentures inc university and ¡unk¿ and mary ¡eos¿ division and va ¡eos¿ ¡eos¿ calif. colorado title and ¡unk¿ and probe and inc and stage ¡unk¿ ¡unk¿ de ¡unk¿ ¡unk¿ ¡unk¿ & & campaigns director & ¡unk¿ inc who de and ¡eos¿ ¡unk¿ and inc ¡unk¿

---

glass of factories scrap bonds laws counterpart properties yen south crude the ¡unk¿ investments N mortgages u.k. bonds bonds N debentures bonds N deposits bonds interests french debentures bonds bonds swiss bonds bonds bonds bond bonds debentures securities bank bonds bonds bonds bonds one-year bonds bonds bonds bonds eurobonds eurobonds freddie ¡unk¿ freddie bonds bonds bonds ¡eos¿ bonds backed bonds notes bonds bonds bonds west certificates bonds bonds notes bonds

---

on whose in b. moreover at and and and assistant and sons ¡unk¿ and ¡unk¿ ¡unk¿ ¡eos¿ ¡eos¿ and ¡unk¿ items rate ¡unk¿ ¡eos¿ walker of ¡unk¿ adjustable n.j d rate notes and ¡eos¿ rate by notes rate notes tax N succeeds ¡eos¿ ¡eos¿ ¡eos¿ rate rate manager rate ¡eos¿ eurobonds and ¡eos¿ rate N adjustable ¡eos¿ rate rate team on ¡unk¿ rate ¡eos¿ notes rate rate rate tentatively adjustable

---

¡eos¿ a by and and and joint and and and and and and and and north and and annual and and and N and and and ¡eos¿ and ¡eos¿ ¡eos¿ and and and and and and transmission ¡eos¿ and ¡eos¿ calif and ¡eos¿ and N and ¡eos¿ co ¡eos¿ annual ¡eos¿ and and N ¡eos¿ N mark and ¡unk¿ ¡eos¿ ¡eos¿ ¡eos¿ and ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿ ¡eos¿

References

[1] The machinists' github repository
https://github.com/faresbs/Representation-Learning