MP4_P2_classification

November 20, 2020

[1]:
```
import os
import time
import math
import glob
import string
import random

import torch
import torch.nn as nn

from lstm.helpers import time_since

%matplotlib inline
```

[2]:
```
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
```

1 Language recognition with an RNN

If you've ever used an online translator you've probably seen a feature that automatically detects the input language. While this might be easy to do if you input unicode characters that are unique to one or a small group of languages (like "" or " "), this problem is more challenging if the input only uses the available ASCII characters. In this case, something like "těší mě" would beome "tesi me" in the ascii form. This is a more challenging problem in which the language must be recognized purely by the pattern of characters rather than unique unicode characters.

We will train an RNN to solve this problem for a small set of languages thta can be converted to romanized ASCII form. For training data it would be ideal to have a large and varied dataset in different language styles. However, it is easy to find copies of the Bible which is a large text translated to different languages but in the same easily parsable format, so we will use 20 different copies of the Bible as training data. Using the same book for all of the different languages will hopefully prevent minor overfitting that might arise if we used different books for each language (fitting to common characteristics of the individual books rather than the language).

[3]:
```
from unidecode import unidecode as unicodeToAscii

all_characters = string.printable
n_letters = len(all_characters)
```

```
print(unicodeToAscii('těší mě'))
```

```
tesi me
```

[4]:
```
# Read a file and split into lines
def readFile(filename):
    data = open(filename, encoding='utf-8').read().strip()
    return unicodeToAscii(data)

def get_category_data(data_path):
    # Build the category_data dictionary, a list of names per language
    category_data = {}
    all_categories = []
    for filename in glob.glob(data_path):
        category = os.path.splitext(os.path.basename(filename))[0].split('_')[0]
        all_categories.append(category)
        data = readFile(filename)
        category_data[category] = data

    return category_data, all_categories
```

The original text is split into two parts, train and test, so that we can make sure that the model is not simply memorizing the train data.

[5]:
```
train_data_path = 'language_data/train/*_train.txt'
test_data_path = 'language_data/test/*_test.txt'

train_category_data, all_categories = get_category_data(train_data_path)
test_category_data, test_all_categories = get_category_data(test_data_path)

n_languages = len(all_categories)

print(len(all_categories))
print(all_categories)
```

```
20
['albanian', 'czech', 'danish', 'english', 'esperanto', 'finnish', 'french',
'german', 'hungarian', 'italian', 'lithuanian', 'maori', 'norwegian',
'portuguese', 'romanian', 'spanish', 'swedish', 'turkish', 'vietnamese',
'xhosa']
```

2 Data processing

[6]:
```
def categoryFromOutput(output):
    top_n, top_i = output.topk(1, dim=1)
    category_i = top_i[:, 0]
    return category_i

# Turn string into long tensor
def stringToTensor(string):
    tensor = torch.zeros(len(string), requires_grad=True).long()
    for c in range(len(string)):
        tensor[c] = all_characters.index(string[c])
    return tensor

def load_random_batch(text, chunk_len, batch_size):
    input_data = torch.zeros(batch_size, chunk_len).long().to(device)
    target = torch.zeros(batch_size, 1).long().to(device)
    input_text = []
    for i in range(batch_size):
        category = all_categories[random.randint(0, len(all_categories) - 1)]
        line_start = random.randint(0, len(text[category])-chunk_len)
        category_tensor = torch.tensor([all_categories.index(category)],␣
 ↪dtype=torch.long)
        line = text[category][line_start:line_start+chunk_len]
        input_text.append(line)
        input_data[i] = stringToTensor(line)
        target[i] = category_tensor
    return input_data, target, input_text
```

3 Implement Model

For this classification task, we can use the same model we implement for the generation task which is located in `rnn/model.py`. See the `MP4_P2_generation.ipynb` notebook for more instructions. In this case each output vector of our RNN will have the dimension of the number of possible languages (i.e. `n_languages`). We will use this vector to predict a distribution over the languages.

In the generation task, we used the output of the RNN at every time step to predict the next letter and our loss included the output from each of these predictions. However, in this task we use the output of the RNN at the end of the sequence to predict the language, so our loss function will use only the predicted output from the last time step.

4 Train RNN

[7]:
```
from lstm.model import LSTM
```

[8]:
```
chunk_len = 25

BATCH_SIZE = 100
n_epochs = 2000
hidden_size = 300
n_layers = 2
learning_rate = 0.001
model_type = 'lstm'

criterion = nn.CrossEntropyLoss()
lstm = LSTM(n_letters, hidden_size, n_languages, model_type=model_type,␣
 ↪n_layers=n_layers).to(device)
```

TODO: Fill in the train function. You should initialize a hidden layer representation using your RNN's `init_hidden` function, set the model gradients to zero, and loop over each time step (character) in the input tensor. For each time step compute the output of the of the RNN and the next hidden layer representation. The cross entropy loss should be computed over the last RNN output scores from the end of the sequence and the target classification tensor. Lastly, call backward on the loss and take an optimizer step.

[9]:
```
def train(lstm, target_tensor, data_tensor, optimizer, criterion,␣
 ↪batch_size=BATCH_SIZE):
    """
    Inputs:
    - rnn: model
    - target_target: target character data tensor of shape (batch_size, 1)
    - data_tensor: input character data tensor of shape (batch_size, chunk_len)
    - optimizer: rnn model optimizer
    - criterion: loss function
    - batch_size: data batch size

    Returns:
    - output: output from RNN from end of sequence
    - loss: computed loss value as python float

    """

    ####################################
    #           YOUR CODE HERE         #
    ####################################
    batch_size = data_tensor.size(0)
    chunk_len = data_tensor.size(1)

    hidden = lstm.init_hidden(batch_size, device = device)
    optimizer.zero_grad()

    for i in range(chunk_len):
```

```
        output, hidden = lstm(data_tensor[:,i], hidden)

    loss = criterion(output, target_tensor.squeeze())
    loss.backward()
    optimizer.step()

    ##########       END      ##########
    return output, loss.data.item()
```

[10]:
```
def evaluate(lstm, data_tensor, seq_len=chunk_len, batch_size=BATCH_SIZE):
    with torch.no_grad():
        data_tensor = data_tensor.to(device)
        hidden = lstm.init_hidden(batch_size, device=device)
        for i in range(seq_len):
            output, hidden = lstm(data_tensor[:,i], hidden)

        return output

def eval_test(lstm, category_tensor, data_tensor):
    with torch.no_grad():
        output = evaluate(lstm, data_tensor)
        loss = criterion(output, category_tensor.squeeze())
        return output, loss.data.item()
```

[11]:
```
n_iters = 10000 #1000 #2000 #100000
print_every = 25
plot_every = 25


# Keep track of losses for plotting
current_loss = 0
current_test_loss = 0
all_losses = []
all_test_losses = []

start = time.time()

optimizer = torch.optim.Adam(lstm.parameters(), lr=learning_rate, weight_decay=0.
 ↪0005)



number_correct = 0
for iter in range(1, n_iters + 1):
    input_data, target_category, text_data =␣
 ↪load_random_batch(train_category_data, chunk_len, BATCH_SIZE)
    output, loss = train(lstm, target_category, input_data, optimizer, criterion)
```

```
    current_loss += loss

    _, test_loss = eval_test(lstm, target_category, input_data)
    current_test_loss += test_loss
    guess_i = categoryFromOutput(output)
    number_correct += (target_category.squeeze()==guess_i.squeeze()).long().sum()

    # Print iter number, loss, name and guess
    if iter % print_every == 0:
        sample_idx = 0
        guess = all_categories[guess_i[sample_idx]]

        category = all_categories[int(target_category[sample_idx])]

        correct = '' if guess == category else ' (%s)' % category
        print('%d %d%% (%s) %.4f %.4f %s / %s %s' % (iter, iter / n_iters * 100,␣
↪time_since(start), loss, test_loss, text_data[sample_idx], guess, correct))
        print('Train accuracy: {}'.format(float(number_correct)/
↪float(print_every*BATCH_SIZE)))
        number_correct = 0

    # Add current loss avg to list of losses
    if iter % plot_every == 0:
        all_losses.append(current_loss / plot_every)
        current_loss = 0
        all_test_losses.append(current_test_loss / plot_every)
        current_test_loss = 0

    if iter % 200 == 0:
        learning_rate = learning_rate * 0.95
```

```
25 0% (0m 2s) 2.6302 2.5436 om du stadig dyrker, han  / english  (norwegian)
Train accuracy: 0.1128
50 0% (0m 4s) 1.9315 1.9308 ad juro Jehova a David, N / italian  (spanish)
Train accuracy: 0.244
75 0% (0m 7s) 1.8365 1.7887 fi groaza, cind vor cadea / spanish  (romanian)
Train accuracy: 0.3108
100 1% (0m 9s) 1.4871 1.4175 te Tama a te tangata. Ka  / maori
Train accuracy: 0.418
125 1% (0m 11s) 1.3415 1.2881 pa haugene. Og Herren slo / german  (norwegian)
Train accuracy: 0.516
150 1% (0m 13s) 1.2430 1.2190 facut Dumnezeu prin El in / romanian
Train accuracy: 0.566
175 1% (0m 16s) 0.8546 0.7799  Herz, das Gebot des HERR / danish  (german)
Train accuracy: 0.6076
200 2% (0m 18s) 0.8518 0.7972 ust: and I cast the dust  / english
Train accuracy: 0.6448
```

```
225 2% (0m 20s) 0.7064 0.6598 oi, tuc la xu co mach nha / vietnamese
Train accuracy: 0.6624
250 2% (0m 22s) 0.7351 0.6344 r i dashuri yt se nje i d / albanian
Train accuracy: 0.6916
275 2% (0m 25s) 0.9068 0.8419 u inaintea Tatalui Meu ca / romanian
Train accuracy: 0.6944
300 3% (0m 27s) 0.7855 0.6943 tuong deu la hu vo, viec  / vietnamese
Train accuracy: 0.7264
325 3% (0m 29s) 0.7073 0.6130 er, da to Ar var gaet, tr / norwegian  (danish)
Train accuracy: 0.7344
350 3% (0m 31s) 0.8281 0.7134 ; ma del bestiame dei fig / italian
Train accuracy: 0.7348
375 3% (0m 34s) 0.7719 0.7207 righteously for God, And  / english
Train accuracy: 0.7404
400 4% (0m 36s) 0.6916 0.5440 kun palvoitte epajumalian / finnish
Train accuracy: 0.7724
425 4% (0m 38s) 0.7177 0.6356 n suuri kansojen keskuude / finnish
Train accuracy: 0.7816
450 4% (0m 40s) 0.5169 0.4447  cevresinde buyuk ve kucu / turkish
Train accuracy: 0.7868
475 4% (0m 42s) 0.5014 0.4037 gnar skola vara sasom en  / swedish
Train accuracy: 0.8008
500 5% (0m 45s) 0.5028 0.4394  la gloria dell'Eterno su / italian
Train accuracy: 0.818
525 5% (0m 47s) 0.4178 0.3477 s are pilda aceasta: ,,Sa / portuguese
(romanian)
Train accuracy: 0.8044
550 5% (0m 49s) 0.6103 0.5098 isht sot? Nuk eshte as di / albanian
Train accuracy: 0.8132
575 5% (0m 51s) 0.5189 0.4235 de. Eliligugu.  Yinene na / turkish  (xhosa)
Train accuracy: 0.8256
600 6% (0m 54s) 0.6201 0.5555  Mert ezt mondja az Ur: N / hungarian
Train accuracy: 0.8176
625 6% (0m 56s) 0.4037 0.3668 rs une lampe devant moi a / french
Train accuracy: 0.8288
650 6% (0m 58s) 0.3868 0.3110  den tho Duc Gie-ho-va va / vietnamese
Train accuracy: 0.8392
675 6% (1m 0s) 0.5496 0.4802 io mi pekis, ke vi transd / maori  (esperanto)
Train accuracy: 0.8472
700 7% (1m 3s) 0.4764 0.3884 i. Zatrubte v Gibeji na p / czech
Train accuracy: 0.8516
725 7% (1m 5s) 0.4954 0.4413 Saule sukure, kad viespat / lithuanian
Train accuracy: 0.8616
750 7% (1m 7s) 0.3368 0.3009 cevremi aydinlatti. Yere  / turkish
Train accuracy: 0.8508
775 7% (1m 9s) 0.3253 0.3342 pi Abjatarille kuningas s / lithuanian  (finnish)
Train accuracy: 0.8448
800 8% (1m 12s) 0.4258 0.3079 i niin kova riita, etta k / finnish
```

```
Train accuracy: 0.8452
825 8% (1m 14s) 0.2873 0.2414 oi den, Chung no ben muon / vietnamese
Train accuracy: 0.8664
850 8% (1m 16s) 0.2703 0.2460 tlerinin arasina yerlesti / turkish
Train accuracy: 0.8572
875 8% (1m 18s) 0.4078 0.3375  lam nguoi day ke tam thu / vietnamese
Train accuracy: 0.8596
900 9% (1m 20s) 0.3922 0.3057 tildekke sig. For en mann / danish  (norwegian)
Train accuracy: 0.8552
925 9% (1m 23s) 0.3587 0.3259 la profetoj estos plenumi / esperanto
Train accuracy: 0.8628
950 9% (1m 25s) 0.3130 0.2564 ke med nadastolen gjorde  / danish  (swedish)
Train accuracy: 0.866
975 9% (1m 27s) 0.3929 0.3013 Viespats pasirode jo tevu / lithuanian
Train accuracy: 0.8724
1000 10% (1m 29s) 0.3153 0.2817 roit a tout son peuple. J / french
Train accuracy: 0.868
1025 10% (1m 32s) 0.2627 0.2583  de glavo, ili falos; al  / esperanto
Train accuracy: 0.8612
1050 10% (1m 34s) 0.3354 0.2667 d thee to scorn; the daug / english
Train accuracy: 0.876
1075 10% (1m 36s) 0.4035 0.3416 undervune Moisiut me fjal / albanian
Train accuracy: 0.8684
1100 11% (1m 38s) 0.3952 0.3332  i nga wahi kuiti. Kei te / maori
Train accuracy: 0.8792
1125 11% (1m 41s) 0.3513 0.2436 niasdesimt, o stipresniuj / lithuanian
Train accuracy: 0.882
1150 11% (1m 43s) 0.3335 0.3421 fericirea Ierusalimului,  / romanian
Train accuracy: 0.8856
1175 11% (1m 45s) 0.3883 0.3425 a cezalandirdim.<br />Cun / turkish
Train accuracy: 0.8684
1200 12% (1m 47s) 0.2314 0.1723 te da mange af Joderne; t / danish
Train accuracy: 0.8812
1225 12% (1m 50s) 0.3316 0.2617  serie di teli; e lo stes / italian
Train accuracy: 0.8648
1250 12% (1m 52s) 0.4299 0.3585 g luc nguoi nu co kinh ng / vietnamese
Train accuracy: 0.878
1275 12% (1m 54s) 0.1955 0.1881 abilono karaliaus kariuom / lithuanian
Train accuracy: 0.8696
1300 13% (1m 56s) 0.2995 0.2688 ciyla demirciyi Babile su / turkish
Train accuracy: 0.8948
1325 13% (1m 58s) 0.4315 0.3712 La entreprenoj de diligen / esperanto
Train accuracy: 0.8772
1350 13% (2m 1s) 0.1962 0.1701 pa al iazului de sus, pe  / romanian
Train accuracy: 0.884
1375 13% (2m 3s) 0.3689 0.2700 antan oni ne povas kalkul / finnish  (esperanto)
Train accuracy: 0.9016
1400 14% (2m 5s) 0.1944 0.1540 schonen Hals; ich will Ep / german
```

```
Train accuracy: 0.8968
1425 14% (2m 7s) 0.3899 0.3171 , khi con nguoi hoi nguoi / vietnamese
Train accuracy: 0.8864
1450 14% (2m 10s) 0.3483 0.2873  for Kongens Livvagt, som / danish
Train accuracy: 0.9024
1475 14% (2m 12s) 0.3037 0.2464 oro, Liajn preskribojn ka / esperanto
Train accuracy: 0.884
1500 15% (2m 14s) 0.2636 0.2372 t -- pitakaa hyvananne! M / finnish
Train accuracy: 0.8912
1525 15% (2m 16s) 0.3284 0.2910 hluli? uThixo wathuma yen / xhosa
Train accuracy: 0.886
1550 15% (2m 19s) 0.2630 0.2315  mi Sian bonecon. Al la h / esperanto
Train accuracy: 0.8952
1575 15% (2m 21s) 0.2606 0.2166 sii de razboi ai Samariei / romanian
Train accuracy: 0.8924
1600 16% (2m 23s) 0.3573 0.3202 a-dan-a-ram. Y-sac khan c / vietnamese
Train accuracy: 0.8824
1625 16% (2m 25s) 0.3243 0.2939  en man hor, ej heller sk / norwegian
(swedish)
Train accuracy: 0.9104
1650 16% (2m 28s) 0.2437 0.2125  kahore hoki tetahi i hae / maori
Train accuracy: 0.8924
1675 16% (2m 30s) 0.2546 0.2072 Iharaira rongo ki nga Amo / maori
Train accuracy: 0.8936
1700 17% (2m 32s) 0.2355 0.1862 masele de leoaica. Mi -a  / romanian
Train accuracy: 0.8904
1725 17% (2m 34s) 0.1859 0.1994 iyileyo uYehova uThixo wa / xhosa
Train accuracy: 0.8956
1750 17% (2m 37s) 0.3773 0.3216 chimb hotarirea, pentruca / romanian
Train accuracy: 0.8972
1775 17% (2m 39s) 0.1771 0.1504 mkani. Wathi ukumkani kuE / xhosa
Train accuracy: 0.8972
1800 18% (2m 41s) 0.2664 0.2089 endos nje beselidhje te r / albanian
Train accuracy: 0.8972
1825 18% (2m 43s) 0.2879 0.2492 a? Na ka tonoa e ia tokor / maori
Train accuracy: 0.892
1850 18% (2m 46s) 0.3401 0.2834  senin hakkinda konusuyor / turkish
Train accuracy: 0.9016
1875 18% (2m 48s) 0.2791 0.2063 de Bogen sammen og gav Tj / danish
Train accuracy: 0.8968
1900 19% (2m 50s) 0.3908 0.2685 ro con Scebna, il segreta / italian
Train accuracy: 0.9024
1925 19% (2m 52s) 0.3238 0.2705 rementes, os marinheiros  / portuguese
Train accuracy: 0.908
1950 19% (2m 55s) 0.2162 0.2055 e le mie leggi che io vi  / italian
Train accuracy: 0.9028
1975 19% (2m 57s) 0.3628 0.2467  ouvriers qui sont avec l / french
Train accuracy: 0.9124
```

```
2000 20% (2m 59s) 0.1831 0.1385 Israel meita tunne. Sina, / finnish
Train accuracy: 0.9044
2025 20% (3m 2s) 0.2561 0.1993 i. Saziaci al mattino del / italian
Train accuracy: 0.9116
2050 20% (3m 4s) 0.3114 0.2407 k bej asgje prej vetvetes / albanian
Train accuracy: 0.9016
2075 20% (3m 6s) 0.2951 0.2713 ue je suis l'Eternel. Moi / french
Train accuracy: 0.894
2100 21% (3m 8s) 0.1763 0.1552 , ka jam Viespats kalbejo / lithuanian
Train accuracy: 0.9016
2125 21% (3m 10s) 0.2177 0.1597 zamosokkal a hajoban hagy / hungarian
Train accuracy: 0.9088
2150 21% (3m 13s) 0.2870 0.1961 os! S'a ispravit cu proor / romanian
Train accuracy: 0.8968
2175 21% (3m 15s) 0.2158 0.1994 Sirayeli? Ngokuba akumnan / xhosa
Train accuracy: 0.9092
2200 22% (3m 17s) 0.2675 0.2177 lii casei lui Israel, car / romanian
Train accuracy: 0.9056
2225 22% (3m 20s) 0.2257 0.1762  Jis man kalbejo: 'Jei ta / lithuanian
Train accuracy: 0.9088
2250 22% (3m 22s) 0.3189 0.2181 menem. Divce, ktera prijd / czech
Train accuracy: 0.9032
2275 22% (3m 24s) 0.3541 0.2967 ue; phai giat quan ao min / vietnamese
Train accuracy: 0.9128
2300 23% (3m 26s) 0.3039 0.2285 em numero, como as estrel / portuguese
Train accuracy: 0.9064
2325 23% (3m 29s) 0.2573 0.1823 o, lukrozo lokuqala ke ol / xhosa
Train accuracy: 0.9016
2350 23% (3m 31s) 0.1825 0.1304 inovo: "Lidsky synu, post / czech
Train accuracy: 0.9004
2375 23% (3m 33s) 0.2856 0.2165 o ne i sviesa. Jis laiko  / lithuanian
Train accuracy: 0.9112
2400 24% (3m 35s) 0.3793 0.2550 n mosiile copiilor lui Is / romanian
Train accuracy: 0.9084
2425 24% (3m 38s) 0.1550 0.0929 i souliers; et ne saluez  / french
Train accuracy: 0.914
2450 24% (3m 40s) 0.1551 0.1121 e Juda habian hecho, y lo / spanish
Train accuracy: 0.9092
2475 24% (3m 42s) 0.2239 0.1882  i te take riri ki ahau,  / maori
Train accuracy: 0.8992
2500 25% (3m 44s) 0.3867 0.3041 :Jen mi kreskigis kaj mul / esperanto
Train accuracy: 0.9156
2525 25% (3m 46s) 0.4387 0.3340 it diesem Volk in das Lan / german
Train accuracy: 0.9072
2550 25% (3m 49s) 0.1823 0.1622 dicsereti! Milyen utalatt / hungarian
Train accuracy: 0.8876
2575 25% (3m 51s) 0.2433 0.1783 omsindo, ubatshabalalise, / xhosa
Train accuracy: 0.8932
```

```
2600 26% (3m 53s) 0.2733 0.2519  Diciendo: Ve a este pueb / spanish
Train accuracy: 0.9048
2625 26% (3m 55s) 0.2316 0.1942 tu as soif, tu iras boire / french
Train accuracy: 0.916
2650 26% (3m 58s) 0.2874 0.2332 n hade gjort lat Gud ock  / swedish
Train accuracy: 0.9024
2675 26% (4m 0s) 0.3631 0.2985  day cach xa mat Ngai. Y- / vietnamese
Train accuracy: 0.8976
2700 27% (4m 2s) 0.2557 0.1544 -en, cho A-ri-en, la than / vietnamese
Train accuracy: 0.9132
2725 27% (4m 4s) 0.1666 0.1557 unu yerine getirdi. Sara  / turkish
Train accuracy: 0.9044
2750 27% (4m 7s) 0.4022 0.2308 aradan, velitel telesne s / hungarian  (czech)
Train accuracy: 0.906
2775 27% (4m 9s) 0.3599 0.2414 gan koyununu kirkmayacaks / turkish
Train accuracy: 0.9128
2800 28% (4m 11s) 0.2376 0.1644  la domoj, en kiuj ili gx / esperanto
Train accuracy: 0.9084
2825 28% (4m 13s) 0.2047 0.1589 a nabuganga bakuwuqiqa. I / xhosa
Train accuracy: 0.9088
2850 28% (4m 16s) 0.3602 0.3278 arets, Et Maaseja, fils d / danish  (french)
Train accuracy: 0.9032
2875 28% (4m 18s) 0.3513 0.2984 rdsprog: Laege, laeg dig  / danish  (norwegian)
Train accuracy: 0.918
2900 28% (4m 20s) 0.2878 0.2407 what Amalek did unto thee / english
Train accuracy: 0.9148
2925 29% (4m 22s) 0.3666 0.2909 tas Lian nomon. Kaj ili e / esperanto
Train accuracy: 0.9144
2950 29% (4m 24s) 0.2690 0.2500 a o tesouro de todos os v / portuguese
Train accuracy: 0.9096
2975 29% (4m 27s) 0.4690 0.2745 karetini bagrimda nasil t / turkish
Train accuracy: 0.9048
3000 30% (4m 29s) 0.2772 0.1826 pedig az o testenek templ / hungarian
Train accuracy: 0.9164
3025 30% (4m 31s) 0.2889 0.2438 Xip-ba hay chia lay dat.  / vietnamese
Train accuracy: 0.916
3050 30% (4m 33s) 0.3194 0.2406 mind; o elotte hajtanak t / hungarian
Train accuracy: 0.9152
3075 30% (4m 36s) 0.2176 0.1715 limin e hosteneve. Keshtu / albanian
Train accuracy: 0.902
3100 31% (4m 38s) 0.2762 0.2631 facuse Isus tina, si -i d / romanian
Train accuracy: 0.9028
3125 31% (4m 40s) 0.1692 0.1082 nai seimai tektu levitu s / finnish
(lithuanian)
Train accuracy: 0.9208
3150 31% (4m 42s) 0.2345 0.1810 m a Fiu, hanem csak az At / hungarian
Train accuracy: 0.9176
3175 31% (4m 45s) 0.2602 0.2489 ; via glavo ekstermis via / esperanto
```

```
Train accuracy: 0.9136
3200 32% (4m 47s) 0.1678 0.1487 le el tomado, y traidolo  / spanish
Train accuracy: 0.9152
3225 32% (4m 49s) 0.3291 0.2819 i minuni; Tu Ti-ai aratat / romanian
Train accuracy: 0.9168
3250 32% (4m 51s) 0.2926 0.2448 skal fortaere den. Jeg ve / norwegian  (danish)
Train accuracy: 0.9056
3275 32% (4m 53s) 0.1715 0.1533  car il offrit la les hol / norwegian  (french)
Train accuracy: 0.9252
3300 33% (4m 56s) 0.3313 0.3126 Bu colde yikima ugrayacak / turkish
Train accuracy: 0.9104
3325 33% (4m 58s) 0.2179 0.1809  ce veneau dupa Isus, str / romanian
Train accuracy: 0.9092
3350 33% (5m 0s) 0.2839 0.2504 get inre doma om vad ratt / swedish
Train accuracy: 0.9032
3375 33% (5m 3s) 0.2533 0.2450 ra!`` Ilie le -a raspuns: / romanian
Train accuracy: 0.918
3400 34% (5m 5s) 0.1919 0.1659 fterkommere 2172, Sjefatj / danish
Train accuracy: 0.9228
3425 34% (5m 7s) 0.3962 0.2957 va scrie cu mina lui: ,Al / romanian
Train accuracy: 0.9116
3450 34% (5m 9s) 0.2666 0.2295 estes, sacudindo contra e / spanish
(portuguese)
Train accuracy: 0.9116
3475 34% (5m 11s) 0.1623 0.1082 dazo de carne, y un frasc / spanish
Train accuracy: 0.914
3500 35% (5m 14s) 0.1883 0.1424 u olarak bir koc getirece / turkish
Train accuracy: 0.9268
3525 35% (5m 16s) 0.2167 0.1955  taro i te whenua; He wai / maori
Train accuracy: 0.9128
3550 35% (5m 18s) 0.2316 0.1881 kemek. Tizenketten vagyun / hungarian
Train accuracy: 0.92
3575 35% (5m 20s) 0.3265 0.1968 o incenso. Entao disse o  / portuguese
Train accuracy: 0.9192
3600 36% (5m 23s) 0.2751 0.2101 or sagde de til ham: "Hvo / danish
Train accuracy: 0.9156
3625 36% (5m 25s) 0.2190 0.1651  grosses Mahl und lud vie / german
Train accuracy: 0.922
3650 36% (5m 27s) 0.2145 0.1713 tisar altisar saf altin m / swedish  (turkish)
Train accuracy: 0.9252
3675 36% (5m 29s) 0.3655 0.3103 inausging, fand er einen  / german
Train accuracy: 0.9104
3700 37% (5m 32s) 0.3106 0.2244  eunuques qui sont nes te / french
Train accuracy: 0.9112
3725 37% (5m 34s) 0.1756 0.1632 m, ukuba ithe yalumka int / xhosa
Train accuracy: 0.9184
3750 37% (5m 36s) 0.3195 0.2523 ncseim kozott? Enyem a bo / hungarian
Train accuracy: 0.9044
```

```
3775 37% (5m 38s) 0.3033 0.2932 have ye not returned unto / english
Train accuracy: 0.92
3800 38% (5m 41s) 0.3314 0.2672 d som Herren har lagt mot / norwegian
Train accuracy: 0.9144
3825 38% (5m 43s) 0.2729 0.2424 agit brevet av sandebuden / swedish
Train accuracy: 0.9208
3850 38% (5m 45s) 0.3328 0.3226 i usurat.`` Saul a raspun / romanian
Train accuracy: 0.9152
3875 38% (5m 47s) 0.2856 0.2580 i Dagenes Haelvt! Dine Ar / danish
Train accuracy: 0.91
3900 39% (5m 49s) 0.1861 0.1762 f] thanksgiving, unto the / english
Train accuracy: 0.9236
3925 39% (5m 52s) 0.1484 0.1215 ah, med ild helt til Medb / danish  (norwegian)
Train accuracy: 0.9272
3950 39% (5m 54s) 0.2207 0.1895 ca conhecestes. Ora, quan / portuguese
Train accuracy: 0.9188
3975 39% (5m 56s) 0.2234 0.2057  dado, y aprestaronlo, e  / spanish
Train accuracy: 0.9224
4000 40% (5m 58s) 0.2276 0.2242  kalosh permes ujerave un / albanian
Train accuracy: 0.9184
4025 40% (6m 1s) 0.1856 0.1464 . Egli recide lo spirito  / italian
Train accuracy: 0.9152
4050 40% (6m 3s) 0.1726 0.1503 s tij nje bekim, nje ofer / albanian
Train accuracy: 0.9268
4075 40% (6m 5s) 0.1603 0.1504  ara e nga kaumatua ratou / maori
Train accuracy: 0.9228
4100 41% (6m 7s) 0.1750 0.1573  gudar. Allenast HERREN,  / swedish
Train accuracy: 0.9236
4125 41% (6m 9s) 0.1563 0.1188 jen juureen ja kuunteli h / finnish
Train accuracy: 0.9132
4150 41% (6m 12s) 0.1846 0.1861 trabate!` Ce! vrei s'apuc / romanian
Train accuracy: 0.9156
4175 41% (6m 14s) 0.3100 0.2138  a te atyadfianak: Atyamf / hungarian
Train accuracy: 0.912
4200 42% (6m 16s) 0.1892 0.1627  ma intet ligge Natten ov / danish
Train accuracy: 0.916
4225 42% (6m 18s) 0.2165 0.1721 con! Pro tio, kiel la lan / esperanto
Train accuracy: 0.9196
4250 42% (6m 21s) 0.2804 0.1988 ice blastama zilele, de c / romanian
Train accuracy: 0.914
4275 42% (6m 23s) 0.2124 0.1776 aci ca sa va scap viata m / romanian
Train accuracy: 0.9152
4300 43% (6m 25s) 0.2768 0.2149  kiun la Eternulo sendos  / esperanto
Train accuracy: 0.9164
4325 43% (6m 27s) 0.2014 0.1630 i Moise, si a zis: ,,Vorb / romanian
Train accuracy: 0.9144
4350 43% (6m 29s) 0.1776 0.1635 du genom detta bliver ore / swedish
Train accuracy: 0.9224
```

```
4375 43% (6m 32s) 0.3957 0.3296 herbeje Baalit dhe te bje / albanian
Train accuracy: 0.9188
4400 44% (6m 34s) 0.1299 0.1034 de los Judios. Este vino  / spanish
Train accuracy: 0.912
4425 44% (6m 36s) 0.2073 0.1819 do mataram a Saul em Gilb / portuguese
Train accuracy: 0.918
4450 44% (6m 38s) 0.1602 0.1401 fanno il male. Egli rende / italian
Train accuracy: 0.922
4475 44% (6m 41s) 0.2407 0.2550 i, mot vai nguoi trong ca / vietnamese
Train accuracy: 0.9264
4500 45% (6m 43s) 0.2079 0.1593 rceiro do reinado do rei  / portuguese
Train accuracy: 0.9216
4525 45% (6m 45s) 0.2380 0.1729  mig bak din rygg; sa bae / norwegian
Train accuracy: 0.9292
4550 45% (6m 47s) 0.2192 0.2191 awawathethayo kukumkani,  / xhosa
Train accuracy: 0.9144
4575 45% (6m 50s) 0.2574 0.1911  mesa, e poras em ordem o / portuguese
Train accuracy: 0.9124
4600 46% (6m 52s) 0.3005 0.2072 'Ka jus turite bendro su  / lithuanian
Train accuracy: 0.9184
4625 46% (6m 54s) 0.2911 0.2397 e uni, niin mina tiedan,  / finnish
Train accuracy: 0.912
4650 46% (6m 56s) 0.4555 0.3102 e mis rebeliones esta lig / spanish
Train accuracy: 0.9196
4675 46% (6m 59s) 0.1670 0.1517  Izraeli is filistinu ir  / lithuanian
Train accuracy: 0.9256
4700 47% (7m 1s) 0.2808 0.2309  spravne v Hospodinovych  / czech
Train accuracy: 0.9196
4725 47% (7m 3s) 0.3275 0.2652  de Jafe: Gomer, Magogue, / danish  (portuguese)
Train accuracy: 0.9156
4750 47% (7m 5s) 0.0791 0.0846 a folyam elott, es ket sz / hungarian
Train accuracy: 0.9308
4775 47% (7m 7s) 0.1704 0.1167 ankas, Tu vis tiek istumt / lithuanian
Train accuracy: 0.926
4800 48% (7m 10s) 0.2806 0.2337  tu, abamphendula nazwi;  / xhosa
Train accuracy: 0.9204
4825 48% (7m 12s) 0.2367 0.2387 ois kaupungista. Myos nam / finnish
Train accuracy: 0.9248
4850 48% (7m 14s) 0.2228 0.1710 iramot, Iehiel, Uni, Elia / romanian
Train accuracy: 0.9296
4875 48% (7m 17s) 0.2161 0.2049 tiu ordono, kiun mi donas / esperanto
Train accuracy: 0.9128
4900 49% (7m 19s) 0.1761 0.1468 khangela komntu; ngokuba  / xhosa
Train accuracy: 0.9192
4925 49% (7m 21s) 0.1216 0.1482 ebusuku, bambe, bathi ke  / xhosa
Train accuracy: 0.9176
4950 49% (7m 23s) 0.2588 0.2169 athe kwindili yakho, ubek / xhosa
Train accuracy: 0.92
```

```
4975 49% (7m 26s) 0.2103 0.1882 i vetem. Prandaj Ati me d / albanian
Train accuracy: 0.9288
5000 50% (7m 28s) 0.3039 0.2367  ca sa vad daca mai traie / vietnamese
(romanian)
Train accuracy: 0.9168
5025 50% (7m 30s) 0.2950 0.2359 jeneres Bon, vi, som gern / danish
Train accuracy: 0.9392
5050 50% (7m 32s) 0.1771 0.1654 n Vidnesbyrdet, skal Aron / danish
Train accuracy: 0.9316
5075 50% (7m 35s) 0.1170 0.1206 bueno para comer: tambien / spanish
Train accuracy: 0.9192
5100 51% (7m 37s) 0.2391 0.1776 a dal je na kamenne dlazd / czech
Train accuracy: 0.932
5125 51% (7m 39s) 0.1892 0.1379 mri, unyana kaImri,  unya / xhosa
Train accuracy: 0.9256
5150 51% (7m 41s) 0.3016 0.2423  para reconciliar en el s / spanish
Train accuracy: 0.934
5175 51% (7m 43s) 0.2025 0.1572 inh cua vua A-sue-ru, dan / vietnamese
Train accuracy: 0.9264
5200 52% (7m 46s) 0.4053 0.3215 okuyigunyazela imini yoku / xhosa
Train accuracy: 0.9208
5225 52% (7m 48s) 0.2298 0.1516 . Sen viha on raadellut l / finnish
Train accuracy: 0.9348
5250 52% (7m 50s) 0.1589 0.1228 e gje qe nuk ekziston me. / albanian
Train accuracy: 0.9252
5275 52% (7m 52s) 0.2048 0.1486 olor del unguento. Y dijo / spanish
Train accuracy: 0.9252
5300 53% (7m 55s) 0.1779 0.1867 ovsangs Toner og indfri d / danish
Train accuracy: 0.9132
5325 53% (7m 57s) 0.3162 0.2442 n kann; er und Oholiab, d / german
Train accuracy: 0.9072
5350 53% (7m 59s) 0.2541 0.2625 cau nguyen cung Duc Chua  / vietnamese
Train accuracy: 0.9152
5375 53% (8m 1s) 0.2508 0.2126 go wentente yokuhlangana; / xhosa
Train accuracy: 0.9212
5400 54% (8m 4s) 0.2507 0.1730 estoupil do sve zahrady k / czech
Train accuracy: 0.9248
5425 54% (8m 6s) 0.3042 0.2389 kulwa. Baya kuhlala elowo / xhosa
Train accuracy: 0.914
5450 54% (8m 8s) 0.1648 0.1607 daton el sia gardistaro,  / esperanto
Train accuracy: 0.9172
5475 54% (8m 10s) 0.3053 0.2266  azert tartottalak fenn t / hungarian
Train accuracy: 0.9284
5500 55% (8m 13s) 0.3343 0.2803 ntendez aujourd'hui ces p / french
Train accuracy: 0.9232
5525 55% (8m 15s) 0.3110 0.2482 s, conforme lhes fora man / portuguese
Train accuracy: 0.9192
5550 55% (8m 17s) 0.1221 0.1063 , Y extiende hacia el med / spanish
```

```
Train accuracy: 0.9276
5575 55% (8m 19s) 0.2493 0.1580  nirgends vorgekommen und / swedish  (german)
Train accuracy: 0.9192
5600 56% (8m 22s) 0.3334 0.3089 rakore: tataitia e ia he  / maori
Train accuracy: 0.9212
5625 56% (8m 24s) 0.2294 0.1830 terno, il re si rallegra  / italian
Train accuracy: 0.9352
5650 56% (8m 26s) 0.1782 0.1679 antanut Edomin, Esaun suv / finnish
Train accuracy: 0.922
5675 56% (8m 29s) 0.2694 0.2230 aikovat tehda?" Han vasta / finnish
Train accuracy: 0.9172
5700 56% (8m 31s) 0.1860 0.1664  ham!" Jeg vil synge din  / norwegian  (danish)
Train accuracy: 0.92
5725 57% (8m 33s) 0.2219 0.1397 te gjalle. Kjo u muar ves / albanian
Train accuracy: 0.926
5750 57% (8m 35s) 0.1545 0.1566 ations n'appartiennent-el / french
Train accuracy: 0.9268
5775 57% (8m 38s) 0.2647 0.2222 as si te shkoje ne qytet. / albanian
Train accuracy: 0.9236
5800 57% (8m 40s) 0.1942 0.1790 dit te Davidit. Keshtu th / albanian
Train accuracy: 0.9264
5825 58% (8m 42s) 0.2120 0.1738  disse Maend Blar i Ojnen / danish
Train accuracy: 0.9176
5850 58% (8m 44s) 0.2452 0.2404 asivat: "Siirrettakoon Is / finnish
Train accuracy: 0.9268
5875 58% (8m 46s) 0.2005 0.1438  kite ai, e kore ano e wh / maori
Train accuracy: 0.9216
5900 59% (8m 49s) 0.1030 0.0869 tria tago iu viro venis e / esperanto
Train accuracy: 0.9236
5925 59% (8m 51s) 0.1699 0.1444 emate nieko gero. Jos pra / lithuanian
Train accuracy: 0.9204
5950 59% (8m 53s) 0.2179 0.2131 ls ho, na sveho pomazaneh / czech
Train accuracy: 0.9284
5975 59% (8m 56s) 0.2957 0.2143  kaluar; dhe Zoti e kisht / albanian
Train accuracy: 0.9188
6000 60% (8m 58s) 0.1433 0.1161 a. A rite tahi ki ta Ihow / maori
Train accuracy: 0.9244
6025 60% (9m 0s) 0.2393 0.1758 , sa at den blev fuldstae / norwegian  (danish)
Train accuracy: 0.9176
6050 60% (9m 2s) 0.2720 0.2356  ngentlawulelo nokwanana, / xhosa
Train accuracy: 0.9296
6075 60% (9m 5s) 0.1882 0.1655 nobubi; Ke lona ilungisa  / xhosa
Train accuracy: 0.9284
6100 61% (9m 7s) 0.2536 0.1786 our brethren, like unto m / english
Train accuracy: 0.9248
6125 61% (9m 9s) 0.1472 0.1359 hjel profetene, og stener / norwegian
Train accuracy: 0.9252
6150 61% (9m 11s) 0.3062 0.2878 minh noi chi, tai ca ba d / vietnamese
```

```
Train accuracy: 0.9224
6175 61% (9m 14s) 0.2099 0.1763 ta ano koutou ki te hunga / maori
Train accuracy: 0.916
6200 62% (9m 16s) 0.1226 0.1025 aciligiyla gonderdigi buy / turkish
Train accuracy: 0.9308
6225 62% (9m 18s) 0.1846 0.1696 ono le offerte che fate!  / italian
Train accuracy: 0.9208
6250 62% (9m 20s) 0.1857 0.1442 nua ja nosti minut jallee / finnish
Train accuracy: 0.9268
6275 62% (9m 22s) 0.2056 0.1843 I Abrahams gjerninger; me / norwegian
Train accuracy: 0.936
6300 63% (9m 25s) 0.3175 0.2424 a mi segunda vez palabra  / romanian  (spanish)
Train accuracy: 0.934
6325 63% (9m 27s) 0.1205 0.0978  bekuar ne perjetesi. Ngj / albanian
Train accuracy: 0.9228
6350 63% (9m 29s) 0.1077 0.1142 t auch keiner uber den Ta / german
Train accuracy: 0.9248
6375 63% (9m 31s) 0.2089 0.1815  kei roto i tona ngakau.  / maori
Train accuracy: 0.932
6400 64% (9m 34s) 0.2346 0.1866 agon annan gud, ty HERREN / swedish
Train accuracy: 0.9376
6425 64% (9m 36s) 0.2721 0.2079 ra Sinaj, fra Seir fremst / norwegian  (danish)
Train accuracy: 0.9304
6450 64% (9m 38s) 0.2554 0.2062 zaron la voz, diciendo: Q / spanish
Train accuracy: 0.9284
6475 64% (9m 40s) 0.1280 0.1332 tonta. Tosi ystava ei jat / finnish
Train accuracy: 0.9276
6500 65% (9m 43s) 0.2882 0.2368 aulus's Sosterson, som ha / danish
Train accuracy: 0.938
6525 65% (9m 45s) 0.1553 0.0917 ne moi si ceice traiesc i / romanian
Train accuracy: 0.9276
6550 65% (9m 47s) 0.2142 0.1886 t igjennem av ekte saed;  / norwegian
Train accuracy: 0.9204
6575 65% (9m 49s) 0.2812 0.2284 atrilo Elimelekovi, i vse / czech
Train accuracy: 0.9264
6600 66% (9m 51s) 0.2046 0.1940  lai noi run bien. Quan n / vietnamese
Train accuracy: 0.9156
6625 66% (9m 54s) 0.1499 0.1468  a Lehi, i Filistei gli s / italian
Train accuracy: 0.9256
6650 66% (9m 56s) 0.2288 0.1771 mme natt kom Guds ord til / norwegian
Train accuracy: 0.9352
6675 66% (9m 58s) 0.3451 0.3228 omme nous, la voix du Die / french
Train accuracy: 0.926
6700 67% (10m 0s) 0.2491 0.2009  est une lampe, l'enseign / french
Train accuracy: 0.9268
6725 67% (10m 3s) 0.1782 0.1455 , kili, gumusu, altini pa / turkish
Train accuracy: 0.928
6750 67% (10m 5s) 0.1552 0.1550 j batalarangxigxis antaux / esperanto
```

```
Train accuracy: 0.9364
6775 67% (10m 7s) 0.2907 0.2686 i uma voz que nao conheci / portuguese
Train accuracy: 0.9324
6800 68% (10m 9s) 0.3121 0.2564 e, jeho predvoj do more v / czech
Train accuracy: 0.918
6825 68% (10m 12s) 0.0894 0.0920 an se. Da sade grannarna  / swedish
Train accuracy: 0.9312
6850 68% (10m 14s) 0.1094 0.1030 na iho e ia te ingoa ko I / maori
Train accuracy: 0.9336
6875 68% (10m 16s) 0.2066 0.2037 muerte o de prision. Mas  / spanish
Train accuracy: 0.9328
6900 69% (10m 19s) 0.2299 0.2437 s; porque virao muitos em / portuguese
Train accuracy: 0.9272
6925 69% (10m 21s) 0.2481 0.2381  omalta tieltaan, hyva ih / finnish
Train accuracy: 0.9268
6950 69% (10m 23s) 0.1057 0.1061 esua, novecentos e setent / portuguese
Train accuracy: 0.926
6975 69% (10m 25s) 0.2381 0.1704 u sa ma omoare din pricin / romanian
Train accuracy: 0.9312
7000 70% (10m 28s) 0.2411 0.2115 ohi, ka miharo ki taua wh / maori
Train accuracy: 0.9276
7025 70% (10m 30s) 0.1340 0.0982 ac se trapit?<< a vse odb / czech
Train accuracy: 0.9372
7050 70% (10m 32s) 0.3167 0.2852 emieres voies de David, s / french
Train accuracy: 0.9388
7075 70% (10m 34s) 0.3572 0.2159  whenua o Ahiria, me te h / maori
Train accuracy: 0.93
7100 71% (10m 37s) 0.2665 0.2227 e peraisin taalta." "Sina / finnish
Train accuracy: 0.9272
7125 71% (10m 39s) 0.2358 0.1986 r - skal de kende, at en  / norwegian
(danish)
Train accuracy: 0.928
7150 71% (10m 41s) 0.1637 0.1577 nri. Gerdekten cikan guve / turkish
Train accuracy: 0.932
7175 71% (10m 43s) 0.2512 0.2224  de fremmede Guder, de fo / danish
Train accuracy: 0.9272
7200 72% (10m 46s) 0.3382 0.2318 natanas mire?" Jaunuolis, / lithuanian
Train accuracy: 0.922
7225 72% (10m 48s) 0.1950 0.1957 avide era fuggito da Keil / portuguese
(italian)
Train accuracy: 0.9208
7250 72% (10m 50s) 0.2805 0.2248 l. U cherubu bylo totiz v / czech
Train accuracy: 0.9224
7275 72% (10m 52s) 0.1215 0.1298 z his son. Now these are  / english
Train accuracy: 0.9352
7300 73% (10m 55s) 0.1713 0.1421 he te parelindurit meshku / albanian
Train accuracy: 0.9268
7325 73% (10m 57s) 0.2711 0.2073 ob, rey de Soba. Y gano D / spanish
```

```
Train accuracy: 0.9256
7350 73% (10m 59s) 0.1501 0.1144 , ci noua, martorilor ale / romanian
Train accuracy: 0.9316
7375 73% (11m 1s) 0.3396 0.2263  sprach zum Konig Josapha / german
Train accuracy: 0.9188
7400 74% (11m 3s) 0.1963 0.1849 Assieds-toi a ma droite,  / french
Train accuracy: 0.9352
7425 74% (11m 6s) 0.2926 0.1923  a ti Istenetek, o az ist / hungarian
Train accuracy: 0.9312
7450 74% (11m 8s) 0.1949 0.1742  gordugunuz gibi RAB onla / turkish
Train accuracy: 0.9364
7475 74% (11m 10s) 0.2420 0.1997  tenei ki no katoa kia pa / maori
Train accuracy: 0.9248
7500 75% (11m 12s) 0.1808 0.1276 aya kuMordekayi endaweni  / xhosa
Train accuracy: 0.9216
7525 75% (11m 15s) 0.2042 0.1826  tekau mano tangata. A, n / maori
Train accuracy: 0.9332
7550 75% (11m 17s) 0.1332 0.1618 menor deles ate o maior,  / portuguese
Train accuracy: 0.9336
7575 75% (11m 19s) 0.2889 0.2202 laronse de que hablaba co / spanish
Train accuracy: 0.9316
7600 76% (11m 22s) 0.2217 0.2330  altigos;   Kaj gxi donos / esperanto
Train accuracy: 0.9308
7625 76% (11m 24s) 0.2506 0.2017 samblene, sic u kishin th / albanian
Train accuracy: 0.9424
7650 76% (11m 26s) 0.2514 0.2491 ,  phambi kwam, lide ixes / xhosa
Train accuracy: 0.928
7675 76% (11m 28s) 0.1713 0.1328 ternes Hovdinger blev vre / danish
Train accuracy: 0.9412
7700 77% (11m 30s) 0.1456 0.1429 alle antanut. Joosua polt / finnish
Train accuracy: 0.926
7725 77% (11m 33s) 0.1308 0.0640 . And Laban said to Jacob / english
Train accuracy: 0.9324
7750 77% (11m 35s) 0.3371 0.2816 t erdhen per te protestua / albanian
Train accuracy: 0.93
7775 77% (11m 37s) 0.1836 0.1782 ugir, e o tremor apoderou / portuguese
Train accuracy: 0.9272
7800 78% (11m 39s) 0.1991 0.1492 oz, mondvan: Vigyazz, ne  / hungarian
Train accuracy: 0.9248
7825 78% (11m 42s) 0.1682 0.1533 m. And Paul, looking sted / english
Train accuracy: 0.9212
7850 78% (11m 44s) 0.1846 0.1204 ouse,) and cry unto thee  / english
Train accuracy: 0.9264
7875 78% (11m 46s) 0.3122 0.2281 behat kone po skalisku? M / czech
Train accuracy: 0.9256
7900 79% (11m 48s) 0.1443 0.1031  delante de sus hermanos  / spanish
Train accuracy: 0.9424
7925 79% (11m 50s) 0.2201 0.2225 -au zis: ,,Atarot, Dibon, / romanian
```

```
Train accuracy: 0.9276
7950 79% (11m 53s) 0.1066 0.0911 is en prison; et maintena / french
Train accuracy: 0.9316
7975 79% (11m 55s) 0.0750 0.0932 a ai o matou ko nga Parih / maori
Train accuracy: 0.9292
8000 80% (11m 57s) 0.1958 0.1820  mai fi pentru casa lui I / romanian
Train accuracy: 0.9212
8025 80% (11m 59s) 0.1872 0.1929 s derlinguose laukuose. T / lithuanian
Train accuracy: 0.9272
8050 80% (12m 2s) 0.2900 0.2061 ? he oranga ngakau ra me  / maori
Train accuracy: 0.93
8075 80% (12m 4s) 0.1027 0.0783 m." Ale Abraham se tim ve / czech
Train accuracy: 0.93
8100 81% (12m 6s) 0.2280 0.1963 e to Dotre, som bor hos d / danish
Train accuracy: 0.9452
8125 81% (12m 8s) 0.2355 0.1741 time entiere; alors on of / french
Train accuracy: 0.9296
8150 81% (12m 11s) 0.2949 0.2456  <<Lutfen onlari yanima g / turkish
Train accuracy: 0.9416
8175 81% (12m 13s) 0.2318 0.1993  en cxi tiu urbo estas Di / esperanto
Train accuracy: 0.934
8200 82% (12m 15s) 0.2367 0.1736 une garcon de Succoth, qu / french
Train accuracy: 0.9288
8225 82% (12m 17s) 0.1586 0.1303 uya kuba yini na umqondis / xhosa
Train accuracy: 0.9304
8250 82% (12m 19s) 0.2323 0.1660 te. Io faro di Rabba un p / italian
Train accuracy: 0.9332
8275 82% (12m 22s) 0.2641 0.2093 grei ao meu nome, lanca-l / portuguese
Train accuracy: 0.9296
8300 83% (12m 24s) 0.2272 0.1467  fists? Who hath bound th / english
Train accuracy: 0.9348
8325 83% (12m 26s) 0.1617 0.1722 la tendoj de Kedar! Tro l / esperanto
Train accuracy: 0.9336
8350 83% (12m 28s) 0.1809 0.1332 n. Jungu la cxevalojn kaj / esperanto
Train accuracy: 0.928
8375 83% (12m 31s) 0.2713 0.1900 o malkvietigxis,  dirante / esperanto
Train accuracy: 0.93
8400 84% (12m 33s) 0.2455 0.2060 tit ne funksion ato dite  / italian
(albanian)
Train accuracy: 0.9276
8425 84% (12m 35s) 0.2600 0.2159 p oturdular. Kutsal Yasa  / turkish
Train accuracy: 0.9384
8450 84% (12m 37s) 0.2073 0.1884 na gelir?>> diye dusundu. / turkish
Train accuracy: 0.936
8475 84% (12m 40s) 0.1501 0.1234 istas alia Dio krom Mi, l / esperanto
Train accuracy: 0.9408
8500 85% (12m 42s) 0.2001 0.1491 Gott mich behutete? als s / german
Train accuracy: 0.9208
```

```
8525 85% (12m 44s) 0.2212 0.1644 dieser unserer Sache auss / german
Train accuracy: 0.9312
8550 85% (12m 46s) 0.2353 0.2384 dessaan Egyptin maata vas / finnish
Train accuracy: 0.9296
8575 85% (12m 49s) 0.1626 0.1348 , perziej beharnat dhe le / albanian
Train accuracy: 0.9308
8600 86% (12m 51s) 0.2550 0.1667 eo den co it nguoi, Duc G / vietnamese
Train accuracy: 0.9332
8625 86% (12m 53s) 0.2171 0.1594  und brachten sie mit sic / german
Train accuracy: 0.9316
8650 86% (12m 55s) 0.2455 0.2069 : "Te kaikki luovutte min / finnish
Train accuracy: 0.92
8675 86% (12m 57s) 0.1789 0.1251 oir salue l'Eglise, il de / french
Train accuracy: 0.926
8700 87% (13m 0s) 0.2523 0.1532 Mikor eljutanak Atad szer / hungarian
Train accuracy: 0.93
8725 87% (13m 2s) 0.2874 0.2561 Wenn ihr es noch einmal t / german
Train accuracy: 0.9212
8750 87% (13m 4s) 0.1507 0.1212  te Kuq. Dhe bijte e Izra / albanian
Train accuracy: 0.9296
8775 87% (13m 6s) 0.1502 0.1026 . Ora, sucedeu no mes de  / portuguese
Train accuracy: 0.9364
8800 88% (13m 9s) 0.2629 0.2122 r. Bundan dolayi Israil h / turkish
Train accuracy: 0.944
8825 88% (13m 11s) 0.2268 0.1756  krejt krahinen malore te / albanian
Train accuracy: 0.932
8850 88% (13m 13s) 0.1753 0.1012 chaumte. Und er fragte se / german
Train accuracy: 0.9348
8875 88% (13m 15s) 0.1416 0.1278 de mao e as lancas; acend / portuguese
Train accuracy: 0.9268
8900 89% (13m 18s) 0.1719 0.1152 im>> dedi. Boylece ikisi  / turkish
Train accuracy: 0.9376
8925 89% (13m 20s) 0.2270 0.1549 tarnauja Dievui. Jie jums / lithuanian
Train accuracy: 0.9324
8950 89% (13m 22s) 0.2937 0.2000 unu bana ayir. Israillile / turkish
Train accuracy: 0.9352
8975 89% (13m 24s) 0.2009 0.1705 orran man holjde over Ham / swedish
Train accuracy: 0.9344
9000 90% (13m 27s) 0.1975 0.1085 en el libro de las cronic / spanish
Train accuracy: 0.9276
9025 90% (13m 29s) 0.1566 0.1498 ricxulojn kaj pereigi la  / esperanto
Train accuracy: 0.9256
9050 90% (13m 31s) 0.2882 0.2144 n dem. Og porten til den  / norwegian
Train accuracy: 0.9272
9075 90% (13m 34s) 0.1816 0.1418  quem a ajunta pouco a po / portuguese
Train accuracy: 0.932
9100 91% (13m 36s) 0.1306 0.1495  med mig selv, men dermed / norwegian
Train accuracy: 0.936
```

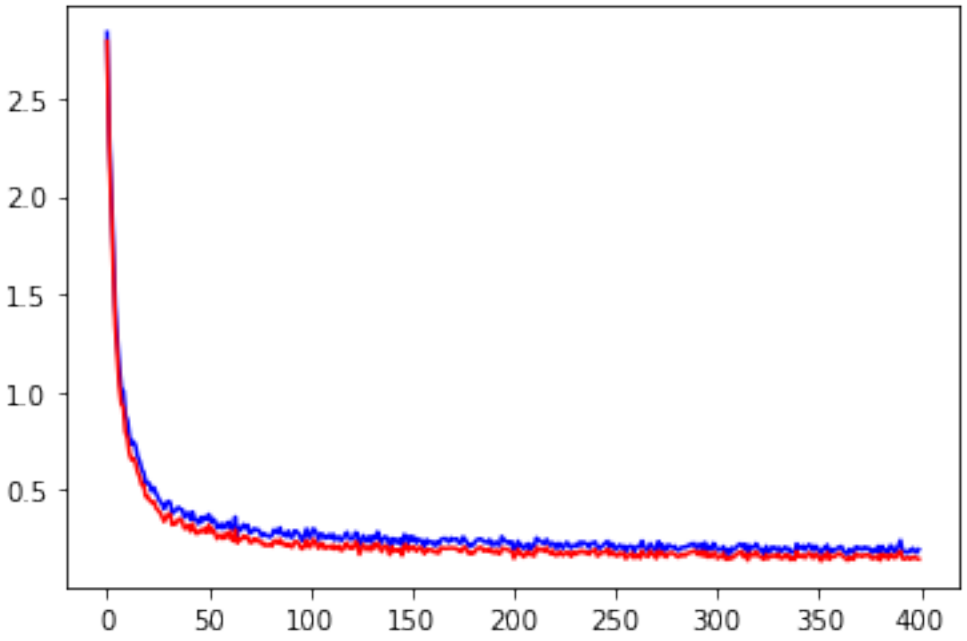
```
9125 91% (13m 38s) 0.1106 0.0607 t aruncat in locuinta mor / romanian
Train accuracy: 0.9404
9150 91% (13m 40s) 0.1958 0.1794 orkemini gorur. Utansin p / turkish
Train accuracy: 0.9308
9175 91% (13m 43s) 0.3340 0.2862 e porte della casa dell'E / italian
Train accuracy: 0.9368
9200 92% (13m 45s) 0.1451 0.1231 eropalcza vala hat singny / hungarian
Train accuracy: 0.9416
9225 92% (13m 47s) 0.3169 0.2983  rog sa nu faceti un lucr / romanian
Train accuracy: 0.9272
9250 92% (13m 49s) 0.3134 0.2665 det ar en vit flack som s / swedish
Train accuracy: 0.9316
9275 92% (13m 51s) 0.2501 0.2184  svek och fortryck; hans  / swedish
Train accuracy: 0.934
9300 93% (13m 54s) 0.0827 0.0776 pats kalbejo Mozei: "Tai  / lithuanian
Train accuracy: 0.93
9325 93% (13m 56s) 0.1874 0.1320 ber stellst mir nach dem  / german
Train accuracy: 0.9288
9350 93% (13m 58s) 0.1925 0.1550 . Gud hjalp ham mot filis / norwegian
Train accuracy: 0.9364
9375 93% (14m 0s) 0.2395 0.1965 ne ekzistas la gloro de M / esperanto
Train accuracy: 0.9276
9400 94% (14m 3s) 0.3174 0.2399 dlnost, pomozte vyvaznout / czech
Train accuracy: 0.932
9425 94% (14m 5s) 0.2416 0.2218 a si de ploaie. Preotul E / romanian
Train accuracy: 0.9352
9450 94% (14m 7s) 0.1186 0.0962 ko. Kaj konstruu tie alta / esperanto
Train accuracy: 0.9368
9475 94% (14m 9s) 0.2076 0.1611 ei a Ihowa; Kei rongo kou / maori
Train accuracy: 0.9288
9500 95% (14m 12s) 0.2354 0.2031 i tij u perhapen, dhe shu / albanian
Train accuracy: 0.9276
9525 95% (14m 14s) 0.2341 0.1750  l'evangelo di Dio e dice / italian
Train accuracy: 0.9416
9550 95% (14m 16s) 0.1931 0.1569 gi jus nezinote? Argi neg / lithuanian
Train accuracy: 0.9308
9575 95% (14m 18s) 0.1975 0.1741 es con sus heredades. Toc / spanish
Train accuracy: 0.9336
9600 96% (14m 21s) 0.1454 0.1024 a: Israelissa oli monta l / finnish
Train accuracy: 0.9324
9625 96% (14m 23s) 0.2527 0.1907 e fare siguri per kedo qe / italian
(albanian)
Train accuracy: 0.9368
9650 96% (14m 25s) 0.2689 0.2130 . And Pilate answered the / english
Train accuracy: 0.9236
9675 96% (14m 27s) 0.2483 0.2072 esti uhrata nuori sonni H / finnish
Train accuracy: 0.9344
9700 97% (14m 30s) 0.2112 0.1783 geentlobo ezine, utsho uY / xhosa
```

```
Train accuracy: 0.9336
9725 97% (14m 32s) 0.2989 0.1962  erfahren, dass ich der H / german
Train accuracy: 0.9376
9750 97% (14m 34s) 0.1627 0.1407 n; mas muerto el, resucit / romanian
(spanish)
Train accuracy: 0.9276
9775 97% (14m 36s) 0.3614 0.2514 l virto. Kaj Li diris al  / esperanto
Train accuracy: 0.914
9800 98% (14m 38s) 0.1616 0.1542 roztrhl sve roucho a pred / czech
Train accuracy: 0.9332
9825 98% (14m 41s) 0.1019 0.1034 edrejaras o tal homem, ou / portuguese
Train accuracy: 0.9316
9850 98% (14m 43s) 0.1706 0.1496 valo Dobre pristavy, neda / czech
Train accuracy: 0.9372
9875 98% (14m 45s) 0.1383 0.1129  duch Hospodinuv a on zat / czech
Train accuracy: 0.9412
9900 99% (14m 47s) 0.2172 0.1047 tettua viinia ja granaatt / finnish
Train accuracy: 0.9336
9925 99% (14m 49s) 0.2011 0.1421 oined battle, Israel was  / english
Train accuracy: 0.9296
9950 99% (14m 52s) 0.2252 0.1877 rcondava la casa d'ogn'in / italian
Train accuracy: 0.934
9975 99% (14m 54s) 0.1286 0.1058 zeben telepitsd le a te a / hungarian
Train accuracy: 0.9428
10000 100% (14m 56s) 0.3415 0.2038 k neznas dilo Boha, ktery / esperanto
(czech)
Train accuracy: 0.9292
```

4.1 Plot loss functions

[12]:
```
import matplotlib.pyplot as plt
import matplotlib.ticker as ticker

plt.figure()
plt.plot(all_losses, color='b')
plt.plot(all_test_losses, color='r')
```

[12]:
```
[<matplotlib.lines.Line2D at 0x28c055b8580>]
```



4.2 Evaluate results

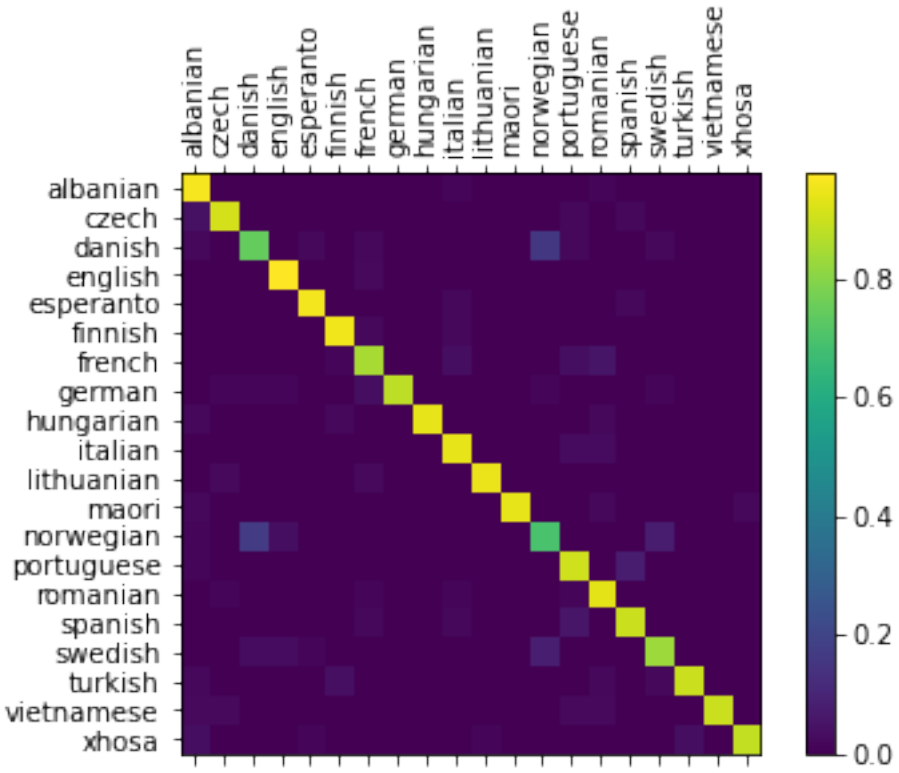
We now vizualize the performance of our model by creating a confusion matrix. The ground truth languages of samples are represented by rows in the matrix while the predicted languages are represented by columns.

In this evaluation we consider sequences of variable sizes rather than the fixed length sequences we used for training.

[13]:
```
eval_batch_size = 1  # needs to be set to 1 for evaluating different sequence␣
 ↪lengths

# Keep track of correct guesses in a confusion matrix
confusion = torch.zeros(n_languages, n_languages)
n_confusion = 1000
num_correct = 0
total = 0

for i in range(n_confusion):
    eval_chunk_len = random.randint(10, 50) # in evaluation we will look at␣
 ↪sequences of variable sizes
    input_data, target_category, text_data =␣
 ↪load_random_batch(test_category_data, chunk_len=eval_chunk_len,␣
 ↪batch_size=eval_batch_size)
    output = evaluate(lstm, input_data, seq_len=eval_chunk_len,␣
 ↪batch_size=eval_batch_size)
```

```
    guess_i = categoryFromOutput(output)
    category_i = [int(target_category[idx]) for idx in␣
 ↪range(len(target_category))]
    for j in range(eval_batch_size):
        category = all_categories[category_i[j]]
        confusion[category_i[j]][guess_i[j]] += 1
        num_correct += int(guess_i[j]==category_i[j])
        total += 1

print('Test accuracy: ', float(num_correct)/float(n_confusion*eval_batch_size))

# Normalize by dividing every row by its sum
for i in range(n_languages):
    confusion[i] = confusion[i] / confusion[i].sum()

# Set up plot
fig = plt.figure()
ax = fig.add_subplot(111)
cax = ax.matshow(confusion.numpy())
fig.colorbar(cax)

# Set up axes
ax.set_xticklabels(['']  + all_categories, rotation=90)
ax.set_yticklabels(['']  + all_categories)

# Force label at every tick
ax.xaxis.set_major_locator(ticker.MultipleLocator(1))
ax.yaxis.set_major_locator(ticker.MultipleLocator(1))

plt.show()
```

```
Test accuracy:  0.896
```



You can pick out bright spots off the main axis that show which languages it guesses incorrectly.

4.3 Run on User Input

Now you can test your model on your own input.

[14]:
```
def predict(input_line, n_predictions=5):
    print('\n> %s' % input_line)
    with torch.no_grad():
        input_data = stringToTensor(input_line).long().unsqueeze(0).to(device)
        output = evaluate(lstm, input_data, seq_len=len(input_line),␣
↪batch_size=1)

    # Get top N categories
    topv, topi = output.topk(n_predictions, dim=1)
    predictions = []

    for i in range(n_predictions):
        topv.shape
        topi.shape
        value = topv[0][i].item()
        category_index = topi[0][i].item()
```

```
        print('(%.2f) %s' % (value, all_categories[category_index]))
        predictions.append([value, all_categories[category_index]])

predict('This is a phrase to test the model on user input')
```

```
> This is a phrase to test the model on user input
(11.74) english
(2.80) albanian
(1.39) french
(0.76) spanish
(0.57) german
```

5 Output Kaggle submission file

Once you have found a good set of hyperparameters submit the output of your model on the Kaggle test file.

[15]:
```
import csv

kaggle_test_file_path = 'kaggle_rnn_language_classification_test.txt'
with open(kaggle_test_file_path, 'r') as f:
    lines = f.readlines()

output_rows = []
for i, line in enumerate(lines):
    sample = line.rstrip()
    sample_chunk_len = len(sample)
    input_data = stringToTensor(sample).unsqueeze(0)
    output = evaluate(lstm, input_data, seq_len=sample_chunk_len, batch_size=1)
    guess_i = categoryFromOutput(output)
    output_rows.append((str(i+1), all_categories[guess_i]))

submission_file_path = 'kaggle_rnn_submission.txt'
with open(submission_file_path, 'w') as f:
    output_rows = [('id', 'category')] + output_rows
    writer = csv.writer(f)
    writer.writerows(output_rows)
```