陽明交通大學 資工系大三 111550177 吳定霖

NLP HW3 Report

Part I (Code Explaination) :

Environment :

* 純粹執行在 Google Colab 中即可。
* 我還是有把 colab 的 packages 寫到 requirements.txt 中，但建議助教直接colab 執行。
* Python version 3.10.12

我使用以下的網頁、Ai 工具來完成我的程式碼 :

1. ChatGPT-4

2. Google Colab Ai

3. Gemini 1.5 pro

4. https://www.freecodecamp.org/news/add-to-dict-in-python/

5. https://stackoverflow.com/questions/43389684/how-can-i-split-a-string-of-a-mathematical-expressions-in-python

6. [https://medium.com/@darriuschen0926/python資料處理領域中不常用但很實用的function及用法整理-c5ce123bfc50](https://medium.com/@darriuschen0926/python%E8%B3%87%E6%96%99%E8%99%95%E7%90%86%E9%A0%98%E5%9F%9F%E4%B8%AD%E4%B8%8D%E5%B8%B8%E7%94%A8%E4%BD%86%E5%BE%88%E5%AF%A6%E7%94%A8%E7%9A%84function%E5%8F%8A%E7%94%A8%E6%B3%95%E6%95%B4%E7%90%86-c5ce123bfc50)

7. https://www.geeksforgeeks.org/adding-new-column-to-existing-dataframe-in-pandas/

8. https://www.runoob.com/python/att-string-isdigit.html

9. https://www.geeksforgeeks.org/python-splitting-operators-in-string/

10. https://codingman.cc/how-to-remove-empty-element-from-list-in-python/

11. https://blog.csdn.net/weixin\_44211968/article/details/128216020

12. <https://pytorch.org/docs/stable/generated/torch.nn.GRU.html>  
我同樣將這些網頁、Ai 工具聲明在我的程式碼中 : 一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

Todo1 :

一張含有 文字, 螢幕擷取畫面, 軟體, 陳列 的圖片

自動產生的描述

我將所有在 df\_train[‘src’] 中的所有字串先將之用 set 取出其中 unique 的 characters，然後再把所有的 set union 得到所有出現在 training data 中的 characters。

接下來就是再把這些 characters 全部和它的 id 存到 char\_to\_id 和 id\_to\_char 的 dictionary 中。

Todo2:

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述  
一張含有 文字, 字型, 螢幕擷取畫面 的圖片

自動產生的描述

我先將 df['src'] 的字串尾加上一個 ‘<eos>’，然後全部透過 char\_to\_id 轉換成 id，就得到 char\_id\_list 了。

再來，我把 char\_id\_list 裡面每一個 sequence 做一個判斷，把 ‘=’ id 之前所有的 id 都變成 ‘<eos>’的 id，得到 label\_id\_list。

Todo3 :

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

我讓 \_\_len\_\_ method return 整個 sequence 的長度。

我讓 \_\_getitem\_\_method return 對應某個 index 的 char\_id\_list 和 label\_id\_list 分別作為 input x 和 target y。

Todo4 :

一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述  
我讓 y = 把 char\_list\_tensor 經過 model forward 之後得到的 predict sequence tensor，我讓 next\_char = 對於整個 sequence 中某個位置所有可能的 characters 中機率最大的那一個用 argmax 取出來，再取出 sequence characters 中最後一個character，也就是 predict 出來的 next character。

Todo5 :



設定好計算 cross entropy loss 的 criterion 還有 optimizer adam。這邊要把 ignore index 設為 ‘<pad>’。

一張含有 文字, 字型, 螢幕擷取畫面 的圖片

自動產生的描述

在 training 的過程中，我用 optimizer.zero\_grad() 先把 gradient 的值清零，然後把對於每一個 batch 的 batch\_x, batch\_x\_lens 丟入 model 中 forward 得到 predict 的 batch\_predict\_y。

接下來把 batch\_predict\_y 和 batch\_y 丟到 criterion 計算出 loss，這邊我是把 batch\_predict\_y[:-1] 和 batch\_y[1:] 進行 loss 計算，這樣才會是 predict ‘=’後的下一個值。，用 loss.backward() 得到 backpropagation 的 gradient，然後經過 gradient clip 把那些梯度爆炸的 gradient 值替換成 1。

最後用 optimizer.step() 更新 model 的權重。

Todo6 :

一張含有 文字, 螢幕擷取畫面 的圖片

自動產生的描述

把 batch\_x, batch\_x\_lens 丟入 model 中計算 forward 後的 predictions y 值，這邊要記得用 torch.no\_grad()，不然下一次 training 的時候會把 evaluation 時 forward 的東西也 backpropagation 回去。然後把所有在 btach 當中， batch\_y 裡面 (t2\_row) 非 ‘<pad>’ id (0) 的值拿出來看看有沒有 match predictions y 中對應的值，如果有的話，count 數量加一，等一個 batch 算完之後再一起加入 match 中，total 直接加整個 batch\_y 的數量，計算出 EM 的值。

Part II (Experiement) :

1. 比較不同模型架構對於模型訓練的影響

一張含有 文字, 繪圖, 行, 圖表 的圖片

自動產生的描述

一張含有 文字, 行, 繪圖, 圖表 的圖片

自動產生的描述

我從 epoch 對於 loss 以及 EM 的圖中觀察到，不同模型架構在訓練結果上的表現差異相當大，反映出它們在處理此類數值運算的 sequence-to-sequence 任務時的優劣。以下是對每個模型表現的分析。

首先，2 Layer LSTM (Baseline) 顯示出最佳的訓練效果。在前10個epoch中，loss快速下降到一個低值，並且在20個epoch後穩定於相對低的水平。從EM曲線來看，2層LSTM的準確率穩步上升並保持較高的穩定性，最後在20個epoch達到0.845左右的值。這顯示2層LSTM在模型深度、記憶保持能力和學習複雜序列依賴方面的平衡度很好。其多層結構的LSTM可以有效捕捉到加減乘序列中的長期依賴關係，避免梯度消失問題。

相比之下，1 Layer LSTM 的表現略低於2層LSTM，loss下降的速度較慢，EM的增長曲線也較為平緩。儘管在20個epoch後接近0.793，但始終不及2層LSTM。這主要是因為1層LSTM的層數較少，模型在學習序列複雜依賴時受限，無法有效捕捉到較深的序列特徵。總體而言，1層LSTM需要更長的訓練時間，並且在達到類似精度方面存在一定瓶頸。

接下來，2 Layer GRU 的表現居於1層LSTM與2層LSTM之間。loss曲線表現出一定震盪，顯示GRU模型在學習時不夠穩定，但在20個epoch後的loss仍降到相對較低的水平。EM曲線也逐步上升，但最高只達到約0.661，顯示2層GRU在處理長期依賴性方面仍不及LSTM。雖然GRU在簡化模型結構上較有優勢，計算效率高，但在這類需要長期記憶的加減乘運算序列上效果稍弱於LSTM。

最後，2 Layer RNN 表現最差，其loss從始至終都在高位徘徊，且EM幾乎無顯著上升，最後穩定在0.051左右，顯示該模型無法有效完成任務。這種結果反映了RNN在長期依賴性方面的先天限制，尤其在梯度消失情況下，RNN難以學習到序列中的複雜運算關係。因此，在這類數值運算序列任務中，RNN不適合作為架構選擇。

綜合來說，2層LSTM (Baseline) 在加減乘運算序列上的表現最佳，能夠高效捕捉序列關聯並避免梯度消失。1層LSTM次之，但需要更多訓練才能達到接近精度，而2層GRU雖能在精度和計算效率間取得平衡，但效果仍稍遜於LSTM。2層RNN則因梯度消失問題難以應用於此類任務。

1. 比較不同 Embedding Size 對於模型訓練的影響

一張含有 繪圖, 行, 圖表, 文字 的圖片

自動產生的描述

一張含有 文字, 繪圖, 行, 圖表 的圖片

自動產生的描述

我從 epoch 對於 loss 以及 EM 的圖中觀察到，不同的 embedding size 對於模型的表現確實有明顯影響。在不同的 embedding size 下，loss 和 EM 的收斂速度以及穩定性都不相同。

首先，當 embedding size 為 64 時 (loss\_list\_1, em\_list\_1)，模型的 loss 一開始下降迅速，並且在前幾個 epoch 內達到了較低的 loss 水平，而 EM 也迅速上升，顯示了 embedding size 為 64 時，模型在初期的收斂速度較快。然而，隨著 epoch 的增長，模型的 loss 有些波動，最終的 loss 和 EM 都略低於較大的 embedding size 結果，可能是因為小的 embedding size 對於複雜的運算符和數字序列表示能力較為有限。

接著，當 embedding size 增加到 128 (loss\_list\_2, em\_list\_2) 和 256 (loss\_list\_3, em\_list\_3) 時，模型在初期的 loss 和 EM 表現同樣良好，而且最終 loss 相對穩定。特別是 embedding size 256，在大部分的 epoch 中模型的 loss 和 EM 保持在較為平滑的狀態下，最後的 EM 甚至接近 0.85 以上，顯示了較大的 embedding size 有助於提升模型在這類數字和符號序列運算的表現。

再來，當 embedding size 進一步增加到 512 (loss\_list\_4, em\_list\_4) 和 1024 (loss\_list\_5, em\_list\_5) 時，模型的 loss 下降速度和最終的 EM 雖然依舊不錯，但並沒有顯著超過 embedding size 為 256 的結果。甚至在某些 epoch，embedding size 512 和 1024 的模型 loss 和 EM 出現輕微的波動。這說明當 embedding size 過大時，模型可能過度擬合，無法有效提取出更多有用的特徵，反而增加了不必要的計算資源需求。

總結來說，在這個數字和符號序列計算任務中，embedding size 為 256 表現最佳，能夠兼顧收斂速度和最終的 EM，達到穩定的學習效果。過小的 embedding size 可能會限制模型的表現，導致準確率降低，而過大的 embedding size 則可能導致模型過度擬合或資源浪費。因此，在這個任務中選擇 embedding size 256 是比較合適的選擇。

1. 比較不同 Gradient Clipping 的值對於模型訓練的影響

一張含有 行, 繪圖, 文字, 圖表 的圖片

自動產生的描述

一張含有 文字, 行, 繪圖, 圖表 的圖片

自動產生的描述

我從 epoch 對於 loss 以及 EM 的圖中觀察到，不同的 Gradient Clipping Threshold 對於模型訓練的影響非常顯著。以下是根據各種不同的 grad\_clip 值進行的分析：

首先，從 grad\_clip = 0.01 的結果來看，模型的 loss 在早期逐漸下降，但在 epoch 10 後變動較大，並且最終的 loss 仍較高，表現出明顯的 loss 不穩定現象。同時 EM 雖然在 epoch 5-10 左右穩步上升，但其整體表現並不如其他較高的 grad\_clip 值。這可能是因為 grad\_clip 值過小，導致梯度更新幅度有限，使得模型的參數更新較慢且容易陷入區域最小值，進而影響模型的收斂速度和準確性。

接著，grad\_clip = 0.1 與 grad\_clip = 1 的曲線表現較為平穩，loss 持續下降，EM 上升相對穩定且達到高準確率。grad\_clip = 0.1 在 epoch 5 以後快速收斂，而 grad\_clip = 1 則呈現出更穩定的趨勢，表明這些 grad\_clip 值較適合當前模型，使模型參數在適當幅度下更新，有助於提升收斂速度和穩定性。

再來，當 grad\_clip = 10 或 100 時，模型的 loss 以及 EM 曲線變得不穩定。特別是 grad\_clip = 100 時，模型 loss 波動劇烈並難以收斂，而 EM 雖然在初期有所上升，但隨著 epoch 增加出現震盪，整體表現不佳。這表示過高的 grad\_clip 會導致過大的梯度更新，進而干擾模型的穩定學習，最終影響收斂效果與準確率。

最後，當不使用 gradient clipping 時，模型的 loss 下降與 EM 上升曲線與較高的 grad\_clip（例如 10 或 100）相同，顯示出梯度過大時可能造成不穩定。由於梯度未被裁切，過大的梯度更新幅度使得模型難以有效收斂。

綜合來看，較小的 grad\_clip 值會減少梯度更新的幅度，使得模型在訓練過程中參數更新較慢，可能影響收斂速度。過高的 grad\_clip 反而會導致梯度過大，增加不穩定性並影響模型準確率。從這些觀察可以推測，適中的 grad\_clip（例如 0.1 或 1）有助於平衡梯度更新幅度，促進模型的穩定學習，提升最終的收斂效果與準確性。因此，在這個任務中選擇 grad\_clip 0.1 是比較合適的選擇。

D. 比較不同 Learning Rate 的值對於模型訓練的影響

一張含有 繪圖, 文字, 行, 圖表 的圖片

自動產生的描述

一張含有 文字, 繪圖, 行, 圖表 的圖片

自動產生的描述

我從 epoch 對於 loss 以及 EM 的圖中觀察到，不同的 Learning Rate 對於模型訓練的收斂速度及效果有明顯影響。

首先，當 Learning Rate 設定為較高的 0.01 時，可以看到 loss 在訓練初期的下降幅度較為顯著，但隨著 epoch 增加，loss 並沒有繼續下降，甚至出現波動。這表明高 Learning Rate 雖然使得模型收斂較快，但會導致訓練的穩定性下降。這樣的高 Learning Rate 會使模型在尋找最優解的過程中出現不穩定現象，因此最終 EM 表現也較差，未能達到很高的精度。

接著，對於 Learning Rate 為 0.003 的情況，loss 的下降趨勢比 0.01 穩定，並且 EM 值隨著 epoch 持續提升，顯示模型在該 Learning Rate 下穩定收斂。儘管相較於 0.01 的高 Learning Rate 收斂速度稍慢，但最終模型在準確度上的提升較顯著。因此，0.003 似乎是比 0.01 更佳的學習率選擇。

再來，對於較低的 Learning Rate（0.001、0.0003 和 0.0001），雖然 loss 逐步下降，並且收斂較穩定，但初期的收斂速度較慢。尤其在 0.0001 時，loss 下降趨勢最為緩慢，顯示模型在此 Learning Rate 下學習過於保守，導致收斂時間延長。不過，低 Learning Rate 仍然可以使模型的 loss 逐步穩定收斂，且 EM 在中後期提升至較高的精度，但這需要較多的訓練時間來達到收斂。在 （0.001、0.0003 和 0.0001）的比較之中，0.0003 是最佳的 learning rate，不僅在 20 epochs 收斂速度夠快，也不會很快就收斂到極限，能更進一步收斂到高精度。

總結來說，從訓練結果可以推斷出中等的 Learning Rate（0.003 或 0.001）在此次訓練中表現較佳，既能穩定收斂，又能有效提升準確度。而過高的 Learning Rate 雖然在初期收斂速度快，但可能會導致模型不穩定，而過低的 Learning Rate 則需要更長的訓練時間才能達到高準確度。因此，在這個任務中選擇 lr 0.0003 是比較合適的選擇。

E. 比較不同 Batch Size 的值對於模型訓練的影響

一張含有 行, 繪圖, 圖表, 文字 的圖片

自動產生的描述

一張含有 文字, 行, 繪圖, 圖表 的圖片

自動產生的描述

我從 epoch 對於 loss 以及 EM 的圖中觀察到，不同的 batch\_size 對模型的訓練結果影響顯著。以下是不同 batch\_size 的分析。

首先，batch\_size 越小（例如 batch\_size = 16 和 32）時，模型的 loss 波動較大，且在初期的訓練 epochs 中，loss 起伏明顯，這是因為較小的 batch\_size 對於每個 mini-batch 的數據分布可能有較大的變異，使得模型的梯度更新方向容易受單一 batch 的影響。而小 batch\_size 也導致模型的 EM 表現相對較低，不過訓練幾個 epochs 後 EM 開始穩定上升，最後趨於穩定。

接著，當 batch\_size 適中（例如 batch\_size = 64 和 128）時， loss 曲線依舊有輕微的波動，這顯示出 batch\_size 較小時，loss 下降的穩定性不足。

再來，batch\_size 增加到 256 時，loss 曲線相對平滑，且下降速度較快，顯示出訓練過程中較穩定的梯度更新。這兩種 batch\_size 在 EM 指標上也表現出較快的提升速度，並且最終的 EM 值也相對較高，說明模型能夠更有效地學習到序列運算的模式。從訓練穩定性和模型效果來看，batch\_size = 256 是較合適的選擇。

還有，Batch size 越大時，訓練速度越快，可以看後面 appendix 的部分，Batch size 16 和 Batch size 256訓練速度可以差到好幾倍。

總結來說，越大的 batch\_size 能夠使模型訓練時 loss 下降的穩定度提高、且更能收斂到足夠小的 loss、訓練速度也越快。因此，在這個任務中選擇 batch\_size 256 是比較合適的選擇。

F. 比較不同 Hidden Size 的值對於模型訓練的影響

一張含有 繪圖, 行, 圖表, 文字 的圖片

自動產生的描述

一張含有 文字, 行, 繪圖, 圖表 的圖片

自動產生的描述

我從 epoch 對於 loss 以及 EM 的圖中觀察到，當 hidden\_dim 提升時，模型的訓練效果在整體上有顯著的改善，尤其是在較高的 hidden\_dim 設置下。

首先，從 loss 曲線來看，隨著 hidden\_dim 從 64 逐漸增大到 1024，模型的 loss 減少幅度相對提升，特別是 hidden\_dim 大於 256 之後的 loss 收斂效果明顯改善，loss 值在接近後期的 epoch (第15-20次) 較為穩定且低。這暗示著較大的 hidden\_dim 有助於模型更有效地學習數學運算的模式與複雜結構，因此在相同訓練次數下達到了較好的結果。

接著，從 EM (精確匹配率) 的表現來看，hidden\_dim = 64 和 128 的情況下，模型的 EM 只能達到大約 0.6 至 0.7 左右，而當 hidden\_dim 增加到 512 或更大時，EM 在第10個 epoch 後就達到了 0.9 以上，甚至接近 0.95，顯示模型有了較好的預測能力並成功提高了精度。這表明隨著 hidden\_dim 增大，LSTM 模型能捕捉到更多的特徵，並有效處理更長的計算序列。

然而，在 hidden\_dim = 1024 的情況下雖然達到了最佳的 EM 和 loss 表現，但訓練時間也較長、需要更多的運算資源。由於 loss 已經相對穩定在低值區域，後續若模型在 validation set 上的表現也持續提升，表示此 hidden\_dim 可確實幫助模型的穩定學習。

這些結果顯示，隨著 hidden\_dim 的增大，模型的學習表現和收斂效果顯著提升，且在數學計算序列的處理上，適度增加 hidden\_dim 是有效的優化方向。因此，在這個任務中選擇 hidden\_size 256 是比較合適的選擇。

G. Best hyperparameters Combination

結合 B-F 的結果，我們得到最佳的 hyperparameters 組合為以下:

batch\_size = 256

epochs = 20

embed\_dim = 64

hidden\_dim = 1024

lr = 0.0003

grad\_clip = 0.1

我用此 hyperparameters 和 Baseline 做圖比較:

一張含有 文字, 行, 繪圖, 圖表 的圖片

自動產生的描述

一張含有 文字, 行, 繪圖, 圖表 的圖片

自動產生的描述

可以看到，使用最佳的 hyperparameters 組合訓練出來的結果非常好，在 epoch 4 EM就已經達到 92% 的 score 了，在 epoch 20 EM 可以達到高達 99% 的 score，Loss 也是很平穩的下降。這證明了這組 hyperparameters 的確是非常好的。

H. 移除某個 number (這裡是 23) 對模型訓練的影響

我自己藉由 chatgpt 產生一個程式碼，生成了一個 training data，其中的 src 不包含任何 23 作為 operand。

一張含有 文字, 軟體, 網頁, 數字 的圖片

自動產生的描述 一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

接著，我用這個 training data 訓練了 Baseline hyperparameters 和 Best hyperparameters 來跟使用原本的 training data 訓練的 Basline hyperparameters 做比較。

一張含有 行, 繪圖, 文字, 圖表 的圖片

自動產生的描述

一張含有 文字, 行, 繪圖, 圖表 的圖片

自動產生的描述

可以看到，相較於擁有完整 training data 的 Baseline，移除掉 23 的 training data 不管是 baseline 還是 best hyperparameters，EM 的 score 都遠低於擁有完整 training data 的 Baseline，可以合理推斷是因為 validation set 裡面包含 operand number 23，但 training set 裡面並沒有，所以 EM score 很低是合理的。在移除掉 23 的 training data 對於 baseline 和 best hyperparameters 進行比較，可以看到，best hyperparameters 在 validation set 得到的 EM score 還是比較好的，代表在有 operand 23 以外的 arithmetic sequence，best hyperparameters 的訓練結果較好。

而在 loss 的部分，移除掉 23 的 training data 不管是 baseline 還是 best hyperparameters，loss 的值都比擁有完整 training data 的 Baseline 還要低，這是因為它們 training data 沒有 23 這個 operand 要訓練，所以比起 擁有完整 training data 的 Baseline，它們更容易訓練。

I. Evaluation on 3 digit

我用 chatgpt 產生了一個程式碼，能夠生成 operand 包含 [0, 500) 的 evaluation data。

一張含有 文字, 字型, 螢幕擷取畫面 的圖片

自動產生的描述 一張含有 文字, 字型, 螢幕擷取畫面, 代數 的圖片

自動產生的描述

接著，我訓練了 Baseline hyperparameters 和 Best hyperparameters 對這個 evaluation set 做 evaluation 看結果。

一張含有 行, 繪圖, 文字, 圖表 的圖片

自動產生的描述

一張含有 行, 繪圖, 圖表, 文字 的圖片

自動產生的描述

可以看到，使用 operand 範圍在 [0, 50) 的 2 digit training data 訓練出來的模型，對於 3 digit 的 evaluation 結果非常差，baseline hyperparameters 的 EM score 僅有不到 0.9 %，best hyperparameters的 EM score 僅有不到 1 %，代表我們的模型只學會在 [0, 50) 範圍 operand 的 arithmetic 問題，如果超出這個範圍，那模型完全無法回答正確。

至於模型在 evaluate 3 digit 的 test set 時的表現到底會如何呢? 可以看以下範例 :

一張含有 文字, 字型, 白色, 行 的圖片

自動產生的描述

一張含有 文字, 字型, 行, 白色 的圖片

自動產生的描述

一張含有 文字, 字型, 行, 白色 的圖片

自動產生的描述

一張含有 文字, 字型, 白色, 行 的圖片

自動產生的描述

一張含有 文字, 字型, 白色, 行 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型, 白色 的圖片

自動產生的描述

一張含有 文字, 字型, 行, 白色 的圖片

自動產生的描述

一張含有 文字, 字型, 螢幕擷取畫面, 白色 的圖片

自動產生的描述

一張含有 文字, 字型, 行, 螢幕擷取畫面 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型, 白色 的圖片

自動產生的描述

從這幾個範例，可以推測，如果是 abc +def，那模型的預測結果看起來像是 (ab) + (de) + f

但再看下面這個範例 :

一張含有 文字, 螢幕擷取畫面, 字型, 白色 的圖片

自動產生的描述

實際上對於每個數，模型的計算方式、預測結果都不一樣，所以無從推斷。但至少可以看出模型有試圖用 2 digit 的方法來推算出 3 digit 的值。

Conclusion :

總結來說，模型的訓練和模型的架構有關，LSTM 有記憶的功能，在 arithmetic 這個問題上有優於 RNN, GRU 的表現。再來，越大的 hidden size 可以讓模型學到更複雜的規則、更能擬合此問題的輸出；因為我們的 embedding 僅有 18 個 characters 要處理，不用太複雜的 embedding，所以越小的 embedding size 較佳。Batch size 的提高可以讓 loss 下降更穩定、訓練速度更快。Learning rate 要在 loss 下降速度和可以達到的 loss 精度之間權衡，0.0003 是我找到的最佳值。Grad Clipping Threshold 可以防止梯度爆炸，可是太小的 Threshold 反而會讓 loss 下降速度變慢，所以權衡後 0.1 是最佳值。把所有最佳的 hyperparameters 合起來訓練，在 20 個 epoch 以內可以訓練到 99% 正確率(EM)。

如果訓練資料內缺少某一個 number，那不管再怎麼訓練，正確率(EM)始終會非常的低。如果訓練在 2 digit 的資料上，但 evaluation 在 3 digit 的資料，那結果會非常差，代表模型只學到 2 digit 的 arithmetic，3 digit 的它完全不會。

Appendix :

Baseline :

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:15<00:00, 189.37it/s, loss=0.367]

Validation epoch 1: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:59<00:00, 69.72it/s, EM=0.568]

0.568338081671415

Train epoch 2: 100%|███████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:15<00:00, 189.74it/s, loss=0.23]

Validation epoch 2: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:58<00:00, 70.21it/s, EM=0.644]

0.6438252611585945

Train epoch 3: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:16<00:00, 188.46it/s, loss=0.235]

Validation epoch 3: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:57<00:00, 71.76it/s, EM=0.704]

0.703954415954416

Train epoch 4: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.17it/s, loss=0.195]

Validation epoch 4: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 72.74it/s, EM=0.719]

0.7192250712250712

Train epoch 5: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.44it/s, loss=0.206]

Validation epoch 5: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 73.37it/s, EM=0.757]

0.7569496676163343

Train epoch 6: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.34it/s, loss=0.126]

Validation epoch 6: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 73.00it/s, EM=0.779]

0.7789591642924977

Train epoch 7: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:11<00:00, 193.00it/s, loss=0.147]

Validation epoch 7: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 72.82it/s, EM=0.788]

0.7876163342830009

Train epoch 8: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 191.90it/s, loss=0.124]

Validation epoch 8: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 73.02it/s, EM=0.785]

0.785272554605888

Train epoch 9: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.03it/s, loss=0.191]

Validation epoch 9: 100%|████████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 72.95it/s, EM=0.8]

0.8000531813865147

Train epoch 10: 100%|█████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.60it/s, loss=0.136]

Validation epoch 10: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 72.81it/s, EM=0.797]

0.7973561253561253

Train epoch 11: 100%|█████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.12it/s, loss=0.173]

Validation epoch 11: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 73.21it/s, EM=0.811]

0.8112706552706552

Train epoch 12: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.15it/s, loss=0.12]

Validation epoch 12: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 72.68it/s, EM=0.816]

0.8157454890788224

Train epoch 13: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.79it/s, loss=0.17]

Validation epoch 13: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 73.01it/s, EM=0.822]

0.8220740740740741

Train epoch 14: 100%|████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.75it/s, loss=0.0756]

Validation epoch 14: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 73.24it/s, EM=0.829]

0.8290636277302944

Train epoch 15: 100%|█████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.33it/s, loss=0.135]

Validation epoch 15: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 73.34it/s, EM=0.824]

0.8242507122507122

Train epoch 16: 100%|████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.26it/s, loss=0.0879]

Validation epoch 16: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 72.87it/s, EM=0.835]

0.8353656220322887

Train epoch 17: 100%|█████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.40it/s, loss=0.101]

Validation epoch 17: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 73.28it/s, EM=0.838]

0.8375042735042735

Train epoch 18: 100%|████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 191.84it/s, loss=0.0749]

Validation epoch 18: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 73.26it/s, EM=0.842]

0.8419639126305793

Train epoch 19: 100%|█████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:12<00:00, 192.53it/s, loss=0.125]

Validation epoch 19: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 72.91it/s, EM=0.839]

0.8389515669515669

Train epoch 20: 100%|█████████████████████████████████████████████████████████████████████████████████| 37020/37020 [03:13<00:00, 191.12it/s, loss=0.123]

Validation epoch 20: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [00:56<00:00, 72.21it/s, EM=0.845]

0.8450522317188984

Using GRU (run\_GRU.py) :

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [09:01<00:00, 68.40it/s, loss=0.347]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [08:26<00:00, 8.12it/s, EM=0.482]

0.48152706552706553

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [10:30<00:00, 58.68it/s, loss=0.366]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [08:11<00:00, 8.38it/s, EM=0.536]

0.5359202279202279

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [11:29<00:00, 53.68it/s, loss=0.311]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [08:01<00:00, 8.55it/s, EM=0.579]

0.5791566951566952

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [11:20<00:00, 54.42it/s, loss=0.264]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [07:35<00:00, 9.02it/s, EM=0.595]

0.5945451092117758

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [11:22<00:00, 54.26it/s, loss=0.225]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [07:53<00:00, 8.69it/s, EM=0.586]

0.5858423551756885

Train epoch 6: 100%|██████████████████████████████████████████████████| 37020/37020 [11:20<00:00, 54.40it/s, loss=0.28]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [08:28<00:00, 8.09it/s, EM=0.621]

0.6213029439696106

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [11:24<00:00, 54.10it/s, loss=0.375]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [08:51<00:00, 7.75it/s, EM=0.631]

0.6310997150997151

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [11:31<00:00, 53.52it/s, loss=0.281]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [09:20<00:00, 7.33it/s, EM=0.619]

0.6187844254510921

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [11:32<00:00, 53.49it/s, loss=0.309]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [08:19<00:00, 8.23it/s, EM=0.634]

0.6342336182336182

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [11:34<00:00, 53.31it/s, loss=0.362]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [08:27<00:00, 8.11it/s, EM=0.638]

0.6379791073124407

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [11:35<00:00, 53.26it/s, loss=0.234]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [08:02<00:00, 8.53it/s, EM=0.653]

0.6531130104463437

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [11:30<00:00, 53.60it/s, loss=0.276]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [07:39<00:00, 8.95it/s, EM=0.655]

0.6550275403608737

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [11:33<00:00, 53.36it/s, loss=0.239]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [07:39<00:00, 8.96it/s, EM=0.648]

0.6484140550807217

Train epoch 14: 100%|████████████████████████████████████████████████| 37020/37020 [11:25<00:00, 54.01it/s, loss=0.235]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [06:20<00:00, 10.82it/s, EM=0.643]

0.6427996201329534

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [09:39<00:00, 63.92it/s, loss=0.358]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [06:28<00:00, 10.60it/s, EM=0.662]

0.6623513770180437

Train epoch 16: 100%|█████████████████████████████████████████████████| 37020/37020 [09:37<00:00, 64.10it/s, loss=0.26]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [06:33<00:00, 10.45it/s, EM=0.667]

0.6669173789173789

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [09:39<00:00, 63.91it/s, loss=0.241]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [06:31<00:00, 10.51it/s, EM=0.653]

0.6525090218423552

Train epoch 18: 100%|████████████████████████████████████████████████| 37020/37020 [09:46<00:00, 63.09it/s, loss=0.211]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [06:50<00:00, 10.03it/s, EM=0.671]

0.67132383665717

Train epoch 19: 100%|████████████████████████████████████████████████| 37020/37020 [09:48<00:00, 62.89it/s, loss=0.226]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [06:32<00:00, 10.49it/s, EM=0.667]

0.6665830959164293

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [08:10<00:00, 75.42it/s, loss=0.223]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [07:03<00:00, 9.71it/s, EM=0.661]

0.6612041785375119

Using RNN (run\_RNN.py) :

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [09:02<00:00, 68.19it/s, loss=0.968]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [06:23<00:00, 10.71it/s, EM=0.121]

0.12147768281101615

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [09:48<00:00, 62.93it/s, loss=0.907]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [07:32<00:00, 9.09it/s, EM=0.129]

0.12878252611585944

Train epoch 3: 100%|██████████████████████████████████████████████████| 37020/37020 [10:25<00:00, 59.20it/s, loss=1.01]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [07:41<00:00, 8.92it/s, EM=0.128]

0.12815574548907882

Train epoch 4: 100%|██████████████████████████████████████████████████| 37020/37020 [10:17<00:00, 59.97it/s, loss=0.94]

Validation epoch 4: 100%|█████████████████████████████████████████████████| 4114/4114 [07:34<00:00, 9.06it/s, EM=0.12]

0.12042165242165243

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [10:19<00:00, 59.74it/s, loss=0.984]

Validation epoch 5: 100%|█████████████████████████████████████████████████| 4114/4114 [07:47<00:00, 8.80it/s, EM=0.12]

0.12030389363722697

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [10:14<00:00, 60.20it/s, loss=0.903]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [07:49<00:00, 8.77it/s, EM=0.115]

0.11508072174738841

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [10:17<00:00, 59.91it/s, loss=0.873]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [07:32<00:00, 9.09it/s, EM=0.109]

0.10857359924026591

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [10:20<00:00, 59.64it/s, loss=0.964]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [07:44<00:00, 8.86it/s, EM=0.101]

0.10107122507122507

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [10:20<00:00, 59.62it/s, loss=0.975]

Validation epoch 9: 100%|███████████████████████████████████████████████| 4114/4114 [07:57<00:00, 8.61it/s, EM=0.0982]

0.0981690408357075

Train epoch 10: 100%|█████████████████████████████████████████████████| 37020/37020 [10:21<00:00, 59.53it/s, loss=1.02]

Validation epoch 10: 100%|██████████████████████████████████████████████| 4114/4114 [08:04<00:00, 8.49it/s, EM=0.0936]

0.0936258309591643

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [10:21<00:00, 59.57it/s, loss=0.968]

Validation epoch 11: 100%|██████████████████████████████████████████████| 4114/4114 [08:07<00:00, 8.44it/s, EM=0.0891]

0.08913200379867046

Train epoch 12: 100%|█████████████████████████████████████████████████| 37020/37020 [10:25<00:00, 59.18it/s, loss=1.06]

Validation epoch 12: 100%|██████████████████████████████████████████████| 4114/4114 [07:32<00:00, 9.09it/s, EM=0.0841]

0.08410636277302944

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [10:24<00:00, 59.26it/s, loss=0.965]

Validation epoch 13: 100%|██████████████████████████████████████████████| 4114/4114 [07:37<00:00, 9.00it/s, EM=0.0806]

0.08056600189933523

Train epoch 14: 100%|█████████████████████████████████████████████████| 37020/37020 [10:31<00:00, 58.61it/s, loss=1.01]

Validation epoch 14: 100%|██████████████████████████████████████████████| 4114/4114 [07:40<00:00, 8.93it/s, EM=0.0741]

0.07411206077872745

Train epoch 15: 100%|█████████████████████████████████████████████████| 37020/37020 [10:08<00:00, 60.79it/s, loss=1.09]

Validation epoch 15: 100%|██████████████████████████████████████████████| 4114/4114 [06:19<00:00, 10.83it/s, EM=0.0716]

0.07163912630579297

Train epoch 16: 100%|█████████████████████████████████████████████████| 37020/37020 [08:55<00:00, 69.19it/s, loss=1.07]

Validation epoch 16: 100%|██████████████████████████████████████████████| 4114/4114 [06:42<00:00, 10.21it/s, EM=0.0656]

0.06555745489078822

Train epoch 17: 100%|██████████████████████████████████████████████████| 37020/37020 [08:48<00:00, 70.06it/s, loss=1.1]

Validation epoch 17: 100%|██████████████████████████████████████████████| 4114/4114 [06:42<00:00, 10.22it/s, EM=0.0638]

0.06377207977207977

Train epoch 18: 100%|██████████████████████████████████████████████████| 37020/37020 [08:52<00:00, 69.58it/s, loss=1.1]

Validation epoch 18: 100%|██████████████████████████████████████████████| 4114/4114 [06:35<00:00, 10.39it/s, EM=0.0569]

0.056881291547958214

Train epoch 19: 100%|██████████████████████████████████████████████████| 37020/37020 [08:57<00:00, 68.93it/s, loss=1.2]

Validation epoch 19: 100%|██████████████████████████████████████████████| 4114/4114 [06:41<00:00, 10.25it/s, EM=0.0551]

0.055114909781576446

Train epoch 20: 100%|█████████████████████████████████████████████████| 37020/37020 [08:57<00:00, 68.86it/s, loss=1.05]

Validation epoch 20: 100%|██████████████████████████████████████████████| 4114/4114 [06:40<00:00, 10.27it/s, EM=0.0515]

0.051464387464387466

Different Embedding size (run\_embed.py) :

Using batch\_size=64, epochs=20,embed\_dim=64, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:06<00:00, 86.71it/s, loss=0.314]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:46<00:00, 11.88it/s, EM=0.623]

0.6229553656220322

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [10:12<00:00, 60.44it/s, loss=0.282]

Validation epoch 2: 100%|█████████████████████████████████████████████████| 4114/4114 [06:12<00:00, 11.05it/s, EM=0.73]

0.7295118708452042

Train epoch 3: 100%|██████████████████████████████████████████████████| 37020/37020 [11:20<00:00, 54.40it/s, loss=0.11]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [07:32<00:00, 9.10it/s, EM=0.793]

0.7933789173789174

Train epoch 4: 100%|████████████████████████████████████████████████| 37020/37020 [11:51<00:00, 52.02it/s, loss=0.0721]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [08:08<00:00, 8.42it/s, EM=0.852]

0.8522659069325736

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [11:52<00:00, 51.97it/s, loss=0.139]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [08:08<00:00, 8.43it/s, EM=0.873]

0.8733219373219373

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [11:50<00:00, 52.10it/s, loss=0.103]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [09:06<00:00, 7.52it/s, EM=0.896]

0.8963000949667617

Train epoch 7: 100%|████████████████████████████████████████████████| 37020/37020 [11:52<00:00, 51.99it/s, loss=0.0442]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [08:21<00:00, 8.21it/s, EM=0.903]

0.9032402659069326

Train epoch 8: 100%|████████████████████████████████████████████████| 37020/37020 [11:51<00:00, 52.06it/s, loss=0.0947]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [08:05<00:00, 8.47it/s, EM=0.908]

0.9082545109211776

Train epoch 9: 100%|████████████████████████████████████████████████| 37020/37020 [11:45<00:00, 52.47it/s, loss=0.0723]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [07:48<00:00, 8.78it/s, EM=0.915]

0.9147198480531814

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [11:41<00:00, 52.81it/s, loss=0.041]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [07:32<00:00, 9.10it/s, EM=0.921]

0.9212535612535613

Train epoch 11: 100%|███████████████████████████████████████████████| 37020/37020 [11:41<00:00, 52.76it/s, loss=0.0421]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [07:52<00:00, 8.70it/s, EM=0.918]

0.9181538461538462

Train epoch 12: 100%|███████████████████████████████████████████████| 37020/37020 [11:52<00:00, 51.98it/s, loss=0.0602]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [07:39<00:00, 8.96it/s, EM=0.928]

0.928022792022792

Train epoch 13: 100%|███████████████████████████████████████████████| 37020/37020 [11:50<00:00, 52.07it/s, loss=0.0188]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [08:34<00:00, 8.00it/s, EM=0.923]

0.9229819563152897

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [12:04<00:00, 51.09it/s, loss=0.0359]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [09:19<00:00, 7.35it/s, EM=0.933]

0.9331054131054131

Train epoch 15: 100%|███████████████████████████████████████████████| 37020/37020 [10:36<00:00, 58.21it/s, loss=0.0297]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [07:59<00:00, 8.58it/s, EM=0.934]

0.9342488129154796

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [10:07<00:00, 60.97it/s, loss=0.0629]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [07:20<00:00, 9.34it/s, EM=0.926]

0.925713200379867

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [10:01<00:00, 61.59it/s, loss=0.0568]

Validation epoch 17: 100%|████████████████████████████████████████████████| 4114/4114 [06:24<00:00, 10.70it/s, EM=0.94]

0.9404368471035137

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [10:05<00:00, 61.17it/s, loss=0.0501]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [06:37<00:00, 10.34it/s, EM=0.935]

0.9352326685660018

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [10:03<00:00, 61.29it/s, loss=0.0223]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [06:37<00:00, 10.36it/s, EM=0.936]

0.9360987654320988

Train epoch 20: 100%|███████████████████████████████████████████████| 37020/37020 [08:49<00:00, 69.88it/s, loss=0.0333]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [06:10<00:00, 11.11it/s, EM=0.942]

0.9417663817663817  
  
  
Using batch\_size=64, epochs=20,embed\_dim=128, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:45<00:00, 79.60it/s, loss=0.405]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:57<00:00, 11.52it/s, EM=0.578]

0.5781272554605887

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [06:59<00:00, 88.24it/s, loss=0.207]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [06:22<00:00, 10.76it/s, EM=0.692]

0.6924520417853751

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [06:59<00:00, 88.27it/s, loss=0.153]

Validation epoch 3: 100%|█████████████████████████████████████████████████| 4114/4114 [06:52<00:00, 9.98it/s, EM=0.77]

0.7704387464387464

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [07:03<00:00, 87.50it/s, loss=0.153]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [05:16<00:00, 13.01it/s, EM=0.813]

0.8132535612535613

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [07:15<00:00, 85.03it/s, loss=0.179]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [05:31<00:00, 12.41it/s, EM=0.849]

0.8490674264007597

Train epoch 6: 100%|████████████████████████████████████████████████| 37020/37020 [07:22<00:00, 83.65it/s, loss=0.0966]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.18it/s, EM=0.833]

0.8330674264007597

Train epoch 7: 100%|████████████████████████████████████████████████| 37020/37020 [07:19<00:00, 84.32it/s, loss=0.0929]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [04:51<00:00, 14.14it/s, EM=0.867]

0.8669933523266856

Train epoch 8: 100%|████████████████████████████████████████████████| 37020/37020 [08:02<00:00, 76.76it/s, loss=0.0512]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [04:42<00:00, 14.55it/s, EM=0.886]

0.8857397910731244

Train epoch 9: 100%|████████████████████████████████████████████████| 37020/37020 [06:39<00:00, 92.59it/s, loss=0.0918]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [04:37<00:00, 14.84it/s, EM=0.901]

0.901283950617284

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [06:49<00:00, 90.30it/s, loss=0.105]

Validation epoch 10: 100%|█████████████████████████████████████████████████| 4114/4114 [05:05<00:00, 13.45it/s, EM=0.9]

0.9001899335232668

Train epoch 11: 100%|███████████████████████████████████████████████| 37020/37020 [06:44<00:00, 91.49it/s, loss=0.0889]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [04:42<00:00, 14.57it/s, EM=0.903]

0.9026514719848053

Train epoch 12: 100%|███████████████████████████████████████████████| 37020/37020 [06:41<00:00, 92.21it/s, loss=0.0814]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [05:04<00:00, 13.49it/s, EM=0.902]

0.9021956315289649

Train epoch 13: 100%|███████████████████████████████████████████████| 37020/37020 [07:02<00:00, 87.71it/s, loss=0.0669]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [05:05<00:00, 13.46it/s, EM=0.912]

0.9121063627730295

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [06:43<00:00, 91.73it/s, loss=0.0436]

Validation epoch 14: 100%|████████████████████████████████████████████████| 4114/4114 [05:13<00:00, 13.12it/s, EM=0.91]

0.9099031339031339

Train epoch 15: 100%|███████████████████████████████████████████████| 37020/37020 [07:01<00:00, 87.80it/s, loss=0.0304]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [05:18<00:00, 12.93it/s, EM=0.918]

0.9177701804368471

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [06:58<00:00, 88.36it/s, loss=0.0487]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [04:43<00:00, 14.50it/s, EM=0.912]

0.9120759734093068

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [07:01<00:00, 87.78it/s, loss=0.0612]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [05:20<00:00, 12.82it/s, EM=0.922]

0.9218765432098766

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [07:07<00:00, 86.63it/s, loss=0.0846]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [04:45<00:00, 14.40it/s, EM=0.916]

0.9157606837606838

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [06:40<00:00, 92.45it/s, loss=0.0539]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [05:04<00:00, 13.51it/s, EM=0.922]

0.9218993352326685

Train epoch 20: 100%|███████████████████████████████████████████████| 37020/37020 [07:12<00:00, 85.50it/s, loss=0.0181]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:03<00:00, 13.56it/s, EM=0.919]

0.9193504273504274  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:11<00:00, 85.85it/s, loss=0.367]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [04:53<00:00, 13.99it/s, EM=0.568]

0.568338081671415

Train epoch 2: 100%|██████████████████████████████████████████████████| 37020/37020 [07:06<00:00, 86.73it/s, loss=0.23]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [05:14<00:00, 13.07it/s, EM=0.644]

0.6438252611585945

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [06:56<00:00, 88.80it/s, loss=0.235]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [04:47<00:00, 14.30it/s, EM=0.704]

0.703954415954416

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [06:39<00:00, 92.68it/s, loss=0.195]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [05:34<00:00, 12.32it/s, EM=0.719]

0.7192250712250712

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [06:56<00:00, 88.98it/s, loss=0.206]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [05:52<00:00, 11.68it/s, EM=0.757]

0.7569496676163343

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [06:43<00:00, 91.86it/s, loss=0.126]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [06:27<00:00, 10.61it/s, EM=0.779]

0.7789591642924977

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [06:53<00:00, 89.60it/s, loss=0.147]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [07:15<00:00, 9.45it/s, EM=0.788]

0.7876163342830009

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [06:53<00:00, 89.51it/s, loss=0.124]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [07:16<00:00, 9.43it/s, EM=0.785]

0.785272554605888

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [06:55<00:00, 89.15it/s, loss=0.191]

Validation epoch 9: 100%|██████████████████████████████████████████████████| 4114/4114 [05:28<00:00, 12.52it/s, EM=0.8]

0.8000531813865147

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [07:03<00:00, 87.45it/s, loss=0.136]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [06:10<00:00, 11.12it/s, EM=0.797]

0.7973561253561253

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [07:15<00:00, 85.02it/s, loss=0.173]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:31<00:00, 12.40it/s, EM=0.811]

0.8112706552706552

Train epoch 12: 100%|█████████████████████████████████████████████████| 37020/37020 [06:51<00:00, 89.93it/s, loss=0.12]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [04:51<00:00, 14.14it/s, EM=0.816]

0.8157454890788224

Train epoch 13: 100%|█████████████████████████████████████████████████| 37020/37020 [07:06<00:00, 86.73it/s, loss=0.17]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [05:24<00:00, 12.69it/s, EM=0.822]

0.8220740740740741

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [06:59<00:00, 88.34it/s, loss=0.0756]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [05:26<00:00, 12.60it/s, EM=0.829]

0.8290636277302944

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [06:44<00:00, 91.56it/s, loss=0.135]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [04:55<00:00, 13.93it/s, EM=0.824]

0.8242507122507122

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [06:57<00:00, 88.77it/s, loss=0.0879]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [04:56<00:00, 13.89it/s, EM=0.835]

0.8353656220322887

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [07:00<00:00, 87.98it/s, loss=0.101]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [05:42<00:00, 12.00it/s, EM=0.838]

0.8375042735042735

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [06:54<00:00, 89.33it/s, loss=0.0749]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [05:49<00:00, 11.76it/s, EM=0.842]

0.8419639126305793

Train epoch 19: 100%|████████████████████████████████████████████████| 37020/37020 [07:06<00:00, 86.75it/s, loss=0.125]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.16it/s, EM=0.839]

0.8389515669515669

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [07:05<00:00, 87.01it/s, loss=0.123]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:30<00:00, 12.46it/s, EM=0.845]

0.8450522317188984  
  
  
Using batch\_size=64, epochs=20,embed\_dim=512, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:41<00:00, 80.17it/s, loss=0.274]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [04:52<00:00, 14.07it/s, EM=0.565]

0.5648395061728395

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [07:51<00:00, 78.52it/s, loss=0.243]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [04:44<00:00, 14.48it/s, EM=0.669]

0.6691737891737892

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [07:16<00:00, 84.78it/s, loss=0.192]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [05:03<00:00, 13.54it/s, EM=0.684]

0.6837378917378918

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [07:34<00:00, 81.51it/s, loss=0.165]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [05:04<00:00, 13.53it/s, EM=0.736]

0.7360911680911681

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [07:04<00:00, 87.29it/s, loss=0.213]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [05:30<00:00, 12.45it/s, EM=0.757]

0.7574396961063627

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [07:07<00:00, 86.59it/s, loss=0.189]

Validation epoch 6: 100%|█████████████████████████████████████████████████| 4114/4114 [05:03<00:00, 13.56it/s, EM=0.77]

0.770005698005698

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [07:52<00:00, 78.41it/s, loss=0.146]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [04:35<00:00, 14.92it/s, EM=0.782]

0.7821918328584996

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [07:38<00:00, 80.81it/s, loss=0.103]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [05:00<00:00, 13.70it/s, EM=0.803]

0.8032972459639126

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [07:29<00:00, 82.30it/s, loss=0.171]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [04:54<00:00, 13.95it/s, EM=0.809]

0.8088015194681861

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [07:26<00:00, 82.85it/s, loss=0.107]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [05:52<00:00, 11.66it/s, EM=0.819]

0.818681861348528

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [07:02<00:00, 87.54it/s, loss=0.167]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [06:00<00:00, 11.40it/s, EM=0.817]

0.8169420702754036

Train epoch 12: 100%|███████████████████████████████████████████████| 37020/37020 [07:33<00:00, 81.67it/s, loss=0.0896]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [05:35<00:00, 12.25it/s, EM=0.836]

0.8355061728395061

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [07:42<00:00, 80.01it/s, loss=0.119]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [05:11<00:00, 13.21it/s, EM=0.836]

0.8355403608736942

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [07:44<00:00, 79.73it/s, loss=0.0963]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [05:19<00:00, 12.89it/s, EM=0.843]

0.8428148148148148

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [07:41<00:00, 80.27it/s, loss=0.119]

Validation epoch 15: 100%|████████████████████████████████████████████████| 4114/4114 [04:49<00:00, 14.21it/s, EM=0.85]

0.8503969610636277

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [07:47<00:00, 79.22it/s, loss=0.0688]

Validation epoch 16: 100%|████████████████████████████████████████████████| 4114/4114 [05:14<00:00, 13.09it/s, EM=0.85]

0.8495840455840455

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [07:27<00:00, 82.80it/s, loss=0.0837]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [05:39<00:00, 12.12it/s, EM=0.848]

0.8478252611585945

Train epoch 18: 100%|████████████████████████████████████████████████| 37020/37020 [07:27<00:00, 82.70it/s, loss=0.235]

Validation epoch 18: 100%|████████████████████████████████████████████████| 4114/4114 [05:44<00:00, 11.96it/s, EM=0.85]

0.850286799620133

Train epoch 19: 100%|█████████████████████████████████████████████████| 37020/37020 [07:17<00:00, 84.69it/s, loss=0.13]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [06:38<00:00, 10.32it/s, EM=0.856]

0.855965811965812

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [07:37<00:00, 80.95it/s, loss=0.114]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [09:28<00:00, 7.23it/s, EM=0.835]

0.8354453941120608  
  
  
Using batch\_size=64, epochs=20,embed\_dim=1024, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [09:01<00:00, 68.41it/s, loss=0.455]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [06:51<00:00, 10.00it/s, EM=0.513]

0.5128015194681861

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [08:32<00:00, 72.22it/s, loss=0.307]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [09:55<00:00, 6.91it/s, EM=0.607]

0.60664767331434

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [08:19<00:00, 74.18it/s, loss=0.247]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [07:28<00:00, 9.17it/s, EM=0.656]

0.655684710351377

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [09:11<00:00, 67.11it/s, loss=0.277]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [08:37<00:00, 7.94it/s, EM=0.678]

0.6784577397910732

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [07:55<00:00, 77.88it/s, loss=0.218]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [07:57<00:00, 8.62it/s, EM=0.698]

0.6979563152896486

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [09:26<00:00, 65.38it/s, loss=0.178]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [05:49<00:00, 11.77it/s, EM=0.701]

0.7007521367521368

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [08:32<00:00, 72.19it/s, loss=0.208]

Validation epoch 7: 100%|█████████████████████████████████████████████████| 4114/4114 [06:43<00:00, 10.19it/s, EM=0.73]

0.729690408357075

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [09:59<00:00, 61.76it/s, loss=0.139]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.18it/s, EM=0.747]

0.7465185185185185

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [09:18<00:00, 66.25it/s, loss=0.171]

Validation epoch 9: 100%|█████████████████████████████████████████████████| 4114/4114 [06:28<00:00, 10.59it/s, EM=0.76]

0.7595023741690409

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [10:00<00:00, 61.66it/s, loss=0.202]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [04:50<00:00, 14.15it/s, EM=0.767]

0.7673314339981007

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [08:15<00:00, 74.74it/s, loss=0.131]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [07:53<00:00, 8.69it/s, EM=0.773]

0.7732535612535613

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [07:28<00:00, 82.51it/s, loss=0.164]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [07:24<00:00, 9.25it/s, EM=0.772]

0.77194301994302

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [09:17<00:00, 66.38it/s, loss=0.136]

Validation epoch 13: 100%|████████████████████████████████████████████████| 4114/4114 [07:48<00:00, 8.79it/s, EM=0.79]

0.7896638176638177

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [08:00<00:00, 77.11it/s, loss=0.0655]

Validation epoch 14: 100%|████████████████████████████████████████████████| 4114/4114 [09:04<00:00, 7.56it/s, EM=0.79]

0.790320987654321

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [09:06<00:00, 67.68it/s, loss=0.145]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [05:58<00:00, 11.49it/s, EM=0.794]

0.7935308641975308

Train epoch 16: 100%|████████████████████████████████████████████████| 37020/37020 [08:44<00:00, 70.57it/s, loss=0.173]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [09:39<00:00, 7.10it/s, EM=0.822]

0.8221120607787274

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [08:14<00:00, 74.92it/s, loss=0.0709]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [07:31<00:00, 9.11it/s, EM=0.812]

0.8118100664767331

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [09:10<00:00, 67.30it/s, loss=0.0726]

Validation epoch 18: 100%|████████████████████████████████████████████████| 4114/4114 [08:31<00:00, 8.04it/s, EM=0.83]

0.8295194681861349

Train epoch 19: 100%|████████████████████████████████████████████████| 37020/37020 [07:08<00:00, 86.46it/s, loss=0.159]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [07:53<00:00, 8.69it/s, EM=0.821]

0.8210750237416904

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [07:52<00:00, 78.30it/s, loss=0.139]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:02<00:00, 13.59it/s, EM=0.824]

0.8238594491927825  
  
  
  
Using Different Gradient Clip (run\_grad.py) :

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=0.01.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:12<00:00, 85.61it/s, loss=0.295]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:38<00:00, 12.15it/s, EM=0.581]

0.5809154795821463

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [09:43<00:00, 63.46it/s, loss=0.198]

Validation epoch 2: 100%|█████████████████████████████████████████████████| 4114/4114 [06:12<00:00, 11.05it/s, EM=0.67]

0.6704957264957265

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [11:09<00:00, 55.30it/s, loss=0.204]

Validation epoch 3: 100%|█████████████████████████████████████████████████| 4114/4114 [07:28<00:00, 9.17it/s, EM=0.72]

0.7204710351377018

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [11:38<00:00, 52.96it/s, loss=0.225]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [08:01<00:00, 8.54it/s, EM=0.747]

0.747031339031339

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [11:50<00:00, 52.11it/s, loss=0.125]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [07:43<00:00, 8.87it/s, EM=0.799]

0.7985109211775878

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [11:46<00:00, 52.39it/s, loss=0.112]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [08:54<00:00, 7.70it/s, EM=0.802]

0.8020208926875594

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [11:46<00:00, 52.36it/s, loss=0.138]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [08:44<00:00, 7.85it/s, EM=0.826]

0.8261386514719848

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [11:41<00:00, 52.74it/s, loss=0.116]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [08:25<00:00, 8.14it/s, EM=0.837]

0.8369306742640076

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [11:45<00:00, 52.46it/s, loss=0.177]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [07:50<00:00, 8.74it/s, EM=0.852]

0.8520721747388414

Train epoch 10: 100%|███████████████████████████████████████████████| 37020/37020 [11:36<00:00, 53.12it/s, loss=0.0984]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [07:27<00:00, 9.19it/s, EM=0.848]

0.8477378917378917

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [11:42<00:00, 52.71it/s, loss=0.131]

Validation epoch 11: 100%|████████████████████████████████████████████████| 4114/4114 [07:32<00:00, 9.09it/s, EM=0.87]

0.8698081671415004

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [11:43<00:00, 52.59it/s, loss=0.159]

Validation epoch 12: 100%|████████████████████████████████████████████████| 4114/4114 [07:23<00:00, 9.28it/s, EM=0.86]

0.8599050332383665

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [11:40<00:00, 52.81it/s, loss=0.127]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [07:34<00:00, 9.05it/s, EM=0.859]

0.8592022792022792

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [11:52<00:00, 51.98it/s, loss=0.0831]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [07:56<00:00, 8.63it/s, EM=0.862]

0.8619525166191833

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [11:06<00:00, 55.51it/s, loss=0.104]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [07:04<00:00, 9.69it/s, EM=0.867]

0.8665792972459639

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [09:56<00:00, 62.08it/s, loss=0.0797]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [08:16<00:00, 8.29it/s, EM=0.859]

0.8589211775878443

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [09:54<00:00, 62.28it/s, loss=0.0717]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [07:52<00:00, 8.71it/s, EM=0.883]

0.8828224121557455

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [09:56<00:00, 62.06it/s, loss=0.0607]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [07:34<00:00, 9.05it/s, EM=0.881]

0.8813447293447293

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [09:40<00:00, 63.73it/s, loss=0.0826]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [07:05<00:00, 9.66it/s, EM=0.879]

0.8793542260208926

Train epoch 20: 100%|███████████████████████████████████████████████| 37020/37020 [08:48<00:00, 70.09it/s, loss=0.0838]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:56<00:00, 11.55it/s, EM=0.884]

0.884269705603039  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=0.1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:43<00:00, 79.82it/s, loss=0.316]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [04:53<00:00, 14.03it/s, EM=0.601]

0.60140930674264

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [06:44<00:00, 91.47it/s, loss=0.218]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [04:30<00:00, 15.19it/s, EM=0.705]

0.7053789173789173

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [06:35<00:00, 93.59it/s, loss=0.176]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [05:04<00:00, 13.51it/s, EM=0.775]

0.7748414055080721

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [06:50<00:00, 90.16it/s, loss=0.155]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [05:06<00:00, 13.42it/s, EM=0.805]

0.8050864197530864

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [06:32<00:00, 94.25it/s, loss=0.124]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [05:02<00:00, 13.61it/s, EM=0.822]

0.8223209876543209

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [06:55<00:00, 89.12it/s, loss=0.106]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [04:58<00:00, 13.80it/s, EM=0.855]

0.8546438746438746

Train epoch 7: 100%|████████████████████████████████████████████████| 37020/37020 [06:57<00:00, 88.57it/s, loss=0.0947]

Validation epoch 7: 100%|█████████████████████████████████████████████████| 4114/4114 [04:17<00:00, 15.99it/s, EM=0.86]

0.8604634377967711

Train epoch 8: 100%|████████████████████████████████████████████████| 37020/37020 [06:52<00:00, 89.68it/s, loss=0.0898]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [04:43<00:00, 14.53it/s, EM=0.869]

0.869037037037037

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [06:58<00:00, 88.52it/s, loss=0.117]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [04:44<00:00, 14.46it/s, EM=0.875]

0.8752060778727445

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [06:42<00:00, 91.92it/s, loss=0.115]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [04:34<00:00, 14.96it/s, EM=0.881]

0.8812839506172839

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [06:52<00:00, 89.64it/s, loss=0.112]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:05<00:00, 13.46it/s, EM=0.881]

0.8805432098765432

Train epoch 12: 100%|███████████████████████████████████████████████| 37020/37020 [06:47<00:00, 90.91it/s, loss=0.0697]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [04:40<00:00, 14.65it/s, EM=0.887]

0.8872744539411206

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [06:44<00:00, 91.55it/s, loss=0.107]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [05:01<00:00, 13.62it/s, EM=0.891]

0.8907236467236467

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [07:04<00:00, 87.22it/s, loss=0.0896]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [05:04<00:00, 13.52it/s, EM=0.906]

0.9056410256410257

Train epoch 15: 100%|███████████████████████████████████████████████| 37020/37020 [06:47<00:00, 90.92it/s, loss=0.0746]

Validation epoch 15: 100%|█████████████████████████████████████████████████| 4114/4114 [05:13<00:00, 13.12it/s, EM=0.9]

0.9002886989553657

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [07:02<00:00, 87.54it/s, loss=0.0689]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [05:16<00:00, 13.01it/s, EM=0.903]

0.9028983855650522

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [07:01<00:00, 87.73it/s, loss=0.092]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [04:44<00:00, 14.45it/s, EM=0.903]

0.9026780626780627

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [07:04<00:00, 87.20it/s, loss=0.0671]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [05:25<00:00, 12.65it/s, EM=0.902]

0.9022981956315289

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [07:09<00:00, 86.17it/s, loss=0.0883]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [04:48<00:00, 14.28it/s, EM=0.908]

0.9084178537511871

Train epoch 20: 100%|███████████████████████████████████████████████| 37020/37020 [06:47<00:00, 90.83it/s, loss=0.0379]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:20<00:00, 12.82it/s, EM=0.902]

0.9015422602089269  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:13<00:00, 85.30it/s, loss=0.367]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:29<00:00, 12.50it/s, EM=0.568]

0.568338081671415

Train epoch 2: 100%|██████████████████████████████████████████████████| 37020/37020 [07:13<00:00, 85.32it/s, loss=0.23]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [05:58<00:00, 11.48it/s, EM=0.644]

0.6438252611585945

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [07:06<00:00, 86.75it/s, loss=0.235]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [06:47<00:00, 10.11it/s, EM=0.704]

0.703954415954416

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [06:56<00:00, 88.80it/s, loss=0.195]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [06:59<00:00, 9.81it/s, EM=0.719]

0.7192250712250712

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [06:52<00:00, 89.81it/s, loss=0.206]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [07:10<00:00, 9.56it/s, EM=0.757]

0.7569496676163343

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [06:43<00:00, 91.78it/s, loss=0.126]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [05:34<00:00, 12.29it/s, EM=0.779]

0.7789591642924977

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [06:57<00:00, 88.66it/s, loss=0.147]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [06:07<00:00, 11.20it/s, EM=0.788]

0.7876163342830009

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [06:47<00:00, 90.85it/s, loss=0.124]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [05:23<00:00, 12.72it/s, EM=0.785]

0.785272554605888

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [06:48<00:00, 90.73it/s, loss=0.191]

Validation epoch 9: 100%|██████████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.17it/s, EM=0.8]

0.8000531813865147

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [06:38<00:00, 92.93it/s, loss=0.136]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [04:52<00:00, 14.06it/s, EM=0.797]

0.7973561253561253

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [07:04<00:00, 87.29it/s, loss=0.173]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:06<00:00, 13.41it/s, EM=0.811]

0.8112706552706552

Train epoch 12: 100%|█████████████████████████████████████████████████| 37020/37020 [07:08<00:00, 86.36it/s, loss=0.12]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [05:07<00:00, 13.38it/s, EM=0.816]

0.8157454890788224

Train epoch 13: 100%|█████████████████████████████████████████████████| 37020/37020 [06:49<00:00, 90.35it/s, loss=0.17]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [04:43<00:00, 14.53it/s, EM=0.822]

0.8220740740740741

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [07:06<00:00, 86.86it/s, loss=0.0756]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [05:13<00:00, 13.14it/s, EM=0.829]

0.8290636277302944

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [06:58<00:00, 88.41it/s, loss=0.135]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [05:21<00:00, 12.80it/s, EM=0.824]

0.8242507122507122

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [06:41<00:00, 92.24it/s, loss=0.0879]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [04:56<00:00, 13.87it/s, EM=0.835]

0.8353656220322887

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [06:57<00:00, 88.76it/s, loss=0.101]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [04:55<00:00, 13.93it/s, EM=0.838]

0.8375042735042735

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [07:01<00:00, 87.82it/s, loss=0.0749]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [05:42<00:00, 12.01it/s, EM=0.842]

0.8419639126305793

Train epoch 19: 100%|████████████████████████████████████████████████| 37020/37020 [06:53<00:00, 89.57it/s, loss=0.125]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [05:52<00:00, 11.67it/s, EM=0.839]

0.8389515669515669

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [07:06<00:00, 86.78it/s, loss=0.123]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:16<00:00, 13.00it/s, EM=0.845]

0.8450522317188984  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=10.

Train epoch 1: 100%|██████████████████████████████████████████████████| 37020/37020 [07:05<00:00, 87.09it/s, loss=0.33]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:41<00:00, 12.04it/s, EM=0.575]

0.5749097815764482

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [07:39<00:00, 80.63it/s, loss=0.232]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [05:16<00:00, 13.02it/s, EM=0.651]

0.6509439696106363

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [07:48<00:00, 79.09it/s, loss=0.208]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [04:31<00:00, 15.13it/s, EM=0.688]

0.687707502374169

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [07:15<00:00, 85.08it/s, loss=0.231]

Validation epoch 4: 100%|██████████████████████████████████████████████████| 4114/4114 [05:03<00:00, 13.54it/s, EM=0.7]

0.6995859449192783

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [07:31<00:00, 82.00it/s, loss=0.135]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [05:02<00:00, 13.60it/s, EM=0.756]

0.7560417853751187

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [07:00<00:00, 88.10it/s, loss=0.122]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [05:46<00:00, 11.86it/s, EM=0.776]

0.7764710351377018

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [07:02<00:00, 87.67it/s, loss=0.165]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [05:42<00:00, 12.02it/s, EM=0.768]

0.7680835707502374

Train epoch 8: 100%|████████████████████████████████████████████████| 37020/37020 [07:46<00:00, 79.39it/s, loss=0.0913]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [05:10<00:00, 13.27it/s, EM=0.806]

0.8057967711301045

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [07:37<00:00, 80.86it/s, loss=0.208]

Validation epoch 9: 100%|█████████████████████████████████████████████████| 4114/4114 [05:28<00:00, 12.51it/s, EM=0.81]

0.8095536562203229

Train epoch 10: 100%|███████████████████████████████████████████████| 37020/37020 [07:27<00:00, 82.65it/s, loss=0.0612]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [05:19<00:00, 12.89it/s, EM=0.813]

0.8131358024691357

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [07:30<00:00, 82.17it/s, loss=0.199]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:05<00:00, 13.48it/s, EM=0.819]

0.819358024691358

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [07:02<00:00, 87.61it/s, loss=0.156]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [06:05<00:00, 11.27it/s, EM=0.823]

0.8230351377018044

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [07:10<00:00, 86.08it/s, loss=0.115]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [06:35<00:00, 10.41it/s, EM=0.835]

0.8352934472934473

Train epoch 14: 100%|████████████████████████████████████████████████| 37020/37020 [07:23<00:00, 83.52it/s, loss=0.132]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [05:56<00:00, 11.53it/s, EM=0.835]

0.8349515669515669

Train epoch 15: 100%|█████████████████████████████████████████████████| 37020/37020 [07:25<00:00, 83.04it/s, loss=0.15]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [07:12<00:00, 9.52it/s, EM=0.853]

0.8531699905033239

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [07:16<00:00, 84.78it/s, loss=0.0861]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [08:16<00:00, 8.29it/s, EM=0.853]

0.8531509971509972

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [07:10<00:00, 86.07it/s, loss=0.113]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [07:12<00:00, 9.51it/s, EM=0.855]

0.8554871794871794

Train epoch 18: 100%|████████████████████████████████████████████████| 37020/37020 [07:05<00:00, 87.05it/s, loss=0.115]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [06:24<00:00, 10.71it/s, EM=0.857]

0.8573257359924027

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [07:17<00:00, 84.69it/s, loss=0.0926]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [05:41<00:00, 12.05it/s, EM=0.867]

0.8672060778727445

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [07:26<00:00, 82.86it/s, loss=0.056]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:36<00:00, 12.22it/s, EM=0.866]

0.8662108262108262  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=100.

Train epoch 1: 100%|██████████████████████████████████████████████████| 37020/37020 [08:44<00:00, 70.56it/s, loss=0.33]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [08:31<00:00, 8.05it/s, EM=0.575]

0.5749097815764482

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [07:29<00:00, 82.33it/s, loss=0.232]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [07:21<00:00, 9.33it/s, EM=0.651]

0.6509439696106363

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [08:44<00:00, 70.61it/s, loss=0.208]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [07:45<00:00, 8.84it/s, EM=0.688]

0.687707502374169

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [07:12<00:00, 85.54it/s, loss=0.231]

Validation epoch 4: 100%|██████████████████████████████████████████████████| 4114/4114 [07:45<00:00, 8.84it/s, EM=0.7]

0.6995859449192783

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [08:57<00:00, 68.93it/s, loss=0.135]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [06:17<00:00, 10.90it/s, EM=0.756]

0.7560417853751187

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [07:55<00:00, 77.80it/s, loss=0.122]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [07:14<00:00, 9.46it/s, EM=0.776]

0.7764710351377018

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [08:57<00:00, 68.85it/s, loss=0.165]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [05:30<00:00, 12.43it/s, EM=0.768]

0.7680835707502374

Train epoch 8: 100%|████████████████████████████████████████████████| 37020/37020 [08:17<00:00, 74.35it/s, loss=0.0913]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [06:44<00:00, 10.17it/s, EM=0.806]

0.8057967711301045

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [09:30<00:00, 64.87it/s, loss=0.208]

Validation epoch 9: 100%|█████████████████████████████████████████████████| 4114/4114 [05:22<00:00, 12.75it/s, EM=0.81]

0.8095536562203229

Train epoch 10: 100%|███████████████████████████████████████████████| 37020/37020 [08:47<00:00, 70.22it/s, loss=0.0612]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [06:27<00:00, 10.61it/s, EM=0.813]

0.8131358024691357

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [09:29<00:00, 64.96it/s, loss=0.199]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:26<00:00, 12.59it/s, EM=0.819]

0.819358024691358

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [07:51<00:00, 78.49it/s, loss=0.156]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [07:41<00:00, 8.91it/s, EM=0.823]

0.8230351377018044

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [07:17<00:00, 84.66it/s, loss=0.115]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [06:04<00:00, 11.27it/s, EM=0.835]

0.8352934472934473

Train epoch 14: 100%|████████████████████████████████████████████████| 37020/37020 [08:52<00:00, 69.49it/s, loss=0.132]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [07:15<00:00, 9.45it/s, EM=0.835]

0.8349515669515669

Train epoch 15: 100%|█████████████████████████████████████████████████| 37020/37020 [07:40<00:00, 80.47it/s, loss=0.15]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [07:07<00:00, 9.63it/s, EM=0.853]

0.8531699905033239

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [08:48<00:00, 70.00it/s, loss=0.0861]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [07:31<00:00, 9.11it/s, EM=0.853]

0.8531509971509972

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [07:19<00:00, 84.30it/s, loss=0.113]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [07:51<00:00, 8.73it/s, EM=0.855]

0.8554871794871794

Train epoch 18: 100%|████████████████████████████████████████████████| 37020/37020 [08:40<00:00, 71.09it/s, loss=0.115]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [06:25<00:00, 10.67it/s, EM=0.857]

0.8573257359924027

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [07:40<00:00, 80.45it/s, loss=0.0926]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [07:56<00:00, 8.63it/s, EM=0.867]

0.8672060778727445

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [08:56<00:00, 68.95it/s, loss=0.056]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:18<00:00, 12.93it/s, EM=0.866]

0.8662108262108262  
  
  
Different Learning Rate (run\_lr.py):

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.01, grad\_clip=1.

Train epoch 1: 100%|██████████████████████████████████████████████████| 37020/37020 [08:05<00:00, 76.27it/s, loss=1.11]

Validation epoch 1: 100%|███████████████████████████████████████████████| 4114/4114 [06:27<00:00, 10.62it/s, EM=0.0592]

0.05922507122507122

Train epoch 2: 100%|██████████████████████████████████████████████████| 37020/37020 [10:24<00:00, 59.25it/s, loss=1.03]

Validation epoch 2: 100%|███████████████████████████████████████████████| 4114/4114 [07:16<00:00, 9.42it/s, EM=0.0708]

0.07084520417853751

Train epoch 3: 100%|█████████████████████████████████████████████████████| 37020/37020 [12:05<00:00, 51.05it/s, loss=1]

Validation epoch 3: 100%|███████████████████████████████████████████████| 4114/4114 [09:32<00:00, 7.19it/s, EM=0.0694]

0.06942830009496676

Train epoch 4: 100%|██████████████████████████████████████████████████| 37020/37020 [12:05<00:00, 51.06it/s, loss=1.01]

Validation epoch 4: 100%|███████████████████████████████████████████████| 4114/4114 [09:16<00:00, 7.39it/s, EM=0.0655]

0.06554985754985755

Train epoch 5: 100%|██████████████████████████████████████████████████| 37020/37020 [12:03<00:00, 51.16it/s, loss=1.01]

Validation epoch 5: 100%|███████████████████████████████████████████████| 4114/4114 [08:35<00:00, 7.98it/s, EM=0.0708]

0.07083760683760684

Train epoch 6: 100%|██████████████████████████████████████████████████| 37020/37020 [11:53<00:00, 51.90it/s, loss=1.04]

Validation epoch 6: 100%|███████████████████████████████████████████████| 4114/4114 [07:53<00:00, 8.69it/s, EM=0.0741]

0.07407407407407407

Train epoch 7: 100%|██████████████████████████████████████████████████| 37020/37020 [11:53<00:00, 51.90it/s, loss=1.06]

Validation epoch 7: 100%|███████████████████████████████████████████████| 4114/4114 [07:42<00:00, 8.90it/s, EM=0.0702]

0.07022222222222223

Train epoch 8: 100%|██████████████████████████████████████████████████| 37020/37020 [11:57<00:00, 51.60it/s, loss=1.03]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [08:18<00:00, 8.26it/s, EM=0.064]

0.06400379867046534

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [11:59<00:00, 51.44it/s, loss=0.963]

Validation epoch 9: 100%|███████████████████████████████████████████████| 4114/4114 [08:40<00:00, 7.90it/s, EM=0.0678]

0.0677644824311491

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [12:05<00:00, 51.01it/s, loss=0.978]

Validation epoch 10: 100%|██████████████████████████████████████████████| 4114/4114 [09:18<00:00, 7.36it/s, EM=0.0696]

0.0695650522317189

Train epoch 11: 100%|█████████████████████████████████████████████████| 37020/37020 [12:07<00:00, 50.86it/s, loss=1.04]

Validation epoch 11: 100%|██████████████████████████████████████████████| 4114/4114 [08:44<00:00, 7.84it/s, EM=0.0697]

0.06973599240265907

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [12:03<00:00, 51.19it/s, loss=0.976]

Validation epoch 12: 100%|██████████████████████████████████████████████| 4114/4114 [07:54<00:00, 8.66it/s, EM=0.0705]

0.07049952516619183

Train epoch 13: 100%|█████████████████████████████████████████████████| 37020/37020 [12:04<00:00, 51.12it/s, loss=1.05]

Validation epoch 13: 100%|██████████████████████████████████████████████| 4114/4114 [07:42<00:00, 8.90it/s, EM=0.0674]

0.06736942070275403

Train epoch 14: 100%|████████████████████████████████████████████████| 37020/37020 [11:56<00:00, 51.64it/s, loss=0.988]

Validation epoch 14: 100%|██████████████████████████████████████████████| 4114/4114 [06:16<00:00, 10.92it/s, EM=0.0719]

0.07187084520417854

Train epoch 15: 100%|█████████████████████████████████████████████████| 37020/37020 [10:01<00:00, 61.50it/s, loss=1.02]

Validation epoch 15: 100%|██████████████████████████████████████████████| 4114/4114 [06:30<00:00, 10.54it/s, EM=0.0706]

0.07061348528015195

Train epoch 16: 100%|████████████████████████████████████████████████| 37020/37020 [10:00<00:00, 61.66it/s, loss=0.981]

Validation epoch 16: 100%|██████████████████████████████████████████████| 4114/4114 [06:44<00:00, 10.17it/s, EM=0.0709]

0.0709059829059829

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [10:11<00:00, 60.59it/s, loss=0.987]

Validation epoch 17: 100%|██████████████████████████████████████████████| 4114/4114 [07:23<00:00, 9.27it/s, EM=0.0745]

0.07445014245014245

Train epoch 18: 100%|█████████████████████████████████████████████████| 37020/37020 [10:14<00:00, 60.28it/s, loss=1.04]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [07:57<00:00, 8.61it/s, EM=0.073]

0.07301044634377968

Train epoch 19: 100%|█████████████████████████████████████████████████| 37020/37020 [10:17<00:00, 59.91it/s, loss=1.07]

Validation epoch 19: 100%|██████████████████████████████████████████████| 4114/4114 [06:24<00:00, 10.71it/s, EM=0.0707]

0.07071225071225071

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [08:34<00:00, 71.89it/s, loss=0.989]

Validation epoch 20: 100%|██████████████████████████████████████████████| 4114/4114 [05:58<00:00, 11.46it/s, EM=0.0682]

0.06823171889838557  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.003, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:12<00:00, 85.57it/s, loss=0.379]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:27<00:00, 12.56it/s, EM=0.501]

0.500957264957265

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [07:04<00:00, 87.15it/s, loss=0.342]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [05:11<00:00, 13.22it/s, EM=0.537]

0.5372383665716999

Train epoch 3: 100%|██████████████████████████████████████████████████| 37020/37020 [06:58<00:00, 88.46it/s, loss=0.38]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [04:48<00:00, 14.27it/s, EM=0.567]

0.5670579297245963

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [06:58<00:00, 88.42it/s, loss=0.295]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [04:45<00:00, 14.43it/s, EM=0.596]

0.5958670465337133

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [07:22<00:00, 83.68it/s, loss=0.262]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [04:48<00:00, 14.24it/s, EM=0.586]

0.5858119658119658

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [07:29<00:00, 82.37it/s, loss=0.323]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [04:39<00:00, 14.74it/s, EM=0.603]

0.602636277302944

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [07:22<00:00, 83.57it/s, loss=0.271]

Validation epoch 7: 100%|█████████████████████████████████████████████████| 4114/4114 [05:20<00:00, 12.85it/s, EM=0.62]

0.6203988603988604

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [07:09<00:00, 86.11it/s, loss=0.256]

Validation epoch 8: 100%|█████████████████████████████████████████████████| 4114/4114 [05:24<00:00, 12.69it/s, EM=0.62]

0.6196087369420703

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [06:56<00:00, 88.92it/s, loss=0.377]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [06:19<00:00, 10.84it/s, EM=0.629]

0.629249762583096

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [07:09<00:00, 86.27it/s, loss=0.269]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [07:13<00:00, 9.48it/s, EM=0.627]

0.6266096866096866

Train epoch 11: 100%|█████████████████████████████████████████████████| 37020/37020 [06:54<00:00, 89.25it/s, loss=0.33]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [06:03<00:00, 11.31it/s, EM=0.635]

0.6351642924976258

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [07:12<00:00, 85.50it/s, loss=0.203]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [05:32<00:00, 12.36it/s, EM=0.632]

0.6318176638176638

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [06:56<00:00, 88.83it/s, loss=0.281]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [04:46<00:00, 14.38it/s, EM=0.637]

0.6372003798670466

Train epoch 14: 100%|████████████████████████████████████████████████| 37020/37020 [07:09<00:00, 86.21it/s, loss=0.203]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [05:17<00:00, 12.96it/s, EM=0.644]

0.6443228869895536

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [07:07<00:00, 86.55it/s, loss=0.308]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [04:58<00:00, 13.80it/s, EM=0.633]

0.6332231718898386

Train epoch 16: 100%|████████████████████████████████████████████████| 37020/37020 [07:09<00:00, 86.17it/s, loss=0.291]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [05:09<00:00, 13.28it/s, EM=0.646]

0.6458043684710352

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [07:16<00:00, 84.76it/s, loss=0.267]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.15it/s, EM=0.644]

0.6441253561253562

Train epoch 18: 100%|████████████████████████████████████████████████| 37020/37020 [07:11<00:00, 85.75it/s, loss=0.212]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [04:50<00:00, 14.18it/s, EM=0.647]

0.647042735042735

Train epoch 19: 100%|████████████████████████████████████████████████| 37020/37020 [07:25<00:00, 83.14it/s, loss=0.241]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [05:01<00:00, 13.62it/s, EM=0.655]

0.6554719848053181

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [07:23<00:00, 83.49it/s, loss=0.251]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [04:30<00:00, 15.23it/s, EM=0.648]

0.6481747388414055  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:17<00:00, 84.64it/s, loss=0.367]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:07<00:00, 13.39it/s, EM=0.568]

0.568338081671415

Train epoch 2: 100%|██████████████████████████████████████████████████| 37020/37020 [07:16<00:00, 84.77it/s, loss=0.23]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [05:06<00:00, 13.40it/s, EM=0.644]

0.6438252611585945

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [06:51<00:00, 89.93it/s, loss=0.235]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [05:07<00:00, 13.36it/s, EM=0.704]

0.703954415954416

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [07:08<00:00, 86.34it/s, loss=0.195]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [04:59<00:00, 13.74it/s, EM=0.719]

0.7192250712250712

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [06:55<00:00, 89.03it/s, loss=0.206]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [04:41<00:00, 14.64it/s, EM=0.757]

0.7569496676163343

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [07:10<00:00, 85.95it/s, loss=0.126]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [05:00<00:00, 13.71it/s, EM=0.779]

0.7789591642924977

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [07:11<00:00, 85.83it/s, loss=0.147]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [04:43<00:00, 14.51it/s, EM=0.788]

0.7876163342830009

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [06:59<00:00, 88.21it/s, loss=0.124]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.15it/s, EM=0.785]

0.785272554605888

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [07:05<00:00, 87.02it/s, loss=0.191]

Validation epoch 9: 100%|██████████████████████████████████████████████████| 4114/4114 [05:28<00:00, 12.54it/s, EM=0.8]

0.8000531813865147

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [07:10<00:00, 86.05it/s, loss=0.136]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [04:41<00:00, 14.63it/s, EM=0.797]

0.7973561253561253

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [07:20<00:00, 83.98it/s, loss=0.173]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:29<00:00, 12.48it/s, EM=0.811]

0.8112706552706552

Train epoch 12: 100%|█████████████████████████████████████████████████| 37020/37020 [07:00<00:00, 87.97it/s, loss=0.12]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [06:05<00:00, 11.26it/s, EM=0.816]

0.8157454890788224

Train epoch 13: 100%|█████████████████████████████████████████████████| 37020/37020 [07:09<00:00, 86.19it/s, loss=0.17]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [05:51<00:00, 11.70it/s, EM=0.822]

0.8220740740740741

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [07:04<00:00, 87.31it/s, loss=0.0756]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [06:39<00:00, 10.30it/s, EM=0.829]

0.8290636277302944

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [06:57<00:00, 88.60it/s, loss=0.135]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [06:55<00:00, 9.91it/s, EM=0.824]

0.8242507122507122

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [06:59<00:00, 88.32it/s, loss=0.0879]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [06:43<00:00, 10.19it/s, EM=0.835]

0.8353656220322887

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [06:58<00:00, 88.39it/s, loss=0.101]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [05:27<00:00, 12.54it/s, EM=0.838]

0.8375042735042735

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [07:18<00:00, 84.48it/s, loss=0.0749]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [05:23<00:00, 12.71it/s, EM=0.842]

0.8419639126305793

Train epoch 19: 100%|████████████████████████████████████████████████| 37020/37020 [07:19<00:00, 84.29it/s, loss=0.125]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [05:04<00:00, 13.49it/s, EM=0.839]

0.8389515669515669

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [07:56<00:00, 77.65it/s, loss=0.123]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [04:57<00:00, 13.81it/s, EM=0.845]

0.8450522317188984  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.0003, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:53<00:00, 78.15it/s, loss=0.396]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:47<00:00, 11.85it/s, EM=0.582]

0.5816220322886989

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [07:26<00:00, 83.00it/s, loss=0.258]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [05:14<00:00, 13.09it/s, EM=0.659]

0.6587692307692308

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [07:37<00:00, 80.92it/s, loss=0.236]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.17it/s, EM=0.666]

0.6656296296296297

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [07:09<00:00, 86.24it/s, loss=0.211]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [05:19<00:00, 12.88it/s, EM=0.725]

0.7251168091168091

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [07:17<00:00, 84.69it/s, loss=0.153]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [04:53<00:00, 14.00it/s, EM=0.765]

0.7654131054131054

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [08:04<00:00, 76.42it/s, loss=0.184]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [04:28<00:00, 15.31it/s, EM=0.791]

0.790659069325736

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [07:44<00:00, 79.73it/s, loss=0.115]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [05:02<00:00, 13.62it/s, EM=0.822]

0.821735992402659

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [07:38<00:00, 80.76it/s, loss=0.105]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [04:52<00:00, 14.05it/s, EM=0.824]

0.8236999050332384

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [07:36<00:00, 81.11it/s, loss=0.103]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [05:45<00:00, 11.90it/s, EM=0.889]

0.8894283000949668

Train epoch 10: 100%|███████████████████████████████████████████████| 37020/37020 [07:10<00:00, 85.92it/s, loss=0.0393]

Validation epoch 10: 100%|████████████████████████████████████████████████| 4114/4114 [05:49<00:00, 11.78it/s, EM=0.89]

0.8898385565052231

Train epoch 11: 100%|███████████████████████████████████████████████| 37020/37020 [07:01<00:00, 87.76it/s, loss=0.0741]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:17<00:00, 12.94it/s, EM=0.871]

0.8711832858499525

Train epoch 12: 100%|███████████████████████████████████████████████| 37020/37020 [07:37<00:00, 81.00it/s, loss=0.0518]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.16it/s, EM=0.909]

0.9085052231718899

Train epoch 13: 100%|███████████████████████████████████████████████| 37020/37020 [07:41<00:00, 80.26it/s, loss=0.0795]

Validation epoch 13: 100%|████████████████████████████████████████████████| 4114/4114 [05:10<00:00, 13.24it/s, EM=0.92]

0.9199506172839507

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [07:38<00:00, 80.81it/s, loss=0.0534]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [05:01<00:00, 13.64it/s, EM=0.931]

0.930716049382716

Train epoch 15: 100%|███████████████████████████████████████████████| 37020/37020 [07:37<00:00, 80.98it/s, loss=0.0988]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [05:39<00:00, 12.11it/s, EM=0.936]

0.9357302943969611

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [07:19<00:00, 84.17it/s, loss=0.0384]

Validation epoch 16: 100%|████████████████████████████████████████████████| 4114/4114 [05:47<00:00, 11.85it/s, EM=0.93]

0.9299487179487179

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [07:18<00:00, 84.39it/s, loss=0.0503]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [05:36<00:00, 12.24it/s, EM=0.935]

0.9353694207027541

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [07:07<00:00, 86.57it/s, loss=0.0331]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [05:48<00:00, 11.79it/s, EM=0.942]

0.9418841405508073

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [07:21<00:00, 83.83it/s, loss=0.0296]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [07:12<00:00, 9.51it/s, EM=0.935]

0.9347388414055081

Train epoch 20: 100%|███████████████████████████████████████████████| 37020/37020 [08:46<00:00, 70.28it/s, loss=0.0497]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [07:34<00:00, 9.04it/s, EM=0.946]

0.9457094017094018  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.0001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:15<00:00, 84.91it/s, loss=0.483]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [08:26<00:00, 8.13it/s, EM=0.498]

0.4984463437796771

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [08:45<00:00, 70.50it/s, loss=0.304]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [06:33<00:00, 10.45it/s, EM=0.548]

0.5476163342830009

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [07:53<00:00, 78.21it/s, loss=0.327]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [09:15<00:00, 7.41it/s, EM=0.578]

0.5784007597340931

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [08:40<00:00, 71.08it/s, loss=0.294]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [06:34<00:00, 10.42it/s, EM=0.629]

0.6292459639126305

Train epoch 5: 100%|██████████████████████████████████████████████████| 37020/37020 [08:43<00:00, 70.68it/s, loss=0.23]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [10:34<00:00, 6.48it/s, EM=0.681]

0.6805090218423552

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [07:41<00:00, 80.24it/s, loss=0.223]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [08:55<00:00, 7.68it/s, EM=0.697]

0.6966153846153846

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [09:28<00:00, 65.15it/s, loss=0.179]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [07:30<00:00, 9.12it/s, EM=0.691]

0.6913922127255461

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [08:29<00:00, 72.72it/s, loss=0.215]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [07:09<00:00, 9.58it/s, EM=0.748]

0.7479126305792972

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [09:27<00:00, 65.24it/s, loss=0.255]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [05:59<00:00, 11.45it/s, EM=0.723]

0.7227996201329535

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [07:48<00:00, 78.98it/s, loss=0.281]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [07:47<00:00, 8.80it/s, EM=0.743]

0.7425982905982906

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [07:18<00:00, 84.34it/s, loss=0.236]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [06:00<00:00, 11.41it/s, EM=0.768]

0.7678176638176638

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [08:52<00:00, 69.50it/s, loss=0.129]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [07:17<00:00, 9.40it/s, EM=0.787]

0.786962962962963

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [07:42<00:00, 80.01it/s, loss=0.218]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [06:54<00:00, 9.93it/s, EM=0.799]

0.7991111111111111

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [08:47<00:00, 70.22it/s, loss=0.0904]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [07:47<00:00, 8.79it/s, EM=0.807]

0.8070693257359924

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [07:15<00:00, 84.99it/s, loss=0.123]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [07:53<00:00, 8.70it/s, EM=0.817]

0.8169496676163343

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [08:39<00:00, 71.21it/s, loss=0.0954]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [06:30<00:00, 10.53it/s, EM=0.851]

0.8510389363722697

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [07:39<00:00, 80.64it/s, loss=0.112]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [07:47<00:00, 8.80it/s, EM=0.846]

0.8461614434947768

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [08:54<00:00, 69.22it/s, loss=0.0669]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [05:17<00:00, 12.96it/s, EM=0.859]

0.8587464387464387

Train epoch 19: 100%|████████████████████████████████████████████████| 37020/37020 [06:56<00:00, 88.94it/s, loss=0.126]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [08:38<00:00, 7.93it/s, EM=0.871]

0.8706400759734093

Train epoch 20: 100%|███████████████████████████████████████████████| 37020/37020 [06:26<00:00, 95.73it/s, loss=0.0678]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [08:20<00:00, 8.22it/s, EM=0.883]

0.8828338081671415  
  
  
Remove 23 from training data (run\_23\_not\_appear.py):

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [09:56<00:00, 62.03it/s, loss=0.127]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [07:17<00:00, 9.40it/s, EM=0.169]

0.16939031339031338

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [10:18<00:00, 59.82it/s, loss=0.117]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [07:01<00:00, 9.77it/s, EM=0.208]

0.20798480531813865

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [10:19<00:00, 59.76it/s, loss=0.102]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [07:52<00:00, 8.71it/s, EM=0.221]

0.22135232668566002

Train epoch 4: 100%|████████████████████████████████████████████████| 37020/37020 [10:40<00:00, 57.80it/s, loss=0.0762]

Validation epoch 4: 100%|█████████████████████████████████████████████████| 4114/4114 [10:07<00:00, 6.77it/s, EM=0.24]

0.2401139601139601

Train epoch 5: 100%|████████████████████████████████████████████████| 37020/37020 [09:53<00:00, 62.41it/s, loss=0.0172]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [07:19<00:00, 9.35it/s, EM=0.239]

0.23924026590693256

Train epoch 6: 100%|████████████████████████████████████████████████| 37020/37020 [10:10<00:00, 60.63it/s, loss=0.0951]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [08:04<00:00, 8.50it/s, EM=0.256]

0.2557606837606838

Train epoch 7: 100%|████████████████████████████████████████████████| 37020/37020 [09:37<00:00, 64.09it/s, loss=0.0549]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [08:21<00:00, 8.20it/s, EM=0.258]

0.25817663817663816

Train epoch 8: 100%|████████████████████████████████████████████████| 37020/37020 [10:24<00:00, 59.32it/s, loss=0.0962]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [07:31<00:00, 9.11it/s, EM=0.259]

0.25889838556505224

Train epoch 9: 100%|████████████████████████████████████████████████| 37020/37020 [09:33<00:00, 64.60it/s, loss=0.0308]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [08:24<00:00, 8.16it/s, EM=0.264]

0.2642507122507122

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [10:46<00:00, 57.29it/s, loss=0.111]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [07:01<00:00, 9.77it/s, EM=0.261]

0.261329534662868

Train epoch 11: 100%|███████████████████████████████████████████████| 37020/37020 [09:08<00:00, 67.48it/s, loss=0.0447]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [04:28<00:00, 15.34it/s, EM=0.268]

0.2682773029439696

Train epoch 12: 100%|███████████████████████████████████████████████| 37020/37020 [07:15<00:00, 85.03it/s, loss=0.0354]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [04:40<00:00, 14.66it/s, EM=0.269]

0.2693333333333333

Train epoch 13: 100%|███████████████████████████████████████████████| 37020/37020 [07:03<00:00, 87.42it/s, loss=0.0492]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [04:50<00:00, 14.15it/s, EM=0.271]

0.27138081671415004

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [07:01<00:00, 87.74it/s, loss=0.0261]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [04:38<00:00, 14.76it/s, EM=0.271]

0.27077682811016146

Train epoch 15: 100%|███████████████████████████████████████████████| 37020/37020 [07:03<00:00, 87.40it/s, loss=0.0931]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [05:09<00:00, 13.30it/s, EM=0.274]

0.27400949667616337

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [06:52<00:00, 89.66it/s, loss=0.0232]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [03:26<00:00, 19.95it/s, EM=0.271]

0.27065906932573597

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [05:01<00:00, 122.97it/s, loss=0.037]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [02:44<00:00, 25.03it/s, EM=0.276]

0.2761823361823362

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [04:51<00:00, 126.84it/s, loss=0.058]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [02:41<00:00, 25.50it/s, EM=0.279]

0.2785679012345679

Train epoch 19: 100%|██████████████████████████████████████████████| 37020/37020 [04:46<00:00, 129.23it/s, loss=0.0602]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [02:35<00:00, 26.46it/s, EM=0.281]

0.2808622981956315

Train epoch 20: 100%|█████████████████████████████████████████████| 37020/37020 [04:14<00:00, 145.37it/s, loss=0.00918]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [01:44<00:00, 39.42it/s, EM=0.281]

0.28065716999050333

Evaluate on three digit (run\_3digit.py) :

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [09:26<00:00, 65.36it/s, loss=0.367]

Validation epoch 1: 100%|██████████████████████████████████████████████| 4114/4114 [05:44<00:00, 11.95it/s, EM=0.00771]

0.007707502374169041

Train epoch 2: 100%|██████████████████████████████████████████████████| 37020/37020 [09:50<00:00, 62.73it/s, loss=0.23]

Validation epoch 2: 100%|██████████████████████████████████████████████| 4114/4114 [07:50<00:00, 8.75it/s, EM=0.00834]

0.008341880341880341

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [10:12<00:00, 60.40it/s, loss=0.235]

Validation epoch 3: 100%|██████████████████████████████████████████████| 4114/4114 [05:39<00:00, 12.11it/s, EM=0.00854]

0.008535612535612535

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [09:50<00:00, 62.71it/s, loss=0.195]

Validation epoch 4: 100%|██████████████████████████████████████████████| 4114/4114 [07:16<00:00, 9.43it/s, EM=0.00842]

0.008421652421652421

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [09:48<00:00, 62.87it/s, loss=0.206]

Validation epoch 5: 100%|██████████████████████████████████████████████| 4114/4114 [06:04<00:00, 11.27it/s, EM=0.00845]

0.00845204178537512

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [09:34<00:00, 64.45it/s, loss=0.126]

Validation epoch 6: 100%|██████████████████████████████████████████████| 4114/4114 [08:47<00:00, 7.80it/s, EM=0.00862]

0.008615384615384615

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [09:50<00:00, 62.74it/s, loss=0.147]

Validation epoch 7: 100%|██████████████████████████████████████████████| 4114/4114 [07:01<00:00, 9.75it/s, EM=0.00858]

0.008577397910731244

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [09:15<00:00, 66.67it/s, loss=0.124]

Validation epoch 8: 100%|██████████████████████████████████████████████| 4114/4114 [07:36<00:00, 9.01it/s, EM=0.00828]

0.008277302943969611

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [10:00<00:00, 61.62it/s, loss=0.191]

Validation epoch 9: 100%|██████████████████████████████████████████████| 4114/4114 [07:26<00:00, 9.21it/s, EM=0.00852]

0.00851661918328585

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [09:23<00:00, 65.70it/s, loss=0.136]

Validation epoch 10: 100%|█████████████████████████████████████████████| 4114/4114 [07:02<00:00, 9.74it/s, EM=0.00866]

0.008664767331433999

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [10:14<00:00, 60.27it/s, loss=0.173]

Validation epoch 11: 100%|█████████████████████████████████████████████| 4114/4114 [06:57<00:00, 9.86it/s, EM=0.00857]

0.00856980056980057

Train epoch 12: 100%|█████████████████████████████████████████████████| 37020/37020 [08:24<00:00, 73.32it/s, loss=0.12]

Validation epoch 12: 100%|█████████████████████████████████████████████| 4114/4114 [05:05<00:00, 13.48it/s, EM=0.00831]

0.008307692307692308

Train epoch 13: 100%|█████████████████████████████████████████████████| 37020/37020 [07:00<00:00, 87.98it/s, loss=0.17]

Validation epoch 13: 100%|██████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.16it/s, EM=0.0085]

0.008497625830959164

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [06:41<00:00, 92.28it/s, loss=0.0756]

Validation epoch 14: 100%|█████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.15it/s, EM=0.00862]

0.008615384615384615

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [06:47<00:00, 90.94it/s, loss=0.135]

Validation epoch 15: 100%|█████████████████████████████████████████████| 4114/4114 [05:30<00:00, 12.45it/s, EM=0.00843]

0.008429249762583095

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [06:51<00:00, 89.97it/s, loss=0.0879]

Validation epoch 16: 100%|█████████████████████████████████████████████| 4114/4114 [05:20<00:00, 12.82it/s, EM=0.00863]

0.008630579297245963

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [05:53<00:00, 104.63it/s, loss=0.101]

Validation epoch 17: 100%|█████████████████████████████████████████████| 4114/4114 [02:54<00:00, 23.59it/s, EM=0.00877]

0.008767331433998101

Train epoch 18: 100%|██████████████████████████████████████████████| 37020/37020 [04:48<00:00, 128.22it/s, loss=0.0749]

Validation epoch 18: 100%|█████████████████████████████████████████████| 4114/4114 [02:38<00:00, 25.92it/s, EM=0.00859]

0.008592592592592593

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [04:43<00:00, 130.54it/s, loss=0.125]

Validation epoch 19: 100%|█████████████████████████████████████████████| 4114/4114 [02:35<00:00, 26.43it/s, EM=0.00845]

0.00844824311490978

Train epoch 20: 100%|███████████████████████████████████████████████| 37020/37020 [04:37<00:00, 133.28it/s, loss=0.123]

Validation epoch 20: 100%|█████████████████████████████████████████████| 4114/4114 [02:42<00:00, 25.38it/s, EM=0.00892]

0.008919278252611585

One LSTM layer (run\_one\_layer.py):

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [06:13<00:00, 98.98it/s, loss=0.381]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:22<00:00, 12.75it/s, EM=0.525]

0.5248433048433049

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [06:30<00:00, 94.82it/s, loss=0.342]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [09:36<00:00, 7.14it/s, EM=0.607]

0.606750237416904

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [07:05<00:00, 87.08it/s, loss=0.275]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [08:32<00:00, 8.03it/s, EM=0.662]

0.6622830009496676

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [07:31<00:00, 81.95it/s, loss=0.214]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [06:55<00:00, 9.89it/s, EM=0.689]

0.6887445394112061

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [07:21<00:00, 83.89it/s, loss=0.148]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [05:33<00:00, 12.32it/s, EM=0.712]

0.7120037986704654

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [07:34<00:00, 81.45it/s, loss=0.185]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [07:41<00:00, 8.91it/s, EM=0.721]

0.7213903133903133

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [07:31<00:00, 81.96it/s, loss=0.185]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [07:12<00:00, 9.52it/s, EM=0.723]

0.7226628679962013

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [06:43<00:00, 91.84it/s, loss=0.118]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [06:58<00:00, 9.84it/s, EM=0.746]

0.745667616334283

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [07:28<00:00, 82.61it/s, loss=0.211]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [08:26<00:00, 8.12it/s, EM=0.744]

0.7441443494776828

Train epoch 10: 100%|█████████████████████████████████████████████████| 37020/37020 [07:12<00:00, 85.68it/s, loss=0.21]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [08:04<00:00, 8.49it/s, EM=0.748]

0.7484064577397911

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [07:45<00:00, 79.60it/s, loss=0.212]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [10:01<00:00, 6.84it/s, EM=0.761]

0.7611851851851852

Train epoch 12: 100%|█████████████████████████████████████████████████| 37020/37020 [07:24<00:00, 83.32it/s, loss=0.14]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [06:18<00:00, 10.88it/s, EM=0.759]

0.7589781576448243

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [07:36<00:00, 81.08it/s, loss=0.171]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [07:52<00:00, 8.71it/s, EM=0.759]

0.7591073124406458

Train epoch 14: 100%|████████████████████████████████████████████████| 37020/37020 [07:50<00:00, 78.75it/s, loss=0.169]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [06:50<00:00, 10.02it/s, EM=0.771]

0.7714226020892687

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [06:26<00:00, 95.73it/s, loss=0.111]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [04:21<00:00, 15.73it/s, EM=0.783]

0.7831035137701804

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [05:10<00:00, 119.15it/s, loss=0.166]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [04:45<00:00, 14.43it/s, EM=0.781]

0.7811813865147198

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [05:01<00:00, 122.90it/s, loss=0.157]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [05:32<00:00, 12.36it/s, EM=0.777]

0.77661918328585

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [04:50<00:00, 127.51it/s, loss=0.166]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [06:09<00:00, 11.13it/s, EM=0.779]

0.7787274453941121

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [05:16<00:00, 116.88it/s, loss=0.119]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [05:51<00:00, 11.70it/s, EM=0.786]

0.7858195631528965

Train epoch 20: 100%|██████████████████████████████████████████████| 37020/37020 [05:00<00:00, 123.03it/s, loss=0.0968]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:16<00:00, 13.01it/s, EM=0.793]

0.793363722697056  
  
  
  
  
No gradient clipping (run\_no\_grad\_clip.py) :

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|████████████████████████████████████████████████████████████████████████████████████| 37020/37020 [07:24<00:00, 83.28it/s, loss=0.33]

Validation epoch 1: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [04:20<00:00, 15.79it/s, EM=0.575]

0.5749097815764482

Train epoch 2: 100%|███████████████████████████████████████████████████████████████████████████████████| 37020/37020 [06:58<00:00, 88.52it/s, loss=0.232]

Validation epoch 2: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [08:42<00:00, 7.87it/s, EM=0.651]

0.6509439696106363

Train epoch 3: 100%|███████████████████████████████████████████████████████████████████████████████████| 37020/37020 [06:34<00:00, 93.93it/s, loss=0.208]

Validation epoch 3: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [08:22<00:00, 8.19it/s, EM=0.688]

0.687707502374169

Train epoch 4: 100%|███████████████████████████████████████████████████████████████████████████████████| 37020/37020 [09:16<00:00, 66.56it/s, loss=0.231]

Validation epoch 4: 100%|████████████████████████████████████████████████████████████████████████████████████| 4114/4114 [05:38<00:00, 12.14it/s, EM=0.7]

0.6995859449192783

Train epoch 5: 100%|███████████████████████████████████████████████████████████████████████████████████| 37020/37020 [09:43<00:00, 63.48it/s, loss=0.135]

Validation epoch 5: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [07:46<00:00, 8.82it/s, EM=0.756]

0.7560417853751187

Train epoch 6: 100%|███████████████████████████████████████████████████████████████████████████████████| 37020/37020 [10:07<00:00, 60.94it/s, loss=0.122]

Validation epoch 6: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [05:39<00:00, 12.11it/s, EM=0.776]

0.7764710351377018

Train epoch 7: 100%|███████████████████████████████████████████████████████████████████████████████████| 37020/37020 [09:42<00:00, 63.57it/s, loss=0.165]

Validation epoch 7: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [07:15<00:00, 9.44it/s, EM=0.768]

0.7680835707502374

Train epoch 8: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [09:39<00:00, 63.86it/s, loss=0.0913]

Validation epoch 8: 100%|██████████████████████████████████████████████████████████████████████████████████| 4114/4114 [05:54<00:00, 11.60it/s, EM=0.806]

0.8057967711301045

Train epoch 9: 100%|███████████████████████████████████████████████████████████████████████████████████| 37020/37020 [09:21<00:00, 65.98it/s, loss=0.208]

Validation epoch 9: 100%|███████████████████████████████████████████████████████████████████████████████████| 4114/4114 [08:16<00:00, 8.29it/s, EM=0.81]

0.8095536562203229

Train epoch 10: 100%|█████████████████████████████████████████████████████████████████████████████████| 37020/37020 [09:41<00:00, 63.71it/s, loss=0.0612]

Validation epoch 10: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [07:31<00:00, 9.12it/s, EM=0.813]

0.8131358024691357

Train epoch 11: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [09:07<00:00, 67.62it/s, loss=0.199]

Validation epoch 11: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [07:11<00:00, 9.53it/s, EM=0.819]

0.819358024691358

Train epoch 12: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [09:50<00:00, 62.64it/s, loss=0.156]

Validation epoch 12: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [07:40<00:00, 8.93it/s, EM=0.823]

0.8230351377018044

Train epoch 13: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [09:06<00:00, 67.80it/s, loss=0.115]

Validation epoch 13: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [07:25<00:00, 9.23it/s, EM=0.835]

0.8352934472934473

Train epoch 14: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [10:10<00:00, 60.59it/s, loss=0.132]

Validation epoch 14: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [06:42<00:00, 10.22it/s, EM=0.835]

0.8349515669515669

Train epoch 15: 100%|███████████████████████████████████████████████████████████████████████████████████| 37020/37020 [08:34<00:00, 71.93it/s, loss=0.15]

Validation epoch 15: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [04:31<00:00, 15.13it/s, EM=0.853]

0.8531699905033239

Train epoch 16: 100%|█████████████████████████████████████████████████████████████████████████████████| 37020/37020 [06:53<00:00, 89.44it/s, loss=0.0861]

Validation epoch 16: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [04:36<00:00, 14.90it/s, EM=0.853]

0.8531509971509972

Train epoch 17: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [06:44<00:00, 91.62it/s, loss=0.113]

Validation epoch 17: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [04:55<00:00, 13.92it/s, EM=0.855]

0.8554871794871794

Train epoch 18: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [06:42<00:00, 91.95it/s, loss=0.115]

Validation epoch 18: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [04:48<00:00, 14.26it/s, EM=0.857]

0.8573257359924027

Train epoch 19: 100%|█████████████████████████████████████████████████████████████████████████████████| 37020/37020 [06:37<00:00, 93.16it/s, loss=0.0926]

Validation epoch 19: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [04:52<00:00, 14.05it/s, EM=0.867]

0.8672060778727445

Train epoch 20: 100%|██████████████████████████████████████████████████████████████████████████████████| 37020/37020 [06:38<00:00, 92.81it/s, loss=0.056]

Validation epoch 20: 100%|█████████████████████████████████████████████████████████████████████████████████| 4114/4114 [04:18<00:00, 15.89it/s, EM=0.866]

0.8662108262108262  
  
  
Run best hyperparameters combination (run\_best\_hyper.py):

Using batch\_size=256, epochs=20,embed\_dim=64, hidden\_dim=1024, lr=0.0003, grad\_clip=0.1.

Train epoch 1: 100%|█████████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:43<00:00, 26.96it/s, loss=0.299]

Validation epoch 1: 100%|██████████████████████████████████████████████████████████████████████████████████| 1029/1029 [01:02<00:00, 16.55it/s, EM=0.591]

0.591278252611586

Train epoch 2: 100%|█████████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:47<00:00, 26.65it/s, loss=0.208]

Validation epoch 2: 100%|██████████████████████████████████████████████████████████████████████████████████| 1029/1029 [01:01<00:00, 16.79it/s, EM=0.746]

0.7464615384615385

Train epoch 3: 100%|████████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:46<00:00, 26.69it/s, loss=0.0831]

Validation epoch 3: 100%|██████████████████████████████████████████████████████████████████████████████████| 1029/1029 [01:02<00:00, 16.41it/s, EM=0.857]

0.856596391263058

Train epoch 4: 100%|████████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:43<00:00, 26.96it/s, loss=0.0496]

Validation epoch 4: 100%|██████████████████████████████████████████████████████████████████████████████████| 1029/1029 [01:02<00:00, 16.58it/s, EM=0.919]

0.9194453941120608

Train epoch 5: 100%|█████████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:41<00:00, 27.08it/s, loss=0.033]

Validation epoch 5: 100%|██████████████████████████████████████████████████████████████████████████████████| 1029/1029 [01:01<00:00, 16.71it/s, EM=0.956]

0.9559202279202279

Train epoch 6: 100%|████████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:41<00:00, 27.13it/s, loss=0.0378]

Validation epoch 6: 100%|███████████████████████████████████████████████████████████████████████████████████| 1029/1029 [01:01<00:00, 16.60it/s, EM=0.94]

0.9402013295346628

Train epoch 7: 100%|██████████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:36<00:00, 27.50it/s, loss=0.02]

Validation epoch 7: 100%|██████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:59<00:00, 17.41it/s, EM=0.962]

0.961982905982906

Train epoch 8: 100%|████████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:35<00:00, 27.61it/s, loss=0.0217]

Validation epoch 8: 100%|██████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:59<00:00, 17.39it/s, EM=0.972]

0.9718518518518519

Train epoch 9: 100%|████████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:36<00:00, 27.51it/s, loss=0.0125]

Validation epoch 9: 100%|██████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:58<00:00, 17.62it/s, EM=0.984]

0.983718898385565

Train epoch 10: 100%|██████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:36<00:00, 27.54it/s, loss=0.00905]

Validation epoch 10: 100%|█████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:59<00:00, 17.40it/s, EM=0.989]

0.9886267806267807

Train epoch 11: 100%|███████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:37<00:00, 27.45it/s, loss=0.0163]

Validation epoch 11: 100%|█████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:59<00:00, 17.37it/s, EM=0.983]

0.9825223171889839

Train epoch 12: 100%|██████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:36<00:00, 27.49it/s, loss=0.00866]

Validation epoch 12: 100%|█████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:58<00:00, 17.57it/s, EM=0.988]

0.987897435897436

Train epoch 13: 100%|████████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:36<00:00, 27.50it/s, loss=0.015]

Validation epoch 13: 100%|█████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:58<00:00, 17.62it/s, EM=0.984]

0.9841595441595442

Train epoch 14: 100%|███████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:36<00:00, 27.49it/s, loss=0.0106]

Validation epoch 14: 100%|█████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:58<00:00, 17.67it/s, EM=0.993]

0.9926457739791074

Train epoch 15: 100%|██████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:36<00:00, 27.51it/s, loss=0.00706]

Validation epoch 15: 100%|██████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:58<00:00, 17.50it/s, EM=0.99]

0.9895384615384616

Train epoch 16: 100%|██████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:38<00:00, 27.37it/s, loss=0.00547]

Validation epoch 16: 100%|█████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:58<00:00, 17.69it/s, EM=0.987]

0.9868679962013296

Train epoch 17: 100%|██████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:38<00:00, 27.37it/s, loss=0.00596]

Validation epoch 17: 100%|█████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:59<00:00, 17.39it/s, EM=0.993]

0.9928243114909782

Train epoch 18: 100%|██████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:36<00:00, 27.48it/s, loss=0.00397]

Validation epoch 18: 100%|█████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:58<00:00, 17.62it/s, EM=0.995]

0.9947996201329534

Train epoch 19: 100%|██████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:37<00:00, 27.45it/s, loss=0.00292]

Validation epoch 19: 100%|█████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:58<00:00, 17.47it/s, EM=0.994]

0.9936296296296296

Train epoch 20: 100%|██████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:37<00:00, 27.41it/s, loss=0.00125]

Validation epoch 20: 100%|█████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:59<00:00, 17.36it/s, EM=0.994]

0.9936296296296296

Train epoch 20: 100%|██████████████████████████████████████████████████████████████████████████████████| 9255/9255 [05:37<00:00, 27.41it/s, loss=0.00125]

Validation epoch 20: 100%|█████████████████████████████████████████████████████████████████████████████████| 1029/1029 [00:59<00:00, 17.36it/s, EM=0.994]

0.9936486229819563

Evaluate 3digit with best hyperparameters (run\_3digit\_best.py):

Using batch\_size=256, epochs=20,embed\_dim=64, hidden\_dim=1024, lr=0.0003, grad\_clip=0.1.

Train epoch 1: 100%|███████████████████████████████████████████████████| 9255/9255 [09:11<00:00, 16.79it/s, loss=0.299]

Validation epoch 1: 100%|██████████████████████████████████████████████| 1029/1029 [02:02<00:00, 8.41it/s, EM=0.00735]

0.007350427350427351

Train epoch 2: 100%|███████████████████████████████████████████████████| 9255/9255 [09:10<00:00, 16.80it/s, loss=0.208]

Validation epoch 2: 100%|██████████████████████████████████████████████| 1029/1029 [02:01<00:00, 8.47it/s, EM=0.00882]

0.00882051282051282

Train epoch 3: 100%|██████████████████████████████████████████████████| 9255/9255 [09:07<00:00, 16.89it/s, loss=0.0831]

Validation epoch 3: 100%|██████████████████████████████████████████████| 1029/1029 [02:17<00:00, 7.46it/s, EM=0.00874]

0.008736942070275403

Train epoch 4: 100%|██████████████████████████████████████████████████| 9255/9255 [09:03<00:00, 17.03it/s, loss=0.0496]

Validation epoch 4: 100%|██████████████████████████████████████████████| 1029/1029 [02:10<00:00, 7.87it/s, EM=0.00897]

0.008972459639126306

Train epoch 5: 100%|███████████████████████████████████████████████████| 9255/9255 [08:58<00:00, 17.19it/s, loss=0.033]

Validation epoch 5: 100%|██████████████████████████████████████████████| 1029/1029 [02:19<00:00, 7.38it/s, EM=0.00955]

0.00954985754985755

Train epoch 6: 100%|██████████████████████████████████████████████████| 9255/9255 [08:48<00:00, 17.50it/s, loss=0.0378]

Validation epoch 6: 100%|██████████████████████████████████████████████| 1029/1029 [02:18<00:00, 7.44it/s, EM=0.00944]

0.009439696106362774

Train epoch 7: 100%|████████████████████████████████████████████████████| 9255/9255 [08:50<00:00, 17.45it/s, loss=0.02]

Validation epoch 7: 100%|██████████████████████████████████████████████| 1029/1029 [02:26<00:00, 7.02it/s, EM=0.00936]

0.00936372269705603

Train epoch 8: 100%|██████████████████████████████████████████████████| 9255/9255 [08:40<00:00, 17.79it/s, loss=0.0217]

Validation epoch 8: 100%|██████████████████████████████████████████████| 1029/1029 [02:31<00:00, 6.79it/s, EM=0.00938]

0.009382716049382716

Train epoch 9: 100%|██████████████████████████████████████████████████| 9255/9255 [08:35<00:00, 17.94it/s, loss=0.0125]

Validation epoch 9: 100%|██████████████████████████████████████████████| 1029/1029 [02:39<00:00, 6.44it/s, EM=0.00944]

0.009439696106362774

Train epoch 10: 100%|████████████████████████████████████████████████| 9255/9255 [08:27<00:00, 18.24it/s, loss=0.00905]

Validation epoch 10: 100%|██████████████████████████████████████████████| 1029/1029 [02:44<00:00, 6.27it/s, EM=0.0098]

0.0098005698005698

Train epoch 11: 100%|█████████████████████████████████████████████████| 9255/9255 [08:23<00:00, 18.39it/s, loss=0.0163]

Validation epoch 11: 100%|█████████████████████████████████████████████| 1029/1029 [02:42<00:00, 6.33it/s, EM=0.00948]

0.009477682811016144

Train epoch 12: 100%|████████████████████████████████████████████████| 9255/9255 [08:20<00:00, 18.48it/s, loss=0.00866]

Validation epoch 12: 100%|█████████████████████████████████████████████| 1029/1029 [02:52<00:00, 5.98it/s, EM=0.00942]

0.00941690408357075

Train epoch 13: 100%|██████████████████████████████████████████████████| 9255/9255 [08:16<00:00, 18.63it/s, loss=0.015]

Validation epoch 13: 100%|█████████████████████████████████████████████| 1029/1029 [02:50<00:00, 6.04it/s, EM=0.00946]

0.009458689458689459

Train epoch 14: 100%|█████████████████████████████████████████████████| 9255/9255 [08:12<00:00, 18.81it/s, loss=0.0106]

Validation epoch 14: 100%|█████████████████████████████████████████████| 1029/1029 [02:45<00:00, 6.21it/s, EM=0.00939]

0.009386514719848053

Train epoch 15: 100%|████████████████████████████████████████████████| 9255/9255 [08:18<00:00, 18.58it/s, loss=0.00706]

Validation epoch 15: 100%|█████████████████████████████████████████████| 1029/1029 [02:52<00:00, 5.96it/s, EM=0.00961]

0.009606837606837606

Train epoch 16: 100%|████████████████████████████████████████████████| 9255/9255 [08:13<00:00, 18.75it/s, loss=0.00547]

Validation epoch 16: 100%|█████████████████████████████████████████████| 1029/1029 [02:53<00:00, 5.93it/s, EM=0.00933]

0.009329534662867997

Train epoch 17: 100%|████████████████████████████████████████████████| 9255/9255 [08:12<00:00, 18.79it/s, loss=0.00596]

Validation epoch 17: 100%|█████████████████████████████████████████████| 1029/1029 [02:53<00:00, 5.92it/s, EM=0.00956]

0.009561253561253561

Train epoch 18: 100%|████████████████████████████████████████████████| 9255/9255 [08:15<00:00, 18.68it/s, loss=0.00397]

Validation epoch 18: 100%|█████████████████████████████████████████████| 1029/1029 [03:02<00:00, 5.63it/s, EM=0.00961]

0.009606837606837606

Train epoch 19: 100%|████████████████████████████████████████████████| 9255/9255 [08:01<00:00, 19.21it/s, loss=0.00292]

Validation epoch 19: 100%|█████████████████████████████████████████████| 1029/1029 [03:02<00:00, 5.63it/s, EM=0.00947]

0.00947008547008547

Train epoch 20: 100%|████████████████████████████████████████████████| 9255/9255 [08:00<00:00, 19.27it/s, loss=0.00125]

Validation epoch 20: 100%|█████████████████████████████████████████████| 1029/1029 [01:27<00:00, 11.78it/s, EM=0.00961]

0.009610636277302943

Remove 23 from training data with best hyperparameters (run\_23\_not\_appear\_best.py) :

Using batch\_size=256, epochs=20,embed\_dim=64, hidden\_dim=1024, lr=0.0003, grad\_clip=0.1.

Train epoch 1: 100%|███████████████████████████████████████████████████| 9255/9255 [09:00<00:00, 17.13it/s, loss=0.101]

Validation epoch 1: 100%|████████████████████████████████████████████████| 1029/1029 [02:02<00:00, 8.38it/s, EM=0.182]

0.1815232668566002

Train epoch 2: 100%|██████████████████████████████████████████████████| 9255/9255 [09:00<00:00, 17.14it/s, loss=0.0474]

Validation epoch 2: 100%|████████████████████████████████████████████████| 1029/1029 [02:01<00:00, 8.45it/s, EM=0.265]

0.26462678062678063

Train epoch 3: 100%|█████████████████████████████████████████████████| 9255/9255 [08:59<00:00, 17.16it/s, loss=0.00982]

Validation epoch 3: 100%|████████████████████████████████████████████████| 1029/1029 [02:18<00:00, 7.40it/s, EM=0.311]

0.3112440645773979

Train epoch 4: 100%|██████████████████████████████████████████████████| 9255/9255 [08:51<00:00, 17.40it/s, loss=0.0161]

Validation epoch 4: 100%|████████████████████████████████████████████████| 1029/1029 [02:11<00:00, 7.82it/s, EM=0.324]

0.3235023741690408

Train epoch 5: 100%|██████████████████████████████████████████████████| 9255/9255 [08:45<00:00, 17.60it/s, loss=0.0413]

Validation epoch 5: 100%|████████████████████████████████████████████████| 1029/1029 [02:18<00:00, 7.45it/s, EM=0.343]

0.3433656220322887

Train epoch 6: 100%|██████████████████████████████████████████████████| 9255/9255 [08:37<00:00, 17.89it/s, loss=0.0221]

Validation epoch 6: 100%|████████████████████████████████████████████████| 1029/1029 [02:25<00:00, 7.07it/s, EM=0.345]

0.34469135802469136

Train epoch 7: 100%|█████████████████████████████████████████████████| 9255/9255 [08:37<00:00, 17.87it/s, loss=0.00695]

Validation epoch 7: 100%|████████████████████████████████████████████████| 1029/1029 [02:27<00:00, 6.99it/s, EM=0.353]

0.35268376068376067

Train epoch 8: 100%|███████████████████████████████████████████████████| 9255/9255 [08:28<00:00, 18.21it/s, loss=0.013]

Validation epoch 8: 100%|████████████████████████████████████████████████| 1029/1029 [02:35<00:00, 6.63it/s, EM=0.349]

0.349306742640076

Train epoch 9: 100%|█████████████████████████████████████████████████| 9255/9255 [08:20<00:00, 18.48it/s, loss=0.00633]

Validation epoch 9: 100%|████████████████████████████████████████████████| 1029/1029 [02:42<00:00, 6.33it/s, EM=0.355]

0.35512250712250715

Train epoch 10: 100%|█████████████████████████████████████████████████| 9255/9255 [08:12<00:00, 18.79it/s, loss=0.0276]

Validation epoch 10: 100%|███████████████████████████████████████████████| 1029/1029 [02:50<00:00, 6.05it/s, EM=0.357]

0.3568015194681861

Train epoch 11: 100%|████████████████████████████████████████████████| 9255/9255 [08:09<00:00, 18.92it/s, loss=0.00141]

Validation epoch 11: 100%|███████████████████████████████████████████████| 1029/1029 [02:52<00:00, 5.95it/s, EM=0.359]

0.35876543209876544

Train epoch 12: 100%|████████████████████████████████████████████████| 9255/9255 [08:07<00:00, 18.98it/s, loss=0.00234]

Validation epoch 12: 100%|████████████████████████████████████████████████| 1029/1029 [02:54<00:00, 5.88it/s, EM=0.37]

0.37017283950617286

Train epoch 13: 100%|████████████████████████████████████████████████| 9255/9255 [07:57<00:00, 19.37it/s, loss=0.00545]

Validation epoch 13: 100%|███████████████████████████████████████████████| 1029/1029 [03:00<00:00, 5.69it/s, EM=0.346]

0.3457245963912631

Train epoch 14: 100%|███████████████████████████████████████████████| 9255/9255 [07:58<00:00, 19.36it/s, loss=0.000684]

Validation epoch 14: 100%|███████████████████████████████████████████████| 1029/1029 [03:06<00:00, 5.53it/s, EM=0.359]

0.35920607787274456

Train epoch 15: 100%|████████████████████████████████████████████████| 9255/9255 [08:03<00:00, 19.15it/s, loss=0.00257]

Validation epoch 15: 100%|███████████████████████████████████████████████| 1029/1029 [03:02<00:00, 5.64it/s, EM=0.358]

0.3582070275403609

Train epoch 16: 100%|████████████████████████████████████████████████| 9255/9255 [08:02<00:00, 19.17it/s, loss=0.00149]

Validation epoch 16: 100%|███████████████████████████████████████████████| 1029/1029 [03:04<00:00, 5.58it/s, EM=0.356]

0.3558214624881292

Train epoch 17: 100%|████████████████████████████████████████████████| 9255/9255 [07:56<00:00, 19.41it/s, loss=0.00719]

Validation epoch 17: 100%|███████████████████████████████████████████████| 1029/1029 [03:07<00:00, 5.48it/s, EM=0.362]

0.36177397910731246

Train epoch 18: 100%|████████████████████████████████████████████████| 9255/9255 [07:54<00:00, 19.50it/s, loss=6.12e-5]

Validation epoch 18: 100%|███████████████████████████████████████████████| 1029/1029 [03:08<00:00, 5.47it/s, EM=0.364]

0.36426210826210825

Train epoch 19: 100%|████████████████████████████████████████████████| 9255/9255 [07:43<00:00, 19.97it/s, loss=0.00556]

Validation epoch 19: 100%|███████████████████████████████████████████████| 1029/1029 [03:15<00:00, 5.27it/s, EM=0.352]

0.35164672364672367

Train epoch 20: 100%|████████████████████████████████████████████████| 9255/9255 [07:43<00:00, 19.96it/s, loss=3.23e-5]

Validation epoch 20: 100%|███████████████████████████████████████████████| 1029/1029 [03:19<00:00, 5.15it/s, EM=0.355]

0.35467426400759733

Different batch size (run\_batch.py):

Using batch\_size=16, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|███████████████████████████████████████████████| 148079/148079 [28:42<00:00, 85.97it/s, loss=0.245]

Validation epoch 1: 100%|██████████████████████████████████████████████| 16454/16454 [09:59<00:00, 27.44it/s, EM=0.567]

0.5665185185185185

Train epoch 2: 100%|███████████████████████████████████████████████| 148079/148079 [39:37<00:00, 62.29it/s, loss=0.215]

Validation epoch 2: 100%|██████████████████████████████████████████████| 16454/16454 [09:04<00:00, 30.23it/s, EM=0.618]

0.6175954415954416

Train epoch 3: 100%|███████████████████████████████████████████████| 148079/148079 [39:11<00:00, 62.96it/s, loss=0.263]

Validation epoch 3: 100%|██████████████████████████████████████████████| 16454/16454 [08:53<00:00, 30.85it/s, EM=0.651]

0.6508072174738841

Train epoch 4: 100%|███████████████████████████████████████████████| 148079/148079 [39:17<00:00, 62.81it/s, loss=0.343]

Validation epoch 4: 100%|██████████████████████████████████████████████| 16454/16454 [09:18<00:00, 29.46it/s, EM=0.681]

0.6805052231718899

Train epoch 5: 100%|███████████████████████████████████████████████| 148079/148079 [39:33<00:00, 62.40it/s, loss=0.429]

Validation epoch 5: 100%|██████████████████████████████████████████████| 16454/16454 [10:25<00:00, 26.30it/s, EM=0.697]

0.6974207027540361

Train epoch 6: 100%|███████████████████████████████████████████████| 148079/148079 [39:19<00:00, 62.75it/s, loss=0.116]

Validation epoch 6: 100%|██████████████████████████████████████████████| 16454/16454 [07:33<00:00, 36.25it/s, EM=0.706]

0.7058309591642925

Train epoch 7: 100%|██████████████████████████████████████████████| 148079/148079 [33:22<00:00, 73.95it/s, loss=0.0981]

Validation epoch 7: 100%|██████████████████████████████████████████████| 16454/16454 [07:47<00:00, 35.21it/s, EM=0.714]

0.714377967711301

Train epoch 8: 100%|███████████████████████████████████████████████| 148079/148079 [33:34<00:00, 73.52it/s, loss=0.135]

Validation epoch 8: 100%|███████████████████████████████████████████████| 16454/16454 [07:26<00:00, 36.83it/s, EM=0.72]

0.7196353276353277

Train epoch 9: 100%|███████████████████████████████████████████████| 148079/148079 [24:55<00:00, 99.02it/s, loss=0.251]

Validation epoch 9: 100%|██████████████████████████████████████████████| 16454/16454 [05:35<00:00, 49.01it/s, EM=0.729]

0.7289420702754036

Train epoch 10: 100%|█████████████████████████████████████████████| 148079/148079 [22:33<00:00, 109.37it/s, loss=0.247]

Validation epoch 10: 100%|█████████████████████████████████████████████| 16454/16454 [06:45<00:00, 40.54it/s, EM=0.728]

0.7277758784425451

Train epoch 11: 100%|█████████████████████████████████████████████| 148079/148079 [23:30<00:00, 104.98it/s, loss=0.305]

Validation epoch 11: 100%|█████████████████████████████████████████████| 16454/16454 [05:54<00:00, 46.45it/s, EM=0.743]

0.74326685660019

Train epoch 12: 100%|██████████████████████████████████████████████| 148079/148079 [23:13<00:00, 106.28it/s, loss=0.29]

Validation epoch 12: 100%|█████████████████████████████████████████████| 16454/16454 [07:02<00:00, 38.90it/s, EM=0.733]

0.7327749287749288

Train epoch 13: 100%|████████████████████████████████████████████| 148079/148079 [22:01<00:00, 112.08it/s, loss=0.0794]

Validation epoch 13: 100%|█████████████████████████████████████████████| 16454/16454 [06:20<00:00, 43.19it/s, EM=0.745]

0.7452003798670466

Train epoch 14: 100%|█████████████████████████████████████████████| 148079/148079 [22:19<00:00, 110.57it/s, loss=0.236]

Validation epoch 14: 100%|█████████████████████████████████████████████| 16454/16454 [06:53<00:00, 39.75it/s, EM=0.746]

0.7457435897435898

Train epoch 15: 100%|████████████████████████████████████████████| 148079/148079 [22:14<00:00, 110.93it/s, loss=0.0617]

Validation epoch 15: 100%|█████████████████████████████████████████████| 16454/16454 [05:29<00:00, 49.95it/s, EM=0.755]

0.755019943019943

Train epoch 16: 100%|██████████████████████████████████████████████| 148079/148079 [22:48<00:00, 108.23it/s, loss=0.28]

Validation epoch 16: 100%|█████████████████████████████████████████████| 16454/16454 [07:36<00:00, 36.03it/s, EM=0.745]

0.7445660018993352

Train epoch 17: 100%|█████████████████████████████████████████████| 148079/148079 [23:02<00:00, 107.15it/s, loss=0.179]

Validation epoch 17: 100%|█████████████████████████████████████████████| 16454/16454 [05:51<00:00, 46.79it/s, EM=0.747]

0.7465375118708452

Train epoch 18: 100%|█████████████████████████████████████████████| 148079/148079 [22:20<00:00, 110.46it/s, loss=0.149]

Validation epoch 18: 100%|█████████████████████████████████████████████| 16454/16454 [06:12<00:00, 44.12it/s, EM=0.753]

0.7525166191832858

Train epoch 19: 100%|████████████████████████████████████████████| 148079/148079 [22:03<00:00, 111.92it/s, loss=0.0915]

Validation epoch 19: 100%|█████████████████████████████████████████████| 16454/16454 [06:57<00:00, 39.41it/s, EM=0.755]

0.7550389363722697

Train epoch 20: 100%|█████████████████████████████████████████████| 148079/148079 [22:06<00:00, 111.59it/s, loss=0.151]

Validation epoch 20: 100%|█████████████████████████████████████████████| 16454/16454 [05:46<00:00, 47.51it/s, EM=0.759]

0.7593732193732193

Using batch\_size=32, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|████████████████████████████████████████████████| 74040/74040 [12:06<00:00, 101.89it/s, loss=0.344]

Validation epoch 1: 100%|████████████████████████████████████████████████| 8227/8227 [06:23<00:00, 21.46it/s, EM=0.607]

0.6066058879392213

Train epoch 2: 100%|████████████████████████████████████████████████| 74040/74040 [12:11<00:00, 101.24it/s, loss=0.271]

Validation epoch 2: 100%|████████████████████████████████████████████████| 8227/8227 [05:19<00:00, 25.72it/s, EM=0.708]

0.7078632478632478

Train epoch 3: 100%|██████████████████████████████████████████████████| 74040/74040 [12:26<00:00, 99.17it/s, loss=0.23]

Validation epoch 3: 100%|████████████████████████████████████████████████| 8227/8227 [06:41<00:00, 20.47it/s, EM=0.763]

0.7632706552706553

Train epoch 4: 100%|████████████████████████████████████████████████| 74040/74040 [12:16<00:00, 100.58it/s, loss=0.212]

Validation epoch 4: 100%|████████████████████████████████████████████████| 8227/8227 [05:28<00:00, 25.07it/s, EM=0.768]

0.767639126305793

Train epoch 5: 100%|██████████████████████████████████████████████| 74040/74040 [12:05<00:00, 102.02it/s, loss=0.00249]

Validation epoch 5: 100%|████████████████████████████████████████████████| 8227/8227 [06:01<00:00, 22.73it/s, EM=0.796]

0.7960075973409306

Train epoch 6: 100%|███████████████████████████████████████████████| 74040/74040 [12:05<00:00, 102.11it/s, loss=0.0982]

Validation epoch 6: 100%|████████████████████████████████████████████████| 8227/8227 [06:16<00:00, 21.83it/s, EM=0.818]

0.8180474833808167

Train epoch 7: 100%|█████████████████████████████████████████████████| 74040/74040 [12:30<00:00, 98.66it/s, loss=0.092]

Validation epoch 7: 100%|████████████████████████████████████████████████| 8227/8227 [05:51<00:00, 23.40it/s, EM=0.824]

0.8238366571699905

Train epoch 8: 100%|██████████████████████████████████████████████████| 74040/74040 [12:29<00:00, 98.77it/s, loss=0.15]

Validation epoch 8: 100%|████████████████████████████████████████████████| 8227/8227 [07:52<00:00, 17.42it/s, EM=0.845]

0.8447255460588794

Train epoch 9: 100%|█████████████████████████████████████████████████| 74040/74040 [12:59<00:00, 94.94it/s, loss=0.111]

Validation epoch 9: 100%|████████████████████████████████████████████████| 8227/8227 [05:22<00:00, 25.55it/s, EM=0.854]

0.853948717948718

Train epoch 10: 100%|███████████████████████████████████████████████| 74040/74040 [13:14<00:00, 93.19it/s, loss=0.0313]

Validation epoch 10: 100%|███████████████████████████████████████████████| 8227/8227 [08:26<00:00, 16.24it/s, EM=0.855]

0.8552174738841406

Train epoch 11: 100%|████████████████████████████████████████████████| 74040/74040 [13:00<00:00, 94.92it/s, loss=0.121]

Validation epoch 11: 100%|███████████████████████████████████████████████| 8227/8227 [05:26<00:00, 25.18it/s, EM=0.855]

0.8547730294396961

Train epoch 12: 100%|███████████████████████████████████████████████| 74040/74040 [12:33<00:00, 98.27it/s, loss=0.0569]

Validation epoch 12: 100%|███████████████████████████████████████████████| 8227/8227 [07:30<00:00, 18.26it/s, EM=0.859]

0.8594150047483381

Train epoch 13: 100%|████████████████████████████████████████████████| 74040/74040 [13:18<00:00, 92.74it/s, loss=0.109]

Validation epoch 13: 100%|████████████████████████████████████████████████| 8227/8227 [06:34<00:00, 20.87it/s, EM=0.87]

0.8698461538461538

Train epoch 14: 100%|███████████████████████████████████████████████| 74040/74040 [13:22<00:00, 92.23it/s, loss=0.0821]

Validation epoch 14: 100%|███████████████████████████████████████████████| 8227/8227 [06:56<00:00, 19.77it/s, EM=0.869]

0.869420702754036

Train epoch 15: 100%|████████████████████████████████████████████████| 74040/74040 [12:46<00:00, 96.58it/s, loss=0.151]

Validation epoch 15: 100%|███████████████████████████████████████████████| 8227/8227 [05:49<00:00, 23.53it/s, EM=0.874]

0.8735574548907882

Train epoch 16: 100%|████████████████████████████████████████████████| 74040/74040 [13:09<00:00, 93.77it/s, loss=0.113]

Validation epoch 16: 100%|███████████████████████████████████████████████| 8227/8227 [05:56<00:00, 23.08it/s, EM=0.868]

0.8684064577397911

Train epoch 17: 100%|███████████████████████████████████████████████| 74040/74040 [13:04<00:00, 94.41it/s, loss=0.0978]

Validation epoch 17: 100%|███████████████████████████████████████████████| 8227/8227 [06:48<00:00, 20.15it/s, EM=0.868]

0.8683988603988604

Train epoch 18: 100%|███████████████████████████████████████████████| 74040/74040 [13:13<00:00, 93.33it/s, loss=0.0708]

Validation epoch 18: 100%|███████████████████████████████████████████████| 8227/8227 [06:22<00:00, 21.49it/s, EM=0.872]

0.8718100664767331

Train epoch 19: 100%|████████████████████████████████████████████████| 74040/74040 [13:08<00:00, 93.88it/s, loss=0.104]

Validation epoch 19: 100%|███████████████████████████████████████████████| 8227/8227 [06:33<00:00, 20.92it/s, EM=0.877]

0.8765811965811966

Train epoch 20: 100%|█████████████████████████████████████████████████| 74040/74040 [12:51<00:00, 95.93it/s, loss=0.03]

Validation epoch 20: 100%|███████████████████████████████████████████████| 8227/8227 [06:11<00:00, 22.17it/s, EM=0.865]

0.8647521367521368

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:08<00:00, 86.47it/s, loss=0.367]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:32<00:00, 12.38it/s, EM=0.568]

0.568338081671415

Train epoch 2: 100%|██████████████████████████████████████████████████| 37020/37020 [07:23<00:00, 83.44it/s, loss=0.23]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [05:39<00:00, 12.12it/s, EM=0.644]

0.6438252611585945

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [08:47<00:00, 70.13it/s, loss=0.235]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [08:16<00:00, 8.28it/s, EM=0.704]

0.703954415954416

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [07:18<00:00, 84.36it/s, loss=0.195]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [07:51<00:00, 8.72it/s, EM=0.719]

0.7192250712250712

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [08:44<00:00, 70.63it/s, loss=0.206]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [07:13<00:00, 9.49it/s, EM=0.757]

0.7569496676163343

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [07:24<00:00, 83.35it/s, loss=0.126]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [07:53<00:00, 8.70it/s, EM=0.779]

0.7789591642924977

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [08:58<00:00, 68.76it/s, loss=0.147]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [05:49<00:00, 11.78it/s, EM=0.788]

0.7876163342830009

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [08:02<00:00, 76.76it/s, loss=0.124]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [07:20<00:00, 9.33it/s, EM=0.785]

0.785272554605888

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [08:57<00:00, 68.84it/s, loss=0.191]

Validation epoch 9: 100%|██████████████████████████████████████████████████| 4114/4114 [05:28<00:00, 12.52it/s, EM=0.8]

0.8000531813865147

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [08:33<00:00, 72.10it/s, loss=0.136]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [07:31<00:00, 9.12it/s, EM=0.797]

0.7973561253561253

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [09:23<00:00, 65.69it/s, loss=0.173]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:09<00:00, 13.30it/s, EM=0.811]

0.8112706552706552

Train epoch 12: 100%|█████████████████████████████████████████████████| 37020/37020 [09:07<00:00, 67.67it/s, loss=0.12]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [07:29<00:00, 9.15it/s, EM=0.816]

0.8157454890788224

Train epoch 13: 100%|█████████████████████████████████████████████████| 37020/37020 [09:17<00:00, 66.36it/s, loss=0.17]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [24:51<00:00, 2.76it/s, EM=0.822]

0.8220740740740741

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [08:10<00:00, 75.47it/s, loss=0.0756]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [09:51<00:00, 6.95it/s, EM=0.829]

0.8290636277302944

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [08:25<00:00, 73.29it/s, loss=0.135]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [06:09<00:00, 11.12it/s, EM=0.824]

0.8242507122507122

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [08:42<00:00, 70.83it/s, loss=0.0879]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [08:35<00:00, 7.97it/s, EM=0.835]

0.8353656220322887

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [07:39<00:00, 80.55it/s, loss=0.101]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [07:06<00:00, 9.66it/s, EM=0.838]

0.8375042735042735

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [08:42<00:00, 70.82it/s, loss=0.0749]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [08:18<00:00, 8.25it/s, EM=0.842]

0.8419639126305793

Train epoch 19: 100%|████████████████████████████████████████████████| 37020/37020 [07:09<00:00, 86.20it/s, loss=0.125]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [07:41<00:00, 8.91it/s, EM=0.839]

0.8389515669515669

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [08:57<00:00, 68.92it/s, loss=0.123]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [06:33<00:00, 10.46it/s, EM=0.845]

0.8450522317188984

Using batch\_size=128, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 18510/18510 [04:01<00:00, 76.51it/s, loss=0.333]

Validation epoch 1: 100%|████████████████████████████████████████████████| 2057/2057 [07:40<00:00, 4.46it/s, EM=0.576]

0.5757720797720798

Train epoch 2: 100%|█████████████████████████████████████████████████| 18510/18510 [04:35<00:00, 67.23it/s, loss=0.244]

Validation epoch 2: 100%|████████████████████████████████████████████████| 2057/2057 [07:00<00:00, 4.89it/s, EM=0.668]

0.6678176638176638

Train epoch 3: 100%|█████████████████████████████████████████████████| 18510/18510 [03:52<00:00, 79.55it/s, loss=0.196]

Validation epoch 3: 100%|████████████████████████████████████████████████| 2057/2057 [10:30<00:00, 3.26it/s, EM=0.711]

0.7109211775878442

Train epoch 4: 100%|████████████████████████████████████████████████| 18510/18510 [02:50<00:00, 108.57it/s, loss=0.184]

Validation epoch 4: 100%|████████████████████████████████████████████████| 2057/2057 [07:45<00:00, 4.42it/s, EM=0.732]

0.732315289648623

Train epoch 5: 100%|█████████████████████████████████████████████████| 18510/18510 [04:40<00:00, 65.88it/s, loss=0.192]

Validation epoch 5: 100%|████████████████████████████████████████████████| 2057/2057 [05:22<00:00, 6.38it/s, EM=0.784]

0.7841291547958215

Train epoch 6: 100%|█████████████████████████████████████████████████| 18510/18510 [04:16<00:00, 72.09it/s, loss=0.112]

Validation epoch 6: 100%|████████████████████████████████████████████████| 2057/2057 [08:13<00:00, 4.17it/s, EM=0.813]

0.8126685660018993

Train epoch 7: 100%|████████████████████████████████████████████████| 18510/18510 [05:31<00:00, 55.85it/s, loss=0.0928]

Validation epoch 7: 100%|████████████████████████████████████████████████| 2057/2057 [07:05<00:00, 4.83it/s, EM=0.819]

0.8194719848053181

Train epoch 8: 100%|█████████████████████████████████████████████████| 18510/18510 [05:43<00:00, 53.85it/s, loss=0.108]

Validation epoch 8: 100%|█████████████████████████████████████████████████| 2057/2057 [08:58<00:00, 3.82it/s, EM=0.87]

0.8702830009496676

Train epoch 9: 100%|█████████████████████████████████████████████████| 18510/18510 [05:57<00:00, 51.81it/s, loss=0.111]

Validation epoch 9: 100%|████████████████████████████████████████████████| 2057/2057 [09:07<00:00, 3.76it/s, EM=0.875]

0.8746552706552706

Train epoch 10: 100%|███████████████████████████████████████████████| 18510/18510 [05:39<00:00, 54.47it/s, loss=0.0735]

Validation epoch 10: 100%|███████████████████████████████████████████████| 2057/2057 [08:09<00:00, 4.20it/s, EM=0.878]

0.8781956315289648

Train epoch 11: 100%|███████████████████████████████████████████████| 18510/18510 [06:02<00:00, 51.10it/s, loss=0.0688]

Validation epoch 11: 100%|███████████████████████████████████████████████| 2057/2057 [07:34<00:00, 4.53it/s, EM=0.894]

0.8938575498575498

Train epoch 12: 100%|███████████████████████████████████████████████| 18510/18510 [05:53<00:00, 52.35it/s, loss=0.0828]

Validation epoch 12: 100%|███████████████████████████████████████████████| 2057/2057 [05:41<00:00, 6.02it/s, EM=0.897]

0.8972915479582146

Train epoch 13: 100%|████████████████████████████████████████████████| 18510/18510 [05:27<00:00, 56.47it/s, loss=0.116]

Validation epoch 13: 100%|███████████████████████████████████████████████| 2057/2057 [06:48<00:00, 5.03it/s, EM=0.904]

0.9041215574548908

Train epoch 14: 100%|███████████████████████████████████████████████| 18510/18510 [05:47<00:00, 53.29it/s, loss=0.0461]

Validation epoch 14: 100%|███████████████████████████████████████████████| 2057/2057 [08:19<00:00, 4.12it/s, EM=0.903]

0.9025489078822412

Train epoch 15: 100%|███████████████████████████████████████████████| 18510/18510 [05:50<00:00, 52.87it/s, loss=0.0969]

Validation epoch 15: 100%|███████████████████████████████████████████████| 2057/2057 [05:49<00:00, 5.88it/s, EM=0.904]

0.9040797720797721

Train epoch 16: 100%|███████████████████████████████████████████████| 18510/18510 [05:36<00:00, 55.08it/s, loss=0.0529]

Validation epoch 16: 100%|█████████████████████████████████████████████████| 2057/2057 [06:56<00:00, 4.94it/s, EM=0.9]

0.8999278252611586

Train epoch 17: 100%|███████████████████████████████████████████████| 18510/18510 [05:56<00:00, 51.87it/s, loss=0.0725]

Validation epoch 17: 100%|███████████████████████████████████████████████| 2057/2057 [07:43<00:00, 4.44it/s, EM=0.912]

0.912011396011396

Train epoch 18: 100%|███████████████████████████████████████████████| 18510/18510 [05:32<00:00, 55.75it/s, loss=0.0548]

Validation epoch 18: 100%|███████████████████████████████████████████████| 2057/2057 [06:46<00:00, 5.06it/s, EM=0.913]

0.9129306742640076

Train epoch 19: 100%|████████████████████████████████████████████████| 18510/18510 [05:49<00:00, 53.02it/s, loss=0.058]

Validation epoch 19: 100%|███████████████████████████████████████████████| 2057/2057 [07:28<00:00, 4.59it/s, EM=0.914]

0.9143969610636278

Train epoch 20: 100%|███████████████████████████████████████████████| 18510/18510 [06:15<00:00, 49.24it/s, loss=0.0236]

Validation epoch 20: 100%|███████████████████████████████████████████████| 2057/2057 [06:18<00:00, 5.43it/s, EM=0.918]

0.9184083570750238

Using batch\_size=256, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|███████████████████████████████████████████████████| 9255/9255 [04:43<00:00, 32.68it/s, loss=0.362]

Validation epoch 1: 100%|████████████████████████████████████████████████| 1029/1029 [05:53<00:00, 2.91it/s, EM=0.554]

0.5535536562203229

Train epoch 2: 100%|███████████████████████████████████████████████████| 9255/9255 [04:06<00:00, 37.61it/s, loss=0.243]

Validation epoch 2: 100%|████████████████████████████████████████████████| 1029/1029 [06:13<00:00, 2.75it/s, EM=0.642]

0.6418309591642924

Train epoch 3: 100%|███████████████████████████████████████████████████| 9255/9255 [03:59<00:00, 38.61it/s, loss=0.232]

Validation epoch 3: 100%|████████████████████████████████████████████████| 1029/1029 [06:03<00:00, 2.83it/s, EM=0.694]

0.6937853751187084

Train epoch 4: 100%|████████████████████████████████████████████████████| 9255/9255 [04:00<00:00, 38.44it/s, loss=0.18]

Validation epoch 4: 100%|████████████████████████████████████████████████| 1029/1029 [05:09<00:00, 3.33it/s, EM=0.756]

0.7558176638176638

Train epoch 5: 100%|███████████████████████████████████████████████████| 9255/9255 [04:01<00:00, 38.29it/s, loss=0.149]

Validation epoch 5: 100%|████████████████████████████████████████████████| 1029/1029 [04:24<00:00, 3.89it/s, EM=0.804]

0.8036011396011395

Train epoch 6: 100%|████████████████████████████████████████████████████| 9255/9255 [04:00<00:00, 38.45it/s, loss=0.13]

Validation epoch 6: 100%|████████████████████████████████████████████████| 1029/1029 [04:57<00:00, 3.46it/s, EM=0.828]

0.8283684710351377

Train epoch 7: 100%|██████████████████████████████████████████████████| 9255/9255 [03:53<00:00, 39.63it/s, loss=0.0817]

Validation epoch 7: 100%|████████████████████████████████████████████████| 1029/1029 [05:09<00:00, 3.33it/s, EM=0.848]

0.8477378917378917

Train epoch 8: 100%|██████████████████████████████████████████████████| 9255/9255 [03:28<00:00, 44.39it/s, loss=0.0859]

Validation epoch 8: 100%|████████████████████████████████████████████████| 1029/1029 [02:37<00:00, 6.55it/s, EM=0.882]

0.8815460588793922

Train epoch 9: 100%|██████████████████████████████████████████████████| 9255/9255 [03:17<00:00, 46.92it/s, loss=0.0583]

Validation epoch 9: 100%|████████████████████████████████████████████████| 1029/1029 [02:38<00:00, 6.49it/s, EM=0.892]

0.8924292497625831

Train epoch 10: 100%|█████████████████████████████████████████████████| 9255/9255 [03:08<00:00, 49.15it/s, loss=0.0694]

Validation epoch 10: 100%|███████████████████████████████████████████████| 1029/1029 [03:01<00:00, 5.67it/s, EM=0.896]

0.896360873694207

Train epoch 11: 100%|█████████████████████████████████████████████████| 9255/9255 [02:51<00:00, 53.90it/s, loss=0.0722]

Validation epoch 11: 100%|███████████████████████████████████████████████| 1029/1029 [03:13<00:00, 5.32it/s, EM=0.896]

0.896034188034188

Train epoch 12: 100%|█████████████████████████████████████████████████| 9255/9255 [02:48<00:00, 55.03it/s, loss=0.0756]

Validation epoch 12: 100%|███████████████████████████████████████████████| 1029/1029 [02:02<00:00, 8.42it/s, EM=0.914]

0.9144197530864198

Train epoch 13: 100%|██████████████████████████████████████████████████| 9255/9255 [02:33<00:00, 60.47it/s, loss=0.106]

Validation epoch 13: 100%|███████████████████████████████████████████████| 1029/1029 [01:37<00:00, 10.50it/s, EM=0.913]

0.9133333333333333

Train epoch 14: 100%|█████████████████████████████████████████████████| 9255/9255 [01:59<00:00, 77.26it/s, loss=0.0814]

Validation epoch 14: 100%|███████████████████████████████████████████████| 1029/1029 [00:51<00:00, 20.02it/s, EM=0.906]

0.9061424501424501

Train epoch 15: 100%|█████████████████████████████████████████████████| 9255/9255 [01:59<00:00, 77.23it/s, loss=0.0462]

Validation epoch 15: 100%|████████████████████████████████████████████████| 1029/1029 [00:51<00:00, 20.02it/s, EM=0.93]

0.9299677113010446

Train epoch 16: 100%|█████████████████████████████████████████████████| 9255/9255 [01:59<00:00, 77.53it/s, loss=0.0584]

Validation epoch 16: 100%|███████████████████████████████████████████████| 1029/1029 [00:53<00:00, 19.06it/s, EM=0.923]

0.9229933523266857

Train epoch 17: 100%|█████████████████████████████████████████████████| 9255/9255 [01:59<00:00, 77.67it/s, loss=0.0702]

Validation epoch 17: 100%|███████████████████████████████████████████████| 1029/1029 [00:51<00:00, 19.99it/s, EM=0.932]

0.9319924026590694

Train epoch 18: 100%|█████████████████████████████████████████████████| 9255/9255 [01:59<00:00, 77.55it/s, loss=0.0533]

Validation epoch 18: 100%|████████████████████████████████████████████████| 1029/1029 [00:51<00:00, 20.11it/s, EM=0.92]

0.919684710351377

Train epoch 19: 100%|█████████████████████████████████████████████████| 9255/9255 [01:59<00:00, 77.45it/s, loss=0.0596]

Validation epoch 19: 100%|███████████████████████████████████████████████| 1029/1029 [00:51<00:00, 19.98it/s, EM=0.924]

0.9242089268755935

Train epoch 20: 100%|██████████████████████████████████████████████████| 9255/9255 [01:59<00:00, 77.15it/s, loss=0.153]

Validation epoch 20: 100%|███████████████████████████████████████████████| 1029/1029 [00:51<00:00, 19.81it/s, EM=0.929]

0.9287027540360874

Different Gradient Clip (run\_grad.py):

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=0.01.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:12<00:00, 85.61it/s, loss=0.295]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:38<00:00, 12.15it/s, EM=0.581]

0.5809154795821463

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [09:43<00:00, 63.46it/s, loss=0.198]

Validation epoch 2: 100%|█████████████████████████████████████████████████| 4114/4114 [06:12<00:00, 11.05it/s, EM=0.67]

0.6704957264957265

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [11:09<00:00, 55.30it/s, loss=0.204]

Validation epoch 3: 100%|█████████████████████████████████████████████████| 4114/4114 [07:28<00:00, 9.17it/s, EM=0.72]

0.7204710351377018

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [11:38<00:00, 52.96it/s, loss=0.225]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [08:01<00:00, 8.54it/s, EM=0.747]

0.747031339031339

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [11:50<00:00, 52.11it/s, loss=0.125]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [07:43<00:00, 8.87it/s, EM=0.799]

0.7985109211775878

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [11:46<00:00, 52.39it/s, loss=0.112]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [08:54<00:00, 7.70it/s, EM=0.802]

0.8020208926875594

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [11:46<00:00, 52.36it/s, loss=0.138]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [08:44<00:00, 7.85it/s, EM=0.826]

0.8261386514719848

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [11:41<00:00, 52.74it/s, loss=0.116]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [08:25<00:00, 8.14it/s, EM=0.837]

0.8369306742640076

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [11:45<00:00, 52.46it/s, loss=0.177]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [07:50<00:00, 8.74it/s, EM=0.852]

0.8520721747388414

Train epoch 10: 100%|███████████████████████████████████████████████| 37020/37020 [11:36<00:00, 53.12it/s, loss=0.0984]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [07:27<00:00, 9.19it/s, EM=0.848]

0.8477378917378917

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [11:42<00:00, 52.71it/s, loss=0.131]

Validation epoch 11: 100%|████████████████████████████████████████████████| 4114/4114 [07:32<00:00, 9.09it/s, EM=0.87]

0.8698081671415004

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [11:43<00:00, 52.59it/s, loss=0.159]

Validation epoch 12: 100%|████████████████████████████████████████████████| 4114/4114 [07:23<00:00, 9.28it/s, EM=0.86]

0.8599050332383665

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [11:40<00:00, 52.81it/s, loss=0.127]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [07:34<00:00, 9.05it/s, EM=0.859]

0.8592022792022792

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [11:52<00:00, 51.98it/s, loss=0.0831]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [07:56<00:00, 8.63it/s, EM=0.862]

0.8619525166191833

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [11:06<00:00, 55.51it/s, loss=0.104]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [07:04<00:00, 9.69it/s, EM=0.867]

0.8665792972459639

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [09:56<00:00, 62.08it/s, loss=0.0797]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [08:16<00:00, 8.29it/s, EM=0.859]

0.8589211775878443

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [09:54<00:00, 62.28it/s, loss=0.0717]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [07:52<00:00, 8.71it/s, EM=0.883]

0.8828224121557455

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [09:56<00:00, 62.06it/s, loss=0.0607]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [07:34<00:00, 9.05it/s, EM=0.881]

0.8813447293447293

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [09:40<00:00, 63.73it/s, loss=0.0826]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [07:05<00:00, 9.66it/s, EM=0.879]

0.8793542260208926

Train epoch 20: 100%|███████████████████████████████████████████████| 37020/37020 [08:48<00:00, 70.09it/s, loss=0.0838]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:56<00:00, 11.55it/s, EM=0.884]

0.884269705603039

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=0.1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:43<00:00, 79.82it/s, loss=0.316]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [04:53<00:00, 14.03it/s, EM=0.601]

0.60140930674264

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [06:44<00:00, 91.47it/s, loss=0.218]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [04:30<00:00, 15.19it/s, EM=0.705]

0.7053789173789173

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [06:35<00:00, 93.59it/s, loss=0.176]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [05:04<00:00, 13.51it/s, EM=0.775]

0.7748414055080721

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [06:50<00:00, 90.16it/s, loss=0.155]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [05:06<00:00, 13.42it/s, EM=0.805]

0.8050864197530864

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [06:32<00:00, 94.25it/s, loss=0.124]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [05:02<00:00, 13.61it/s, EM=0.822]

0.8223209876543209

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [06:55<00:00, 89.12it/s, loss=0.106]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [04:58<00:00, 13.80it/s, EM=0.855]

0.8546438746438746

Train epoch 7: 100%|████████████████████████████████████████████████| 37020/37020 [06:57<00:00, 88.57it/s, loss=0.0947]

Validation epoch 7: 100%|█████████████████████████████████████████████████| 4114/4114 [04:17<00:00, 15.99it/s, EM=0.86]

0.8604634377967711

Train epoch 8: 100%|████████████████████████████████████████████████| 37020/37020 [06:52<00:00, 89.68it/s, loss=0.0898]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [04:43<00:00, 14.53it/s, EM=0.869]

0.869037037037037

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [06:58<00:00, 88.52it/s, loss=0.117]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [04:44<00:00, 14.46it/s, EM=0.875]

0.8752060778727445

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [06:42<00:00, 91.92it/s, loss=0.115]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [04:34<00:00, 14.96it/s, EM=0.881]

0.8812839506172839

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [06:52<00:00, 89.64it/s, loss=0.112]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:05<00:00, 13.46it/s, EM=0.881]

0.8805432098765432

Train epoch 12: 100%|███████████████████████████████████████████████| 37020/37020 [06:47<00:00, 90.91it/s, loss=0.0697]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [04:40<00:00, 14.65it/s, EM=0.887]

0.8872744539411206

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [06:44<00:00, 91.55it/s, loss=0.107]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [05:01<00:00, 13.62it/s, EM=0.891]

0.8907236467236467

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [07:04<00:00, 87.22it/s, loss=0.0896]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [05:04<00:00, 13.52it/s, EM=0.906]

0.9056410256410257

Train epoch 15: 100%|███████████████████████████████████████████████| 37020/37020 [06:47<00:00, 90.92it/s, loss=0.0746]

Validation epoch 15: 100%|█████████████████████████████████████████████████| 4114/4114 [05:13<00:00, 13.12it/s, EM=0.9]

0.9002886989553657

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [07:02<00:00, 87.54it/s, loss=0.0689]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [05:16<00:00, 13.01it/s, EM=0.903]

0.9028983855650522

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [07:01<00:00, 87.73it/s, loss=0.092]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [04:44<00:00, 14.45it/s, EM=0.903]

0.9026780626780627

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [07:04<00:00, 87.20it/s, loss=0.0671]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [05:25<00:00, 12.65it/s, EM=0.902]

0.9022981956315289

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [07:09<00:00, 86.17it/s, loss=0.0883]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [04:48<00:00, 14.28it/s, EM=0.908]

0.9084178537511871

Train epoch 20: 100%|███████████████████████████████████████████████| 37020/37020 [06:47<00:00, 90.83it/s, loss=0.0379]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:20<00:00, 12.82it/s, EM=0.902]

0.9015422602089269  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:13<00:00, 85.30it/s, loss=0.367]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:29<00:00, 12.50it/s, EM=0.568]

0.568338081671415

Train epoch 2: 100%|██████████████████████████████████████████████████| 37020/37020 [07:13<00:00, 85.32it/s, loss=0.23]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [05:58<00:00, 11.48it/s, EM=0.644]

0.6438252611585945

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [07:06<00:00, 86.75it/s, loss=0.235]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [06:47<00:00, 10.11it/s, EM=0.704]

0.703954415954416

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [06:56<00:00, 88.80it/s, loss=0.195]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [06:59<00:00, 9.81it/s, EM=0.719]

0.7192250712250712

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [06:52<00:00, 89.81it/s, loss=0.206]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [07:10<00:00, 9.56it/s, EM=0.757]

0.7569496676163343

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [06:43<00:00, 91.78it/s, loss=0.126]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [05:34<00:00, 12.29it/s, EM=0.779]

0.7789591642924977

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [06:57<00:00, 88.66it/s, loss=0.147]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [06:07<00:00, 11.20it/s, EM=0.788]

0.7876163342830009

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [06:47<00:00, 90.85it/s, loss=0.124]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [05:23<00:00, 12.72it/s, EM=0.785]

0.785272554605888

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [06:48<00:00, 90.73it/s, loss=0.191]

Validation epoch 9: 100%|██████████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.17it/s, EM=0.8]

0.8000531813865147

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [06:38<00:00, 92.93it/s, loss=0.136]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [04:52<00:00, 14.06it/s, EM=0.797]

0.7973561253561253

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [07:04<00:00, 87.29it/s, loss=0.173]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:06<00:00, 13.41it/s, EM=0.811]

0.8112706552706552

Train epoch 12: 100%|█████████████████████████████████████████████████| 37020/37020 [07:08<00:00, 86.36it/s, loss=0.12]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [05:07<00:00, 13.38it/s, EM=0.816]

0.8157454890788224

Train epoch 13: 100%|█████████████████████████████████████████████████| 37020/37020 [06:49<00:00, 90.35it/s, loss=0.17]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [04:43<00:00, 14.53it/s, EM=0.822]

0.8220740740740741

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [07:06<00:00, 86.86it/s, loss=0.0756]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [05:13<00:00, 13.14it/s, EM=0.829]

0.8290636277302944

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [06:58<00:00, 88.41it/s, loss=0.135]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [05:21<00:00, 12.80it/s, EM=0.824]

0.8242507122507122

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [06:41<00:00, 92.24it/s, loss=0.0879]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [04:56<00:00, 13.87it/s, EM=0.835]

0.8353656220322887

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [06:57<00:00, 88.76it/s, loss=0.101]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [04:55<00:00, 13.93it/s, EM=0.838]

0.8375042735042735

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [07:01<00:00, 87.82it/s, loss=0.0749]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [05:42<00:00, 12.01it/s, EM=0.842]

0.8419639126305793

Train epoch 19: 100%|████████████████████████████████████████████████| 37020/37020 [06:53<00:00, 89.57it/s, loss=0.125]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [05:52<00:00, 11.67it/s, EM=0.839]

0.8389515669515669

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [07:06<00:00, 86.78it/s, loss=0.123]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:16<00:00, 13.00it/s, EM=0.845]

0.8450522317188984  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=10.

Train epoch 1: 100%|██████████████████████████████████████████████████| 37020/37020 [07:05<00:00, 87.09it/s, loss=0.33]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:41<00:00, 12.04it/s, EM=0.575]

0.5749097815764482

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [07:39<00:00, 80.63it/s, loss=0.232]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [05:16<00:00, 13.02it/s, EM=0.651]

0.6509439696106363

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [07:48<00:00, 79.09it/s, loss=0.208]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [04:31<00:00, 15.13it/s, EM=0.688]

0.687707502374169

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [07:15<00:00, 85.08it/s, loss=0.231]

Validation epoch 4: 100%|██████████████████████████████████████████████████| 4114/4114 [05:03<00:00, 13.54it/s, EM=0.7]

0.6995859449192783

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [07:31<00:00, 82.00it/s, loss=0.135]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [05:02<00:00, 13.60it/s, EM=0.756]

0.7560417853751187

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [07:00<00:00, 88.10it/s, loss=0.122]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [05:46<00:00, 11.86it/s, EM=0.776]

0.7764710351377018

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [07:02<00:00, 87.67it/s, loss=0.165]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [05:42<00:00, 12.02it/s, EM=0.768]

0.7680835707502374

Train epoch 8: 100%|████████████████████████████████████████████████| 37020/37020 [07:46<00:00, 79.39it/s, loss=0.0913]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [05:10<00:00, 13.27it/s, EM=0.806]

0.8057967711301045

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [07:37<00:00, 80.86it/s, loss=0.208]

Validation epoch 9: 100%|█████████████████████████████████████████████████| 4114/4114 [05:28<00:00, 12.51it/s, EM=0.81]

0.8095536562203229

Train epoch 10: 100%|███████████████████████████████████████████████| 37020/37020 [07:27<00:00, 82.65it/s, loss=0.0612]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [05:19<00:00, 12.89it/s, EM=0.813]

0.8131358024691357

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [07:30<00:00, 82.17it/s, loss=0.199]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:05<00:00, 13.48it/s, EM=0.819]

0.819358024691358

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [07:02<00:00, 87.61it/s, loss=0.156]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [06:05<00:00, 11.27it/s, EM=0.823]

0.8230351377018044

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [07:10<00:00, 86.08it/s, loss=0.115]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [06:35<00:00, 10.41it/s, EM=0.835]

0.8352934472934473

Train epoch 14: 100%|████████████████████████████████████████████████| 37020/37020 [07:23<00:00, 83.52it/s, loss=0.132]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [05:56<00:00, 11.53it/s, EM=0.835]

0.8349515669515669

Train epoch 15: 100%|█████████████████████████████████████████████████| 37020/37020 [07:25<00:00, 83.04it/s, loss=0.15]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [07:12<00:00, 9.52it/s, EM=0.853]

0.8531699905033239

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [07:16<00:00, 84.78it/s, loss=0.0861]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [08:16<00:00, 8.29it/s, EM=0.853]

0.8531509971509972

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [07:10<00:00, 86.07it/s, loss=0.113]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [07:12<00:00, 9.51it/s, EM=0.855]

0.8554871794871794

Train epoch 18: 100%|████████████████████████████████████████████████| 37020/37020 [07:05<00:00, 87.05it/s, loss=0.115]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [06:24<00:00, 10.71it/s, EM=0.857]

0.8573257359924027

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [07:17<00:00, 84.69it/s, loss=0.0926]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [05:41<00:00, 12.05it/s, EM=0.867]

0.8672060778727445

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [07:26<00:00, 82.86it/s, loss=0.056]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:36<00:00, 12.22it/s, EM=0.866]

0.8662108262108262  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=100.

Train epoch 1: 100%|██████████████████████████████████████████████████| 37020/37020 [08:44<00:00, 70.56it/s, loss=0.33]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [08:31<00:00, 8.05it/s, EM=0.575]

0.5749097815764482

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [07:29<00:00, 82.33it/s, loss=0.232]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [07:21<00:00, 9.33it/s, EM=0.651]

0.6509439696106363

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [08:44<00:00, 70.61it/s, loss=0.208]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [07:45<00:00, 8.84it/s, EM=0.688]

0.687707502374169

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [07:12<00:00, 85.54it/s, loss=0.231]

Validation epoch 4: 100%|██████████████████████████████████████████████████| 4114/4114 [07:45<00:00, 8.84it/s, EM=0.7]

0.6995859449192783

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [08:57<00:00, 68.93it/s, loss=0.135]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [06:17<00:00, 10.90it/s, EM=0.756]

0.7560417853751187

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [07:55<00:00, 77.80it/s, loss=0.122]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [07:14<00:00, 9.46it/s, EM=0.776]

0.7764710351377018

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [08:57<00:00, 68.85it/s, loss=0.165]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [05:30<00:00, 12.43it/s, EM=0.768]

0.7680835707502374

Train epoch 8: 100%|████████████████████████████████████████████████| 37020/37020 [08:17<00:00, 74.35it/s, loss=0.0913]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [06:44<00:00, 10.17it/s, EM=0.806]

0.8057967711301045

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [09:30<00:00, 64.87it/s, loss=0.208]

Validation epoch 9: 100%|█████████████████████████████████████████████████| 4114/4114 [05:22<00:00, 12.75it/s, EM=0.81]

0.8095536562203229

Train epoch 10: 100%|███████████████████████████████████████████████| 37020/37020 [08:47<00:00, 70.22it/s, loss=0.0612]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [06:27<00:00, 10.61it/s, EM=0.813]

0.8131358024691357

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [09:29<00:00, 64.96it/s, loss=0.199]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:26<00:00, 12.59it/s, EM=0.819]

0.819358024691358

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [07:51<00:00, 78.49it/s, loss=0.156]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [07:41<00:00, 8.91it/s, EM=0.823]

0.8230351377018044

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [07:17<00:00, 84.66it/s, loss=0.115]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [06:04<00:00, 11.27it/s, EM=0.835]

0.8352934472934473

Train epoch 14: 100%|████████████████████████████████████████████████| 37020/37020 [08:52<00:00, 69.49it/s, loss=0.132]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [07:15<00:00, 9.45it/s, EM=0.835]

0.8349515669515669

Train epoch 15: 100%|█████████████████████████████████████████████████| 37020/37020 [07:40<00:00, 80.47it/s, loss=0.15]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [07:07<00:00, 9.63it/s, EM=0.853]

0.8531699905033239

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [08:48<00:00, 70.00it/s, loss=0.0861]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [07:31<00:00, 9.11it/s, EM=0.853]

0.8531509971509972

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [07:19<00:00, 84.30it/s, loss=0.113]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [07:51<00:00, 8.73it/s, EM=0.855]

0.8554871794871794

Train epoch 18: 100%|████████████████████████████████████████████████| 37020/37020 [08:40<00:00, 71.09it/s, loss=0.115]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [06:25<00:00, 10.67it/s, EM=0.857]

0.8573257359924027

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [07:40<00:00, 80.45it/s, loss=0.0926]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [07:56<00:00, 8.63it/s, EM=0.867]

0.8672060778727445

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [08:56<00:00, 68.95it/s, loss=0.056]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:18<00:00, 12.93it/s, EM=0.866]

0.8662108262108262  
  
  
  
  
Different Hidden Size (run\_hidden.py):

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=64, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:01<00:00, 87.84it/s, loss=0.486]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:38<00:00, 12.14it/s, EM=0.393]

0.3934093067426401

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [08:52<00:00, 69.48it/s, loss=0.542]

Validation epoch 2: 100%|█████████████████████████████████████████████████| 4114/4114 [06:14<00:00, 10.99it/s, EM=0.47]

0.46966381766381765

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [09:38<00:00, 63.96it/s, loss=0.327]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [07:32<00:00, 9.10it/s, EM=0.521]

0.5214093067426401

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [10:07<00:00, 60.90it/s, loss=0.403]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [07:38<00:00, 8.98it/s, EM=0.524]

0.5242849002849003

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [10:01<00:00, 61.54it/s, loss=0.357]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [07:37<00:00, 8.99it/s, EM=0.524]

0.5236201329534663

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [10:02<00:00, 61.41it/s, loss=0.391]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [07:51<00:00, 8.72it/s, EM=0.567]

0.5674795821462488

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [09:58<00:00, 61.83it/s, loss=0.325]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [07:50<00:00, 8.75it/s, EM=0.552]

0.551661918328585

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [10:02<00:00, 61.44it/s, loss=0.303]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [07:26<00:00, 9.21it/s, EM=0.584]

0.5843912630579298

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [10:04<00:00, 61.26it/s, loss=0.266]

Validation epoch 9: 100%|█████████████████████████████████████████████████| 4114/4114 [07:39<00:00, 8.96it/s, EM=0.59]

0.5899753086419753

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [10:04<00:00, 61.29it/s, loss=0.331]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [07:43<00:00, 8.88it/s, EM=0.572]

0.5719962013295347

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [10:04<00:00, 61.24it/s, loss=0.255]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [07:23<00:00, 9.28it/s, EM=0.604]

0.6038100664767332

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [10:03<00:00, 61.39it/s, loss=0.348]

Validation epoch 12: 100%|████████████████████████████████████████████████| 4114/4114 [07:43<00:00, 8.88it/s, EM=0.59]

0.5902792022792023

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [10:08<00:00, 60.82it/s, loss=0.334]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [07:36<00:00, 9.01it/s, EM=0.585]

0.5850940170940171

Train epoch 14: 100%|████████████████████████████████████████████████| 37020/37020 [10:07<00:00, 60.89it/s, loss=0.382]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [07:35<00:00, 9.03it/s, EM=0.597]

0.5972839506172839

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [10:14<00:00, 60.26it/s, loss=0.304]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [07:52<00:00, 8.70it/s, EM=0.599]

0.5987958214624881

Train epoch 16: 100%|█████████████████████████████████████████████████| 37020/37020 [10:06<00:00, 61.08it/s, loss=0.21]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [07:13<00:00, 9.48it/s, EM=0.629]

0.6287787274453941

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [08:45<00:00, 70.41it/s, loss=0.335]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [06:51<00:00, 10.00it/s, EM=0.586]

0.5860094966761633

Train epoch 18: 100%|████████████████████████████████████████████████| 37020/37020 [08:37<00:00, 71.58it/s, loss=0.301]

Validation epoch 18: 100%|████████████████████████████████████████████████| 4114/4114 [06:50<00:00, 10.01it/s, EM=0.61]

0.6098575498575498

Train epoch 19: 100%|████████████████████████████████████████████████| 37020/37020 [08:38<00:00, 71.33it/s, loss=0.351]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [06:19<00:00, 10.85it/s, EM=0.624]

0.6237872744539411

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [08:46<00:00, 70.37it/s, loss=0.238]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [06:41<00:00, 10.26it/s, EM=0.586]

0.5860322886989554

Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=128, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [09:14<00:00, 66.75it/s, loss=0.384]

Validation epoch 1: 100%|█████████████████████████████████████████████████| 4114/4114 [06:47<00:00, 10.09it/s, EM=0.49]

0.49042355175688507

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [08:24<00:00, 73.40it/s, loss=0.365]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [05:53<00:00, 11.65it/s, EM=0.528]

0.5282393162393162

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [07:36<00:00, 81.16it/s, loss=0.268]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [04:47<00:00, 14.33it/s, EM=0.567]

0.5670579297245963

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [06:41<00:00, 92.10it/s, loss=0.293]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [04:39<00:00, 14.71it/s, EM=0.614]

0.6138765432098765

Train epoch 5: 100%|██████████████████████████████████████████████████| 37020/37020 [06:34<00:00, 93.84it/s, loss=0.29]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [05:00<00:00, 13.69it/s, EM=0.646]

0.6461234567901235

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [06:25<00:00, 95.94it/s, loss=0.241]

Validation epoch 6: 100%|█████████████████████████████████████████████████| 4114/4114 [04:48<00:00, 14.26it/s, EM=0.66]

0.660360873694207

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [06:29<00:00, 95.00it/s, loss=0.214]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [04:38<00:00, 14.76it/s, EM=0.676]

0.6764862298195632

Train epoch 8: 100%|██████████████████████████████████████████████████| 37020/37020 [06:51<00:00, 89.91it/s, loss=0.24]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [04:49<00:00, 14.20it/s, EM=0.681]

0.6809534662867996

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [06:56<00:00, 88.82it/s, loss=0.253]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [04:49<00:00, 14.21it/s, EM=0.681]

0.6808167141500475

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [06:52<00:00, 89.70it/s, loss=0.277]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [04:48<00:00, 14.24it/s, EM=0.671]

0.6711718898385565

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [06:37<00:00, 93.02it/s, loss=0.206]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.17it/s, EM=0.686]

0.686343779677113

Train epoch 12: 100%|████████████████████████████████████████████████| 37020/37020 [06:27<00:00, 95.62it/s, loss=0.196]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [04:44<00:00, 14.44it/s, EM=0.698]

0.697982905982906

Train epoch 13: 100%|████████████████████████████████████████████████| 37020/37020 [06:42<00:00, 92.05it/s, loss=0.225]

Validation epoch 13: 100%|████████████████████████████████████████████████| 4114/4114 [05:15<00:00, 13.02it/s, EM=0.71]

0.7095536562203228

Train epoch 14: 100%|████████████████████████████████████████████████| 37020/37020 [06:26<00:00, 95.90it/s, loss=0.208]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [04:50<00:00, 14.15it/s, EM=0.725]

0.7252459639126306

Train epoch 15: 100%|█████████████████████████████████████████████████| 37020/37020 [06:23<00:00, 96.62it/s, loss=0.13]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [04:57<00:00, 13.81it/s, EM=0.706]

0.7064805318138652

Train epoch 16: 100%|████████████████████████████████████████████████| 37020/37020 [06:43<00:00, 91.85it/s, loss=0.197]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [04:58<00:00, 13.77it/s, EM=0.726]

0.7258499525166192

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [06:25<00:00, 95.93it/s, loss=0.0925]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [04:40<00:00, 14.68it/s, EM=0.729]

0.7287673314339981

Train epoch 18: 100%|████████████████████████████████████████████████| 37020/37020 [06:35<00:00, 93.59it/s, loss=0.261]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [05:22<00:00, 12.75it/s, EM=0.728]

0.7279202279202279

Train epoch 19: 100%|████████████████████████████████████████████████| 37020/37020 [06:32<00:00, 94.36it/s, loss=0.207]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [05:18<00:00, 12.91it/s, EM=0.726]

0.7262564102564103

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [06:37<00:00, 93.22it/s, loss=0.141]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:32<00:00, 12.38it/s, EM=0.731]

0.7309477682811016  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=256, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [07:13<00:00, 85.32it/s, loss=0.367]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [05:26<00:00, 12.59it/s, EM=0.568]

0.568338081671415

Train epoch 2: 100%|██████████████████████████████████████████████████| 37020/37020 [07:10<00:00, 86.06it/s, loss=0.23]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [04:54<00:00, 13.97it/s, EM=0.644]

0.6438252611585945

Train epoch 3: 100%|█████████████████████████████████████████████████| 37020/37020 [07:23<00:00, 83.55it/s, loss=0.235]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [05:08<00:00, 13.32it/s, EM=0.704]

0.703954415954416

Train epoch 4: 100%|█████████████████████████████████████████████████| 37020/37020 [07:20<00:00, 83.97it/s, loss=0.195]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [04:35<00:00, 14.93it/s, EM=0.719]

0.7192250712250712

Train epoch 5: 100%|█████████████████████████████████████████████████| 37020/37020 [07:17<00:00, 84.54it/s, loss=0.206]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.15it/s, EM=0.757]

0.7569496676163343

Train epoch 6: 100%|█████████████████████████████████████████████████| 37020/37020 [07:15<00:00, 85.06it/s, loss=0.126]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [05:10<00:00, 13.26it/s, EM=0.779]

0.7789591642924977

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [06:50<00:00, 90.08it/s, loss=0.147]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [05:09<00:00, 13.30it/s, EM=0.788]

0.7876163342830009

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [07:07<00:00, 86.69it/s, loss=0.124]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [04:59<00:00, 13.73it/s, EM=0.785]

0.785272554605888

Train epoch 9: 100%|█████████████████████████████████████████████████| 37020/37020 [06:55<00:00, 89.02it/s, loss=0.191]

Validation epoch 9: 100%|██████████████████████████████████████████████████| 4114/4114 [04:38<00:00, 14.75it/s, EM=0.8]

0.8000531813865147

Train epoch 10: 100%|████████████████████████████████████████████████| 37020/37020 [07:09<00:00, 86.13it/s, loss=0.136]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [05:01<00:00, 13.65it/s, EM=0.797]

0.7973561253561253

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [07:11<00:00, 85.82it/s, loss=0.173]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [04:42<00:00, 14.56it/s, EM=0.811]

0.8112706552706552

Train epoch 12: 100%|█████████████████████████████████████████████████| 37020/37020 [06:58<00:00, 88.55it/s, loss=0.12]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [05:05<00:00, 13.46it/s, EM=0.816]

0.8157454890788224

Train epoch 13: 100%|█████████████████████████████████████████████████| 37020/37020 [07:04<00:00, 87.24it/s, loss=0.17]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [05:12<00:00, 13.18it/s, EM=0.822]

0.8220740740740741

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [07:11<00:00, 85.79it/s, loss=0.0756]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [04:28<00:00, 15.31it/s, EM=0.829]

0.8290636277302944

Train epoch 15: 100%|████████████████████████████████████████████████| 37020/37020 [07:18<00:00, 84.45it/s, loss=0.135]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [04:57<00:00, 13.84it/s, EM=0.824]

0.8242507122507122

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [07:01<00:00, 87.87it/s, loss=0.0879]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [05:01<00:00, 13.63it/s, EM=0.835]

0.8353656220322887

Train epoch 17: 100%|████████████████████████████████████████████████| 37020/37020 [07:11<00:00, 85.87it/s, loss=0.101]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [04:38<00:00, 14.75it/s, EM=0.838]

0.8375042735042735

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [07:03<00:00, 87.39it/s, loss=0.0749]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [05:13<00:00, 13.10it/s, EM=0.842]

0.8419639126305793

Train epoch 19: 100%|████████████████████████████████████████████████| 37020/37020 [06:52<00:00, 89.85it/s, loss=0.125]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [05:36<00:00, 12.21it/s, EM=0.839]

0.8389515669515669

Train epoch 20: 100%|████████████████████████████████████████████████| 37020/37020 [07:01<00:00, 87.89it/s, loss=0.123]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:15<00:00, 13.04it/s, EM=0.845]

0.8450522317188984  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=512, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [08:52<00:00, 69.59it/s, loss=0.186]

Validation epoch 1: 100%|█████████████████████████████████████████████████| 4114/4114 [06:45<00:00, 10.14it/s, EM=0.68]

0.6795555555555556

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [08:49<00:00, 69.92it/s, loss=0.173]

Validation epoch 2: 100%|████████████████████████████████████████████████| 4114/4114 [06:34<00:00, 10.43it/s, EM=0.817]

0.8172307692307692

Train epoch 3: 100%|████████████████████████████████████████████████| 37020/37020 [08:50<00:00, 69.77it/s, loss=0.0848]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [05:11<00:00, 13.19it/s, EM=0.862]

0.8620854700854701

Train epoch 4: 100%|████████████████████████████████████████████████| 37020/37020 [09:32<00:00, 64.61it/s, loss=0.0621]

Validation epoch 4: 100%|█████████████████████████████████████████████████| 4114/4114 [04:31<00:00, 15.16it/s, EM=0.88]

0.8804786324786325

Train epoch 5: 100%|████████████████████████████████████████████████| 37020/37020 [09:41<00:00, 63.62it/s, loss=0.0956]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [04:39<00:00, 14.71it/s, EM=0.892]

0.8924368471035138

Train epoch 6: 100%|████████████████████████████████████████████████| 37020/37020 [09:04<00:00, 67.99it/s, loss=0.0716]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [06:10<00:00, 11.11it/s, EM=0.908]

0.908022792022792

Train epoch 7: 100%|█████████████████████████████████████████████████| 37020/37020 [09:20<00:00, 66.06it/s, loss=0.108]

Validation epoch 7: 100%|████████████████████████████████████████████████| 4114/4114 [07:33<00:00, 9.07it/s, EM=0.913]

0.913025641025641

Train epoch 8: 100%|████████████████████████████████████████████████| 37020/37020 [08:56<00:00, 69.04it/s, loss=0.0435]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [05:20<00:00, 12.82it/s, EM=0.917]

0.9172687559354226

Train epoch 9: 100%|████████████████████████████████████████████████| 37020/37020 [09:38<00:00, 64.02it/s, loss=0.0187]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [04:47<00:00, 14.31it/s, EM=0.926]

0.9262336182336183

Train epoch 10: 100%|███████████████████████████████████████████████| 37020/37020 [09:22<00:00, 65.76it/s, loss=0.0189]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [04:47<00:00, 14.31it/s, EM=0.925]

0.9246077872744539

Train epoch 11: 100%|███████████████████████████████████████████████| 37020/37020 [09:17<00:00, 66.35it/s, loss=0.0806]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:41<00:00, 12.03it/s, EM=0.931]

0.9307996201329535

Train epoch 12: 100%|███████████████████████████████████████████████| 37020/37020 [09:17<00:00, 66.37it/s, loss=0.0524]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [06:04<00:00, 11.30it/s, EM=0.933]

0.9325128205128205

Train epoch 13: 100%|███████████████████████████████████████████████| 37020/37020 [08:49<00:00, 69.86it/s, loss=0.0296]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [06:43<00:00, 10.19it/s, EM=0.933]

0.9328281101614435

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [08:57<00:00, 68.90it/s, loss=0.0685]

Validation epoch 14: 100%|████████████████████████████████████████████████| 4114/4114 [06:16<00:00, 10.93it/s, EM=0.94]

0.9397113010446344

Train epoch 15: 100%|███████████████████████████████████████████████| 37020/37020 [08:57<00:00, 68.89it/s, loss=0.0864]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [05:22<00:00, 12.75it/s, EM=0.937]

0.9366685660018993

Train epoch 16: 100%|███████████████████████████████████████████████| 37020/37020 [09:05<00:00, 67.92it/s, loss=0.0234]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [04:48<00:00, 14.28it/s, EM=0.938]

0.9384387464387465

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [09:09<00:00, 67.37it/s, loss=0.0173]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [05:07<00:00, 13.38it/s, EM=0.942]

0.9423399810066476

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [08:51<00:00, 69.62it/s, loss=0.0555]

Validation epoch 18: 100%|███████████████████████████████████████████████| 4114/4114 [06:19<00:00, 10.83it/s, EM=0.941]

0.9413485280151946

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [08:43<00:00, 70.70it/s, loss=0.0268]

Validation epoch 19: 100%|███████████████████████████████████████████████| 4114/4114 [06:47<00:00, 10.10it/s, EM=0.945]

0.944934472934473

Train epoch 20: 100%|███████████████████████████████████████████████| 37020/37020 [08:39<00:00, 71.28it/s, loss=0.0415]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [05:57<00:00, 11.52it/s, EM=0.945]

0.9447787274453942  
  
  
Using batch\_size=64, epochs=20,embed\_dim=256, hidden\_dim=1024, lr=0.001, grad\_clip=1.

Train epoch 1: 100%|█████████████████████████████████████████████████| 37020/37020 [22:03<00:00, 27.97it/s, loss=0.169]

Validation epoch 1: 100%|████████████████████████████████████████████████| 4114/4114 [06:20<00:00, 10.80it/s, EM=0.742]

0.7419221272554606

Train epoch 2: 100%|█████████████████████████████████████████████████| 37020/37020 [21:53<00:00, 28.19it/s, loss=0.163]

Validation epoch 2: 100%|█████████████████████████████████████████████████| 4114/4114 [06:29<00:00, 10.55it/s, EM=0.83]

0.8302905982905983

Train epoch 3: 100%|████████████████████████████████████████████████| 37020/37020 [22:00<00:00, 28.03it/s, loss=0.0768]

Validation epoch 3: 100%|████████████████████████████████████████████████| 4114/4114 [06:07<00:00, 11.20it/s, EM=0.866]

0.866286799620133

Train epoch 4: 100%|████████████████████████████████████████████████| 37020/37020 [22:32<00:00, 27.37it/s, loss=0.0684]

Validation epoch 4: 100%|████████████████████████████████████████████████| 4114/4114 [06:05<00:00, 11.25it/s, EM=0.895]

0.8952592592592593

Train epoch 5: 100%|████████████████████████████████████████████████| 37020/37020 [23:40<00:00, 26.07it/s, loss=0.0962]

Validation epoch 5: 100%|████████████████████████████████████████████████| 4114/4114 [05:39<00:00, 12.10it/s, EM=0.905]

0.9048243114909782

Train epoch 6: 100%|████████████████████████████████████████████████| 37020/37020 [24:15<00:00, 25.43it/s, loss=0.0377]

Validation epoch 6: 100%|████████████████████████████████████████████████| 4114/4114 [05:13<00:00, 13.10it/s, EM=0.913]

0.9130674264007598

Train epoch 7: 100%|████████████████████████████████████████████████| 37020/37020 [17:02<00:00, 36.22it/s, loss=0.0492]

Validation epoch 7: 100%|█████████████████████████████████████████████████| 4114/4114 [06:10<00:00, 11.10it/s, EM=0.92]

0.9198670465337132

Train epoch 8: 100%|█████████████████████████████████████████████████| 37020/37020 [22:20<00:00, 27.62it/s, loss=0.051]

Validation epoch 8: 100%|████████████████████████████████████████████████| 4114/4114 [06:15<00:00, 10.95it/s, EM=0.929]

0.9288167141500475

Train epoch 9: 100%|████████████████████████████████████████████████| 37020/37020 [22:09<00:00, 27.85it/s, loss=0.0343]

Validation epoch 9: 100%|████████████████████████████████████████████████| 4114/4114 [06:12<00:00, 11.05it/s, EM=0.929]

0.9292611585944919

Train epoch 10: 100%|███████████████████████████████████████████████| 37020/37020 [21:52<00:00, 28.20it/s, loss=0.0545]

Validation epoch 10: 100%|███████████████████████████████████████████████| 4114/4114 [06:27<00:00, 10.63it/s, EM=0.936]

0.9361899335232668

Train epoch 11: 100%|████████████████████████████████████████████████| 37020/37020 [22:17<00:00, 27.68it/s, loss=0.101]

Validation epoch 11: 100%|███████████████████████████████████████████████| 4114/4114 [05:44<00:00, 11.94it/s, EM=0.935]

0.934636277302944

Train epoch 12: 100%|███████████████████████████████████████████████| 37020/37020 [16:23<00:00, 37.64it/s, loss=0.0256]

Validation epoch 12: 100%|███████████████████████████████████████████████| 4114/4114 [04:24<00:00, 15.58it/s, EM=0.938]

0.9384919278252611

Train epoch 13: 100%|███████████████████████████████████████████████| 37020/37020 [13:16<00:00, 46.48it/s, loss=0.0322]

Validation epoch 13: 100%|███████████████████████████████████████████████| 4114/4114 [03:15<00:00, 21.02it/s, EM=0.941]

0.9406381766381766

Train epoch 14: 100%|███████████████████████████████████████████████| 37020/37020 [14:03<00:00, 43.88it/s, loss=0.0246]

Validation epoch 14: 100%|███████████████████████████████████████████████| 4114/4114 [04:14<00:00, 16.14it/s, EM=0.942]

0.9424615384615385

Train epoch 15: 100%|███████████████████████████████████████████████| 37020/37020 [19:00<00:00, 32.47it/s, loss=0.0389]

Validation epoch 15: 100%|███████████████████████████████████████████████| 4114/4114 [06:15<00:00, 10.94it/s, EM=0.942]

0.9418689458689459

Train epoch 16: 100%|██████████████████████████████████████████████| 37020/37020 [29:07<00:00, 21.19it/s, loss=0.00997]

Validation epoch 16: 100%|███████████████████████████████████████████████| 4114/4114 [05:59<00:00, 11.44it/s, EM=0.946]

0.9462943969610637

Train epoch 17: 100%|███████████████████████████████████████████████| 37020/37020 [29:00<00:00, 21.27it/s, loss=0.0184]

Validation epoch 17: 100%|███████████████████████████████████████████████| 4114/4114 [07:23<00:00, 9.27it/s, EM=0.947]

0.9472592592592592

Train epoch 18: 100%|███████████████████████████████████████████████| 37020/37020 [28:34<00:00, 21.59it/s, loss=0.0377]

Validation epoch 18: 100%|████████████████████████████████████████████████| 4114/4114 [06:30<00:00, 10.53it/s, EM=0.95]

0.9496714150047484

Train epoch 19: 100%|███████████████████████████████████████████████| 37020/37020 [28:22<00:00, 21.74it/s, loss=0.0122]

Validation epoch 19: 100%|████████████████████████████████████████████████| 4114/4114 [06:24<00:00, 10.69it/s, EM=0.95]

0.9499411206077872

Train epoch 20: 100%|███████████████████████████████████████████████| 37020/37020 [29:23<00:00, 20.99it/s, loss=0.0142]

Validation epoch 20: 100%|███████████████████████████████████████████████| 4114/4114 [07:48<00:00, 8.79it/s, EM=0.949]

0.9488243114909781