

$$|\{\hat{y}_{1:t}^{(k)} \in S_t \mid f(\hat{y}_{1:t}^{(k)}, \hat{h}_{t-1}^{(k)}) > f(\hat{y}_t^{(K)}, \hat{h}_{t-1}^{(K)})\}| = K - 1$$

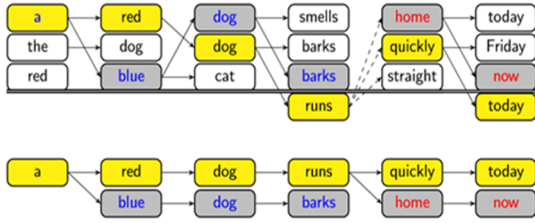
với  $\hat{y}_t^{(k)}$  là token thứ  $t$  trong  $\hat{y}_{1:t}^{(k)}$ ,  $\hat{h}_{t-1}^{(k)}$  là trạng thái ẩn của RNN tương ứng ở bước thứ  $t - 1$ .

Hàm  $\mathcal{L}(f)$  thể hiện lỗi tại mỗi lần điểm số của chuỗi  $y_{1:t}$  không vượt quá điểm số của chuỗi  $\hat{y}_{1:t}^{(K)}$  tính bằng công thức:

$$\mathcal{L}(f) = \sum_{t=1}^T \Delta(\hat{y}_{1:t}^{(K)}) [1 - f(y_t, h_{t-1}) + f(\hat{y}_t^{(K)}, \hat{h}_{t-1}^{(K)})]$$

với  $\Delta(\hat{y}_{1:t}^{(K)})$  biểu thị độ lỗi cụ thể của hàm lỗi; nó trả về giá trị 0 khi các yêu cầu lệ thỏa mãn và ngược lại trả về một số dương. Hàm lỗi được tối ưu ở hai bước tiếp theo.

b. *Forward: (Find Violations) xây dựng tập chuỗi  $S_t$  và tìm kiếm vi phạm lệ mô tả ở Hình 4.*



Hình 4: Beam search với vi phạm lệ [7]

Để tối ưu hóa hàm lỗi, ta xét các trường hợp như sau:

Nếu không có vi phạm lệ ở bước  $t - 1$  thì  $S_t$  được xây dựng từ giải thuật Beam Search tiêu chuẩn.

Nếu có vi phạm lệ,  $S_t$  được xây dựng từ  $K$  chuỗi, bắt đầu bằng chuỗi mục tiêu  $y_{1:t-1}$ .

Giả sử hàm  $\text{succ}$  ánh xạ một chuỗi  $w_{1:t-1} \in \mathcal{V}^{t-1}$  vào một danh sách những chuỗi có độ dài  $t$ , điều này có thể làm bằng cách bổ sung cho nó một từ hợp lệ  $w \in \mathcal{V}$ . Trong trường hợp đơn giản, ta có:

$$\text{succ}(w_{1:t-1}) = \{w_{1:t-1}, w \mid w \in \mathcal{V}\}$$

Sau khi xác định một hàm  $\text{succ}$  thích hợp, chúng ta khai báo một tập hợp:

$$S_t = \text{top}K \left\{ \begin{array}{ll} \text{succ}(y_{1:t-1}) & \text{vi phạm tại } t - 1 \\ \bigcup_{k=1}^K \text{succ}(\hat{y}_{1:t}^{(k)}) & \text{ngược lại} \end{array} \right.$$

ta có một vi phạm lệ tại  $t - 1$  nếu  $f(y_{t-1}, h_{t-2}) < f(\hat{y}_{t-1}^{(K)}, \hat{h}_{t-2}^{(K)}) + 1$  và  $\text{top}K$  để xem xét điểm số cho bởi hàm  $f$ .

c. *Backward: (Merge Sequences) sử dụng lan truyền ngược để tối ưu hàm lỗi của mô hình RNN.*

Giả sử một vi phạm lệ tại bước thứ  $t$  giữa chuỗi dự đoán  $\hat{y}_{1:t}^{(K)}$  và chuỗi nhãn  $y_{1:t}$ . Mô hình Sequence-to-Sequence chuẩn thực hiện lan truyền ngược để tối ưu hàm lỗi thông qua các từ mục tiêu; tuy nhiên ở đây chúng ta có thêm một gradient cho các từ dự đoán sai trước đó. Để lan truyền ngược hàm lỗi thông qua một RNN, chúng ta sử dụng một thủ tục đệ quy *BRNN* – tại mỗi bước thứ  $t$ , nó chứa gradients của bước kế tiếp và lỗi trong tương lai đối với  $h_t$ . Ta có:

$$\nabla_{h_t} \mathcal{L} \leftarrow \text{BRNN}(\nabla_{h_t} \mathcal{L}_{t+1}, \nabla_{h_{t+1}} \mathcal{L})$$

với  $\mathcal{L}_{t+1}$  là lỗi tại bước  $t + 1$ . Chạy hàm *BRNN* từ  $t = T - 1$  đến  $t = 0$  được xem như lan truyền ngược theo thời gian (back-propagation through time).

Giải thuật Beam Search sử dụng cho mô hình Sequence-to-Sequence được tóm tắt ở Hình 5:

#### Algorithm 1 Seq2seq Beam-Search Optimization

```

1: procedure BSO( $x, K_{tr}, \text{succ}$ )
2:   /*FORWARD*/
3:   Init empty storage  $\hat{y}_{1:T}$  and  $\hat{h}_{1:T}$ ; init  $S_1$ 
4:    $r \leftarrow 0$ ;  $\text{violations} \leftarrow \{0\}$ 
5:   for  $t = 1, \dots, T$  do
6:      $K = K_{tr}$  if  $t \neq T$  else  $\arg \max_{k: \hat{y}_{1:t}^{(k)} \neq y_{1:t}} f(\hat{y}_t^{(k)}, \hat{h}_{t-1}^{(k)})$ 
7:     if  $f(y_t, h_{t-1}) < f(\hat{y}_t^{(K)}, \hat{h}_{t-1}^{(K)}) + 1$  then
8:        $\hat{h}_{r:t-1} \leftarrow \hat{h}_{r:t-1}^{(K)}$ 
9:        $\hat{y}_{r+1:t} \leftarrow \hat{y}_{r+1:t}^{(K)}$ 
10:      Add  $t$  to  $\text{violations}$ 
11:       $r \leftarrow t$ 
12:       $S_{t+1} \leftarrow \text{top}K(\text{succ}(y_{1:t}))$ 
13:    else
14:       $S_{t+1} \leftarrow \text{top}K(\bigcup_{k=1}^K \text{succ}(\hat{y}_{1:t}^{(k)}))$ 
15:   /*BACKWARD*/
16:    $\text{grad}_{h_T} \leftarrow 0$ ;  $\text{grad}_{h_T} \leftarrow 0$ 
17:   for  $t = T - 1, \dots, 1$  do
18:      $\text{grad}_{h_t} \leftarrow \text{BRNN}(\nabla_{h_t} \mathcal{L}_{t+1}, \text{grad}_{h_{t+1}})$ 
19:      $\text{grad}_{\hat{h}_t} \leftarrow \text{BRNN}(\nabla_{\hat{h}_t} \mathcal{L}_{t+1}, \text{grad}_{\hat{h}_{t+1}})$ 
20:     if  $t - 1 \in \text{violations}$  then
21:        $\text{grad}_{h_t} \leftarrow \text{grad}_{h_t} + \text{grad}_{\hat{h}_t}$ 
22:        $\text{grad}_{\hat{h}_t} \leftarrow 0$ 

```

Hình 5: Sequence-to-Sequence as

Beam Search Optimization [8]

## 3 KẾT QUẢ THỰC NGHIỆM

### 3.1 Xây dựng tập dữ liệu

Để xây dựng thành công mô hình trên tập dữ liệu tiếng Việt, trong bài báo này, chúng tôi thu thập các đoạn văn bản lấy từ các trang báo trực tuyến ở Việt Nam (Hình 6). Với đoạn văn bản cần tóm tắt là cả