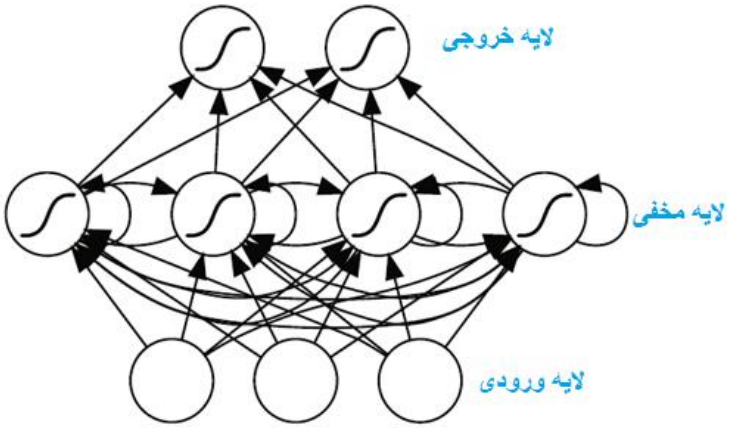
فصل سوم

شبکه‌های عصبی بازگشتی
(Recurrent Neural Networks)

۳- شبکه‌های عصبی بازگشتی(RNN)

در بخش‌های قبلی شبکه‌های عصبی را مورد بررسی قرار دادیم که در اتصالات آن‌ها حلقه[۴۱] وجود نداشت. اگر از این حلقه‌ها در اتصال‌های شبکه استفاده کنیم شبکه‌های عصبی بازگشتی را تشکیل می‌دهیم. مدل‌های مختلفی از شبکه‌های بازگشتی ارائه شده است؛ اما در اینجا ساده‌ترین مدل را مورد بررسی قرار می‌دهیم، مدلی که طبق شکل۳- ۱ تنها یک لایه مخفی دارد که آن لایه به خودش حلقه دارد. همانطور که در شکل مشخص است لایه ورودی شامل سه نورون، لایه مخفی شامل ۴ نورون و لایه خروجی شامل ۲ نورون می‌باشد که در لایه مخفی آن، هر نورون به خودش یک حلقه دارد و به تمام نورون‌های دیگر در لایه خودش هم یک اتصال دارد.

مقدار تاثیر ورودی شبکه در نورون‌های لایه مخفی، با فرض اینکه تعداد نورون‌های لایه مخفی H باشد به صورت رابطه ۱- ۳ به دست می‌آید. بقیه‌ی قسمت‌های الگوریتم‌های پیش‌رو مانند تابع فعالساز و لایه‌ی خروجی مشابه گذشته می‌باشد. اما قسمت آموزش مربوط به قاعده زنجیره‌ای در رابطه ۱- ۳ به شکل رابطه ۲- ۳ به دست می‌آید.



شکل۳- ۱- یک شبکه‌ی RNN [۵]

$$a_h^t=\sum_{i=1}^{I} w_{ih}x_i^t + \sum_{h'=1}^{H} w_{h'h}b_{h'}^{t-1} \quad ۱-۳$$

$$\frac{\partial L}{\partial a_j^t}=\theta'(a_h^t)(\sum_{k=1}^{K} \delta_k^t w_{hk}+\sum_{h'=1}^{H} \delta_{h'}^{t+1} w_{hh'}) \quad ۲-۳$$

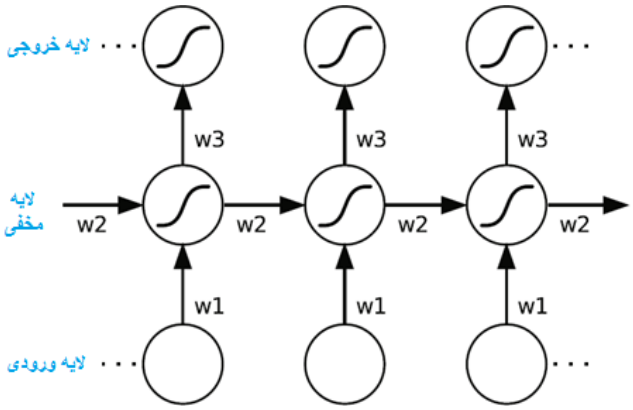
۴۱ Cycle

با توجه به اینکه متغیر a_k ، اولین متغیر بعد از وزن‌های لایه مخفی می‌باشد آموزش وزن‌ها به صورت رابطه ۳- ۳ به دست می‌آید، که در آن W_{ij} نشان دهنده وزن بین نورون i ام به نورون j ام می‌باشد. [۶,۵]

$$\frac{\partial L}{\partial w_{i,j}}=\sum_{t=1}^{T}\frac{\partial L}{\partial a_j^t}\frac{\partial a_j^t}{\partial w_{i,j}}=\sum_{t=1}^{T}\delta_j^t b_i^t \qquad ۳-۳$$

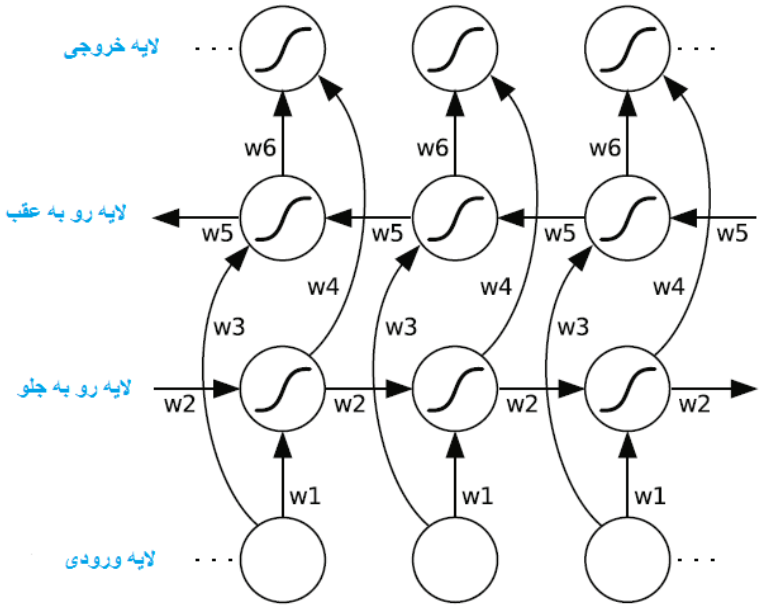
۳-۱-شبکه‌های دو طرفه[۴۲]

در بسیاری از کارهای دنباله برچسب‌گذاری دستیابی به اطلاعات آینده همانند اطلاعات گذشته مورد اهمیت قرار دارد و می‌تواند مفید باشد. در بازشناسی گفتار پیوسته هم اینگونه است. ایده‌ی کلی آن به این صورت است که در لایه مخفی دو دسته نورون وجود دارد: یکی رو به جلو و یکی رو به عقب، که هر دوی آن‌ها به یک نورون خروجی متصل هستند. شکل۳- ۳ روند این شبکه‌ی دو طرفه را نشان می‌دهد، همانطور که در شکل نمایان است یک لایه مخفی برای رو به جلو و یک لایه مخفی برای رو به عقب تعبیه شده است و هر نورون خروجی از یک لایه مخفی رو به جلو و از یک لایه مخفی رو به عقب اتصال می‌گیرد-برای مقایسه می‌توانید آن را با شکل۳-۲***شکل۳- ۱*** که روند یک شبکه‌ی یک طرفه را نشان می‌دهد مقایسه نمایید-. روند کار به این صورت است که بعد از اینکه تمام دنباله داده‌ها –T داده زمانی- به شبکه داده شدند یک لایه مخفی از زمان ۱ تا T داده را رو به جلو محاسبه می‌کند و همینطور یک لایه مخفی دیگر از زمان T تا زمان ۱ داده‌ها را رو به عقب محاسبه می‌کند و در پایان یک بار با حرکت پیش‌رو و استفاده از هر دو داده‌های به دست آمده در سمت جلو و سمت عقب، مقادیر خروجی را محاسبه می‌کنیم. الگوریتم این روش در شکل۳- ۴ نشان داده شده است.



شکل۳- ۲- نمونه‌ای از شبکه‌ی بازگشتی یک طرفه[۵]

۴۲ Bidirectional



شکل۳- ۳ – نمونه‌ای از شبکه‌ی بازگشتی دو طرفه[۵]

for $t = 1$ to T **do**
Forward pass for the forward hidden layer, storing activations at each timestep
for $t = T$ to 1 **do**
Forward pass for the backward hidden layer, storing activations at each timestep
for all t, in any order **do**
Forward pass for the output layer, using the stored activations from both hidden layers

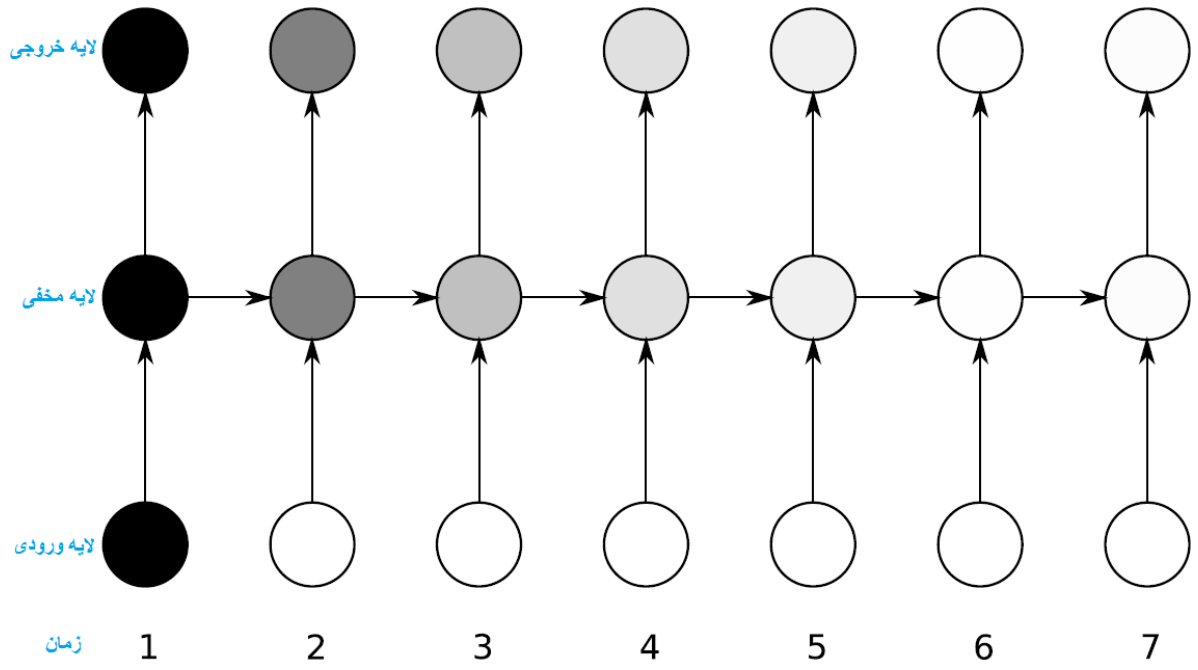
شکل۳- ۴- مرحله پیش‌رو در BRNN [۵]

برای آموزش شبکه، فرمول‌ها و روش آموزش مشابه گذشته است با این تفاوت که ابتدا خطای محلی تمام داده‌ها را محاسبه می‌کنیم سپس عکس مرحله‌ی پیش‌رو، برای بخش رو به جلو از زمان T تا زمان ۱ الگوریتم پس انتشار خطا با استفاده از خطای محلی به دست آمده در مرحله‌ی قبل را به کار می‌بریم و وزن‌های مربوط به لایه مخفی آن را آموزش می‌دهیم، همچنین برای بخش رو به عقب از زمان ۱ تا زمان T الگوریتم پس انتشار

خطا با استفاده از خطای محلی به دست آمده در دو مرحله‌ی قبل را به کار می‌بریم و وزن‌های مربوط به لایه مخفی آن را آموزش می‌دهیم. الگوریتم این روش در شکل۳- ۵ نشان داده شده است.[۶,۵]

for all t, in any order **do**
  Backward pass for the output layer, storing δ terms at each timestep
for $t = T$ to 1 **do**
  BPTT backward pass for the forward hidden layer, using the stored δ terms from the output layer
for $t = 1$ to T **do**
  BPTT backward pass for the backward hidden layer, using the stored δ terms from the output layer

شکل۳- ۵- آموزش **BRNN** [۵]



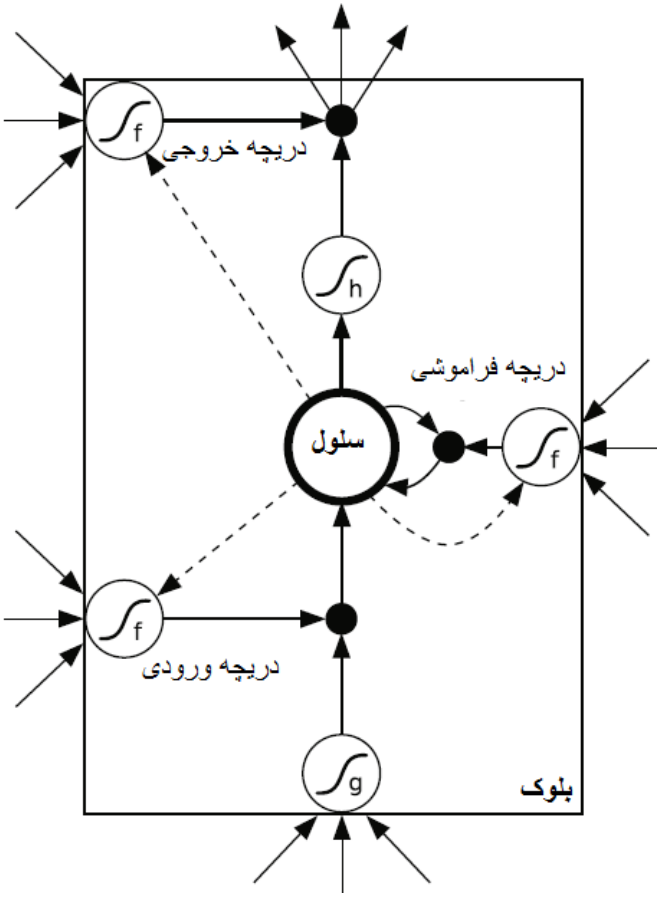
شکل۳- ۶- مشکل محو شدن گرادیان در **RNN** [۵]

۳-۲- Long Short-Term Memory (LSTM)

در ساختار RNN استاندارد، محدوده‌ی محتوایی قابل دسترس در عمل بسیار محدود می‌باشد. مشکل این است که تاثیر یک ورودی داده شده بر لایه مخفی و در نتیجه بر خروجی شبکه به صورت نمایی پایین می‌آید و

از بین می‌رود. این مشکل به نام مشکل محو شدن گرادیان[۴۳] شناخته می‌شود. این مشکل به صورت شماتیک در شکل۳- ۶ نشان داده شده است.

ساختار LSTM (معرفی شده در سال ۱۹۹۷) شامل یک مجموعه‌ی زیر شبکه‌ی متصل بازگشتی، به نام بلوک‌های حافظه می‌باشد. هر بلوک شامل یک یا تعداد بیشتری سلول حافظه‌ی خود بازگشتی و سه واحد ضرب- دریچه ورودی، خروجی و فراموشی- که آنالوگ‌های مداوم نوشتن، خواندن و تنظیم مجدد عملکردهای سلول‌ها را ارائه می‌دهند.[۶,۵,۱۰]



شکل۳- ۷- بلوک LSTM با یک سلول[۵]

۴۳ Vanishing gradient problem

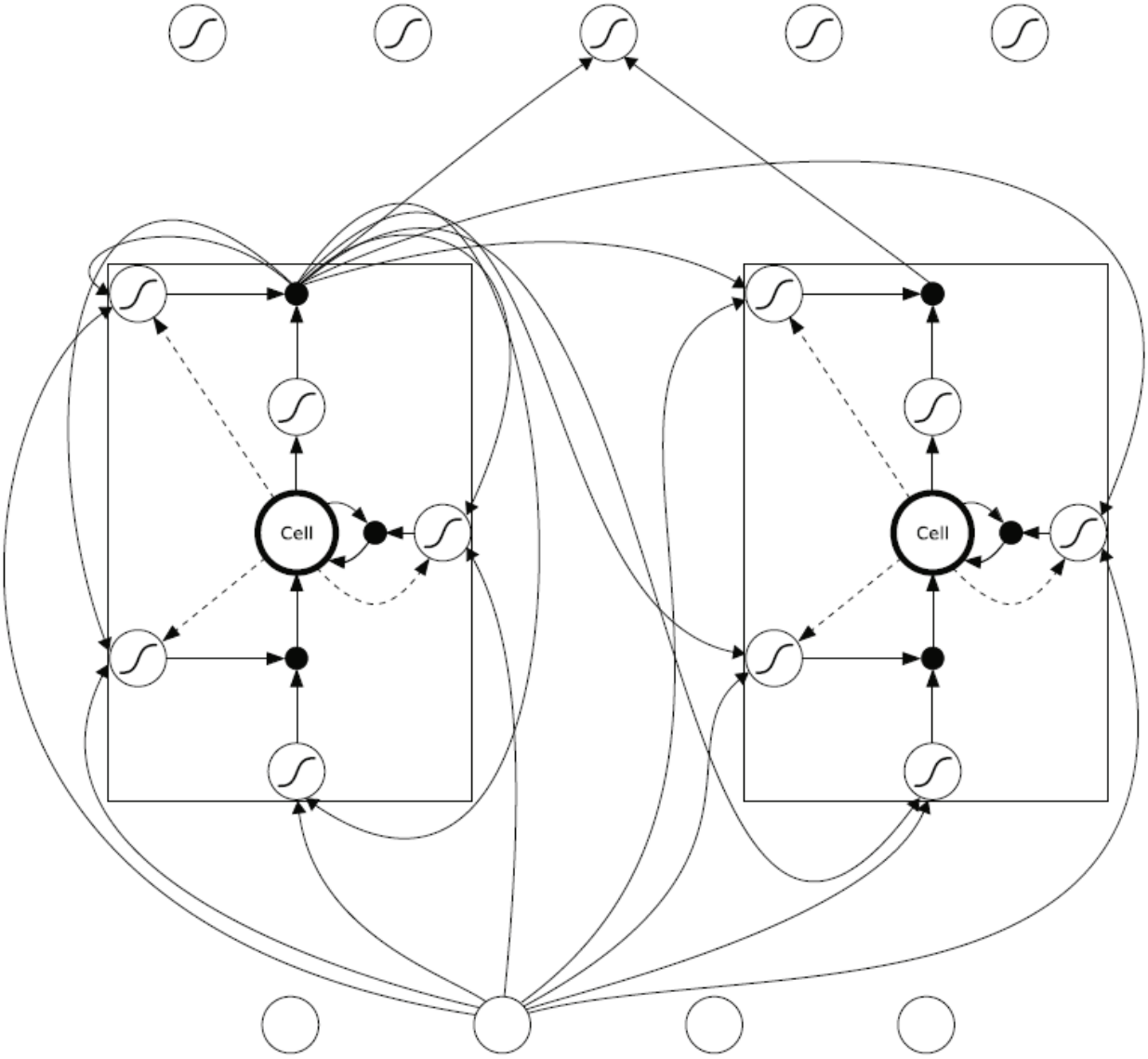
شکل۳- ۷ یک نمونه از بلوک LSTM با یک سلول را نشان می‌دهد. یک شبکه‌ی LSTM مشابه یک RNN استاندارد است، با این تفاوت که واحدهای جمع کننده(مقدار درونی) نورون‌ها در لایه مخفی توسط بلوک‌های حافظه جایگزین می‌شوند.

در شکل۳- ۸ یک نمونه از این شبکه نشان داده شده است. دریچه‌های ضربی به سلول‌های حافظه‌ی LSTM امکان نگهداری و دسترسی به اطلاعات در دوره‌های زمانی دراز مدت را می‌دهد. بنابراین مشکل محو شدن گرادیان کاهش می یابد. به عنوان مثال تا زمانی که دریچه ورودی بسته باقی بماند(یعنی تابع فعالیت نزدیک ۰ داشته باشد)، تابع فعالیت سلول توسط ورودی‌های جدید که به شبکه می‌رسند باز نوشته نخواهد شد، در نتیجه با باز کردن دریچه خروجی، در لحظات طولانی مدت آینده‌ی دنباله، برای شبکه قابل دسترس است.

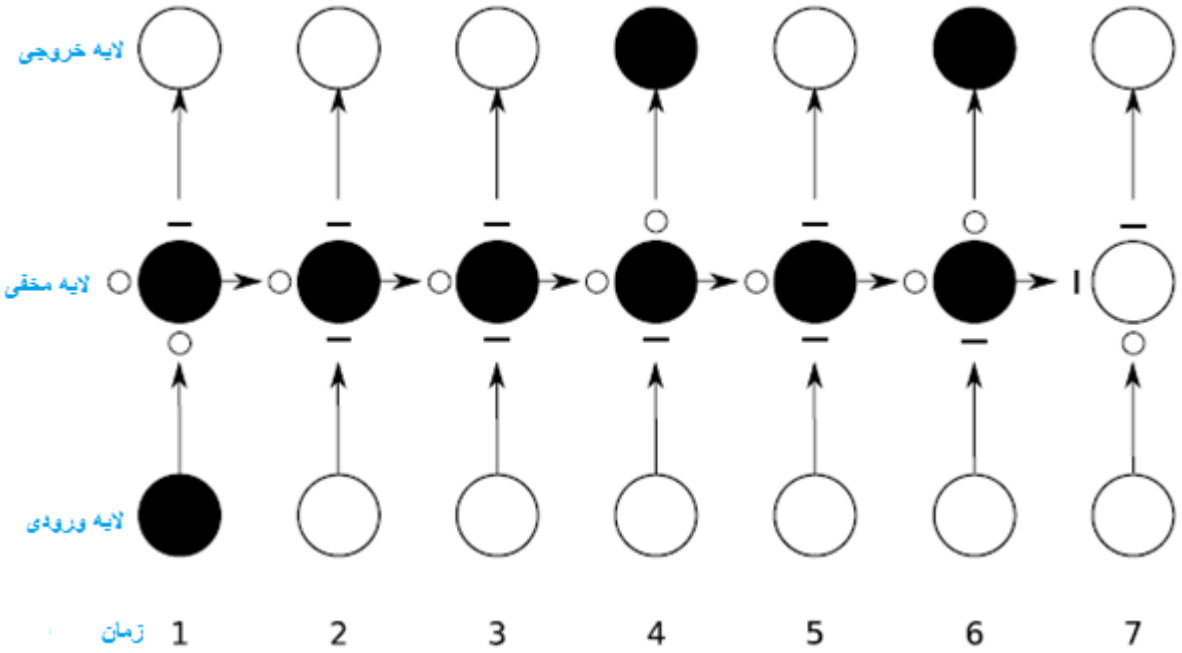
حفظ اطلاعات گرادیان در طول زمان توسط LSTM در شکل۳- ۹ نشان داده شده است.

در فرم اصلی، LSTM تنها شامل دریچه‌های ورودی و خروجی می‌باشد. دریچه‌ی فراموشی به همراه ضرایب روزنه[۴۴] که دریچه‌ها را به سلول حافظه متصل می‌کند بعدها برای گسترش LSTM اضافه شده است. هدف دریچه‌های فراموشی فراهم آوردن امکان بازنشانی سلول‌های حافظه می‌باشد که زمانی که نیاز به فراموشی ورودی‌های گذشته باشد بسیار مفید می‌باشد. [۶،۵]

۴۴ peephole



شکل۳- ۸ - یک شبکه‌ی LSTM [۵]



شکل۳- ۹ حفظ اطلاعات گرادیان توسط LSTM : خط‌های تیره نشان‌دهنده‌ی تابع فعالیت یعنی بسته بودن دریچه می‌باشد و دایره تو خالی نشان‌دهنده باز بودن دریچه می‌باشد. منظور از نورون‌های تو خالی عدم تاثیر ورودی مورد نظر بر روی آن‌ها می‌باشد. [۵]

۳-۳-LSTM دو طرفه (BLSTM)

استفاده از ساختار LSTM در شبکه‌ی RNN دو طرفه موجب تولید شبکه‌ی دو طرفه‌ی LSTM (به اختصار BLSTM) می‌شود.[۶]

۳-۳-۱- محاسبات در شبکه

همانند گذشته، w_{ij} وزن اتصالات از نورون i به نورون j ام ، a^t_j ورودی شبکه به نورون j ام در لحظه t و b^t_j خروجی تابع فعالیت نورون j در لحظه t می‌باشد. فرمول‌های LSTM برای یک بلوک حافظه نوشته شده‌اند. برای چند بلوکی محاسبات به سادگی برای هر بلوک تکرار می‌شوند. عبارات در فرمول‌ها به صورت : l دریچه‌ی ورودی، ϕ دریچه‌ی فراموشی ، w دریچه‌ی خروجی و C سلول‌های حافظه می‌باشند. وزنی که سلول را به دریچه‌های ورودی، فراموشی و خروجی خودش متصل می‌کند به ترتیب به صورت W_{CQ}, W_{CL} و W_{CW} نشان داده شده‌اند. S^t_c نشان دهنده‌ی حالت سلول C در لحظه t می‌باشد. f تابع فعالیت دریچه‌هاست ،g و h نشان دهنده‌ی توابع فعالیت ورودی و خروجی سلول‌ها می‌باشند.

I تعداد ورودی‌ها، K تعداد خروجی‌ها و H تعداد سلول‌های لایه مخفی می‌باشد. توجه شود که تنها خروجی‌های سلول b^t_c به دیگر بلوک‌های آن لایه متصل است. دیگر مقادیر مانند ورودی سلول، حالت سلول و ... درون خود سلول قابل دسترس است. از عبارت h برای اندیس‌گذاری خروجی‌های سلول از دیگر بلوک‌های لایه مخفی استفاده می‌شود. همانند RNN استاندارد الگوریتم پیش‌رو برای یک دنباله ورودی x از لحظه ۱ تا لحظه T ادامه می‌یابد و برعکس آن برای آموزش از لحظه T شروع و تا لحظه ۱ باز خواهیم گشت و آن را آموزش خواهیم داد. مشتقات وزنی نهایی با جمع زدن بر روی مشتق‌ها در هر لحظه‌ی زمانی به دست می‌آید. یادآوری می‌شود که خطای محلی هم به صورت رابطه ۳- ۴ به دست می‌آید که در آن L نشان‌دهنده تابع خطا می‌باشد.

$$\delta_j^t \stackrel{\text{def}}{=} \frac{\partial L}{\partial a_j^t} \qquad ۴-۳$$

نکته‌ی قابل توجه اینست که همانند RNN استاندارد، تمام حالات و توابع فعالیت در زمان شروع t=0 - با 0 مقدار دهی اولیه می‌شوند و تمام عبارات خطای محلی در زمان t=T+1 نیز 0 می‌باشند.

۳-۳-۱-۱- مرحله پیش‌رو

دریچه‌ی ورودی

$$a_l^t=\sum_{i=1}^{I} w_{il}x_i^t + \sum_{h=1}^{H} w_{hl}b_h^{t-1} + \sum_{c=1}^{C} w_{cl}s_c^{t-1} \qquad ۵-۳$$

$$b_l^t=\mathrm{f}(a_l^t) \qquad ۶-۳$$

دریچه‌ی فراموشی

$$a_{\phi}^t=\sum_{i=1}^{I} w_{i\phi}x_i^t+\sum_{h=1}^{H} w_{h\phi}b_h^{t-1}+\sum_{c=1}^{C} w_{c\phi}s_c^{t-1} \qquad ۷-۳$$

$$b_{\phi}^{t}=\mathrm{f}(a_{\phi}^{t}) \quad ۸-۳$$

سلول‌ها

$$a_{c}^{t}=\sum_{i=1}^{I} w_{ic}x_{i}^{t}+\sum_{h=1}^{H} w_{hc}b_{h}^{t-1} \quad ۹-۳$$

$$s_{c}^{t}=b_{\phi}^{t}s_{c}^{t-1}+b_{l}^{t}g(a_{c}^{t}) \quad ۱۰-۳$$

$$a_{w}^{t}=\sum_{i=1}^{I} w_{iw}x_{i}^{t}+\sum_{h=1}^{H} w_{hw}b_{h}^{t-1}+\sum_{c=1}^{C} w_{cw}s_{c}^{t} \quad ۱۱-۳$$

$$b_{w}^{t}=\mathrm{f}(a_{w}^{t}) \quad ۱۲-۳$$

خروجی سلول

$$b_{c}^{t}=b_{w}^{t}\mathrm{h}(s_{c}^{t}) \quad ۱۳-۳$$

۳-۳-۱-۲- مرحله پس‌رو (آموزش)

با فرض $\epsilon_{s}^{t} \stackrel{\text{def}}{=} \frac{\partial L}{\partial s_{c}^{t}}$, $\epsilon_{c}^{t} \stackrel{\text{def}}{=} \frac{\partial L}{\partial b_{c}^{t}}$ خواهیم داشت:

خروجی سلول

$$\epsilon_{c}^{t}=\sum_{k=1}^{K} w_{ck}\delta_{k}^{t}+\sum_{h=1}^{H} w_{ch}\delta_{h}^{t+1} \quad ۱۴-۳$$

دریچه ی خروجی

$$\delta_w^t=\mathrm{f}'(a_w^t)\sum_{c=1}^{C} h(s_c^t)\epsilon_c^t \quad ۱۵ -۳$$

حالت

$$\epsilon_s^t=b_w^t h'(s_c^t)\epsilon_c^t+b_{\emptyset}^{t+1}\epsilon_s^{t+1}+w_{cl}\delta_l^{t+1}+w_{c\emptyset}\delta_{\emptyset}^{t+1}+w_{cw}\delta_w^t \quad ۱۶ -۳$$

سلول‌ها

$$\delta_c^t=b_l^t g'(a_c^t)\epsilon_s^t \quad ۱۷ -۳$$

$$\delta_{\emptyset}^t=f'(a_{\emptyset}^t)\sum_{c=1}^{C} s_c^{t-1}\epsilon_s^t \quad ۱۸ -۳$$

$$\delta_l^t = f'(a_l^t)\sum_{c=1}^{C} g(a_c^t)\epsilon_s^t \quad ۱۹ -۳$$

۳-۴-مقایسه‌ی ساختارهای شبکه

در این بخش یک مقایسه‌ی عملی بین ساختارهای مختلف شبکه عصبی را در عمل کلاس‌بندی واج به صورت قاب‌بندی شده ارائه می‌دهیم. کلاس‌بندی واج‌ها در سطح قاب یک نمونه از عمل کلاس‌بندی segment است. در ادامه توانایی یک الگوریتم را در بخش‌بندی و تشخیص اجزای تشکیل دهنده یک سیگنال گفتاری، که به اطلاعات محتوایی نیاز دارد، و می‌تواند به عنوان اولین مرحله در بازشناسی گفتار پیوسته شناخته شود مورد آزمایش قرار می‌دهیم.

محتوا در بازشناسی گفتار اهمیت فراوانی دارد، در بسیاری موارد تشخیص یک بخش از واج بدون دانستن واج‌هایی که قبل یا بعد آن می‌آیند بسیار سخت است. دلیل آن ماهیت گفتار می‌باشد که برخی واج‌ها مشابه هم می‌باشند و برخی هرگز به دنبال دیگری نمی‌آیند و برخی هم با احتمال بیشتری به دنبال هم می‌آیند. بنابراین در بازشناسی گفتار پیوسته، واج‌های قبل و بعد از قاب مورد نظر اهمیت فراوانی دارند. در نتیجه ساختارهای شبکه‌ای