

① CNM ها برای این آموزش داده می شود که بهترین ویژگی ها را از تصویر تشخیص دهند و باعث آن ها با فرض

locality ویژگی ها می تواند شبکه های خلی پستی را آموزش دهند. (spatial coherence)

② ~~همچنین~~ scale می تواند در حالی که یک شبکه FC به علت تعداد پارامترها املا در یک سری مدل قابل استفاده نیست

③ Parameter sharing را دارند که باعث می شود تعداد پارامترها خلی خلی کمتر شود

۱.
۳

در واقع فرض می شود یک شبکه داریم و آنرا در این حالت می بینیم
وزن استفاده می شود چون حتماً از یک حافظه برای ذخیره سازی ~~حافظه~~ و آپدیت با استرهای اشتراکی
استفاده می شود تقسیمی می شود که ~~حافظه~~ است. در واقع برای یک ~~حافظه~~ 2×2 برای سلا و درص
 10×10 حتماً از ~~حافظه~~ 4 وزن یک استفاده می کنیم و آن را ~~حافظه~~ $input$ می نامند و نظریه دیگر
پس $back$ نیز یک وزن را آپدیت می کند $prop$

۳ الف

در واقع استقلال از خود و درسی خردی (Reluz) به ندرت مساعی صرف است

عدد پیرامتری منفرجه α و $\beta \geq 0.5$ در هر دو معنی معیار است.

این سبک املا قابل یادگیری نیست، صبحگاه ~~خطی~~ است به از 0.5 مینی توان کرد

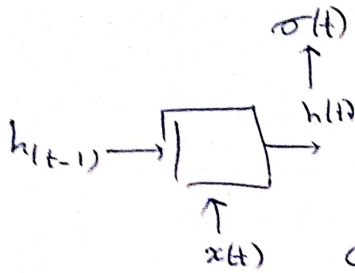
مکانات $\text{vanishing gradient}$ نیزه عقب‌نشینی خواص کم‌کم ضعیف می‌شود نسبت به اصل شبکه عصبی است

~~improved~~ RNN (بهبود یافته) RNN

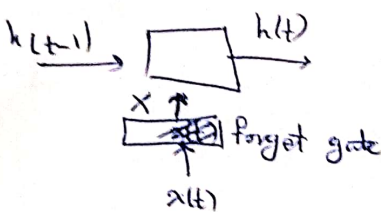
Gated Recurrent Unit

GRU. (ع) (الف)

~~کتابخانه~~ reset gate
update gate

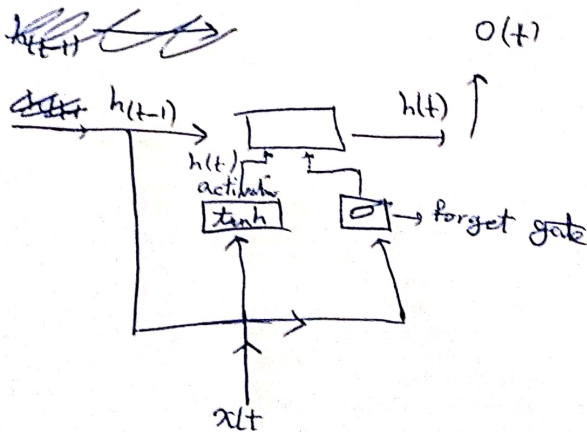


دلیل تفاوت اصلی GRU، RNN است زیرا که RNN کنترلی روی مقادیر نداشته و فقط از طریق gate برای reset کردن یا forget کردن دارد.



دقیق Forget gate فعال است یا هیچ اطلاعاتی رد و بدل نمی شود

~~کتابخانه~~



④ ~~یک روش دیگر~~ یک راه حل LSTM یا GRU است زیرا هر کدام forget gate دارند در دو نوع اطلاعات اضافی را حذف می کنند که باعث می شود مشکل طولانی بودن ورودی و وابستگی های بلند مدت حل شود

یک روش دیگر adaptive computation time است زیرا این روش ~~یک روش دیگر~~ RNN های با طول مختلف در نظری می گیرد مشکل یکسان حل می شود

یک روش دیگر attention است که در واقع به سمت های خاصی توجه می کند و آن ها را در نظر می گیرد و سمت های خاصی را در نظر می گیرد

روش های $Eco state$ و $Liquid state$ نیز اینها را بهبود دارد.