

1- در این معماری encoder معنی جمله را به یک hidden state با طول ثابت تبدیل میکند که انتظار می‌رود این بازنمایی یک خلاصه خوب از تمام جمله ورودی باشد و سپس این بردار که آخرین state از encoder میباشد را به به decoder میدهد تا ترجمه متناظر را تولید کند. اشکال واضح این معماری همین hidden state با طول ثابت است که قابلیت به خاطر سپردن جملات بلند را ندارد و قسمت های اولیه جمله را فراموش میکند درحالیکه ما برای ترجمه یک کلمه نیاز داریم به کلمات مرتبط در جمله ورودی توجه کنیم. (که ممکن است این کلمه موثر در ترجمه، اوایل جمله ورودی باشد).

2- Context vector ای که استفاده میکند اطلاعات ورودی را طوری خلاصه میکند که قسمت های اولیه یا آخر جمله ورودی فراموش نشوند. این context vector شامل تمامی hidden state میباشد که میتوانند وزن مساوی یا متفاوت داشته باشند. اگر وزن مساوی داشته باشند تمامی کلمات جمله ورودی سهم مساوی در ترجمه خروجی دارند و اگر وزن داشته باشند کلمات مرتبط سهم بیشتری خواهند داشت. در هر صورت دیگر دچار gradient vanishing نمیشویم.

3- Teacher forcing یک روش سریع و کارآمد برای آموزش شبکه های عصبی بازگشتی میباشد که در decoder از جواب صحیح (ground truth) تایپ استپ قبلی برای ورودی تایم استپ فعلی استفاده میکند و به همین دلیل است که آموزش مدل سریعتر و کارآمدتر میشود.