



**دانشکده مهندسی کامپیوتر**

# **یادگیری توالی به توالی با شبکه های عصبی**

Sequence to Sequence Learning with Neural Networks

**پروژه درس پردازش زبان های طبیعی**

**دانشجو:**

**مرتضی ذاکری**

**استاد درس:**

**دکتر بهروز مینایی**

**دی ماه ۱۳۹۶**

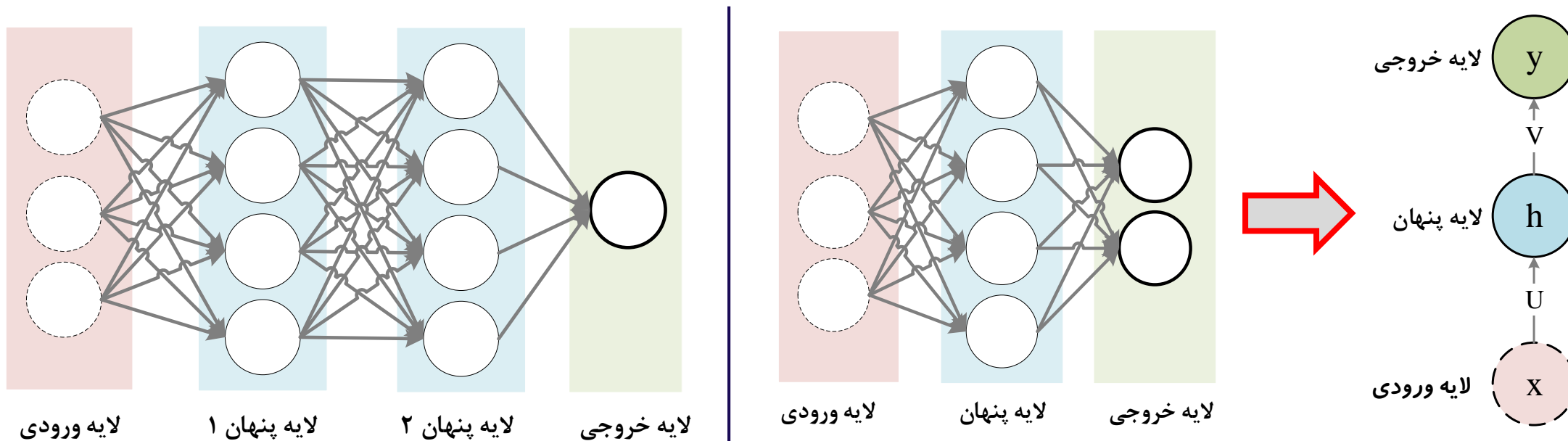
# رئوس مطالب

- مقدمه
- یادگیری ژرف
- شرح مسئله و اهمیت موضوع
- مفاهیم اولیه
- مدل زبانی
- شبکه‌های عصبی مکرر
- کارهای مرتبط
- مدل کدگذار - کدگشا
- مدل توالی به توالی به توالی
- آموزش و جزئیات مدل
- آزمایش‌ها
- مجموعه داده‌ها
- وارون‌سازی جمله‌های مبدأ
- نتایج
- نتیجه‌گیری
- مسائل باز و کارهای آتی



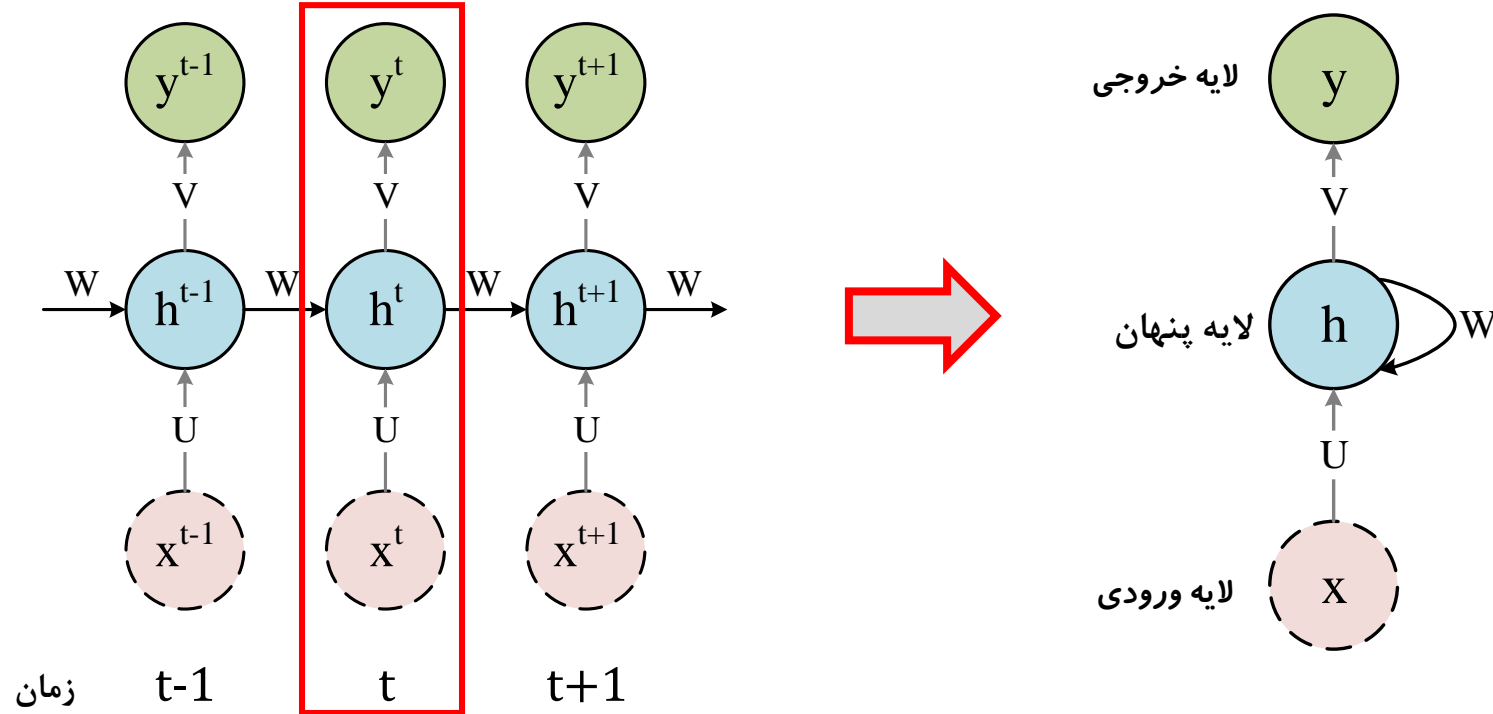
# یادگیری ژرف

- استفاده از گراف‌های محاسباتی با عمق بیشتر از یک لایه در شبکه‌های عصبی



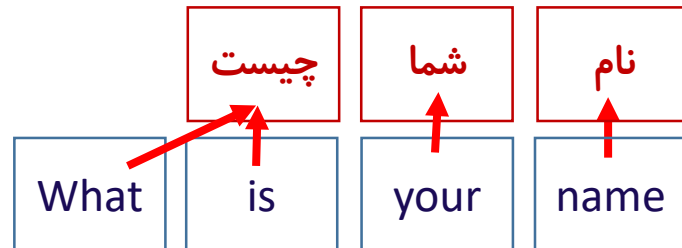
# یادگیری ژرف

- شبکه‌های عصبی مکرر: یادگیری وظایف مبتنی بر توالی (Goodfellow et al. 2016).
- یادگیری وظایفی که ترتیب ورودی در آنها مهم است: **اکثر وظایف حوزه NLP**.



## شرح مسئله

- در شبکه عصبی مکرر، طول توالی ورودی و خروجی ثابت و برابر است.
- این مدل برای توالی‌های ورودی و خروجی با طول متفاوت و متغیر مناسب نیست.
- نمونه‌هایی از وظایف نگاشت توالی‌به‌توالی:



- ترجمه ماشینی،
- تبدیل گفتار به نوشتار،
- سیستم پرسش و پاسخ،
- و غیره.

- نیاز به یک مدل برای یادگیری نگاشت توالی‌به‌توالی، مستقل از دامنه مسئله (وظیفه).

(Ilya Sutskever, Oriol Vinyals et al. 2014)



# مدل زبانی

• یک مفهوم پایه در NLP (Jurafsky & Martin 2009)

• امکان پیش‌بینی نشانه بعدی در یک توالی را فراهم می‌کند.

$$x = \langle x^{(1)}, x^{(2)}, \dots, x^{(m)} \rangle \Rightarrow p(x) = \prod_{t=1}^m p(x^{(t)} | x^{(<t)})$$

• مدل‌های سنتی n-gram (Luong 2016):

• محدودسازی رابطه‌ی بالا به تنها n-1 نشانه‌ی قبل، با استناد به فرض مارکوف.

• نارسایی مدل در توالی‌های طولانی بر اثر فرض ساده‌کننده.

• مدل‌های زبانی عصبی (Luong 2016):

• استفاده از شبکه‌های عصبی ژرف برای استنباط احتمال بالا.



# شبکه عصبی مکرر

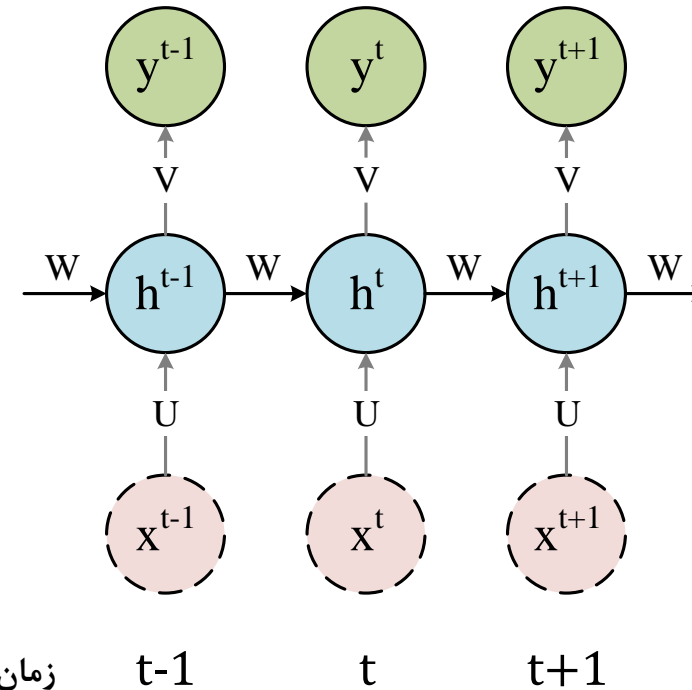
• یک معماری جدید برای مدل زبانی

$$1) a^{(t)} = Ux^{(t)} + Wh^{(t-1)} + b,$$

$$2) h^{(t)} = \Phi(a^{(t)}),$$

$$3) y^{(t)} = Vh^{(t)} + c,$$

$$4) \hat{y}^{(t)} = \text{softmax}(y^{(t)})$$



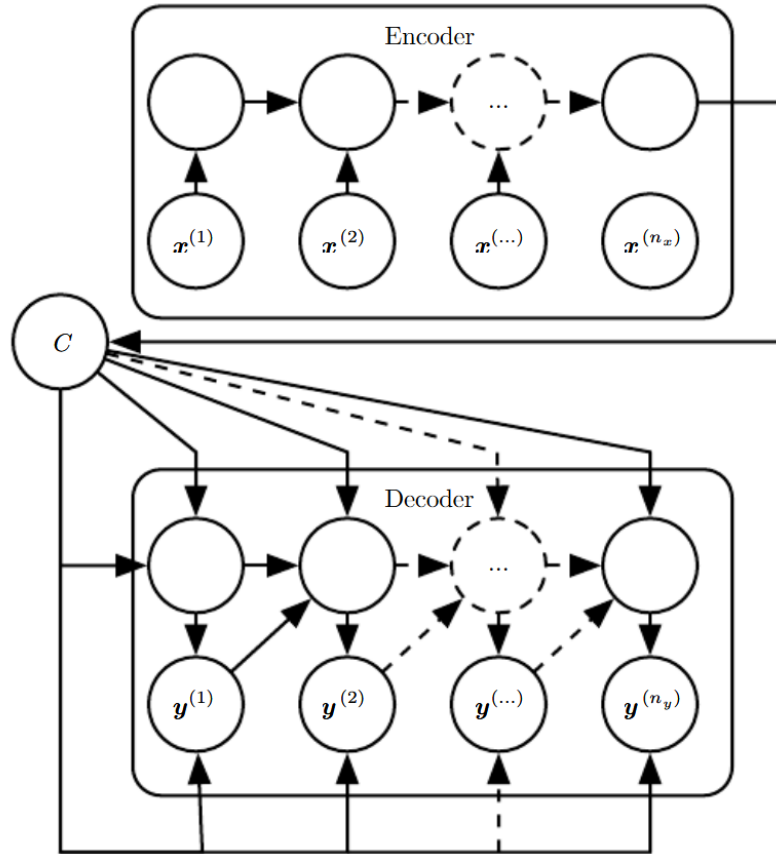
# کارهای مرتبط

- ایده اصلی استفاده شده در کارهای قبلی:
- فشردن اطلاعات **توالی مبدأ (ورودی)** در یک بردار با بعد ثابت،
- استخراج توالی **مقصد (خروجی)** از روی بردار حاصل شده در مرحله قبل.
- Kalchbrenner و Blunsom (Kalchbrenner & Blunsom 2013):
- استفاده از شبکه عصبی پیچشی برای کدگذاری،
- **عدم حفظ ترتیب نشانه‌ها (واژه‌ها) در توالی.**
- مدل کدگذار - کدگشا (Cho et al. 2014):
- استفاده از شبکه عصبی مکرر با سلول حافظه غیر LSTM،
- **مشکل در یادگیری توالی‌های طولانی.**

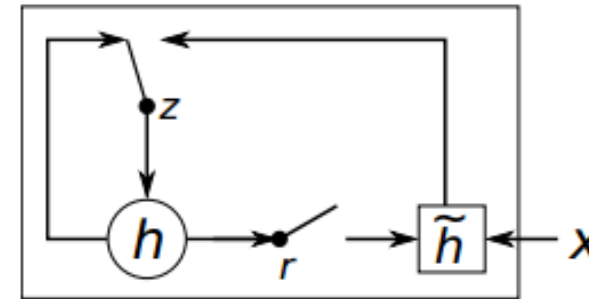




# مدل کدگذار - کدگشا



مدل کدگذار - کدگشا (Cho et al. 2014).



هسته سلول حافظه استفاده شده در لایه پنهان  
مدل کدگذار - کدگشا (Cho et al. 2014).



# مدل توالی به توالی

• هدف: تخمین احتمال شرطی

$$p(\langle y^{(1)}, y^{(2)}, \dots, y^{(m)} \rangle | \langle x^{(1)}, x^{(2)}, \dots, x^{(n)} \rangle)$$

• با استفاده از رابطه

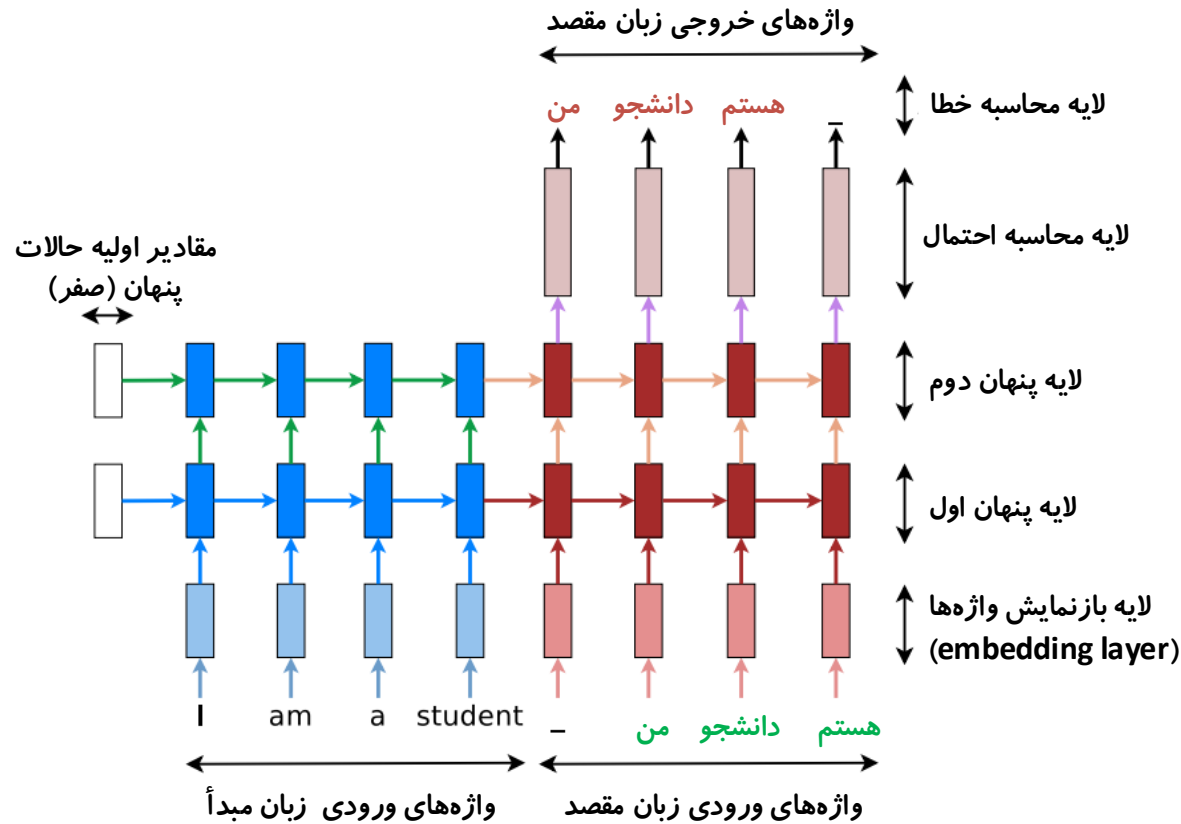
$$p(\langle y^{(1)}, \dots, y^{(m)} \rangle | \langle x^{(1)}, \dots, x^{(n)} \rangle) = \prod_{t=1}^m p(y^{(t)} | v, y^{(1)}, \dots, y^{(t-1)})$$

• **وظیفه انتخاب شده: ترجمه ماشینی عصبی (ترجمه انگلیسی به فرانسوی)**

(Ilya Sutskever, Oriol Vinyals et al. 2014)



# آموزش و آزمون مدل



ترجمه ماشینی عصبی با شبکه عصبی مکرر ژرف (Luong 2016)



# جزئیات مدل

- حافظه کوتاه‌مدت بلند (LSTM) با ژرفای ۴ لایه و ۱۰۰۰ سلول در هر لایه،
- مقداردهی اولیه تصادفی پارامترها با توزیع یکنواخت،
- یادگیری به روش **کاهش گرادیان تصادفی (SGD)** با نرخ یادگیری **۰.۷**،
- پیاده‌سازی اولیه در C++ (Ilya Sutskever, Oriol Vinyals et al. 2014)،
- پیاده‌سازی‌های مشابه در tensorflow (Luong et al. 2017) و keras (Keras-Team 2017).

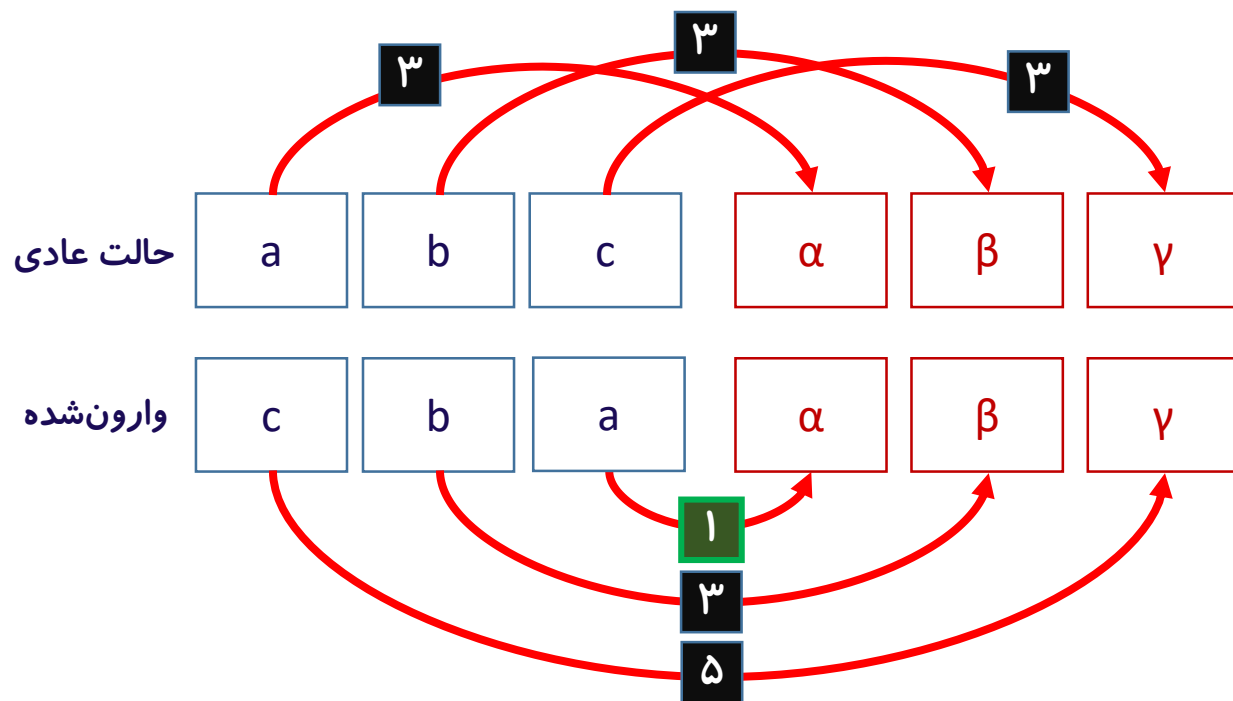


# مجموعه داده

- مجموعه داده ترجمه انگلیسی به فرانسوی WMT'14 (Pecina 2014):
  - زیرمجموعه ۱۲ میلیون جمله‌ای،
  - ۳۴۸ میلیون واژه فرانسوی و ۳۴۰ میلیون واژه انگلیسی.
- مجموعه آموزشی کوچک تر (Kelly 2017)
  - برای آموزش در سطح کاراکتر.
  - شامل مجموعه داده نمونه ترجمه انگلیسی به فارسی.



# وارون سازی جمله های مبدأ



- در این روش جمله مبدأ وارون شده و سپس جمله مقصد به آن الصاق می شود.

- کاهش تأخیر زمانی کمینه

- کاهش مقدار سرگشتگی مدل از ۵.۸ به ۴.۷

- افزایش امتیاز BLEU کسب شده توسط مدل از ۲۵.۹ به ۳۰.۶

- عدم تغییر میانگین تأخیر زمانی

- اثر این پدیده بر زبان های از راست به چپ؟؟؟



# امتیازهای BLEU حاصل شده

امتیاز BLEU (ntst14)	روش
۲۸.۴۵	Bahdanau و همکاران (Bahdanau et al. 2014)
۲۶.۱۷	یک LSTM روبه جلو، اندازه پرتوی ۱۲
۳۰.۵۹	یک LSTM با ورودی و ارون، اندازه پرتوی ۱۲
۳۳.۰۰	پنج LSTM با ورودی و ارون، اندازه پرتوی ۱
۳۳.۲۷	دو LSTM با ورودی و ارون، اندازه پرتوی ۱۲
۳۴.۵۰	پنج LSTM با ورودی و ارون، اندازه پرتوی ۲۱
۳۴.۸۱	پنج LSTM با ورودی و ارون، اندازه پرتوی ۱۲



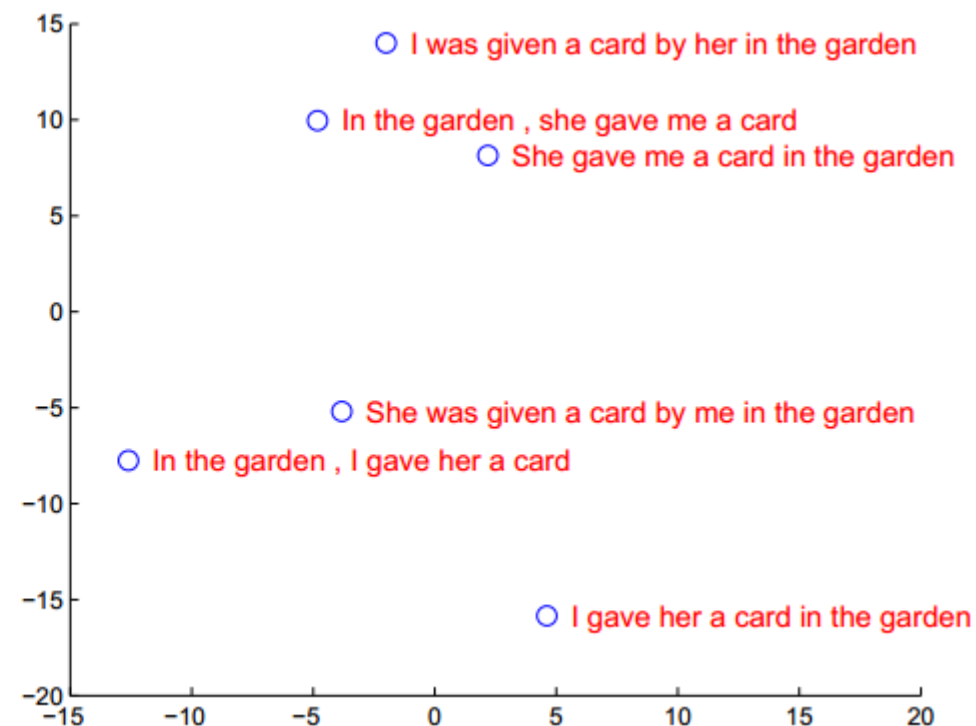
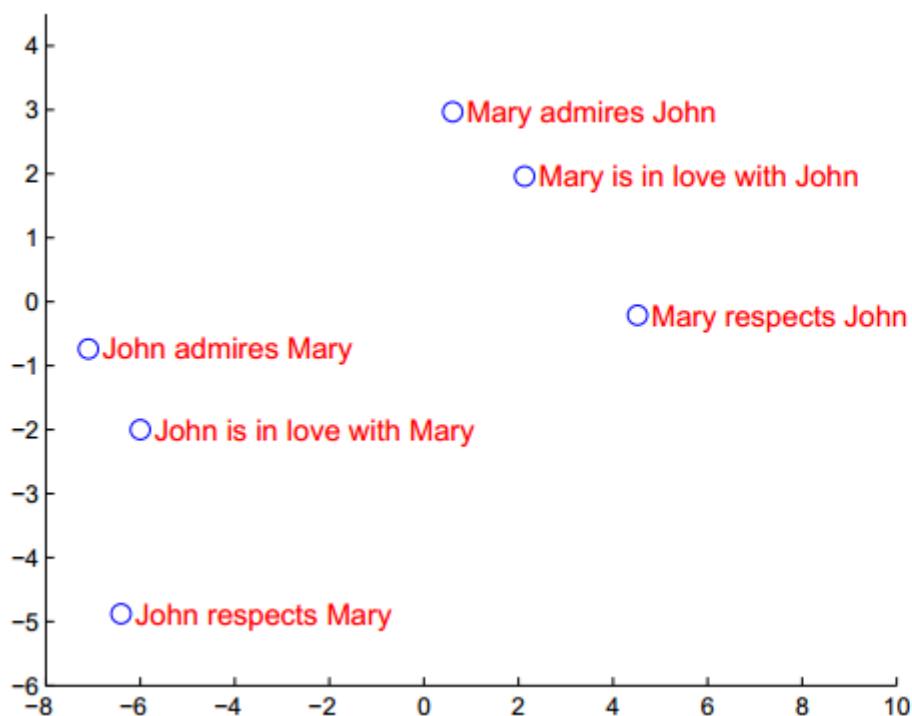
# کارآمدی روی جمله‌های طولانی

نوع	جمله
مدل	Ulrich UNK , membre du conseil d' administration du constructeur automobile Audi , affirme qu' il s' agit d' une pratique courante depuis des années pour que les téléphones portables puissent être collectés avant les réunions du conseil d' administration afin qu' ils ne soient pas utilisés comme appareils d' écoute à distance .
ترجمه صحیح	Ulrich Hackenberg , membre du conseil d' administration du constructeur automobile Audi , déclare que la collecte des téléphones portables avant les réunions du conseil , afin qu' ils ne puissent pas être utilisés comme appareils d' écoute à distance , est une pratique courante depuis des années .
مدل	“ Les téléphones cellulaires , qui sont vraiment une question , non seulement parce qu' ils pourraient potentiellement causer des interférences avec les appareils de navigation , mais nous savons , selon la FCC , qu' ils pourraient interférer avec les tours de téléphone cellulaire lorsqu' ils sont dans l' air ” , dit UNK .
ترجمه صحیح	“ Les téléphones portables sont véritablement un problème , non seulement parce qu' ils pourraient éventuellement créer des interférences avec les instruments de navigation , mais parce que nous savons , d' après la FCC , qu' ils pourraient perturber les antennes-relais de téléphonie mobile s' ils sont utilisés à bord ” , a déclaré Rosenker .
مدل	Avec la crémation , il y a un “ sentiment de violence contre le corps d' un être cher ” , qui sera “ réduit à une pile de cendres ” en très peu de temps au lieu d' un processus de décomposition “ qui accompagnera les étapes du deuil ” .
ترجمه صحیح	Il y a , avec la crémation , “ une violence faite au corps aimé ” , qui va être “ réduit à un tas de cendres ” en très peu de temps , et non après un processus de décomposition , qui “ accompagnerait les phases du deuil ” .





# یادگیری بازنمایی



خوشه بندی عبارات با توجه به معنای آنها در دو بعد ثابت و تأثیر ترتیب آمدن واژه ها در عبارت



# نتیجه‌گیری و کارهای آتی

- ارایه یک مدل مستقل از دامنه، برای وظایف مبتنی بر نگاشت توالی‌ها به یکدیگر.
- یک شبکه LSTM ژرف قادر به شکست سیستم‌های ترجمه ماشینی سنتی است.
- وارون سازی توالی مبدأ سبب افزایش دقت و بهبود کارآمدی مدل می‌شود.
- LSTM در یادگیری توالی‌های طولانی خیلی خوب عمل می‌کند.
- افزایش دقت مدل با روش‌های مختلف مثل افزایش مجموعه داده آموزش
- استفاده از این مدل برای سایر وظایف مبتنی بر نگاشت توالی به توالی



- hdanau, D., Cho, K. & Bengio, Y., 2014. **Neural machine translation by jointly learning to align and translate.** , pp.1–15. Available at: <http://arxiv.org/abs/1409.0473>
- Cho, K. et al., 2014. **Learning phrase representations using RNN encoder-decoder for statistical machine translation.** Available at: <http://arxiv.org/abs/1406.1078>.
- Goodfellow, I., Bengio, Y. & Courville, A., 2016. **Deep learning**, MIT Press. Available at: <http://www.deeplearningbook.org/>.
- Ilya Sutskever, Oriol Vinyals, Q.V. Le et al., 2014. **Sequence to sequence learning with neural networks.** Nips, pp.1–9.
- Jurafsky, D. & Martin, J.H., 2009. **Speech and language processing (2nd edition)**, Upper Saddle River, NJ, USA: Prentice-Hall, Inc.
- Kalchbrenner, N. & Blunsom, P., 2013. **Recurrent continuous translation models.** Emnlp, (October), pp.1700–1709.



# مراجع (ادامه)

---

- Kelly, C., 2017. **Tab-delimited bilingual sentence pairs from the Tatoeba project (good for anki and similar flashcard applications)**. Available at: <http://www.manythings.org/anki/> [Accessed November 13, 2017].
- Keras-Team, 2017. **Sequence to sequence example in Keras (character-level)**. *Keras*. Available at: [https://github.com/fchollet/keras/blob/master/examples/lstm\\_seq2seq.py](https://github.com/fchollet/keras/blob/master/examples/lstm_seq2seq.py) [Accessed November 13, 2017].
- Luong, M.-T., Brevdo, E. & Zhao, R., 2017. **Neural machine translation (seq2seq) tutorial**. <https://github.com/tensorflow/nmt>.
- Luong, M.T., 2016. *Neural machine translation*. Stanford university. Available at: <https://github.com/lmthang/thesis>.
- Pecina, P., 2014. **ACL 2014 ninth workshop on statistical machine translation**. Available at: <http://www.statmt.org/wmt14/medical-task/index.html> [Accessed November 13, 2017].



**با تشکر از توجه شما**

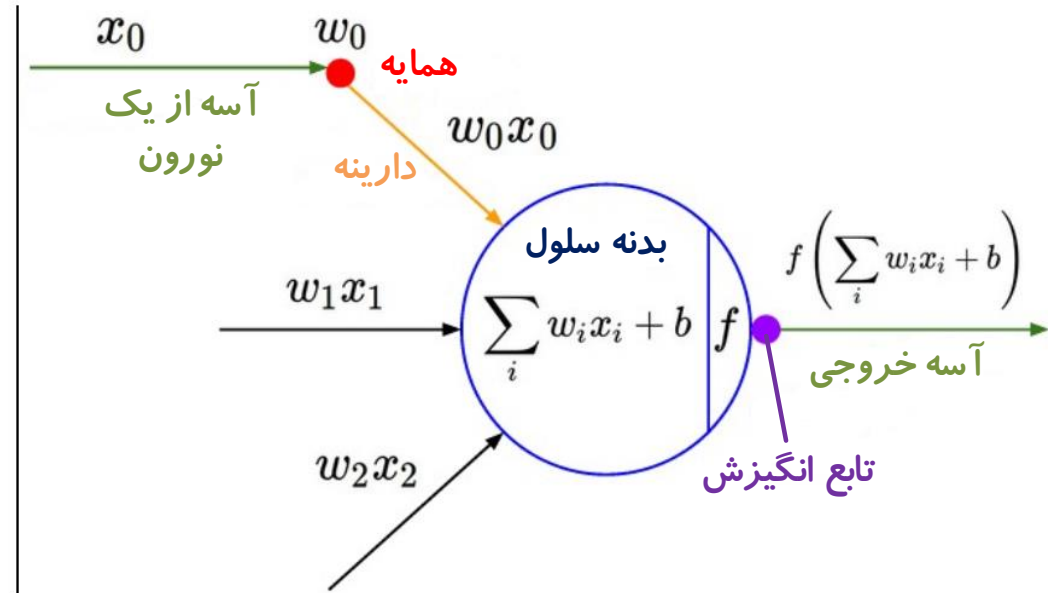
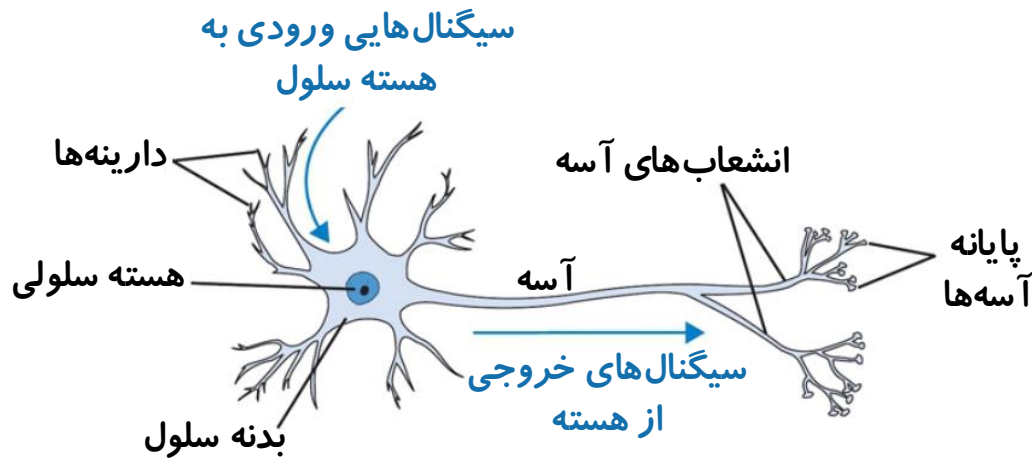


**M-ZAKERI@LIVE.COM**



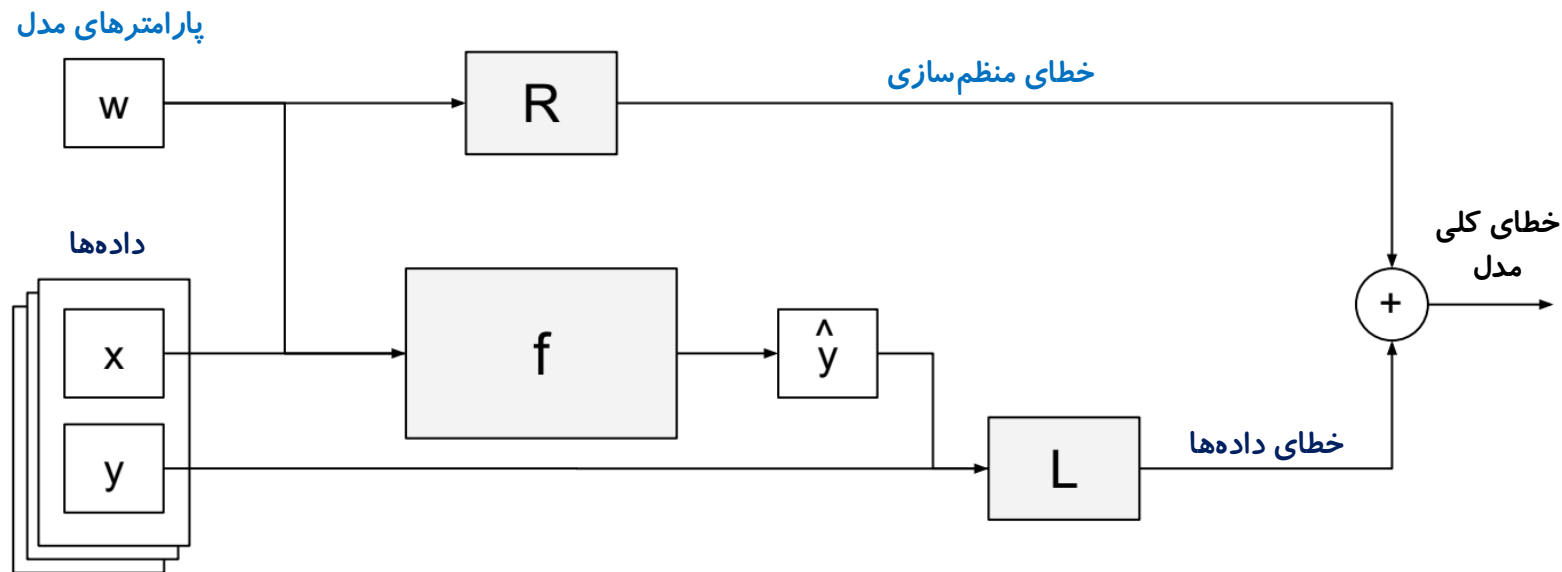
# مدل محاسباتی عصب (نورون)

- شبکه‌های عصبی مصنوعی؛ الهام گرفته شده از شبکه عصبی مغز انسان



# جریان داده در یادگیری بانظارت

• محاسبه خطا و آموزش

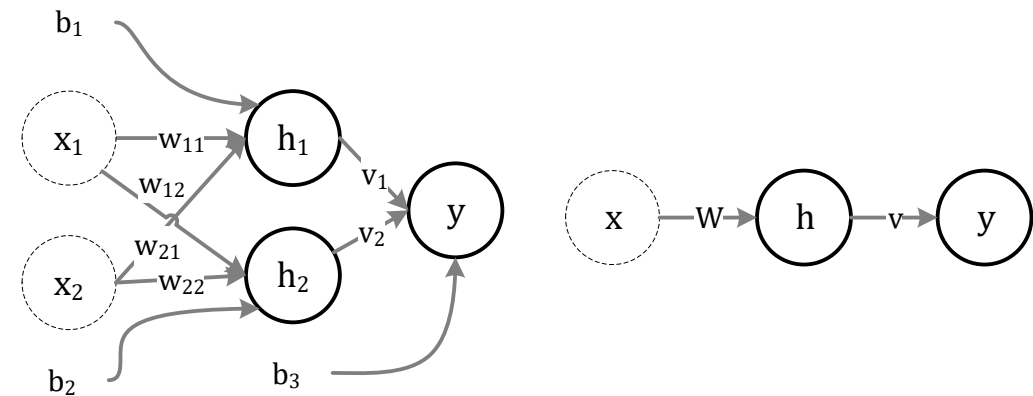


# محاسبات شبکه عصبی ژرف

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

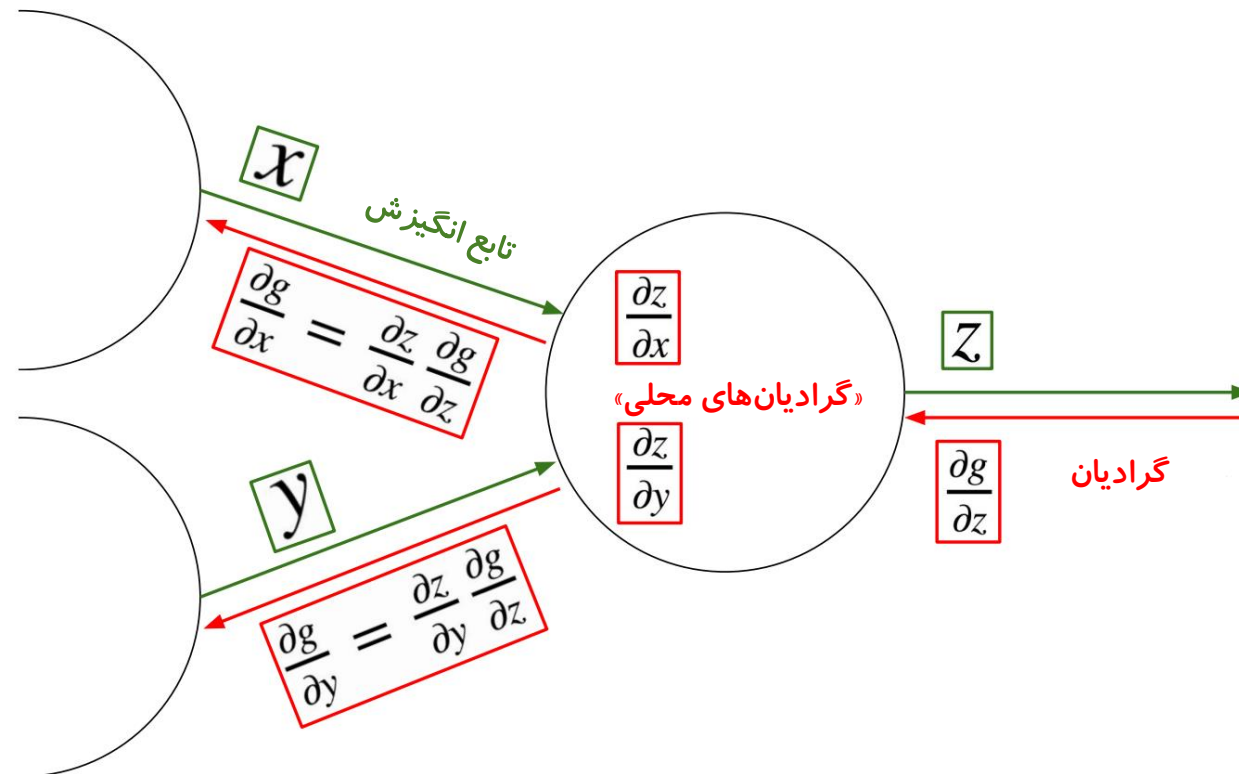
$$h = W^T x + b$$

$$= \begin{bmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} w_{11}x_1 + w_{21}x_2 + b_1 \\ w_{12}x_1 + w_{22}x_2 + b_2 \end{bmatrix}$$

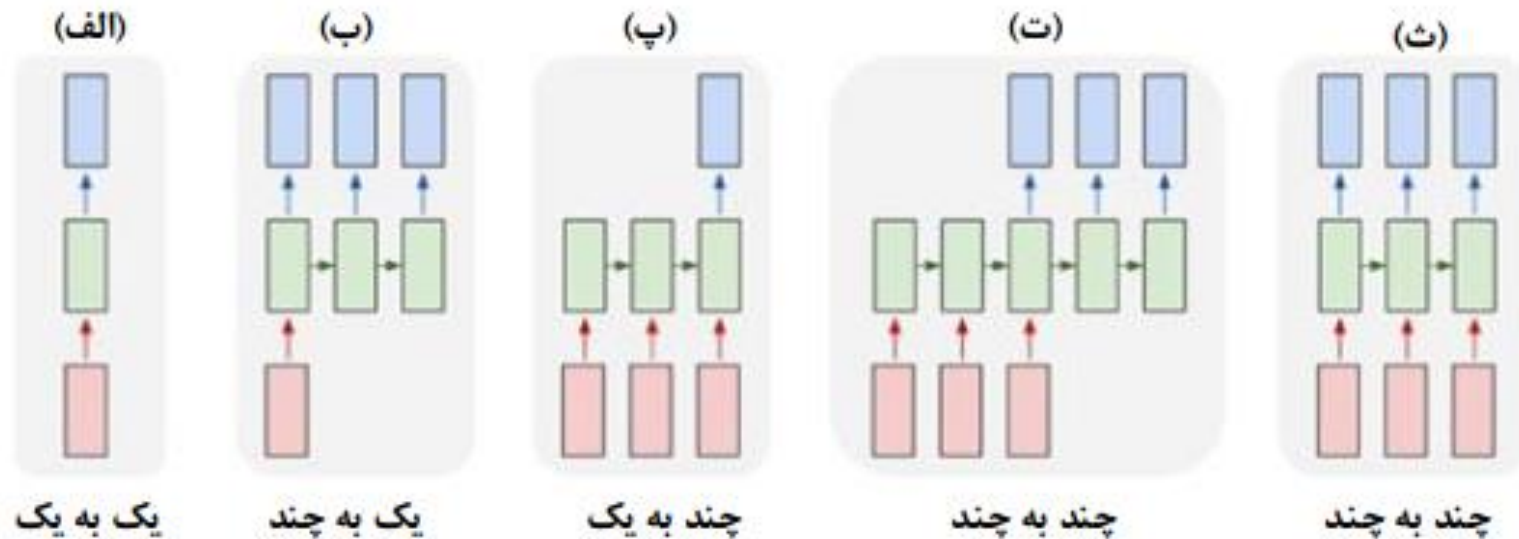




# یادگیری به روش پس انتشار (انتشار پس رو)



# انواع بهم‌بست شبکه‌های عصبی مکرر



# اثر پدیده وارونگی بر جمله‌های فارسی

• اثر پدیده وارون سازی در ترجمه انگلیسی به فارسی؟

چیست	شما	نام	name	your	is	What
------	-----	-----	------	------	----	------

وارون شده

چیست	شما	نام	What	is	your	name
------	-----	-----	------	----	------	------



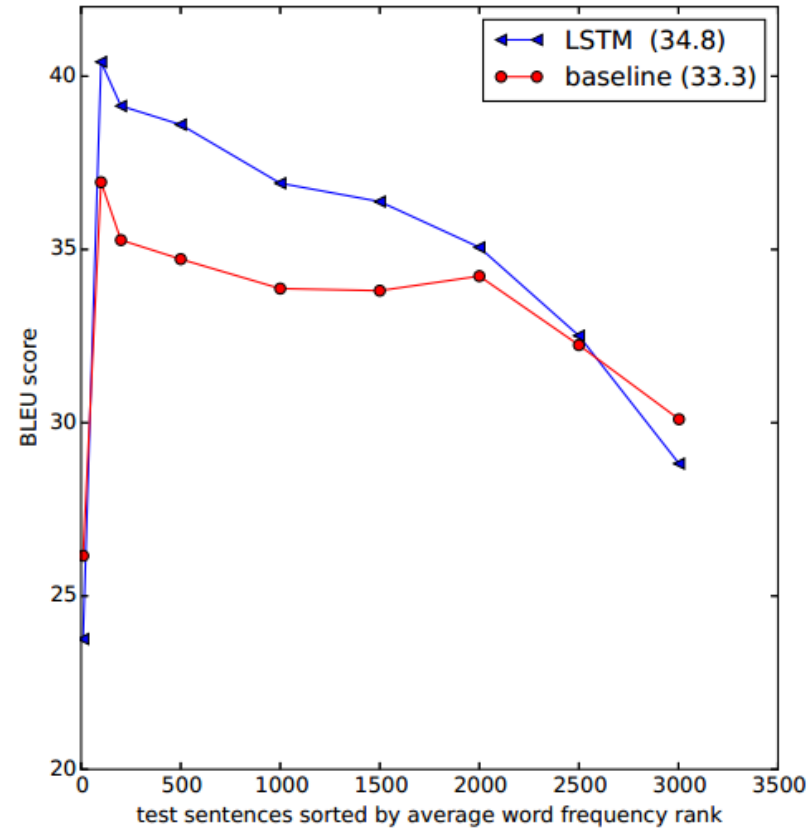
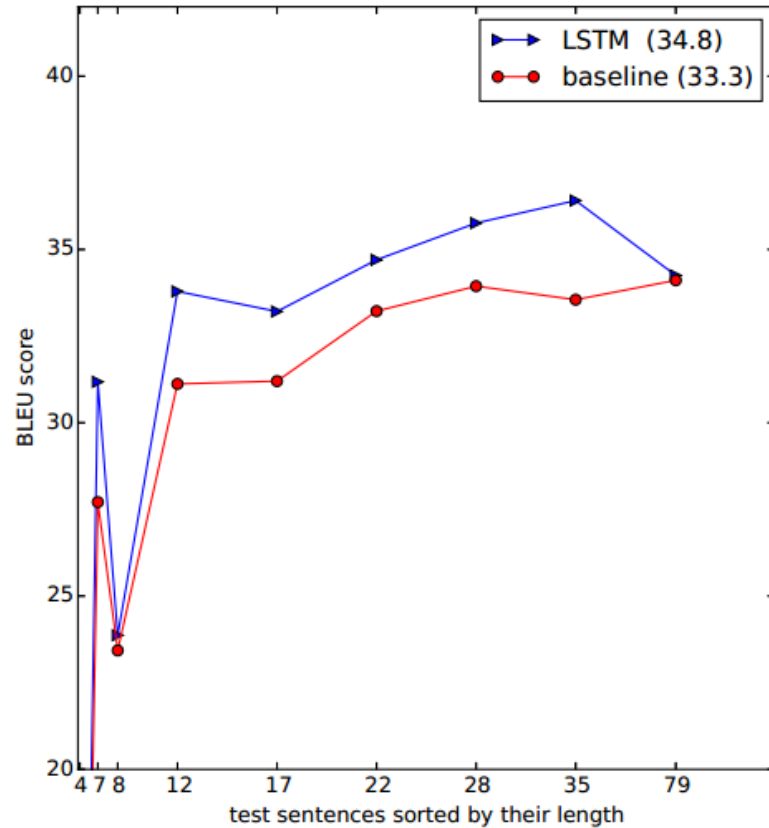
# ترکیب ترجمه سنتی با ترجمه ماشینی عصبی

• امتیاز دهی  $n$  فهرست بهتر

امتیاز BLEU (ntst14)	روش
۳۴.۵۴	چو و همکاران (Cho et al. 2014)
۳۵.۶۱	امتیازدهی مجدد ۱۰۰۰ فهرست بهتر با یک LSTM روبه جلو
۳۵.۸۵	امتیازدهی مجدد ۱۰۰۰ فهرست بهتر با یک LSTM وارون
۳۶.۵۰	امتیازدهی مجدد ۱۰۰۰ فهرست بهتر با پنج LSTM وارون
~۴۵	پیش گویی امتیازدهی مجدد ۱۰۰۰ فهرست بهتر



# کارآمدی روی جمله‌های طولانی



# پایان

- برای اطلاعات بیشتر به فایل کامل پروژه در **تارنمای بوته** رجوع فرمایید.  
<http://www.boute.ir/iust-nlp-96/290> •

