

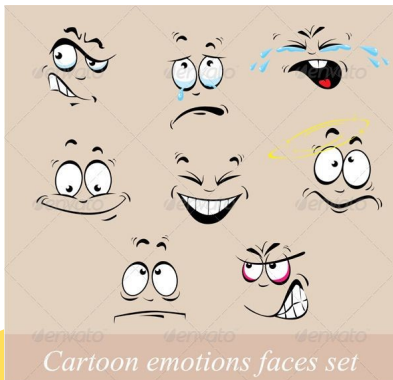


New Neuromarketing

Created 17/03/2021 at 20:59

تشخیص احساسات از روی متن فارسی و نگاهی کوتاه به ابزارهای تست پروژه های هوش مصنوعی

(Recognize **emotions** from **Persian text** and a brief look at the **testing tools** of **artificial intelligence** projects)



گرداوردنده:

سهیل سبزواری
برای گروه
NeuroMarketing

خرداد ۱۴۰۰

Table 2: Comparative results of different methods on Persian dataset

Model	F-Score	Precision	Recall
Model1: NBSVM	44%	70.7%	31.9%
Model 2: Naive Bayes	62.5%	65.4%	61.7%
Model 3: logistic regression	67.1%	68.9%	65.4%
Model 4: Bidirectional-LSTM	53.8%	54.3%	55%
Model 5: CNN	67%	66%	67%
Proposed Model: CNN-LSTM	85%	86.5%	86%
CNN-GRU	85.8%	85%	82%
GRU	74%	77.5%	75.6%

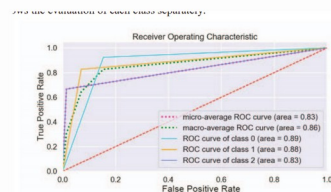


Fig. 5: ROC curve.

نگاهی کوتاه به مفهوم احساس

- احساس به معنی **دریافتن**، **آگاه شدن** و درک چیزی با یکی از **حواس** **می باشد** و بازتاب ذهنی تأثیرات مادی (فیزیکی) است که **شالوده** **همه ادراکات** به شمار می رود. (ویکی پدیا)

About Emotion Estimation And Technologies

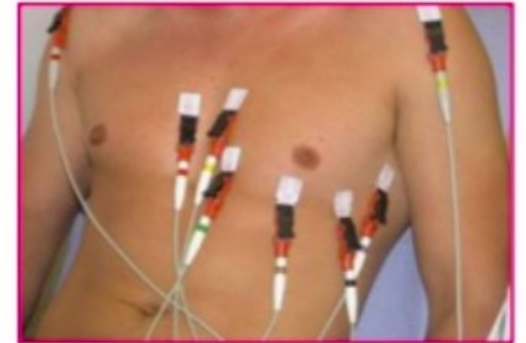
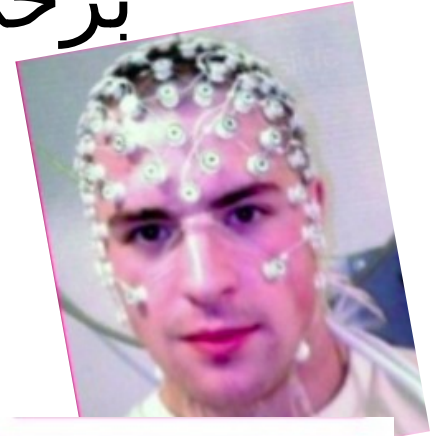


جستارهای وابسته [ویرایش]

- خودآگاهی
- علوم اعصاب شناختی
- هیجان
- مسئله دشوار خودآگاهی
- شهود
- آزمون شخصیت مایرز-بریگز
- ادراک
- کیفیات ذهنی
- حس

برخی از روشهای تشخیص احساسات (۱)

- EEG + Brain Computer Interfaces
-
- ECG + Cardiovascular signals
-
- Electrodermal activity



برخی از روشهای تشخیص احساسات (2)

- **Speech** + **Voice** intonation



- **Facial** expressions



- **Body** language



<https://psycatgames.com>

[s/would-you-rather/](https://www.would-you-rather.com/)

<https://www.slideshare.net/franzonadiman/applications-of-emotions-recognition>

کاربردهای تشخیص احساسات

نظارت

آموزش

پزشکی

بزاریابی

قانون

تفریحی

کاربردهای تشخیص احساسات در پزشکی

8

1. Medicine

Health care → patients' feelings about treatment



baysidecarecenter.com



iavrpt.org



degreequery.com



masainternational.com

rg

کاربردهای تشخیص احساسات در آموزش

2

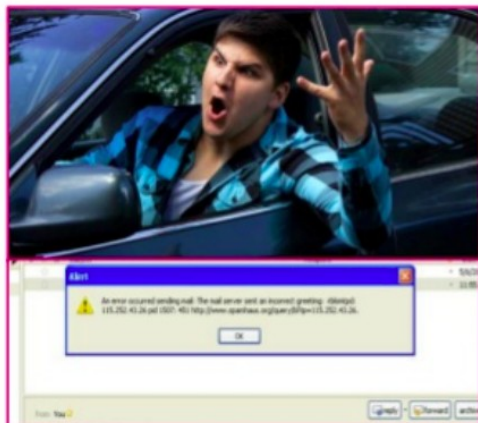
2. E-learning

- adjust the presentation **style** of an **online tutor**
- detect the **state** of the learner



کاربردهای تشخیص احساسات در نظارت

3. Monitoring



sbtjapan.com

ATM not dispensing money when **scared**



manage.grabweb.net

s3.amazonaws.com

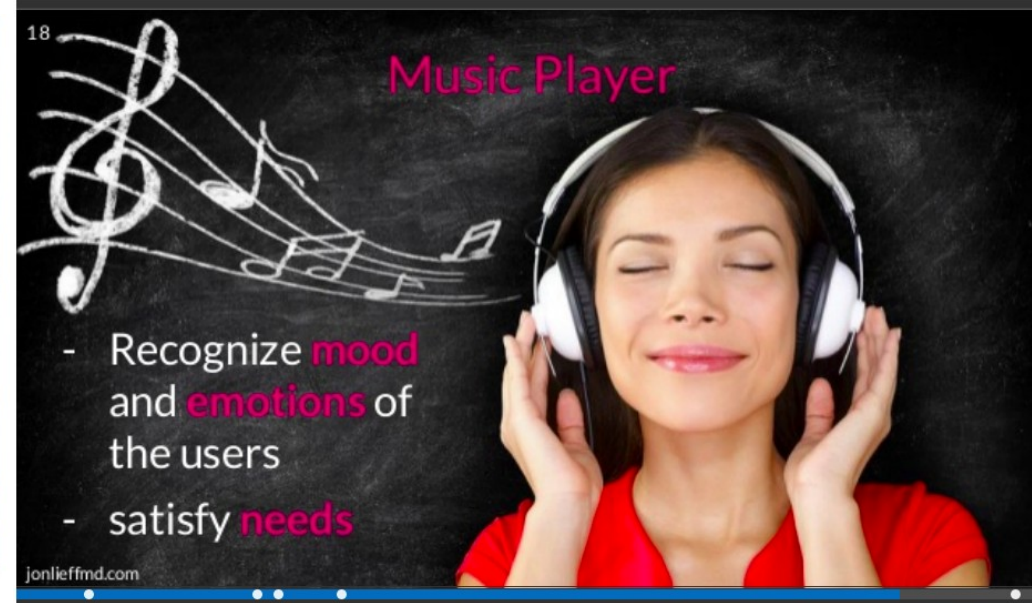
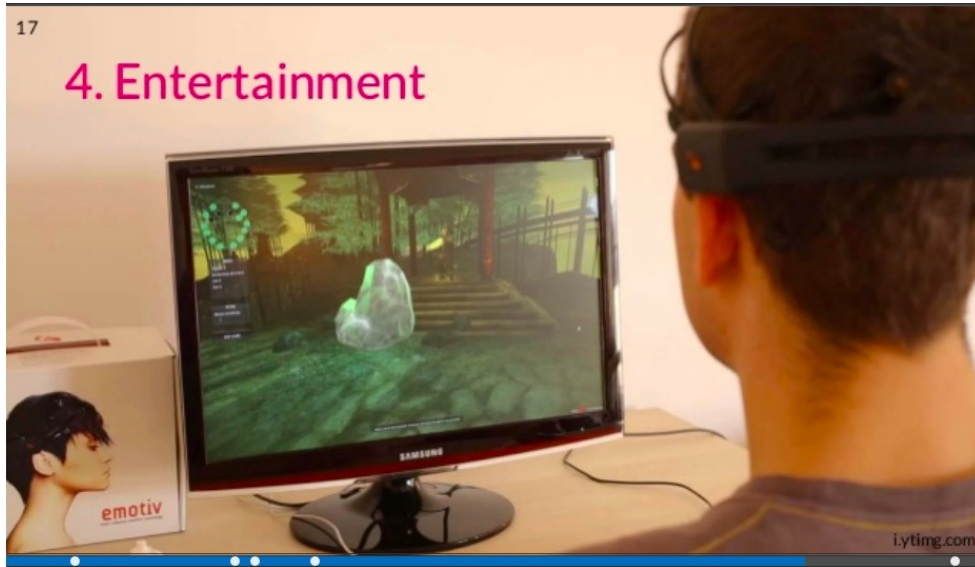
16

Remote Love

- globalization → mobility
- **long distance** relations
- recognize partner's **true** emotional **state**
- **heartbeat** pulsing rings
- phone detecting **affection**



کاربردهای تشخیص احساسات در سرگرمی



کاربردهای تشخیص احساسات در **قانون**



کاربردهای تشخیص احساسات در بازاریابی

Be the first to clip this slide

20

6. Marketing

- impact of **ads**
- attention and engagement
- **emotion** vital in **purchasing** decisions
→ improve **sales**



redorbit.com

ابزارهای تست پروژه های هوش مصنوعی



نحوه استفاده از ابزارهای آنلاین برای کار روی پروژه های هوش مصنوعی در زمینه های مختلف

- (1) جستجوی تگ های انگلیسی در مقالات و پست های اینترنتی مربوط به هوش مصنوعی
- (2) اجرا کردن کدهای مقالات یا پستهای اینترنتی (از سایت GitHub و ...) در محیط ادیتور Google Colab و یا سایت Kaggle
- (3) کار بیشتر روی پروژه ها از طریق :

(A) برقراری ارتباط با متخصصان رشته برنامه نویسی و هوش مصنوعی از طریق طرح پرسش در انجمن های انگلیسی مربوط و یا ارسال ISSU برای پروژه های سایت GitHub

(B) ارسال ایمیل برای درخواست راهنمایی به نویسندگان پست ها یا پروژه های بالا

(C) تشکیل گروه های کاری دانشجویی کوچک (۵ نفره) برای کار بیشتر و رشد دسته جمعی در این زمینه

در رابطه با جستجوی تگ های انگلیسی مربوط به هوش مصنوعی

▪ استفاده از سایت های حاوی مقالات و کدهای آماده مانند:

- <https://paperswithcode.com/>
- <http://github.com>

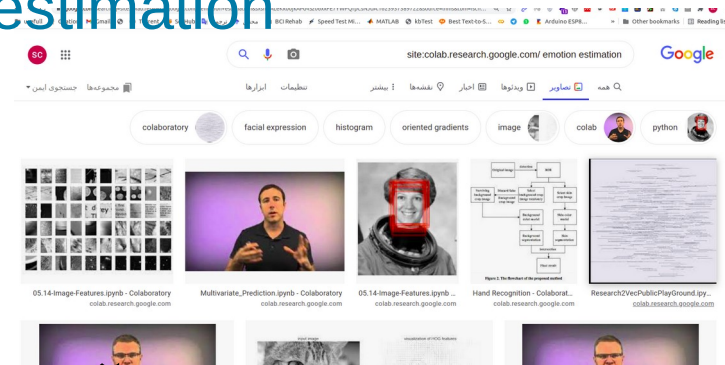
▪ مشاهده ویدیو های Youtube.

▪ مطالعه پست های وبلاگ ها و ...:

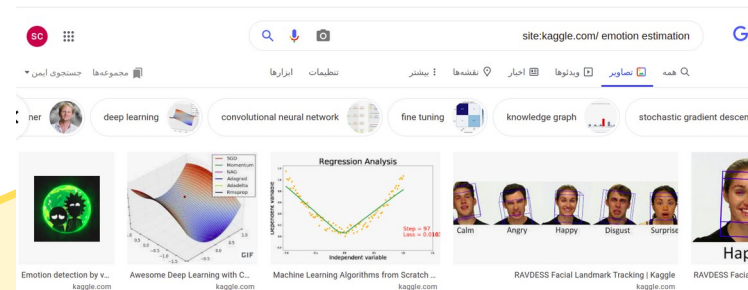
- <https://medium.com/>
- <https://virgool.io/>

اجرا کردن کدهای هوش مصنوعی

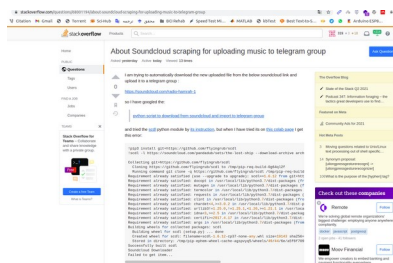
- جستجوی کدهای آماده در صفحات **Google Colab**:
[site:colab.research.google.com/ emotion estimation](https://colab.research.google.com/emotion-estimation)



- جستجوی کدهای آماده در صفحات سایت **Kaggle**:
[site:kaggle.com/ emotion estimation](https://www.kaggle.com/emotion-estimation)

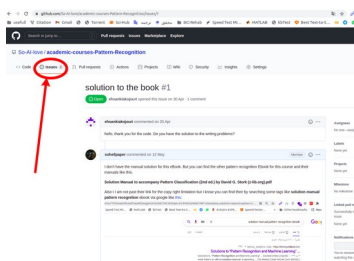


طرح پرسش در انجمن های انگلیسی



- طرح سوال در زیر شاخه های انجمن Stackexhcnage:

- طرح پرسش یا Issu بر روی پروژه های متن باز، هوش مصنوعی مربوط به سایت GitHub:



- ارسال ایمیل به دارندگان پروژه ها یا نویسندگان پست های مربوط:

تشکیل گروه های کاری دانشجویی کوچک (۵ نفره)

۱۰ دلیل برای تشکیل گروه های دانشجویی (لینک):

1. درک بهتری از موضوع بهتر بدست می آورید. 🧐🧐
2. نمرات بهتری را کسب می کنید. 😊
3. بینش کاملاً خوبی نسبت به مطالب درسی بدست می آورید. 😊
- 4- ویژگی های شخصی خود را حفظ خواهید نمود. (حفظ بیشتر زمان و سبک مناسب شخصی برای مطالعه دروس)
5. کسب تجربه کار تیمی 🤗🤗
6. به تأخیر انداختن و دستیابی بیشتر (در ارتباط قرار گرفتن با دانشجویان با انگیزه تر و کاهش تمایل به تأخیر انداختن و ...) 😊
7. مهارت های حل مسئله خود را پیشرفت دهید. 😊
- 8- چرا ما باید در گروه ها مطالعه کنیم؟ (امکان دیدار در یک مکان عمومی و سایر امکانات خلاقانه پیش روی شما برای لذت بردن از کار تیمی و ... 🤗🤗)
9. قرار ملاقات گذاشتن در فرمت های مورد علاقه مانند : رفتن به نزدیکترین کافه، و لذت بردن از تجربه مطالعه و روابط انسانی ، احاطه شدن توسط متخصصین ، دانشجویان و انتقال انرژی مثبت گروه خود به محیط اطراف، سایر افراد محلی و ... 🤗🤗
10. انگیزه و الهام بخشیدن به یکدیگر 🤗🤗

بررسی یک نمونه پروژه هوش مصنوعی در زمینه تشخیص احساسات از روی متن فارسی

پروژه جالب تشخیص احساسات در متن فارسی (کاری از Parsiya

(Keshavarz

با درود و سپاس

روزمگی های گذر گایز
اینجا جاییه که ما برنامه نویسن...

دنبال کردن



در ادامه آموزش های هوش مصنوعی با گوگل گولب (Google Cloba Notebook) ، در اینجا با معرفی کاری از Parsiya Keshavarz (متن باز در سایت GitHub) ، با شما همراه هستم.

توضیحات مربوط به پروژه را می توانید در ویدیوی زیر مشاهده بفرمایید:

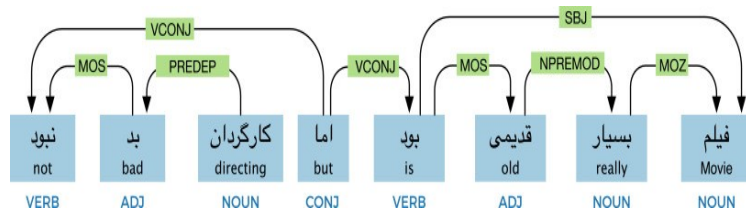


تشخیص احساسات در متن

اشتراک دیدگاه ها در وب سایت های ارتباط جمعی، راهی مناسب برای انتقال پیام ها به کاربران نهایی در یک موضوع خاص است .

آنالیز احساسی شاخه ای از پردازش زبان طبیعی (NLP) است، که در زمینه های مختلفی مانند موارد زیر کاربرد دارد :

fastText



- (1) بازاریابی
- (2) خدمات مشتری
- (3) و غیره

مدل به کار رفته :

- Hazm برای تحلیل اولیه متن
- fastText (برای نمایش کلمه)
- CNN با لایه هایی از LSTM

نتایج تجربی اثربخشی مدل پیشنهادی را با دقت ۱۵٪ نشان می دهد .

A COMBINED DEEP LEARNING MODEL FOR
PERSIAN SENTIMENT ANALYSIS

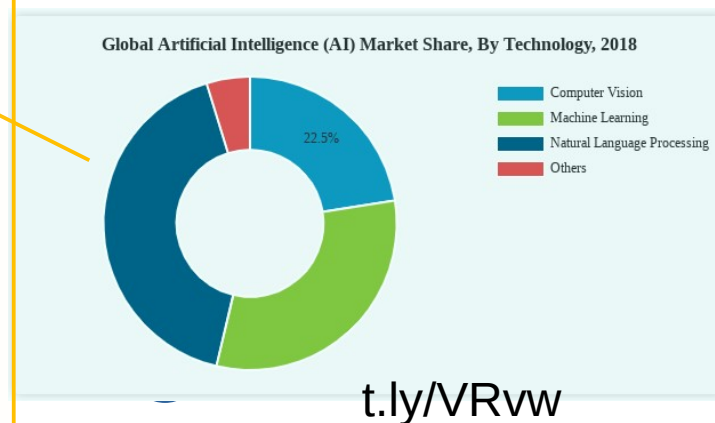
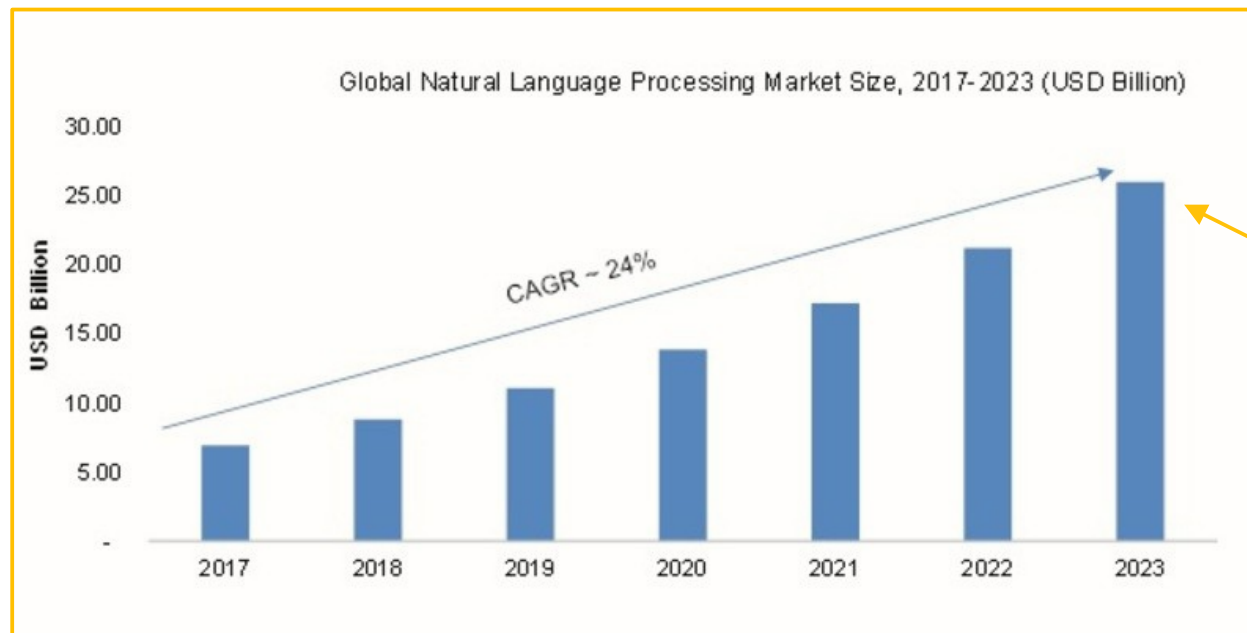
ZAHRA BOKAEE NEZHAD¹ AND MOHAMMAD ALI DEIHIMI²

¹Department of Information and Computer Science, Zand university, Shiraz, Iran.

²Department of Electrical and Electronics Engineering, Bahaonar University, Shiraz, Iran.

نکته: دقت بدست آمده در نتایج مربوط به مقاله تشخیص احساسات در متن فارسی با الگوریتم Word2vec به جای fastText می باشد.

پیش بینی اندازه و سهم بازار هوش مصنوعی در بخش پردازش های زبانی تا سال ۲۰۲۳



t.ly/rGmK

مدل شبکه عصبی

مراحل انجام کار:

من پارسال خریدم اصلا پوشش دهی خوبی نداره نمیدونم فکر کنم به خاطر لایت بودنش چون تا حالا لایت استفاده نکرده بودم چند بار بیشتر استفاده نکردم پوست من رو که اصلا کوچک نکرد که بعد از ردن گرم رنگی کمتر به چشم بیان فقط ماندگاری آرایش رو برد بالا کانترس آریں خیلی بهتره و چروکها و منافذ رو کم میکنه کوچک و ولی واقعا غایبه تمام معایب صورت رو میپوشونه منظورم از معایب سوراخ هایی که روی پوست های چربه هست منگه راضی ام ولی ای کاش ورژن هم نوشته میشد و سفارش دادم ولی وقتی رسید دیدم انقدر برارش پرداخت کردم خیلی تو دوقم خورد چون فکر میکردم ابعادش خیلی بزرگتر باشه کاش دیجی درمورد ابعادشم مینوشت کارایی خاصی ازش ندیدم از خریدش خیلی راضی نیستم ولی اگه کسی به صورت حرفه ای آرایش میکنه شاید بدرخش بخوره بن بنیک سالهاست که استفاده میکنم و خیلی راضی ام روی پوست خوب میشینه و اصلا جوش نرزم خیلی خوبه که دیجی کالا اینو آورد خیلی خوشحال شدم و خدمش

word embedding , convolutional , pooling , dropout , RNIN , and fully connected.

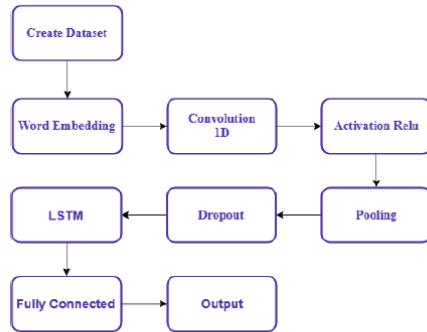
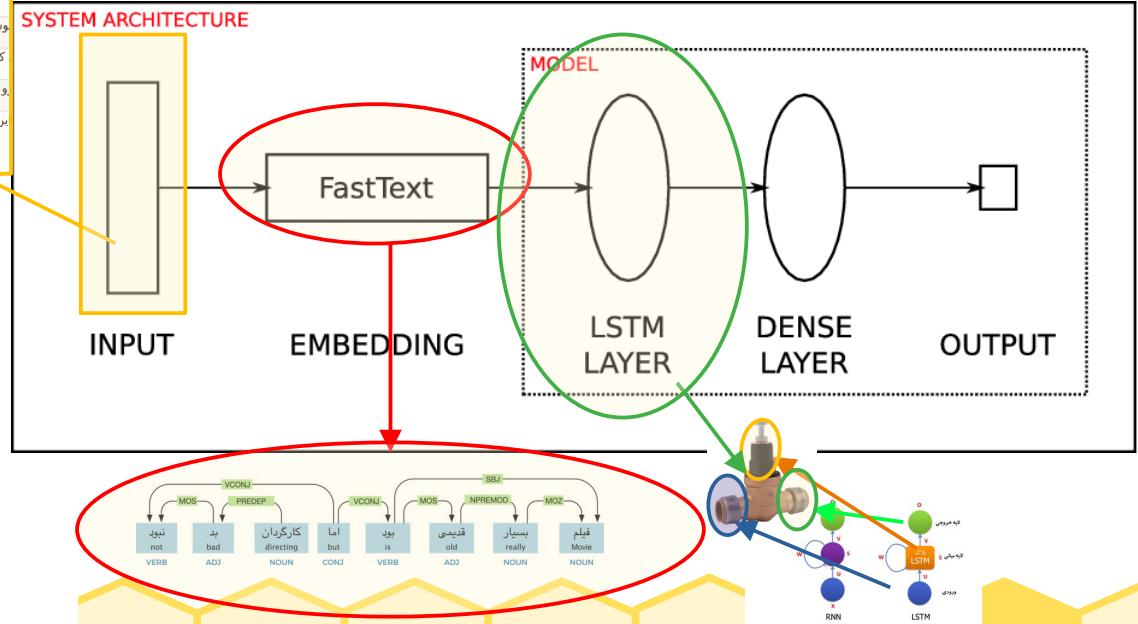


Fig. 1: System architecture of the Proposed Model.



دیتا بیس

- مجموعه داده‌های فارسی مختلف را از Twiter و digikala به عنوان مشهورترین سایت دیجیتال جمع‌آوری شده و به ۳ دسته بندی **مثبت** ، منفی و یا **خنثی** تقسیم بندی شده (Tutorial_Dataset.csv)
- تصویر زیر بخشی از فایل **دیتابیس** را محیط کولب نشان می دهد:

Tutorial_Dataset.csv

✓

RAM

Disk

Editing

Text	Score	Suggestion
من پارسال خریدم اصلا پوشش دهی خوبی نداره نمیدونم فکر کنم به خاطر لایت بودنش چون تا حالا لایت استفاده نکرده بودم چند بار بیشتر استفاده نکردم والا منافذ باز پوست من رو که اصلا کوچک نکرد که بعد از زدن کرم رنگی کمتر به چشم بیان فقط ماندگاری ارایشم رو برد بالا کاترپس ازین خیلی بهتره و جروکها و منافذ رو کم میکنه	60	3
سایزش خیلی کوچیکه ولی واقعا عالیه تمام معایب صورت رو میپوشونه منظورم از معایب سوراخ هایی که روی پوست های چربه هست منگه راضی ام ولی ای کاش وزنش هم نوشته میشد	52	3
این کالارو سفارش دادم ولی وقتی رسید دیدم انقدر براش پرداخت کردم خیلی تو ذوقم خورد چون فکر میکردم ابعادش خیلی بزرگتر باشه کاش دیجی درمورد ابعادشم مینوشت کارایی خاصی ازش ندیدم از خریدش خیلی راضی نیستم ولی اگه کسی به صورت حرفه ای ارایش میکنه شاید بدردش بخوره من از این پنکیک سالهاست که استفاده میکنم و خیلی راضی ام روی پوست خوب میشینه و اصلا جوش نزدنم خیلی خوبه که دیجی کالا اینو آورد خیلی خوشحال شدم و خریدمش	88	1
من از صبح تا شب اینو رو پوستم دارم خیلی سبکه لطیف و مخملیه پوشانندگی بالایی نداره ولی حس خوبی بهت میده پوشش دهیش عالیه ماندگاریشم روی پوست چرب خوبه طوری که بدون فیکساتور از صبح تا شب تقریبا ثابت میمونه فقط اگه مدت طولانی قراره بمونه روی پوستتون و پوستتون چربه نباید دست بزنید بهش چون پاک میشه	48	2
واقعا عالی بوی خوبی داره تمیز کنندگی چشم گیری داره چربی پوست رو کم میکنه تو اولین استفاده به وضوح تفاوتش مشخصه	84	1
من این محصول رو برای خرید عروسی ام به صورت خیلی اتفاقی خریدم هیچ بویی نداره اما قدرت پاک کنندگی اش مخصوصا برای ریمل عالیه سخت ترین و سنگین ترین مدادها و ریمل ها و سایه هارو بدون دردسر و راحت پاک میکنه جالب اینه که اگه اتفاقی موقع پاک کردن چشمتان یکم داخل چشم بشه اصلا چشم رو نمی سوزاند و اذیت نمیشوید من برای پاک کردن رزلب هم ازش استفاده کردم واسه اونم عالی جواب داد تنها نقطه ضعیفش اینه که قبل از استفاده باید تکان اش بدهید و بعدش یکم بریزید روی پنبه و به چشمتان بکشید	100	1
د. کلا. محصول. خیلی. خوبه. فقط. من. با. وجود. این. که. پوست. چرب. دارم. بعد. از. استفاده. روی. چشمام. حتما. باید. با. یک. پنبه. یا. تکه. تکه. میسوز. ه. پاک. کنم. و. به. چشمتان. بکشید.	64	2
	88	1
	80	1
	96	1

دانلود Databse در فایل Google Colab NoteBooks

Step B) Preparing opinion dataset

in this step we going to collect a dataset that crawled by @minasmz it's not good and I only used 450 pos and 450 neg reviews from it.anyway here we will download the dataset and split it to train and test (I created Train and Test then I filled it with data)

```
#@title Download and prepare Dataset
!wget https://raw.githubusercontent.com/ashalogic/Persian-Sentiment-Analyzer/master/Tutorial_Dataset.csv
```

Tutorial_Dataset.csv

991 to 1000 of 3261 entries

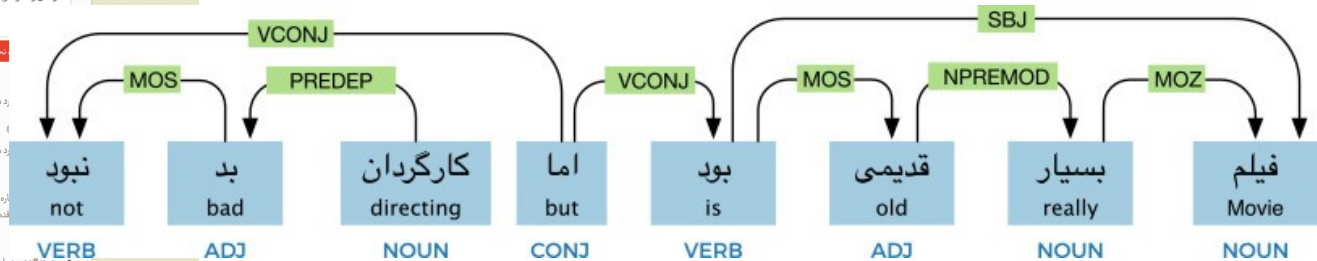
Text	Score	Suggestion
من پارسال خریدم اصلاً پوشش دهی خوب نداره نمیدونم فکر کنم به خاطر لایت بودنش چون تا حالا لایت استفاده نکرده بودم چند بار بیشتر استفاده نکردم والا منافذ بار پوست من رو که اصلاً کوچک نکرد که بعد از ردن کرم رنگی کمتر به چشم بیان فقط ماندگاری ارایشم رو برد بالا کانس آبرن خیلی بهتره و چروکها و منافذ رو کم میکنه	60	3
سایزش خیلی کوچیکه ولی واقعا عالیه تمام معایب صورت رو میپوشونه منظورم از معایب سوراخ های که روی پوست های چربه هست منکه راضی ام ولی ای کاش وزنش هم نوشته میشد	52	3
این کالارو سفارش دادم ولی وقتی رسید دیدم انقدر براش پرداخت کردم خیلی تو ذوقم خورد چون فکر میکردم ابعادش خیلی بزرگتر باشه کاش بجهی درمورد ابعادشم میپوشنت کارایی خاصی ازش ندیدم از خریدش خیلی راضی نیستم ولی اگه کسی به صورت جرقه ای ارایش میکنه شاید بدردش بخوره من از این پنکیک سالهاست که استفاده میکنم و خیلی راضی ام روی پوست خوب میشینه و اصلاً جوش نزنم خیلی خوبه که بجهی کالا اینو آورد خیلی خوشحال شدم و خریدمش	88	1
من از صبح تا شب اینو رو پوستم دارم خیلی سبکه لطیف و مخملیه پوشاندننگی بالایی نداره ولی حس خوبی بهت مده پوشش دهیش عالیه ماندگاریشم روی پوست چرب خوبه طوری که بدون فیکساتور از صبح تا شب تقریباً ثابت میمونه فقط اگه مدت طولانی قراره بمونه روی پوستتون و پوستتون چربه باید دست بزنید بهش چون پاک میشه	48	2
واقعا عالی بوی خوبی داره تمیز کنندگی چشم گیری داره چربی پوست رو کم میکنه تو اولین استفاده به وضوح تفاوتش مشخصه من این محصول رو برای خرید غروب ام به صورت خیلی اتفاقی خریدم هیچ بوی نداره اما قدرت پاک کنندگی اش مخصوصاً برای ریمل عالیه سبک ترین و سبکترین موداها و ریمل ها و سایه هارو بدون دردسر و راحت پاک میکنه جالب اینه که اگه اتفاقی موقع پاک کردن چشمتان یکم داخل چشم بشه اصلاً چشم رو نم سوزاند و اذیت نمیشود من برای پاک کردن رزب هم ازش استفاده کردم واسه اونم عالی جواب داد تنها نقطه ضعیفش اینه که قبل از استفاده باید تکان اش بدهید و بعدش یکم بریزید روی پنبه و به چشمتان بکشید در کل محصول خیلی خوبیه فقط من با وجود این که پوست چرب دارم بعد از استفاده روی چشمتام حتما باید سریعاً روغن بزنم وگرنه میسوره	84	1
	100	1
	64	2
	88	1
	80	1
	96	1

Show 10 per page

آماده سازی دیتابیس

- همچنین در این پروژه با استفاده از ماژول **Hazm** در محیط پایتون ([لینک فرادرس در صورت نیاز](#))، جملات فارسی مانند نمونه زیر تحلیل شده اند و به نحوی برای ورود به عنوان داده های آموزشی و تست در شبکه عصبی استفاده گردیده است:

The screenshot shows the Hazm library website. At the top, there's a navigation bar with links like 'درباره', 'نویسندگان', and 'مشارکت'. Below the navigation bar, there's a section titled 'متن ورودی' (Input Text) with a text area containing the sentence: 'در می‌زند و دوباره وارد میشود. همه می‌گویند قدمش پرپرکت است.' (It hits and enters again. Everyone says its step is prapakt). Below the text area, there's a 'تحلیل صرفی' (Morphological Analysis) section showing the analysis of the sentence. The analysis is presented in a table-like format with columns for the word, its part of speech, and its Hazm tag. The words are: 'نبود' (not), 'بد' (bad), 'کارگردان' (directing), 'اما' (but), 'بود' (is), 'قدیمی' (old), 'بسیار' (really), and 'فیلم' (Movie). The parts of speech are: VERB, ADJ, NOUN, CONJ, VERB, ADJ, NOUN, and NOUN. The Hazm tags are: MOS, PREDEP, VCONJ, MOS, NPREDMOD, and MOZ. The website also has a footer with contact information and a license notice.



آموزش فرادرس در مورد HAZM

بخش Hazm در فایل Google Colab NoteBooks

Installing HAZM

Step B) Preparing opinion dataset

in this step we going to collect a dataset that crawled by @minasmz it's not good and I only used 450 pos and 450 neg reviews from it.anyway here we will download the dataset and split it to train and test (I created Train and Test then I filled it with data)

```
#title Download and prepare Dataset
!wget https://raw.githubusercontent.com/ashalogic/Persian-Sentiment-Analyzer/master/Tutorial_Dataset.csv

!pip install hazm

import random
import numpy
import hazm

def CleanPersianText(text):
    _normalizer = hazm.Normalizer()
    text = _normalizer.normalize(text)
    return text

csv_dataset = pandas.read_csv("/content/Tutorial_Dataset.csv")
revlist = list(map(lambda x: [CleanPersianText(x[0]),x[1]],zip(csv_dataset['Text'],csv_dataset['Suggestion'])))
pos=list(filter(lambda x: x[1] == 1,revlist))
nat=list(filter(lambda x: x[1] == 2,revlist))
neg=list(filter(lambda x: x[1] == 3,revlist))

revlist_shuffle = pos[:450] + neg[:450]
random.shuffle(revlist_shuffle)

print("Posetive count {}".format(len(pos)))
print("Negetive count {}".format(len(neg)))
print("Natural count {}".format(len(nat)))
print()
print("Total count {}".format(len(revlist)))
print()

print(" ", "\n", pos[random.randrange(1, len(pos))])
print(" ", "\n", neg[random.randrange(1, len(neg))])
print(" ", "\n", nat[random.randrange(1, len(nat))])
print(" ", "\n", format(len(revlist_shuffle)))
```

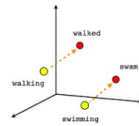
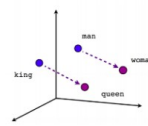
Output of Hazm
(Input the CSV
Database)

Tutorial_Dataset.csv X		991 to 1000 of 3261 entries		Filter
		Score	Suggestion	
Test				
60	3			
52	3			
88	1			
48	2			
84	1			
100	1			
64	2			
88	1			
80	1			
96	1			

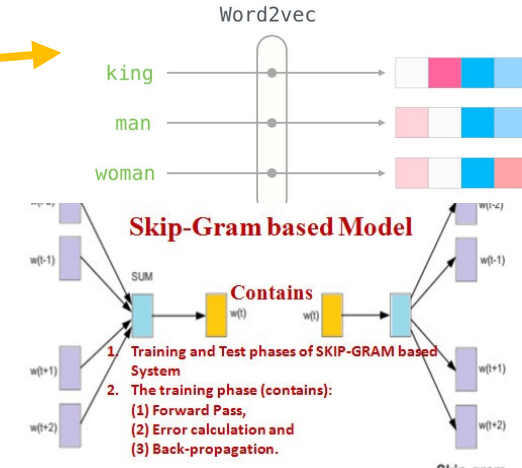
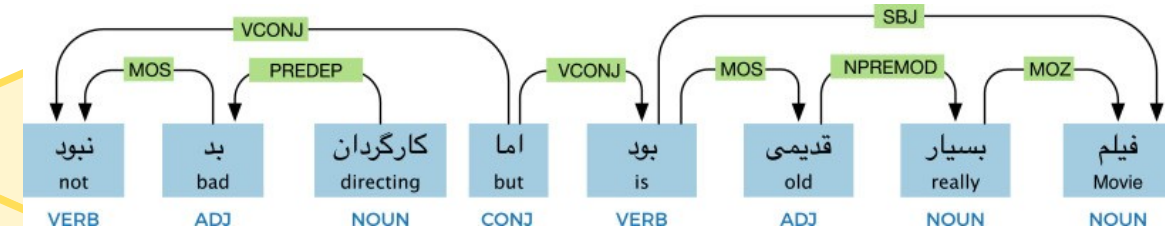
انواع روشهای تعبیه کلمه یا Text Embeddings

- Word2Vec
- Skip-gram Model
- CBOW Model
- GloVe
- fastText

Usage



fastText



دانلود و کار با Fasttext در فایل Google Colab NoteBooks

if you Want to know more then read [this article from Thomas Wolf](#) but now we gonna use Fasttext because it's Pretrained by Facebook and we can use it (there is nothing to worry about this model it's pretty easy to train it by your self or your corpus facebook used Persian Wikipedia and some other staff as dataset for this model so it's just very simpler for us 🤓)

```
#@title Download, extract and load Fasttext word embedding model

!wget https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.fa.300.bin.gz
!gunzip /content/cc.fa.300.bin.gz
!pip install fasttext

import fasttext

%time
model = fasttext.load_model("/content/cc.fa.300.bin")
```

--2020-10-18 09:09:04-- <https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.fa.300.bin.gz>
Resolving dl.fbaipublicfiles.com (dl.fbaipublicfiles.com)... 104.22.75.142, 172.67.9.4, 104.22.74.142, ...
Connecting to dl.fbaipublicfiles.com (dl.fbaipublicfiles.com)|104.22.75.142|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 4502524724 (4.2G) [application/octet-stream]
Saving to: 'cc.fa.300.bin.gz'

cc.fa.300.bin.gz 100%[=====>] 4.19G 17.2MB/s in 4m 12s

Dowr

تنظیمات اولیه شبکه عصبی در فایل Google Colab NoteBooks

Prepare Train & Test Data

تنظیم طول بردار کلمه (vector_size) و
بیشترین تعداد دسته بندی مجاز برای
کلمات (max_no_tokens)

```
▶ #@title Prepare Train & Test Data
vector_size = 300 #@param {type:"integer"}
max_no_tokens = 20 #@param {type:"integer"}
import numpy as np
import keras.backend as K
train_size = int(0.9*(len(revlist_shuffle)))
test_size = int(0.1*(len(revlist_shuffle)))

indexes = set(np.random.choice(len(revlist_shuffle), train_size + test_size, replace=False))

x_train = np.zeros((train_size, max_no_tokens, vector_size), dtype=K.floatx())
y_train = np.zeros((train_size, 2), dtype=np.int32)

x_test = np.zeros((test_size, max_no_tokens, vector_size), dtype=K.floatx())
y_test = np.zeros((test_size, 2), dtype=np.int32)
```


داده های تست و آموزش

```
▶ #@title Prepare Train & Test Data
vector_size = 300 #@param {type:"integer"}
max_no_tokens = 20 #@param {type:"integer"}
import numpy as np
import keras.backend as K
train_size = int(0.9*(len(revlist_shuffle)))
test_size = int(0.1*(len(revlist_shuffle)))
```

90% of Data for Train

10% of Data for test


Fill X_Train, X_Test, Y_Train, Y_Test with Dataset

```
#@title Fill X_Train, X_Test, Y_Train, Y_Test with Dataset
for i, index in enumerate(indexes):
    text_words = hazm.word_tokenize(revlist_shuffle[index][0])
    for t in range(0, len(text_words)):
        if t >= max_no_tokens:
            break

        if text_words[t] not in model.words:
            continue
        if i < train_size:
            x_train[i, t, :] = model.get_word_vector(text_words[t])
        else:
            x_test[i - train_size, t, :] = model.get_word_vector(text_words[t])

    if i < train_size:
        y_train[i, :] = [1.0, 0.0] if revlist_shuffle[index][1] == 3 else [0.0, 1.0]
    else:
        y_test[i - train_size, :] = [1.0, 0.0] if revlist_shuffle[index][1] == 3 else [0.0, 1.0]

x_train.shape, x_test.shape, y_train.shape, y_test.shape
```

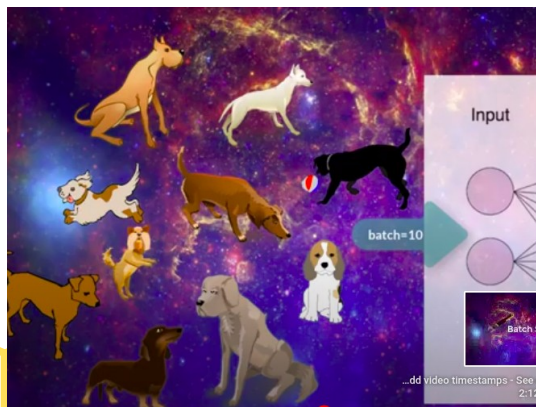
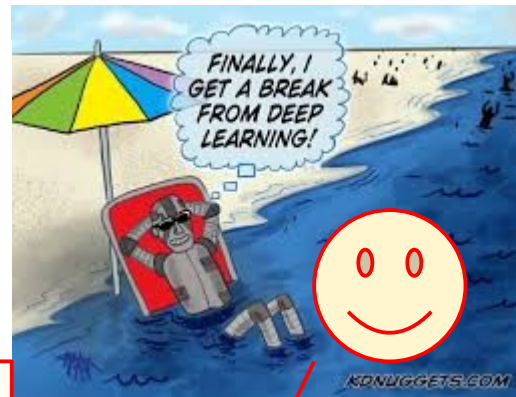
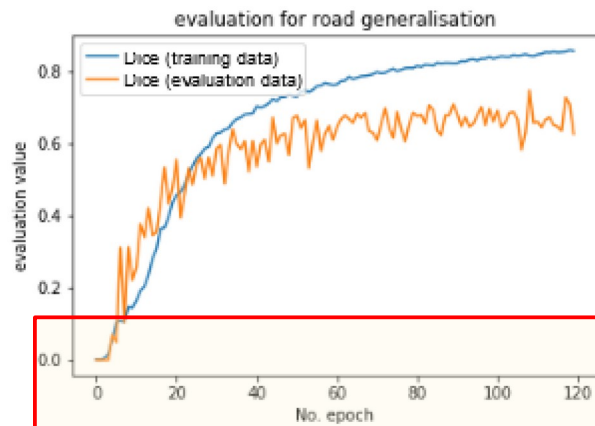


Epochs & batch_size

Step C) Preparing LSTM model

Now we will create our LSTM model then feed it our Train data and b

```
#@title Set batchSize and epochs
batch_size = 500 #@param {type:"integer"}
no_epochs = 200 #@param {type:"integer"}
w2v_model = model
del model
```



(a)

دوره‌ها (Epochs) نشانگر تعداد دفعاتی هستند که مجموعه داده کامل به شبکه عصبی مصنوعی پاس داده خواهد شد. $batch_size$ تعداد داده‌های نمونه‌ای است که در یک دوره به کار می‌رود

لایه های شبکه عصبی

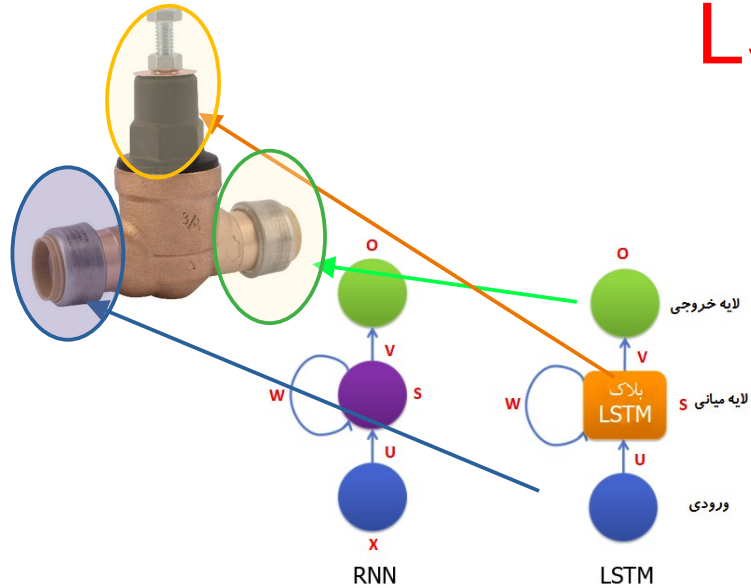
- در این پروژه لایه های متفاوتی به صورت زیر به کار رفته است:
- Conv1D, Dropout, Dense, Flatten, **LSTM**, MaxPooling1D, Bidirectional

```
##title Building LSTM Model
from keras.models import Sequential
from keras.layers import Conv1D, Dropout, Dense, Flatten, LSTM, MaxPooling1D, Bidirectional
from keras.optimizers import Adam
from keras.callbacks import EarlyStopping, TensorBoard
```

Building LSTM Model

یک واحد LSTM

یک واحد LSTM مشترک از موارد زیر میباشد:



(1) یک سلول

(2) یک گیت ورودی

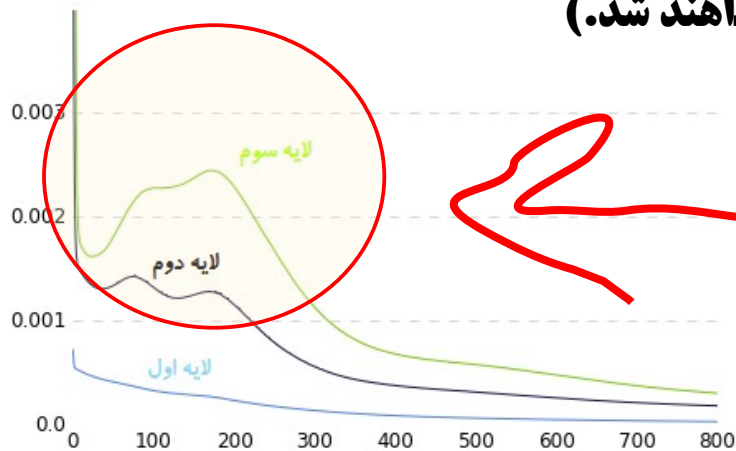
(3) یک گیت فراموشی

(4) یک گیت خروجی

گیت ورودی مشخص می‌کند که کدام اطلاعات باید وارد وضعیت سلول شوند . گیت فراموشی ، مشخص می‌کند که **کدام اطلاعات باید از حالات سلول قبلی حذف شوند** تا فقط اطلاعات مربوطه را حفظ کنند . گیت خروجی مشخص می‌کند که چه مقدار از وضعیت داخلی باید در معرض لایه‌های بالاتر قرار گیرد .

معماری لایه پیچیده LSTM

- در یک شبکه عصبی معمولی تمام ورودی ها و خروجی ها مستقل از یکدیگر هستند. در صورتی که شبکه بتواند روابط بین کلمات را یاد بگیرد مسلماً نمی تواند کلمه بعدی را به درستی پیش بینی کند.
- بزرگترین ویژگی LSTM امکان یادگیری وابستگی بلند مدت است که توسط شبکه های عصبی بازگشتی امکان پذیر نبود.
- امکان وقوع مشکل انفجار گرادیان کماکان در این نوع شبکه ها وجود دارد. (پارمترهای ناصحیح لایه اول منجر به تحت تاثیر قرار گرفتن سایر لایه ها و انفجار گرادیان خواهند شد).



تنظیمات لایه ها برای شبکه عصبی در فایل Google Colab NoteBooks

+ Code + Text Copy to Drive

no_epochs: 200

```
#@title Building LSTM Model
from keras.models import Sequential
from keras.layers import Conv1D, Dropout, Dense, Flatten, LSTM, MaxPooling1D, Bidirectional
from keras.optimizers import Adam
from keras.callbacks import EarlyStopping, TensorBoard

model = Sequential()

model.add(Conv1D(32, kernel_size=3, activation='elu', padding='same',
                 input_shape=(max_no_tokens, vector_size)))
model.add(Conv1D(32, kernel_size=3, activation='elu', padding='same'))
model.add(Conv1D(32, kernel_size=3, activation='relu', padding='same'))
model.add(MaxPooling1D(pool_size=3))

model.add(Bidirectional(LSTM(512, dropout=0.2, recurrent_dropout=0.3)))

model.add(Dense(512, activation='sigmoid'))
model.add(Dropout(0.2))
model.add(Dense(512, activation='sigmoid'))
model.add(Dropout(0.25))
model.add(Dense(512, activation='sigmoid'))
model.add(Dropout(0.25))

model.add(Dense(2, activation='softmax'))

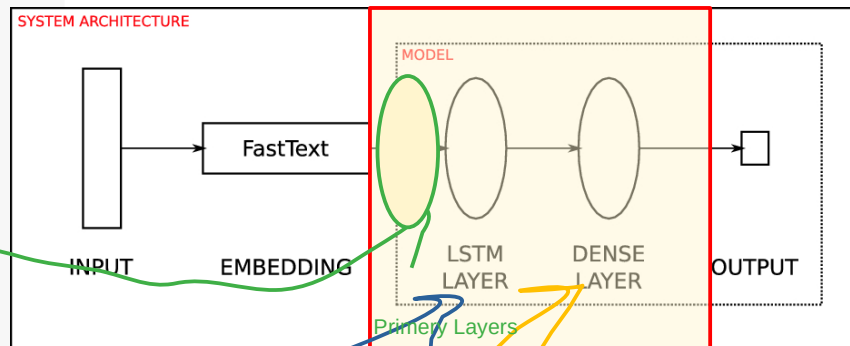
model.compile(loss='categorical_crossentropy', optimizer=Adam(lr=0.0001, decay=1e-6), metrics=['accuracy'])

# tensorboard = TensorBoard(log_dir='logs/', histogram_freq=0, write_graph=True, write_images=True)

model.summary()
```

Building LSTM Model

وارد کردن لایه های مورد استفاده در شبکه عصبی



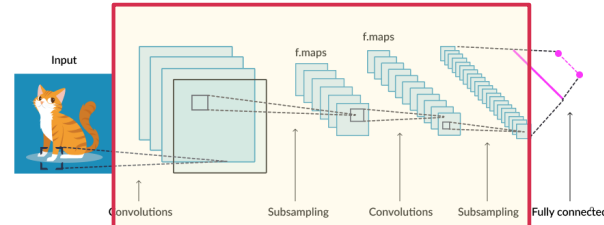
خروجی تنظیمات لایه ها برای شبکه عصبی در فایل Google Colab NoteBooks

WARNING:tensorflow:Layer lstm will not use cuDNN kernel since it doesn't meet the cuDNN kernel criteria. It will use generic GPU kernel as fallback when running on GPU
WARNING:tensorflow:Layer lstm will not use cuDNN kernel since it doesn't meet the cuDNN kernel criteria. It will use generic GPU kernel as fallback when running on GPU
WARNING:tensorflow:Layer lstm will not use cuDNN kernel since it doesn't meet the cuDNN kernel criteria. It will use generic GPU kernel as fallback when running on GPU
Model: "sequential"

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 20, 32)	28832
conv1d_1 (Conv1D)	(None, 20, 32)	3104
conv1d_2 (Conv1D)	(None, 20, 32)	3104
max_pooling1d (MaxPooling1D)	(None, 6, 32)	0
bidirectional (Bidirectional)	(None, 1024)	2232320
dense (Dense)	(None, 512)	524800
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 512)	262656
dropout_1 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 512)	262656
dropout_2 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 2)	1026
Total params: 3,318,498		
Trainable params: 3,318,498		
Non-trainable params: 0		

لایه های شبکه عصبی به
ترتیب

Output Shape
(None, 20, 32)
(None, 20, 32)
(None, 20, 32)
(None, 6, 32)
(None, 1024)
(None, 512)
(None, 512)
(None, 512)
(None, 512)
(None, 512)
(None, 512)
(None, 512)
(None, 2)



ابعاد ماتریس لایه ها

تعداد پارامترهای متغیر در
هر لایه !!

هشدار های قابل
توجه در رابطه با
عدم انجام برخی
تنظیمات و بالا
بودن بار مصرفی
CPU در شبکه

رشد شبکه ساخته شده با داده های تست و آموزش در فایل Google Colab NoteBooks

خروجی اجرای دستورات (قابل توجه
برای تنظیمات بهتر جهت **دقت بیشتر**
خروجی ها و رفع **خطاهای احتمالی**

تمرین شبکه عصبی با داده های Train (epochs, batch_size)

تمرین شبکه عصبی با بلادرده های **Test** (**epochs**, **batch_size**)

ذخیره مدل برای اجرای در زمان ورودی های جدید مانند درخواست کاربر برای

نتایج و مقایسه (LSTM & Word2vec)

نتایج تجربی اثربخشی مدل پیشنهادی را با دقت ۸۵٪ نشان می‌دهد.

A COMBINED DEEP LEARNING MODEL FOR PERSIAN SENTIMENT ANALYSIS

ZAHRA BOKAEE NEZHAD¹ AND MOHAMMAD ALI DEIHIMI²

¹Department of Information and Computer Science, Zand university, Shiraz, Iran.

²Department of Electrical and Electronics Engineering, Bahonar University, Shiraz, Iran.

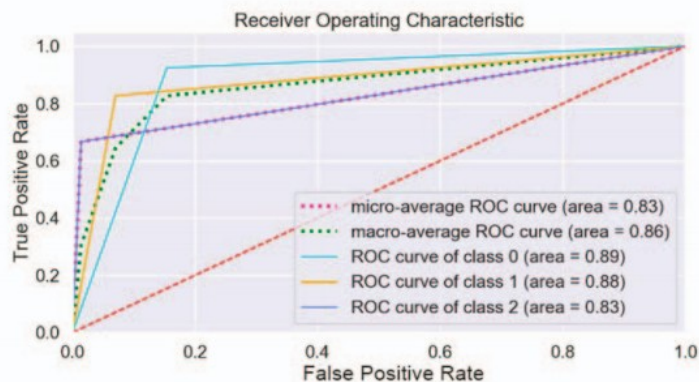


Fig. 5: ROC curve.

Table 2: Comparative results of different methods on Persian dataset

Model	F-Score	Precision	Recall
Model1: NBSVM	44%	70.7%	31.9%
Model 2: Naïve Bayes	62.5%	65.4%	61.7%
Model 3: logistic regression	67.1%	68.9%	65.4%
Model 4: Bidirectional-LSTM	53.8%	54.3%	55%
Model 5: CNN	67%	66%	67%
Proposed Model: CNN-LSTM	85%	86.5%	86%
CNN-GRU	85.8%	85%	82%
GRU	74%	77.5%	75.6%

مقایسه با برخی از سایر روشهای بکار رفته در تحلیل احساسات در زبان انگلیسی

جدول زیر مقایسه ای بین دقت مدل‌های مختلف شبکه عصبی برای پردازش کلمات نشان می دهد:

	Travel Scheduling	Ask Ubuntu	Web Applications
NB	.959	.726	.502
SVM FastText average	.958	.812	.771
SVM FastText sum	.968	.800	.658
SVM GloVe average	.946	.805	.591
SVM GloVe sum	.957	.729	.692
SVM Word2Vec average	.979	.742	.698
SVM Word2Vec sum	.946	.742	.680
LSTM FastText	.968	.644	.465
BiLSTM FastText	.979	.646	.549
LSTM GloVe	.945	.665	.546
BiLSTM GloVe	.979	.667	.635
LSTM Word2Vec	.989	.631	.395
BiLSTM Word2Vec	.989	.710	.443
LUIS	.979	.743	.690
Watson	.968	.819	.630
APLai	.931	.782	.628
RASA	.979	.708	.494

Intent Classification for Dialogue Utterances

همچنین امکان استفاده از مدل های دیگر، شبکه عصبی مانند GRU نیز مطابق معماری هایی از قبیل ارایه شده در این مقاله و عکس های زیر، نیز برای این کار، موجود می باشد:

پروژه های مشابه

- A-Determination-of-Effective-Word-Embedding-of-Bengali-Language-using-text-classification

```
In [0]: bengali_news_after_preprocessing.tail()
```

Out[0]:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
39995	অতল	অবোধের	করণে	ইউরোপের	ক্ষেত্রে	জ্যেট	আরও	অন	ফায়ার	আন্ত	বিভিৎ	সেফট	ঈ	বাংলাদেশের	পোশাক	করাধনা	পা
39996	সড়কসড়ক	ছোট	বড়	অসং	গত	পিচ	লোপাই	অভিহি	দ্ব্যাবাস	উভয়ে	সমানে	গাড়ি	লোচল	মুহ	ওয়েট	চলচল	ক
39997	দুবিব	বাং	দেশ	বেহ	সিবেছিল	কিলোগের	বাগাশির	মফিল	বাপ্তুত	একু	কিলোগের	আবেগ	অবৃত্তিক	আগম	মত	বশি	বা
39998	তেরুখবি	ইন্টিনাস	অধ্যাপক	আমিসুজ্ঞানসে	তম	জন্মাবিসি	উপলক্ষে	শফির	আতীয়	জাফর	মূল	সিদ্ধান্ত	সংকর্ষা	অপ্টেল	ওভেল	জানতে	অ
39999	টাকা	হেতালের	করণে	বতর	পাইকারি	মোকাম	মহাশাল	বাগার	ঢাকা	টাইগ্রাস	সারা	দেশে	সবজির	সবজির	বহ	পাইকারি	বা

E reane o 1974 redunant

```

translate_map = dict.fromkeys(translate_map)
wordlist = sentence.translate(translate_map).split()
global c,x;
c+=1
if c>X:
    print(x,end=' ')
    x=100
return list(filter(lambda x: x not in stops, wordlist))

```

```
!tk.word_tokenize(text) for text in title]
```

```
[sentence to wordlist(text) for text in content]
l.Dataframe(filter text)
pickle('/content/gdrive/My Drive/Projects/Bengali Text Classification/Bengali_Text_after_preprocessing.pkl')
```

eprocessed pickle text file from drive. Change the path file according to local path file.

```
after_preprocessing = pd.read_pickle('/content/gdrive/My Drive/Projects/Bengali Text Classification/Bengali_Text_preprocessing.pkl')
```

```
after_preprocessing.head()
```

[illegible]

5 rows x 42/1 column

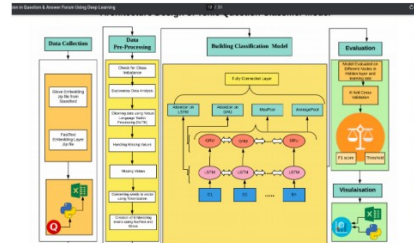
```
In [0]: bengali_news_after_preprocessing.tail()
```

Out[0]:

[illegible]

5 rows x 4271 columns

- **GRU** یک مکانیزم **gating** در **RNN**، شبیه **LSTM**، با پارامترهای کمتر - <آموزش آسان می باشد.



پایان

با تشکر از:

توجه و نظرات شما

