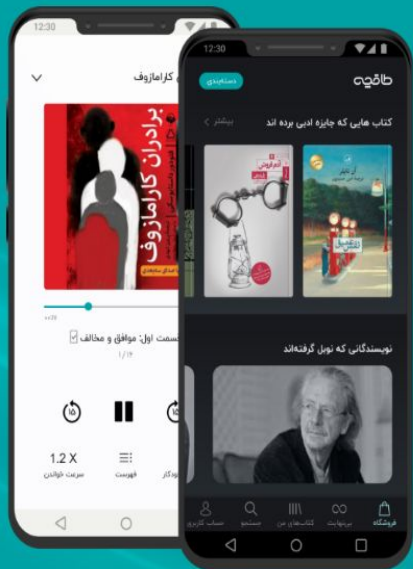

Sentiment Analysis

Using Ensemble of CNN and Bi-LSTM Models



August 2020



لذت خواندن و شنیدن کتاب‌ها

دانلود قانونی بیش از ۵۵۰۰۰ کتاب متنی و صوتی

نصب اپلیکیشن طاقچه



دلار

تومان

نمایش قیمت‌ها در فروشگاه

Dataset

	date	comment	bookname	rate	bookID	like
	Comment published date	Text of comment	The name of the book whose comment is recorded	User rating to the book	One ID is assigned to each book	The num who like
	<p>1398/09/01 1%</p> <p>1398/09/02 1%</p> <p>Other (1850) 98%</p>	<p>عاشی 1%</p> <p>عاشی بود 1%</p> <p>Other (64187) 98%</p>	<p>11006</p> <p>unique values</p>			
1	1395/11/14	<p>No one writes to the Colonel ترجمش همیشه کسی به سرهنگ نامه نمی‌نویسد. مترجمی که اسم کتاب لشکر ترجمه می‌کند دیگه چه توقعی از ترجمه ی من کتاب بده داشته بشیم. این هم کار... مارکز که هست جریان</p>	سرهنگ کسی ندارد. برایش نامه بنویسد	8.0	3	
2	1395/11/14	<p>مقاله عزیزانم کتاب "کسی به سرهنگ نامه نمی‌نویسد" است بسیار من شادمانم! خدا تصحیح کند</p>	سرهنگ کسی ندارد. برایش نامه بنویسد	5.0	3	
3	1394/06/06	<p>بظنرم این اثر مارکز خیلی از همه سال تلوایی که بخاطرش نوشتن ادبیات گرفت. زیاده و روان تره. این داستان یاد شاهکاره بی نظیره. سرگشت سرهنگی که به جویزه و جایزه و همدانیه. به خورایی بد مای جان... دلیشو</p>	سرهنگ کسی ندارد. برایش نامه بنویسد	5.0	3	
4	1393/09/02	<p>به نظر کتاب خوبی میوه اما من از ترجمش خورتم نبود. لطفاً طوری نبود که کتاب تا آخر</p>	سرهنگ کسی ندارد. برایش نامه بنویسد	2.0	3	

70K Comments



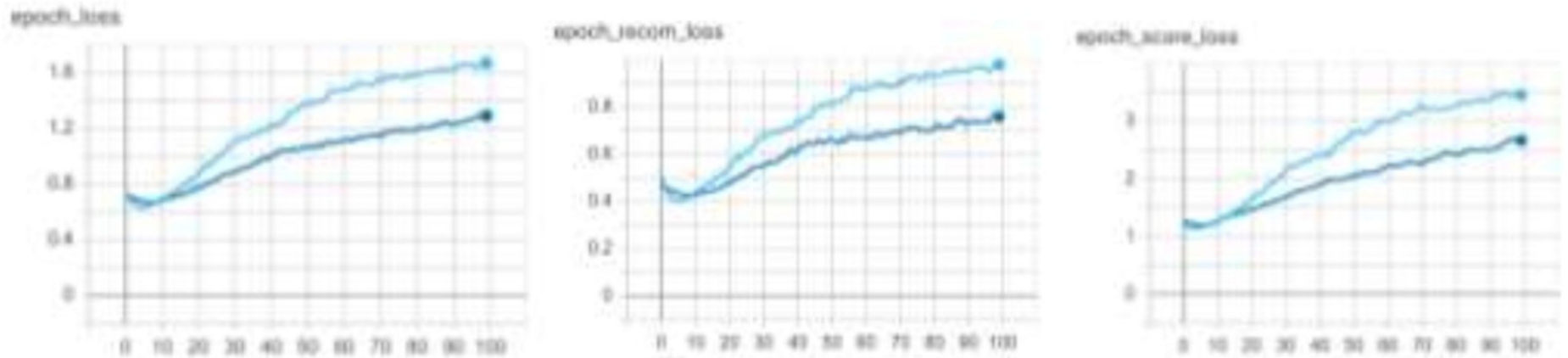
1. Algorithm Structure

- **Pre-processing**
- **Embedding**
Mapping words to lower dimension
- **CNN**
Convolutional neural network to feature extraction
- **LSTM**
Recurrent neural network to use of sequential information
- **Ensemble**
A methods to use multiple learning algorithms to obtain better result

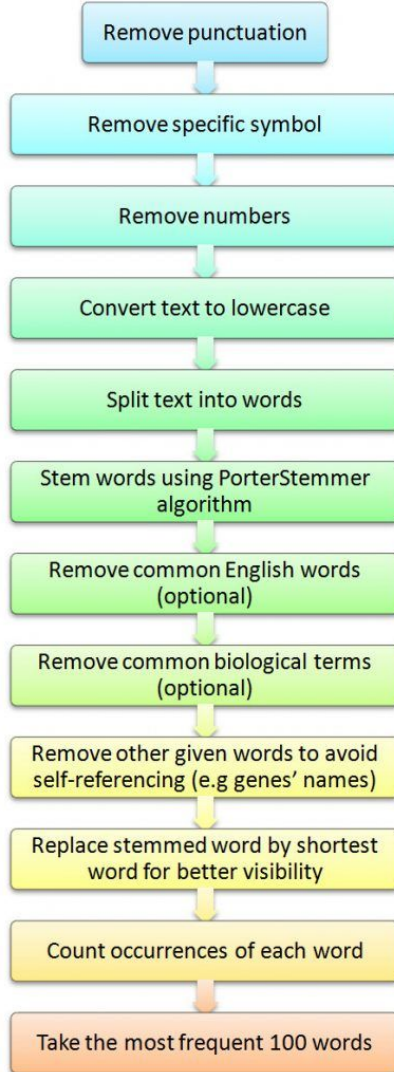
مترجمی که اسم کتاب اشتباه ترجمه میکنه
دیگه چه توقعی از ترجمه ی متن کتاب باید
. اون هم آثار مارکز که همش 😊 داشته باشیم
جریان سیال ذهنه و ترجمش کار هر کسی
😊🙄🙄 نیست



Emoji-vector reduce loss



Light-blue: single text input; dark-blue: text input concatenated with emoji-vector



Pre-Processing

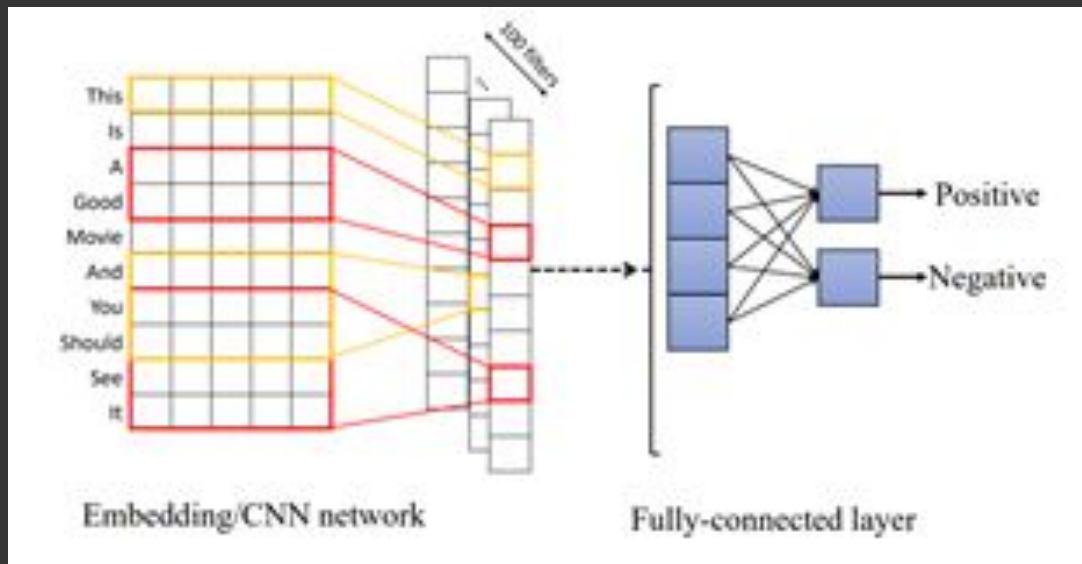
```
>>> normalizer = Normalizer()
>>> normalizer.normalize('اصلاح نویسه ها و استفاده از نیمفاصله پردازش را آسان می کند')
'اصلاح نویسه ها و استفاده از نیمفاصله پردازش را آسان می کند'

>>> sent_tokenize('ما هم برای وصل کردن آمدیم! ولی برای پردازش، جدا بهتر نیست؟')
['ما هم برای وصل کردن آمدیم!', 'ولی برای پردازش، جدا بهتر نیست?']
>>> word_tokenize('ولی برای پردازش، جدا بهتر نیست؟')
['ولی', 'برای', 'پردازش', 'و', 'جدا', 'و', 'بهتر', 'و', 'نیست', 'و', '?']

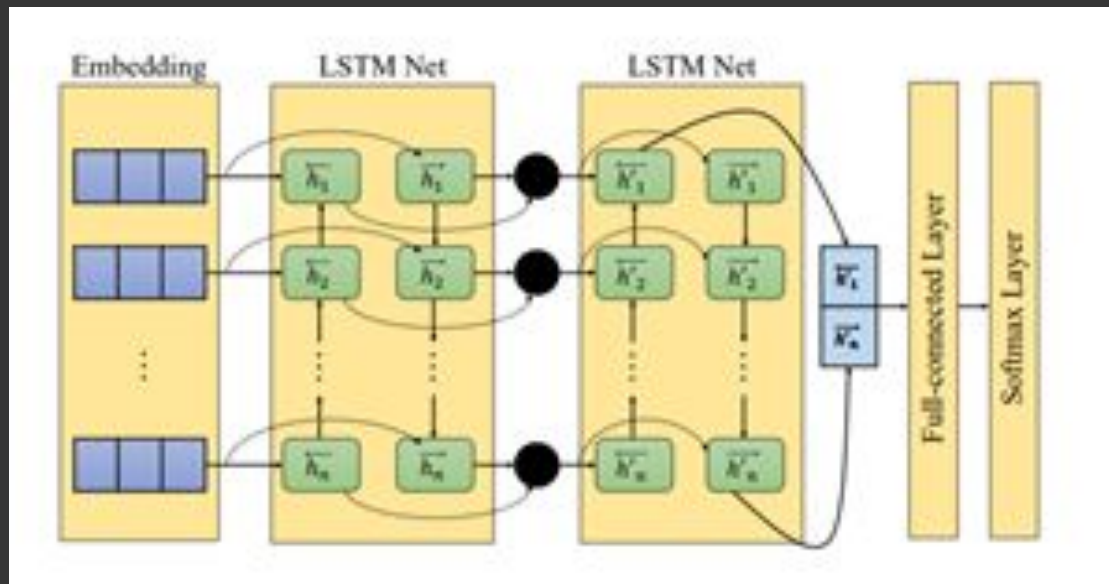
>>> stemmer = Stemmer()
>>> stemmer.stem('کتابها')
'کتاب'

>>> lemmatizer = Lemmatizer()
>>> lemmatizer.lemmatize('میروم')
'رفت#رو'
```

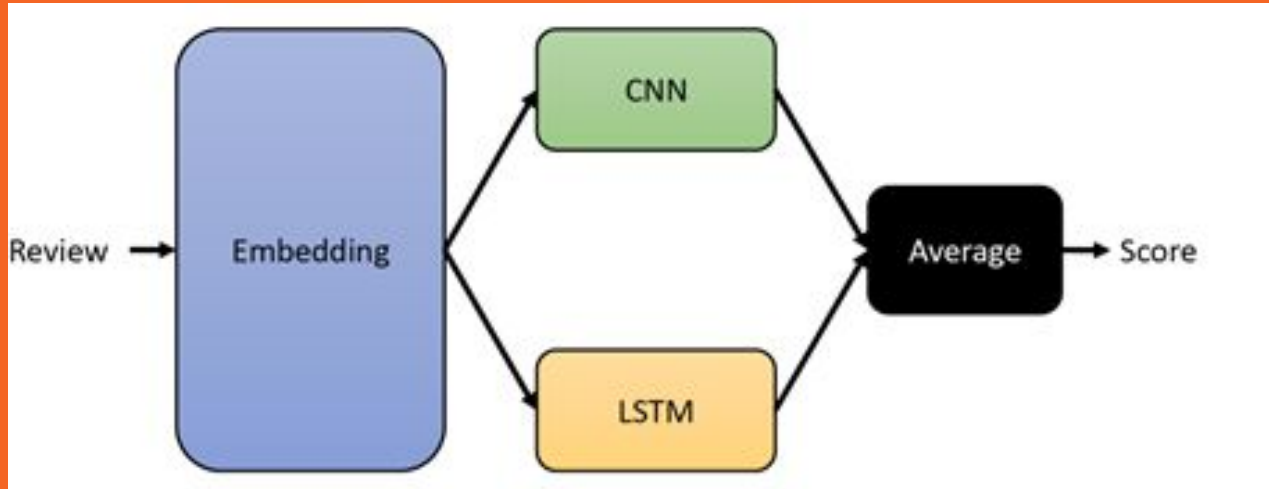
CNN



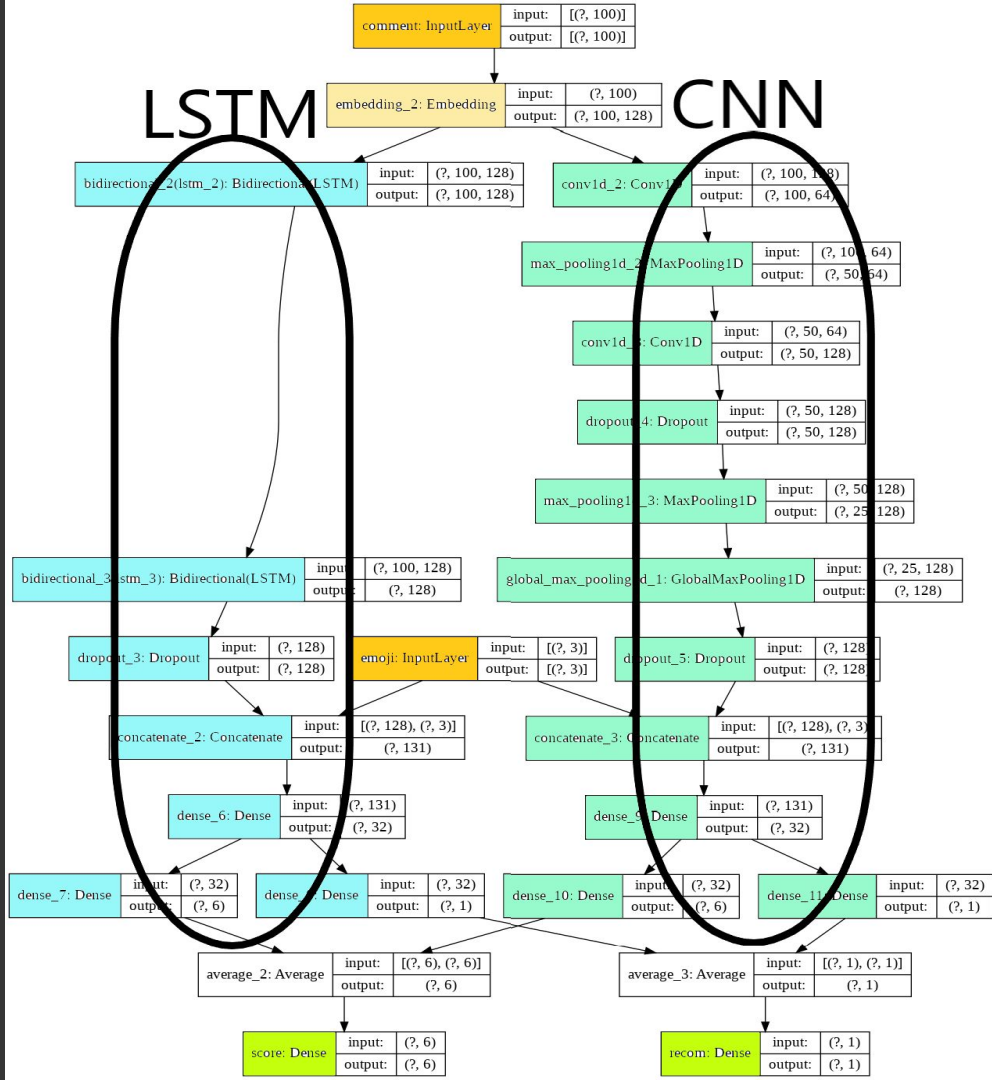
LSTM



Ensemble



Model



Inputs

→ Comment

مترجمی که اسم کتاب اشتباه ترجمه میکنه دیگه چه <<
توقعی از ترجمه ی متن کتاب باید داشته باشیم. اون هم
آثار مارکز که همش جریان سیال ذهنه و ترجمش کار
>> 😞 هرکسی نیست

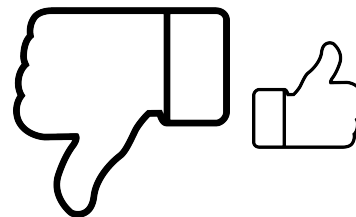
→ Emojie vector

Sense vector: ([3, 61, 108])



Outputs

→ Is recommended or not



→ 5-star rating












Emoji Vector

Emoji Sentiment Analysis

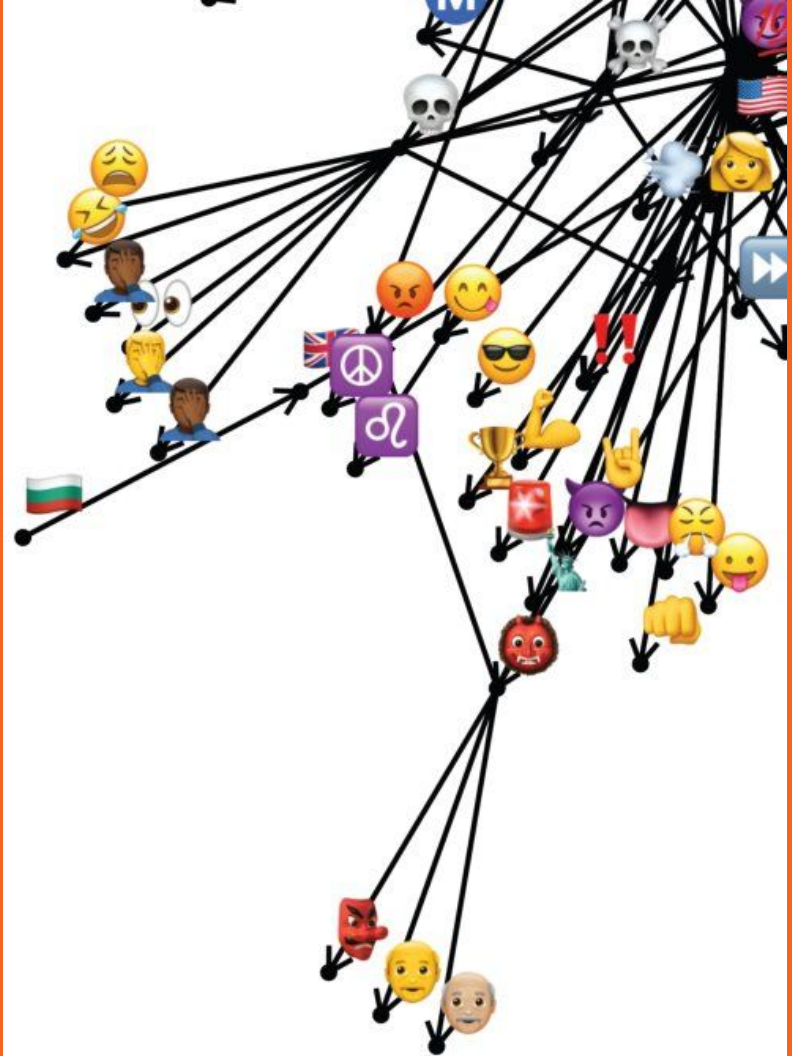
- Emoji sentiment data
- 70k labeled tweets
- weighted average

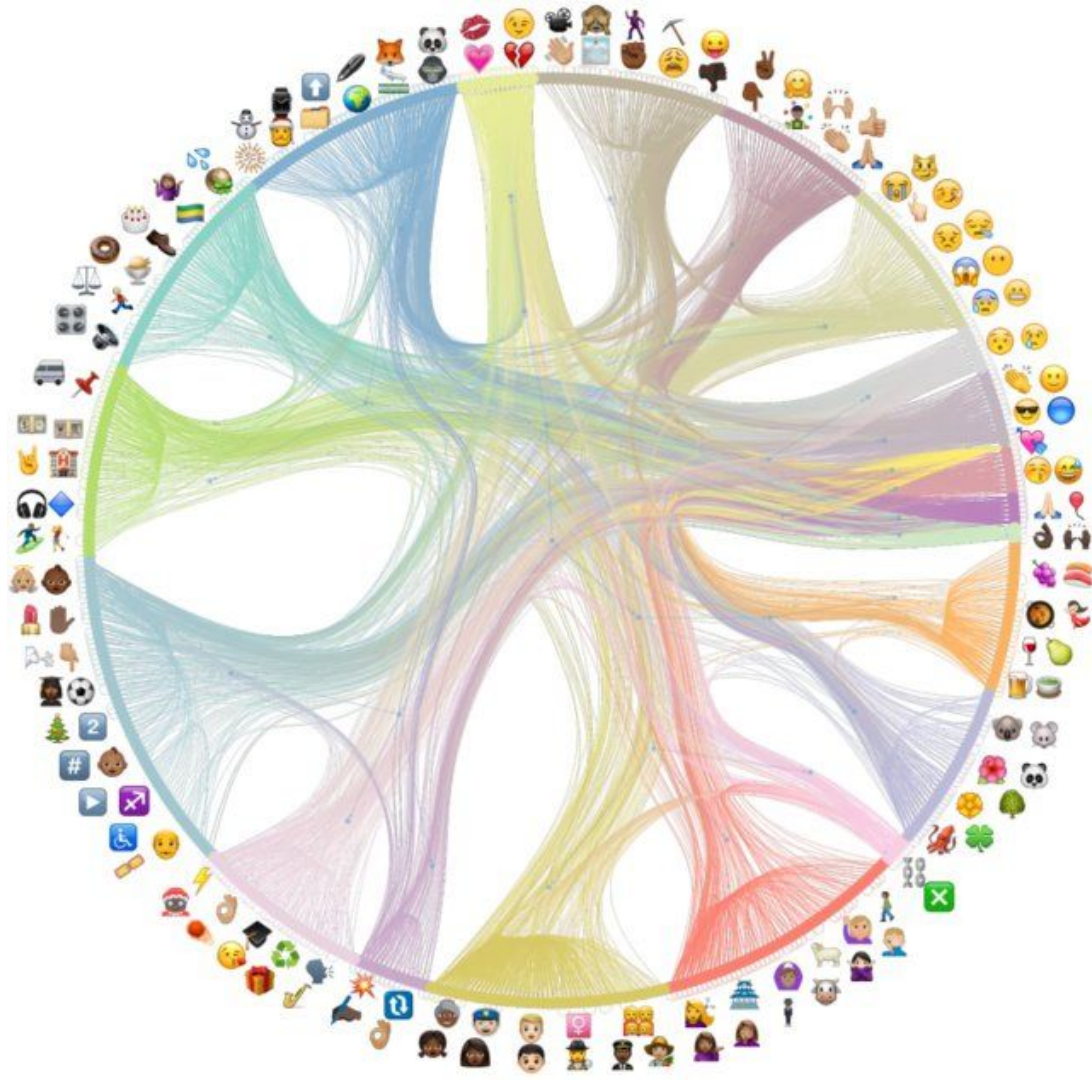
- Emoji vector:   

Image [tweemoji]	Unicode codepoint	Occurrences [5...max]	Position [0...1]	Neg [0...1]	Neut [0...1]	Pos [0...1]	Sentiment score [-1...+1]	Sentiment bar (c.i. 95%)	Unicode name
	0x1f602	14622	0.805	0.247	0.285	0.468	0.221		FACE WITH TEARS OF JOY
	0x2764	8050	0.747	0.044	0.166	0.790	0.746		HEAVY BLACK HEART
	0x2665	7144	0.754	0.035	0.272	0.693	0.657		BLACK HEART SUIT
	0x1f60d	6359	0.765	0.052	0.219	0.729	0.678		SMILING FACE WITH HEART-SHAPED EYES
	0x1f62d	5526	0.803	0.436	0.220	0.343	-0.093		LOUDLY CRYING FACE
	0x1f618	3648	0.854	0.053	0.193	0.754	0.701		FACE THROWING A KISS
	0x1f60a	3186	0.813	0.060	0.237	0.704	0.644		SMILING FACE WITH SMILING EYES
	0x1f44c	2925	0.805	0.094	0.249	0.657	0.563		OK HAND SIGN
	0x1f495	2400	0.766	0.042	0.285	0.674	0.632		TWO HEARTS
	0x1f44f	2336	0.787	0.104	0.271	0.624	0.520		CLAPPING HANDS SIGN
	0x1f601	2189	0.796	0.127	0.296	0.577	0.449		GRINNING FACE WITH SMILING EYES
	0x263a	2062	0.799	0.062	0.218	0.720	0.657		WHITE SMILING FACE

Emoji net

```
1:{7 items
  "category":string"Miscellaneous
  Symbols And Pictographs -> Emoji
  modifiers"
  "keywords":[... ]3 items
  "definition":string"The Female Guard,
  Type-6 emoji is a sequence of the ??
  Guardsman, ?? Emoji Modifier
  Fitzpatrick Type-6, ? Female Sign
  and ? Variation Select
  http://emojipedia.org/female-guard-t
  ype-6/"
  "unicode":string"U+1F482 U+1F3FF
  U+200D U+2640 U+FE0F"
  "name":string"woman guard: dark skin
  tone"
  "shortcode":NULL
  "senses":{... }3 items
```

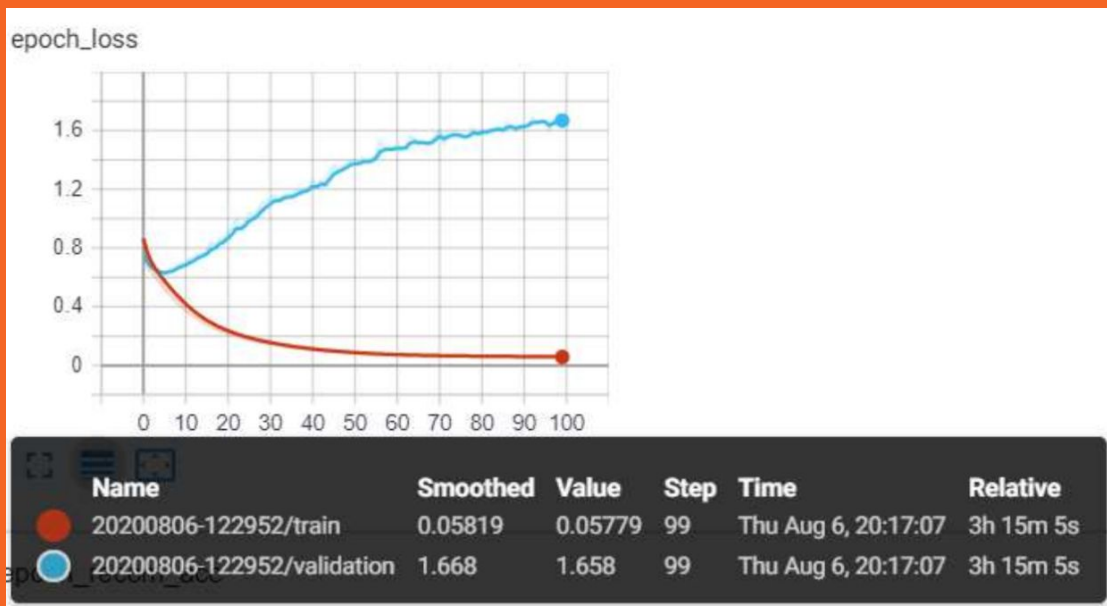




Noise comments

1. هر کس کد تخفیف داره لطفا به من بده مهدیس ۱۱ ساله
 2. میلاد حضرت علی (ع) و روز پدر پیشاپیش مبارک باد"
 3. "برای شفای عاجل تمام بیماران دعا کنیم اللهم اشفع کل مریض بحق الحسین ع
 4. لولو دروغه لالا قشنگه لالا میره با لولو بجنگه
 5. بارالها توفیق عمل به این فرامین را روزی ما بگردان آمین بحق هشت و چهارت آمین یا رب العالمین 30/تیرماه/سال جهاد اقتصادی 90__ روح الله
-

Overfitting



Final result



Improvement

1. Liked comments with more affect
 2. Pre-trained embeddings
 3. Emoji embedding
 4. Named Entity Recognition (NER)
-

Thanks
