한국정보과학회 KSC 2017

Social media based event detection using multi word embeddings and deep learning

SCLAB GROUP

Table of Contents

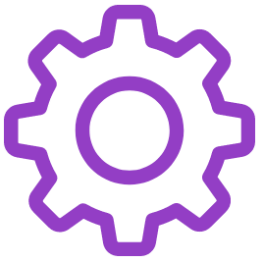
Our proposal

Motivations

Twitter is one of the most popular social networks, in where there is much data to explore useful information

Social based event detection system respond in acceptable time delay

Contributions



This work proposed a novel approach for temporal event detection by monitoring social media

Convolutional Neural Network augmented with multiple word embeddings is used to filtering noise data from raw input data

Event detection is supported by Recurrent Neural Network (LSTM) in proposed algorithm

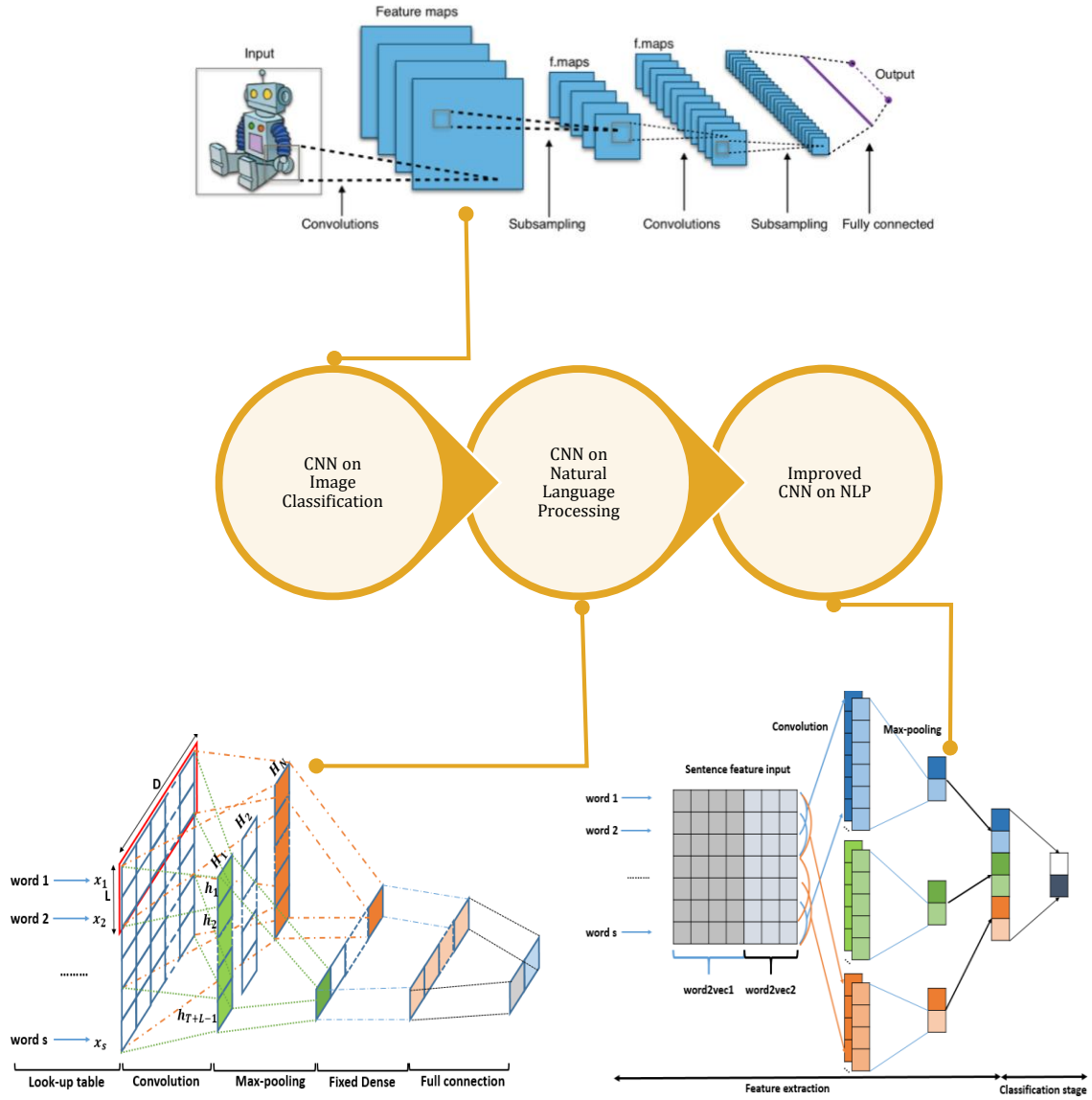
Demonstration

We implemented the demonstration to illustrate our approach

Collect data → preprocessing → identifying informative → temporal event detection

Details of Proposal

Filter informative data



Above figure illustrate applications of Convolutional Neural Network from Image Processing to Natural Language Processing. And last one is improved version, our architecture is inspired by considering different word embedding in deep CNN. We investigated to consideration of CNN approach over multiple pre-trained word embeddings with different size to support for event detection system. This phase is to filter noise messages from a collected dataset, identify informative data and moving them to next phase

Performance

CNN vs Improved CNN (AUC measure)

Disaster Name (year)	# positive	# negative	Total
Philippines floods (2012)	760	145	905
Colorado floods (2013)	768	157	925
Queensland floods (2013)	728	191	919
Manila floods (2013)	628	293	921
Chile earthquakes (2014)	1447	287	1734
Australia fires (2013)	704	245	949

Feature	*Philippines floods*	*Colorado floods*	*Queensland floods*	*Manila floods*	*Chile earthquake*	*Australia fire*
Glove (**200**)	90.02	90.73	89.47	91.65	88.9	86.87
word2vec (**300**)	90.23	91.69	90.36	88.37	88.25	87.92
Glove-word2vec(**500**)	**91.26**	**92.77**	**90.63**	**93.05**	**90.01**	**88.95**

Good points

CNN based classifier can automatically learn features without feature hand-engineering compare to traditional machine learning methods

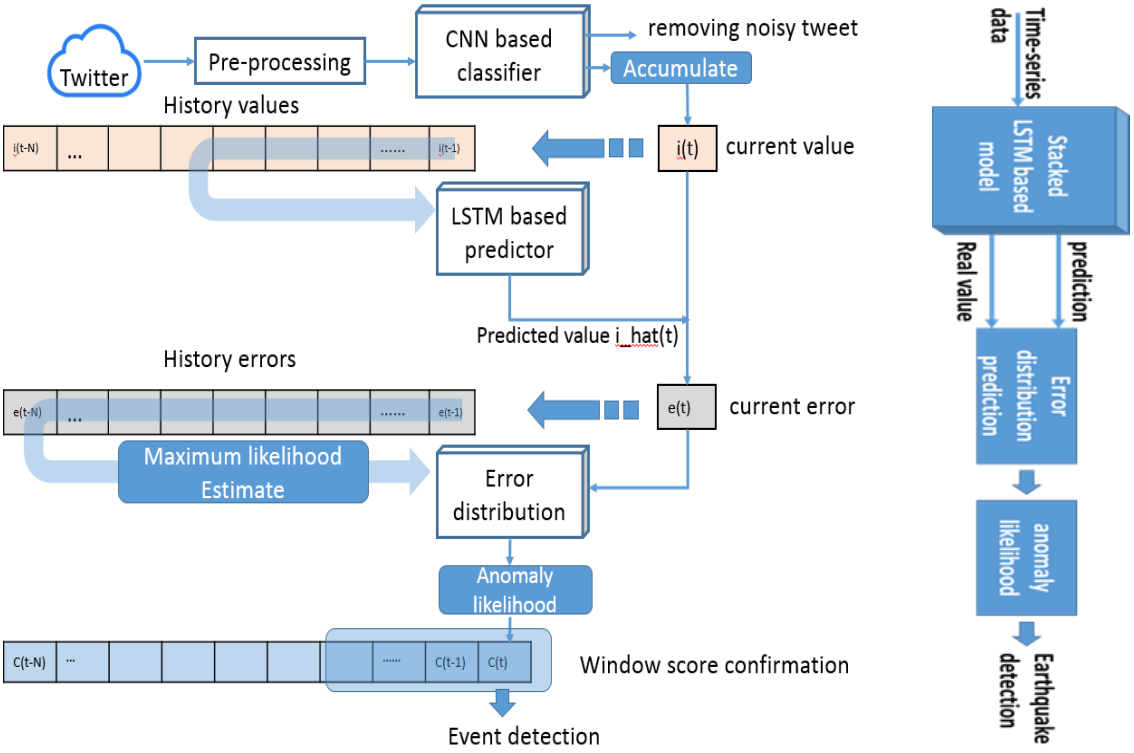
CNN on multi word-embedding overcome limitation of the previous studies in term of various size dimensionality of embeddings, flexible domain, easily implement,..

A frequent token can have more representation in input matrix or learned feature on penultimate layer instead of only one time → more available information

A rare token can be missed in some embedding version, it will be supplemented by other

Details of Proposal

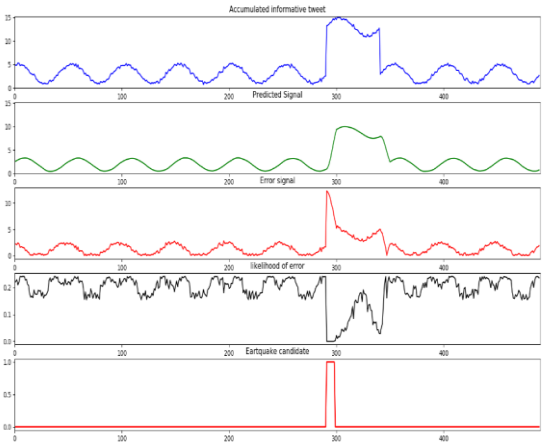
Temporal event detection



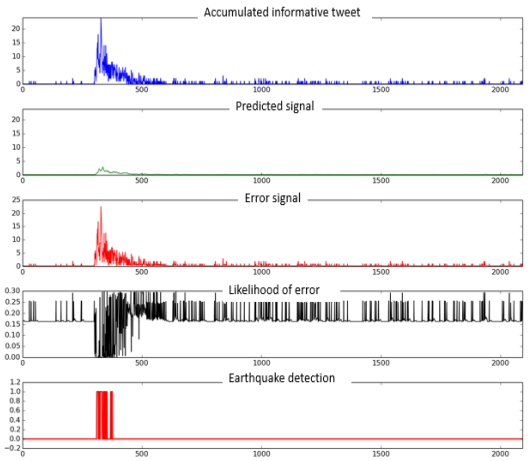
Output of CNN based model will be accumulated, what play a role as the inputs of event detection algorithm. Typically, this time series data look like underlying discrete signal. Value of each signal is accumulated frequency of tweets in given interval. The moving average can be used to filter noise in time series data. Our problem is similar to anomaly detection, novel detection, that also makes sense for the case of disaster event detection. The LSTM based prediction model refer to given current and past data to estimate next time point in the time-series. Then, the errors distribution is used to event detection model which are built by anomaly likelihood as above figure

Performance

Synthetic data



Real-world data



We conduct the experiment in some tweet dataset we collected. The true information and time response of our system are compared to prove the tolerant delay as bellows.

	Earthquake	Our proposal	Delay
Domestic earthquake (Korea)	2016/09/19 11:33:58 (UTC)	11:35:25	87sec
	2016/09/12 10:44:33 (UTC)	10:46:30	117sec
International earthquake	2016-12-27 23:20:55 (UTC)	23:22:52	117sec
	2016-12-28 08:18:00 (UTC)	08:19:16	76sec
	2016-12-28 12:38:49 (UTC)	12:39:57	68sec
	2017-06-06 18:12:28 (UTC)	18:19:14	406sec
	2017-06-08 17:01:19 (UTC)	17:02:49	90sec

Good points

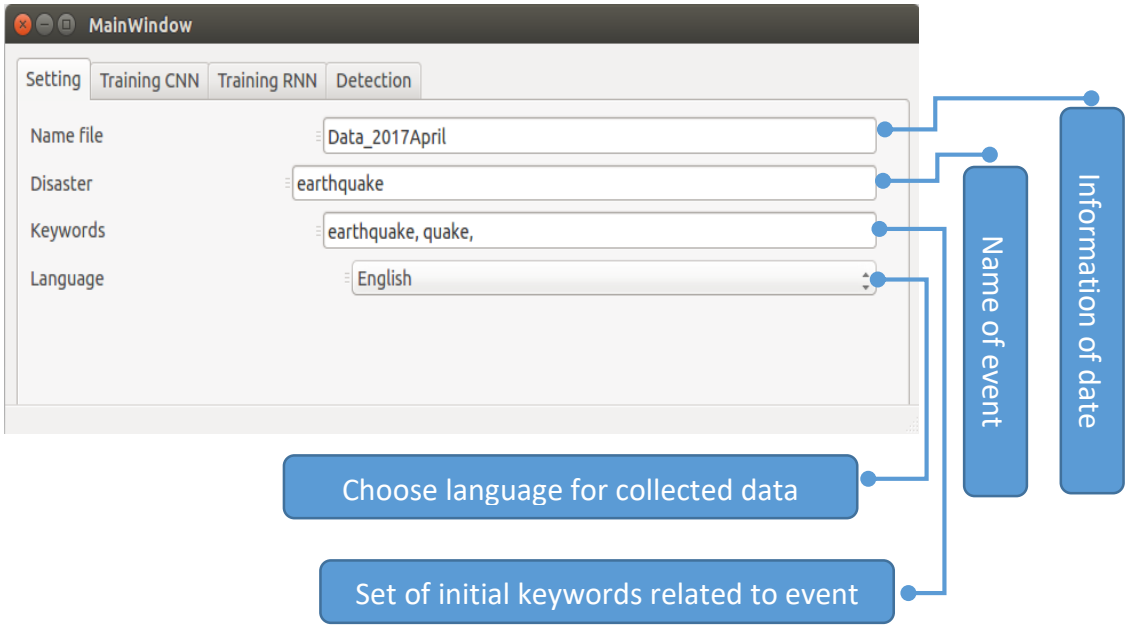
Threshold is easily predefined as compared to other methods

This event detection using LSTM can be extended to many applications

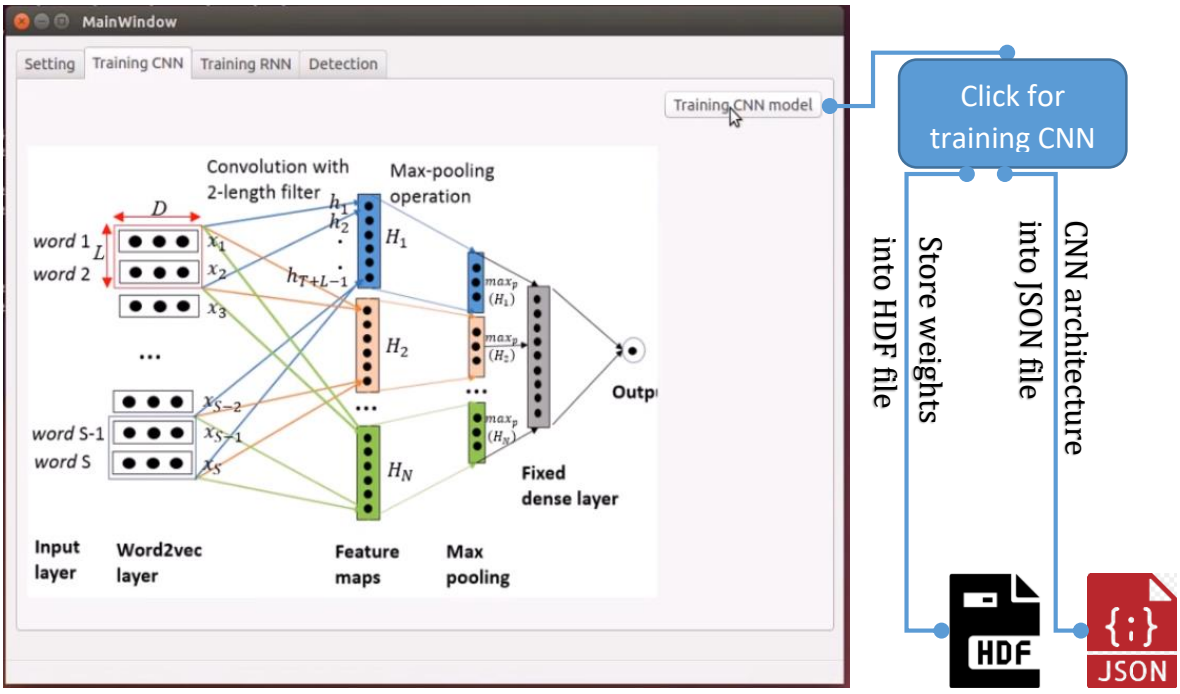
Demonstration

We introduce you step by step to run our demonstration

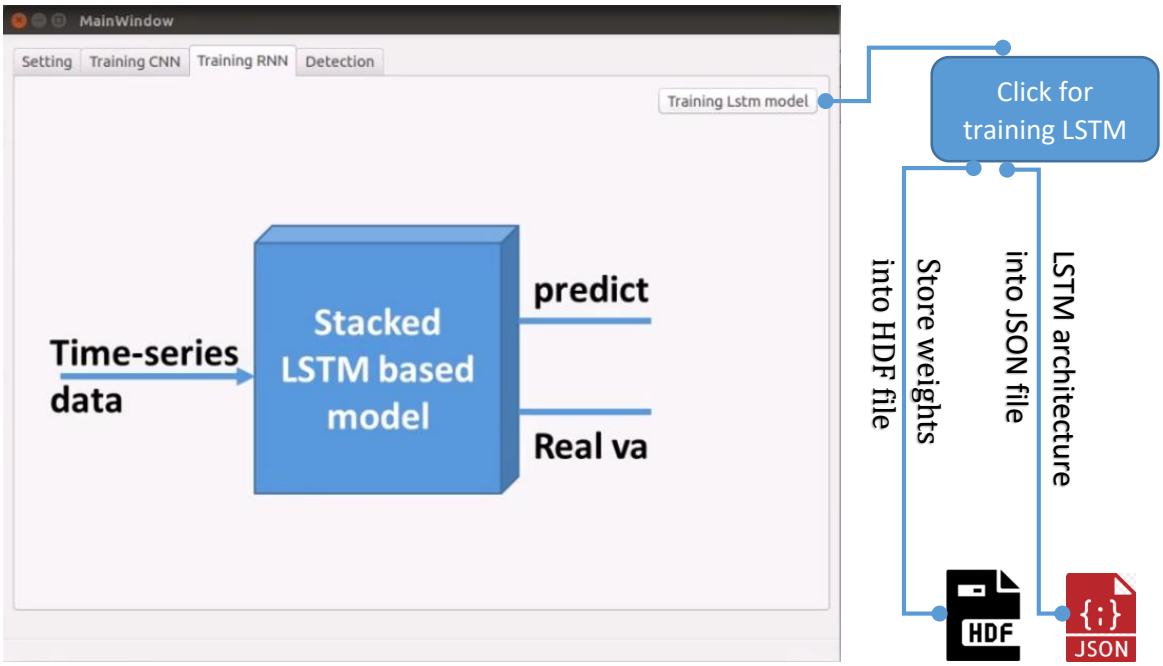
Setting



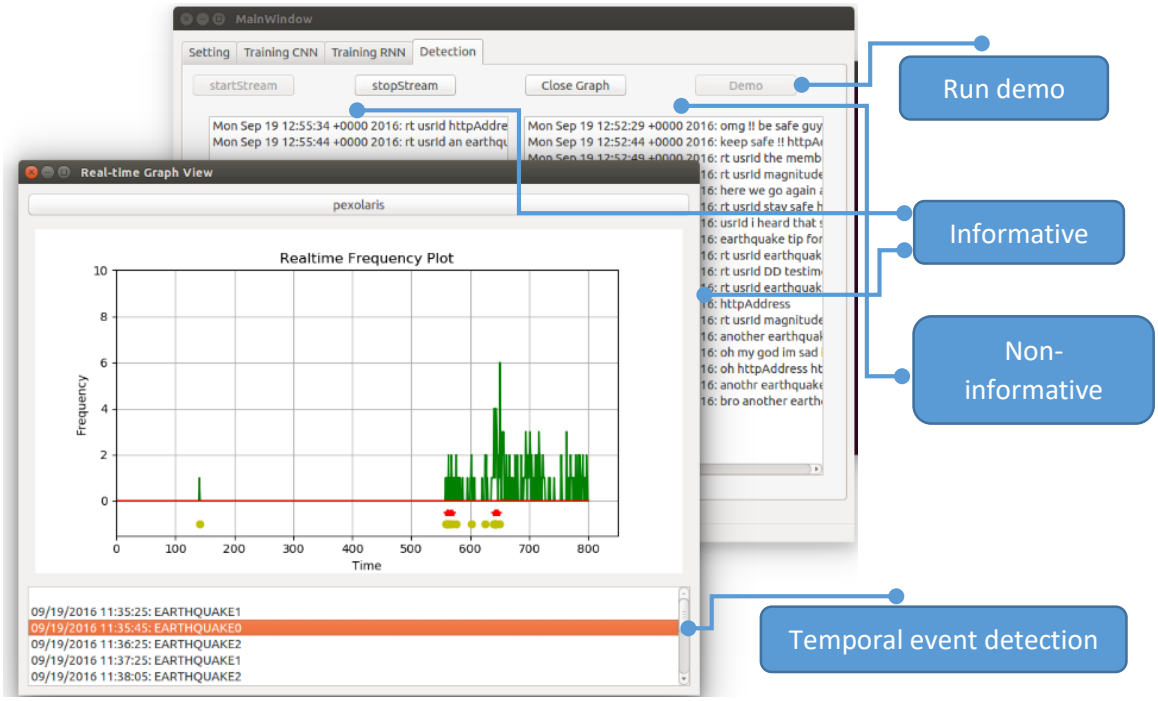
Training CNN model



Training LSTM for event detection



Running in 2 modes: Streaming or demo



We designed a GUI to illustrate our approach on the temporal detection in case of earthquake situation.

Applications

This demonstration has just been focus on Temporal event detection using Social media. Our proposal takes advantages of Deep Learning techniques.

- CNN based classifier identifies informative data from noise environment
- LSTM model used in the event detection to estimate time of occurrence

Contact Us

If you have any question, please contact us via an available email.