Advancing Spatial and Textual Analysis with GeoAI

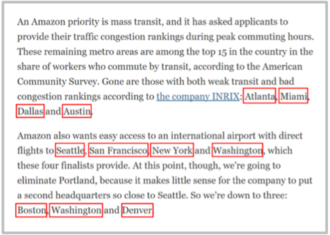
Yingjie Hu

GeoAI Lab, Department of Geography, University at Buffalo, SUNY

Spatial Tech @ UCSB

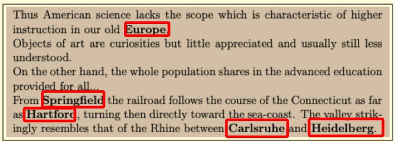
Extracting Locations from Texts

- Geoparsing: Recognizing and geo-locating locations from natural language texts

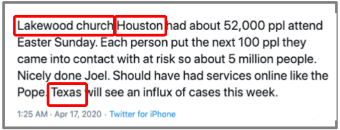

An Amazon priority is mass transit, and it has asked applicants to provide their traffic congestion rankings during peak commuting hours. These remaining metro areas are among the top 15 in the country in the share of workers who commute by transit, according to the American Community Survey. Gone are those with both weak transit and bad congestion rankings according to the company INRIX: Atlanta, Miami, Dallas and Austin.

Amazon also wants easy access to an international airport with direct flights to Seattle, San Francisco, New York and Washington, which these four finalists provide. At this point, though, we're going to eliminate Portland, because it makes little sense for the company to put a second headquarters so close to Seattle. So we're down to three: Boston, Washington and Denver

News articles


Thus American science lacks the scope which is characteristic of higher instruction in our old Europe.
Objects of art are curiosities but little appreciated and usually still less understood.
On the other hand, the whole population shares in the advanced education provided for all...
From Springfield the railroad follows the course of the Connecticut as far as Hartford, turning then directly toward the sea-coast. The valley strikingly resembles that of the Rhine between Carlsruhe and Heidelberg.

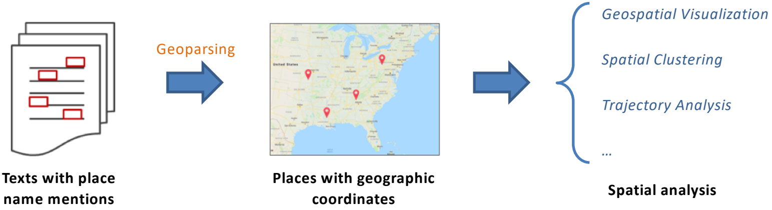
Historical archives


Lakewood church Houston had about 52,000 ppl attend Easter Sunday. Each person put the next 100 ppl they came into contact with at risk so about 5 million people. Nicely done Joel. Should have had services online like the Pope. Texas will see an influx of cases this week.
1:25 AM · Apr 17, 2020 · Twitter for iPhone

Social media messages

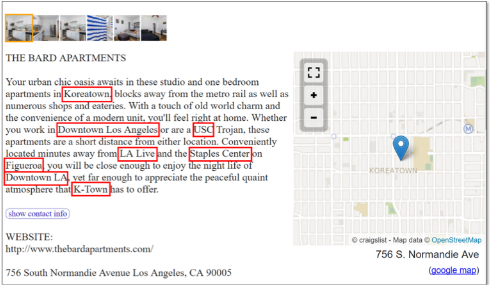
Extracting Locations from Texts

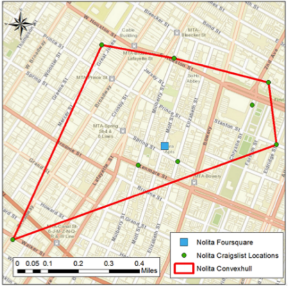
- Enabling spatial analysis on textual data



Applications of Geoparsing

- Understanding local place names

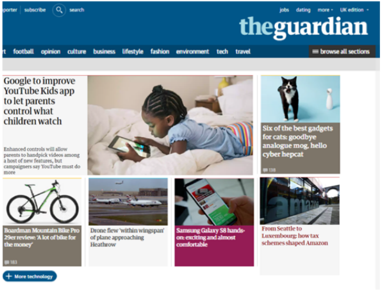


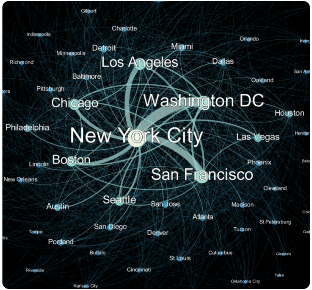


[Hu, Y., Mao, H. and McKenzie, G., 2019. *International Journal of Geographical Information Science*, 33(4), pp.714-738.]

Applications of Geoparsing

- Understanding place relations through the lens of news articles

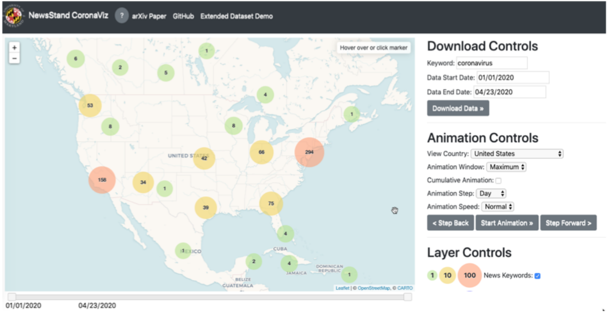




[Hu, Y., Ye, X. and Shaw, S.L., 2017. *International Journal of Geographical Information Science*, *31*(12), pp.2427-2451.]

Applications of Geoparsing

- Indexing textual documents using geographic locations



Kastner, J., Wei, H. and Samet, H., 2020. Viewing the Progression of the Novel Corona Virus (COVID-19) with NewsStand. *arXiv preprint arXiv:2003.00107*.

Applications of Geoparsing

- Location extraction for digital humanities

3. Major-Generals Grant and Buell will retain the immediate command of their respective armies in the field.

By command of Major-General Halleck:

N. H. McLEAN,

Assistant Adjutant-General.

HEADQUARTERS DEPARTMENT OF THE MISSISSIPPI,
Pittsburg, Tenn., April 14, 1862.

Major General U. S. GRANT,

Commanding District and Army in the Field:

Immediate and active measures must be taken to put your command in condition to resist another attack by the enemy. Fractions of batteries will be united temporarily under competent officers, supplied with ammunition, and placed in position for service. Divisions and brigades should, where necessary, be reorganized and put in position, and all stragglers returned to their companies and regiments. Your army is not now in condition to resist an attack. It must be made so without delay. Staff officers must be sent out to obtain returns from division commanders and assist in supplying all deficiencies.

H. W. HALLECK,

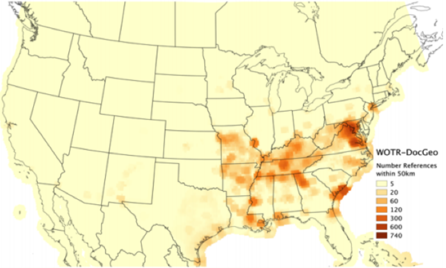


Figure 1: Distribution of References in WoTR-DocGeo

DeLozier, G., Wing, B., Baldridge, J. and Nesbit, S., 2016. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)* (pp. 188-198).

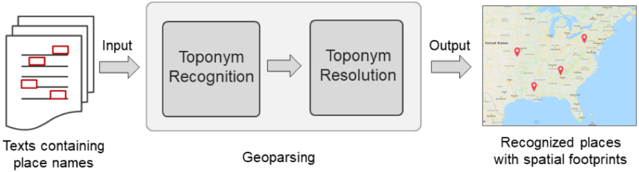
Applications of Geoparsing

- Voice and text-based location extraction for intelligent personal assistants

Geoparsers and their comparison

- A number of geoparsers have already been developed

- All of them divide the process of geoparsing into two steps: toponym recognition and toponym resolution



Existing geoparsers

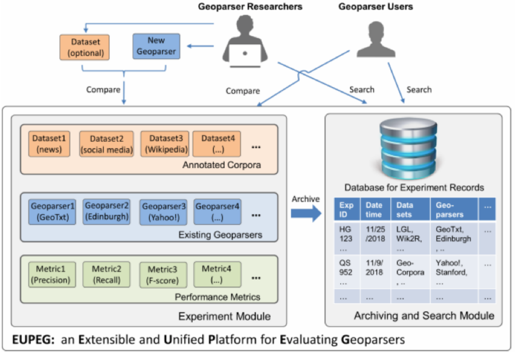
Geoparser	Toponym Recognition	Toponym Resolution	Gazetteer
GeoTxt (Version 2.0)	Stanford NER	Heuristic rules	GeoNames (July 2017)
Edinburgh (Version 1.1)	LT-TTT2	Heuristic rules	GeoNames (Online)
TopoCluster (Nov. 2016)	Stanford NER	Geo-profiles of words	GeoNames+ Natural Earth (Nov. 2016)
CLAVIN (Version 2.1.0)	Apache OpenNLP	Heuristic rules	GeoNames (Apr. 2019)
Yahoo! PlaceSpotter (Online)	Proprietary	Proprietary	WOEID (Where on Earth ID) (Online)
CamCoder (Sept. 2018)	spaCy NER	CNNs+Map-based word vectors	GeoNames (July 2018)
Stanford NER + Population (Version 3.9.2)	Stanford NER	Highest population	GeoNames (Online)
spaCy NER + Population (Version 2.0.18)	spaCy NER	Highest population	GeoNames (Online)
DBpedia Spotlight (Version 1.0.0)	LingPipe Exact Dictionary Chunker	Context similarity	DBpedia (Online)

Wang, J. and Hu, Y., 2019. *Transactions in GIS*, *23*(6), pp.1393-1419.

- Existing geoparsers are usually tested on different datasets using different metrics
- There is a lack of comparison among the geoparsers on the same datasets

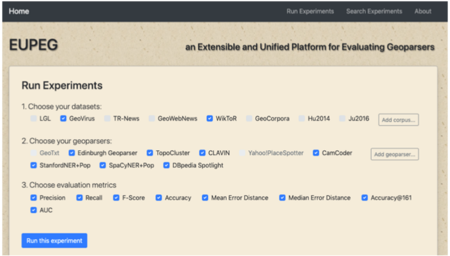
EUPEG: an extensible and unified platform for evaluating geoparsers

- Eight datasets
- Nine geoparsers
- Eight metrics



Wang, J. and Hu, Y., 2019. *Transactions in GIS*, *23*(6), pp.1393-1419.

EUPEG: an extensible and unified platform for evaluating geoparsers



Experiment ID: FJUD8DANXXE457EL

Please remember this ID. You will need to input this ID when searching for this experiment.

Results: Download

Performances based on the dataset: GeoVirus

Geoparser Name	Precision	Recall	F-Score	Accuracy	Mean (km)	Median (km)	Accuracy@161	AUC
Edinburgh parser_version: version 1.1 gaze_version: R202006	0.860	0.559	0.678	0.559	435.799	33.187	0.807	0.319
TopoCluster parser_version: updated Nov. 2016 gaze_version: updated Nov. 2016	0.877	0.813	0.844	0.813	599.632	63.858	0.673	0.407
CLAVIN parser_version: version 2.1.0 gaze_version: updated Apr. 2019	0.913	0.637	0.750	0.637	522.176	35.503	0.786	0.320
CamCoder parser_version: updated Sept. 2018 gaze_version: updated July 2018	0.940	0.802	0.866	0.802	619.397	33.945	0.770	0.336
StanfordNER parser_version: version 3.9.2 gaze_version: R202006	0.927	0.903	0.915	0.903	791.296	48.676	0.655	0.378
SpaCyNER parser_version: version 2.0.18 gaze_version: R202006	0.721	0.382	0.499	0.382	788.231	40.653	0.698	0.367
DBpedia parser_version: version 1.0.0 gaze_version: 2020-06	0.792	0.616	0.693	0.616	1272.937	122.314	0.533	0.406

https://geoai.geog.buffalo.edu/EUPEG/

https://github.com/geoai-lab/EUPEG

EUPEG: an extensible and unified platform for evaluating geoparsers

- No clear winner: a geoparser that excels in one metric may not be as good as another in some other metrics

- Different running speeds

- Lacking ability to handle case insensitive texts
  - Edinburgh
  - CLAVIN
  - …

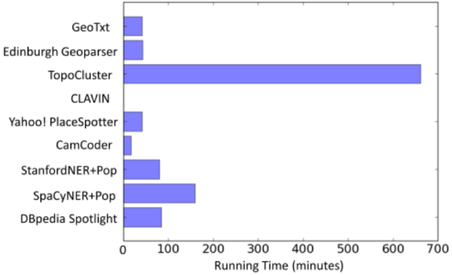


Figure 7: Running time of different geoparsers on GeoCorpora.

SemEval-2019 Task 12: Toponym Resolution in Scientific Papers

Davy Weissenbacher, Arjun Magge, Karen O'Connor, Matthew Scotch, Graciela Gonzalez-Hernandez

PDF

BibTeX

Search

Abstract

We present the SemEval-2019 Task 12 which focuses on toponym resolution in scientific articles. Given an article from PubMed, the task consists of detecting mentions of names of places, or toponyms, and mapping the mentions to their corresponding entries in GeoNames.org, a database of geospatial locations. We proposed three subtasks. In Subtask 1, we asked participants to detect all toponyms in an article. In Subtask 2, given toponym mentions as input, we asked participants to disambiguate them by linking them to entries in GeoNames. In Subtask 3, we asked participants to perform both the detection and the disambiguation steps for all toponyms. A total of 29 teams registered, and 8 teams submitted a system run. We summarize the corpus and the tools created for the challenge. They are freely available at https://competitions.codalab.org/competitions/19948. We also analyze the methods, the results and the errors made by the competing systems with a focus on toponym disambiguation.

Anthology ID:	S19-2155
Volume:	Proceedings of the 13th International Workshop on Semantic Evaluation
Month:	June
Year:	2019
Address:	Minneapolis, Minnesota, USA
Venue:	*SEMEVAL
SIG:	SIGLEX
Publisher:	Association for Computational Linguistics

https://www.aclweb.org/anthology/S19-2155/

Are we there yet?

- Top 3 winning teams all used deep neural network based models (e.g., BiLSTM) for toponym recognition

- #1: DM_NLP: over 90% precision, recall, and F-score

- Are we there yet?
  - The competition was based on a single dataset with 45 research articles in Biomedicine
  - How would the winners perform on other datasets?

SemEval-2019 Task 12: Toponym Resolution in Scientific Papers

Davy Weissenbacher, Arjun Magge, Karen O'Connor, Matthew Scotch, Graciela Gonzalez-Hernandez

Abstract

We present the SemEval-2019 Task 12 which focuses on toponym resolution in scientific articles. Given an article from PubMed, the task consists of detecting mentions of names of places, or toponyms, and mapping the mentions to their corresponding entries in GeoNames.org, a database of geospatial locations. We proposed three subtasks. In Subtask 1, we asked participants to detect all toponyms in an article. In Subtask 2, given toponym mentions as input, we asked participants to disambiguate them by linking them to entries in GeoNames. In Subtask 3, we asked participants to perform both the detection and the disambiguation steps for all toponyms. A total of 29 teams registered, and 8 teams submitted a system run. We summarize the corpus and the tools created for the challenge. They are freely available at https://competitions.codalab.org/competitions/19948. We also analyze the methods, the results and the errors made by the competing systems with a focus on toponym disambiguation.

Wang, J. and Hu, Y., 2019. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Geospatial Humanities* (pp. 1-6).

Are we there yet?

- The winning models indeed have good performance on well-formatted texts, such as news articles
- But so does a simple off-the-shelf Stanford NER

Geoparser	precision	recall	f_score	mean	median	acc@161	AUC
DM_NLP+Pop	0.917	0.916	**0.917**	770.337	48.676	0.655	0.378
StanfordNER	0.927	0.903	0.915	791.296	48.676	0.655	0.378
UniMelb+Pop	0.882	**0.936**	0.908	777.234	48.466	0.657	0.379
UArizona	0.887	0.859	0.873	769.810	55.635	0.640	0.386
CamCoder	**0.940**	0.802	0.866	619.397	33.945	0.770	0.336
TopoCluster	0.877	0.813	0.844	599.632	63.858	0.673	0.407
GeoTxt	0.857	0.726	0.786	487.874	36.255	0.787	0.338
CLAVIN	0.913	0.637	0.750	522.176	35.503	0.786	0.320
DBpedia	0.792	0.616	0.693	1272.937	122.314	0.533	0.406
Edinburgh	0.860	0.559	0.678	**435.799**	**33.187**	**0.807**	**0.319**
SpaCyNER	0.721	0.382	0.499	788.231	40.653	0.698	0.367

Performances on the GeoVirus corpus

Are we there yet?

- The winning models have only fair performances on ill-formatted texts, such as social media messages

Geoparser	precision	recall	f_score	mean	median	acc@161	AUC
DM_NLP+Pop	0.888	**0.669**	**0.763**	1249.865	0.000	0.661	0.288
UniMelb+Pop	0.852	0.661	0.745	1245.992	0.000	0.659	0.289
UArizona	0.892	0.598	0.716	1079.012	0.000	0.668	0.278
GeoTxt	**0.926**	0.521	0.667	714.94	0.000	0.876	0.116
StanfordNER	0.899	0.526	0.664	1063.473	0.000	0.676	0.270
CamCoder	0.904	0.503	0.647	1024.723	0.000	0.820	0.163
TopoCluster	0.882	0.506	0.643	575.225	32.948	0.698	0.361
DBpedia	0.865	0.500	0.633	669.105	33.816	0.654	0.352
Edinburgh	0.832	0.505	0.628	958.401	0.000	0.848	0.139
SpacyNER	0.705	0.467	0.562	982.137	0.000	0.752	0.224
CLAVIN	0.907	0.341	0.496	**373.563**	0.000	**0.913**	**0.084**

Performances on the GeoCorpora corpus (tweets)

Social Media and Natural Disasters

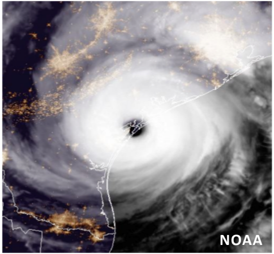
- Social media platforms have been increasingly used by people in natural disasters to request for help and share information

- A few tweets from 2017 Hurricane Harvey

"12 Y/O BOY NEEDs RESCUED! 8100 Cypresswood Dr Spring TX 77379 They are trapped on second story! #houstonflood"

"Anyone with a boat in the Meyerland area! A pregnant lady named Nisa is stranded near Airport blvd & station dr #Harvey"

"Rescue needed: 2907 Trinity Drive, Pearland, Tx. Need boat rescue 3 people, 2 elderly, one is 90 not steady in her feet & cant swim. #Harvey"



Hurricane Harvey in 2017

Social Media and Natural Disasters

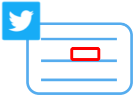
- Effectively extracting locations described in social media messages can help first responders reach the people in need

- Geotagged locations vs. described locations

- GPS location (or general area) attached to a tweet
- Where the tweet was sent
- Available in the metadata
- About 1% of the total tweets

Geotagged locations

(Twitter removed precise geotagging in June 2019)

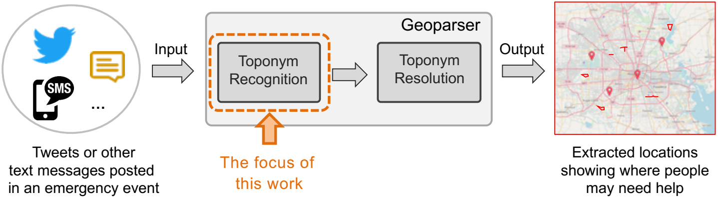


- Location described in the content of a tweet
- Where the tweet talked about
- Needs to be extracted
- Over 10% of the total tweets

Described locations

Extracting Locations from Social Media Messages

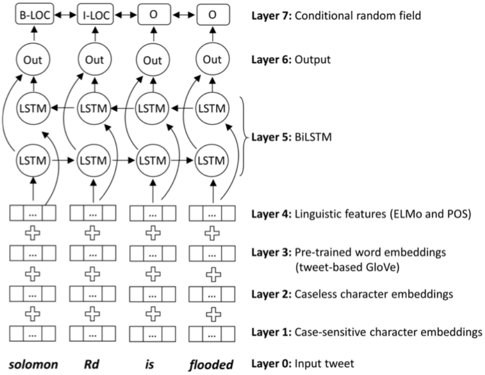
- Two steps: Toponym recognition and toponym resolution



Extracting Locations from Social Media Messages

- Challenges in recognizing locations from social media messages
  - Misspellings (e.g., "*Californa*")
  - Inconsistent upper and lower cases (e.g., "*there is a HUGE fire near camino and springbrook rd*")
  - Language abbreviations (e.g., "ppl", "pls", "@", …)
  - Informal sentence structures

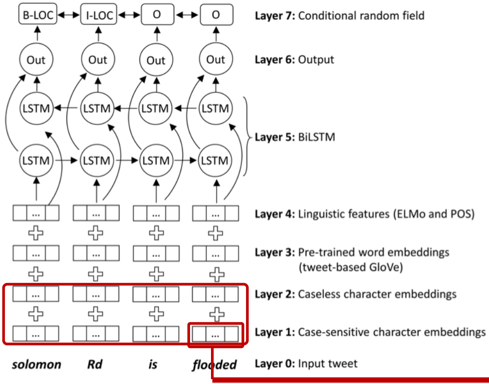
NeuroTPR: a Neuro-net ToPonym Recognition model



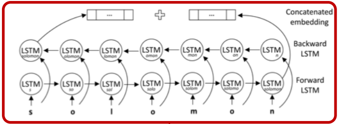
- Based on a Bidirectional Long Short-Term Memory (BiLSTM) architecture with seven layers

[Wang, J., Hu, Y., & Joseph, K. (2020). Transactions in GIS, 24(3), 719-735.]

NeuroTPR: a Neuro-net ToPonym Recognition model

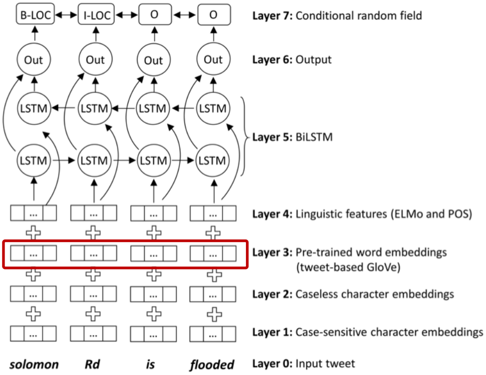


- Character embeddings help handle misspellings in tweets
- Two layers of caseless and case-sensitive embeddings help handle ill-capitalizations



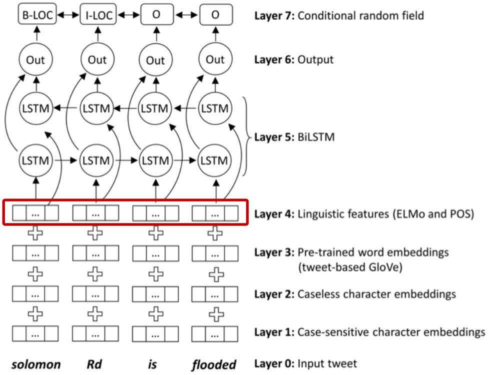
Wang, J., Hu, Y., & Joseph, K. (2020). Transactions in GIS, 24(3), 719-735.

NeuroTPR: a Neuro-net ToPonym Recognition model



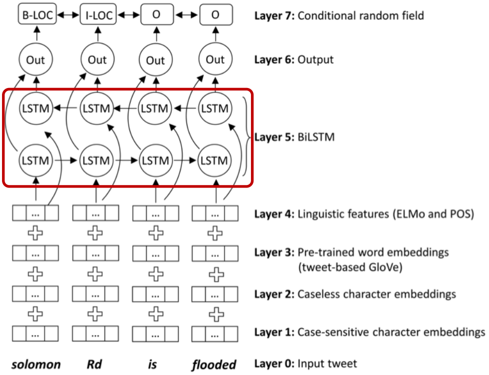
- Word embeddings GloVe pre-trained on 2 billion tweets
- Tweet-based GloVe capture the semantics of the informal words and abbreviations often used in tweets

NeuroTPR: a Neuro-net ToPonym Recognition model



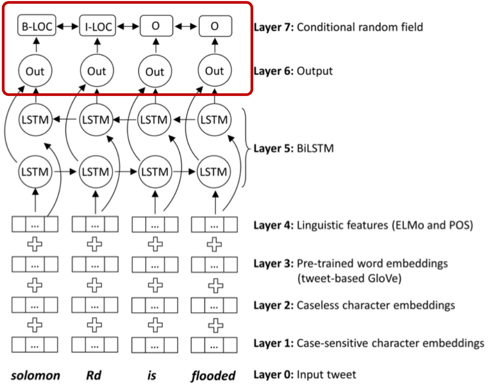
- POS (Part of Speech): informs the model about the type of a word
- ELMo: Deep contextualized word embeddings that capture the change of word meaning based on usage context

NeuroTPR: a Neuro-net ToPonym Recognition model



- Forward LSTM layer: Captures the context of a word from its left side
- Backward LSTM layer: Captures the context of a word from its right side

NeuroTPR: a Neuro-net ToPonym Recognition model



- Fully connected output layer: Concatenates the output from the previous two layers
- Conditional random field layer: Infers the type of a word by considering sequential dependence

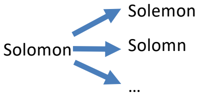
How to train NeuroTPR?

- Obtaining a sufficient amount of labeled training data is often a bottleneck for training a deep learning model

- Two datasets for training NeuroTPR:
  - 599 human annotated tweets from *WNUT 2017 Shared Task on Novel and Emerging Entity Recognition*
  - A dataset automatically generated from Wikipedia articles using a proposed workflow

How to train NeuroTPR?

- A workflow for generating training datasets from Wikipedia articles
  - Take the first paragraphs of Wikipedia articles
  - Keep only locations from the hyperlink annotations
  - Dive a paragraph into short sentences
  - Random flipping to simulate misspellings



How to train NeuroTPR?

- Initial effort based on geotagged Wikipedia articles
- Geotagged Wikipedia articles are not always about locations

Normandy landings

From Wikipedia, the free encyclopedia

Coordinates: 49°20′N 0°34′W

"D-Day" and "Operation Neptune" redirect here. For other uses, see D-Day (disambiguation) and Operation Neptune (disambiguation).

The **Normandy landings** were the landing operations on Tuesday, 6 June 1944 of the Allied invasion of Normandy in Operation Overlord during World War II. Codenamed **Operation Neptune** and often referred to as **D-Day**, it was the largest seaborne invasion in history. The operation began the liberation of German-occupied France (and later western Europe) and laid the foundations of the Allied victory on the Western Front.

Planning for the operation began in 1943. In the months leading up to the invasion, the Allies conducted a substantial military deception, codenamed Operation Bodyguard, to mislead the Germans as to the date and location of the main Allied landings. The weather on D-Day was far from ideal, and the operation had to be delayed 24 hours; a further postponement would have meant a delay of at least two weeks, as the invasion planners had requirements for the phase of the moon, the tides, and the time of day that meant only a few days each month were deemed suitable. Adolf Hitler placed Field Marshal Erwin Rommel in command of German forces and of developing fortifications along the Atlantic Wall in anticipation of an Allied invasion.

The amphibious landings were preceded by extensive aerial and naval bombardment and an airborne assault—the landing of 24,000 American, British, and Canadian airborne troops shortly after midnight. Allied infantry and armoured divisions began landing on the coast of France at 06:30. The target 50-mile

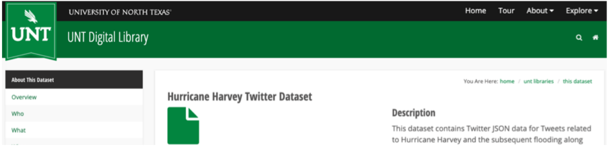
Normandy landings

Part of Operation Overlord, Invasion of Normandy, Western Front of World War II

Men of the 16th Infantry Regiment, US 1st Infantry Division wading ashore on Omaha Beach on the morning of 6 June 1944

Experiments

- Test datasets:
  - Three test datasets but we will focus on a dataset on Hurricane Harvey
  - 7,041,866 tweets retrieved during Harvey; collected by the libraries of University of North Texas
  - 1,000 tweets are extracted using a regular expression focusing on location-related terms and then random selection; These 1000 tweets are manually annotated and are used as the ground truth



https://digital.library.unt.edu/ark:/67531/metadc993940

Experiments

- Evaluation criteria:

$$Precision = \frac{tp}{tp + fp}$$
$$Recall = \frac{tp}{tp + fn}$$
$$F\text{-}score = 2 \cdot \frac{Precision \times Recall}{Precision + Recall}$$

- Baseline models:
  - Stanford NER
  - Caseless Stanford NER
  - SpaCy NER
  - Basic BiLSTM+CRF (Lample et al., 2016)
  - DM_NLP (Wang et al., 2019)

Using existing NER models is not straightforward:

- Stanford NER outputs three types of entities, which are Location, Organization, and Person
- Using Location only will miss schools and churches which are often used as shelters
- Using both Location and Organization will include false positives.

Experiments

Similar difficulty happens with spaCy NER:

- Location only refers to natural geographic features, such as rivers and mountains
- All geography-related entities: FACILITY (e.g., buildings, airports, and highways), ORG (e.g., companies, agencies, and institutions), GPE (e.g., countries, cities, and states), and LOC (e.g., non-GPE locations, mountain ranges, and bodies of water)

Baseline models:

- Stanford NER (narrow)
- Stanford NER (broad)
- Stanford NER (re-trained)
- Caseless Stanford NER (narrow)
- Caseless Stanford NER (broad)
- SpaCy NER (narrow)
- SpaCy NER (broad)
- Basic BiLSTM+CRF (Lample et al., 2016)
- DM_NLP (Wang et al., 2019)

Experiments

Evaluating different training strategies

Training Strategy	*Precision*	*Recall*	*F-score*
S1: WNUT2017 Only	0.687	0.633	0.656
S2: 1000 Wikipedia articles	0.551	0.392	0.458
S3: 3000 Wikipedia articles	0.573	0.468	0.516
S4: 5000 Wikipedia articles	0.547	0.481	0.512
S5: 1000 Wikipedia articles + random flipping	0.558	0.324	0.410
S6: 3000 Wikipedia articles + random flipping	0.566	0.359	0.439
S7: 5000 Wikipedia articles + random flipping	0.520	0.410	0.459
S8: 3000 Wikipedia articles + WNUT2017	**0.787**	**0.678**	**0.728**

- More training data is not always better, especially when the training data contain noise
- A combination of a moderate-size generated training dataset and real tweet dataset works the best

Experiments

Comparing NeuroTPR with the baseline models on Hurricane Harvey tweets

Model	*Precision*	*Recall*	*F-score*
Stanford NER (narrow location)	**0.828**	0.399	0.539
Stanford NER (broad location)	0.729	0.440	0.548
Retrained Stanford NER	0.604	0.410	0.489
Caseless Stanford NER (narrow location)	0.803	0.320	0.458
Caseless Stanford NER (broad location)	0.721	0.336	0.460
spaCy NER (narrow location)	0.575	0.024	0.046
spaCy NER (broad location)	0.461	0.304	0.366
Basic BiLSTM+CRF (Lample et al., 2016)	0.703	0.600	0.649
DM_NLP (toponym recognition) (Wang et al., 2019)	0.729	**0.680**	0.703
NeuroTPR	0.787	0.678	**0.728**

- Default Stanford NER achieves the best precision but very low recall
- DM_NLP achieves the highest recall, which is only slightly higher than NeuroTPR
- NeuroTPR achieves the most balanced performance with the highest F1-score

Experiments

Comparing NeuroTPR with the baseline models on two additional test datasets

Model	*Precision*	*Recall*	*F-score*
Stanford NER (narrow location)	**0.899**	0.526	0.664
Stanford NER (broad location)	0.751	0.553	0.637
Retrained Stanford NER	0.590	0.364	0.450
Caseless Stanford NER (narrow location)	0.898	0.487	0.631
Caseless Stanford NER (broad location)	0.774	0.503	0.610
spaCy NER (narrow location)	0.503	0.037	0.069
spaCy NER (broad location)	0.579	0.453	0.508
Basic BiLSTM+CRF (Lample et al., 2016)	0.631	0.527	0.574
DM_NLP (toponym recognition) (Wang et al., 2019)	0.797	0.650	0.715
NeuroTPR	0.800	**0.761**	**0.780**

GeoCorpora (Wallgrün et al. 2018)

Model	*Accuracy*
Stanford NER (narrow location)	0.010
Stanford NER (broad location)	0.012
Retrained Stanford NER	0.078
Caseless Stanford NER (narrow location)	0.460
Caseless Stanford NER (broad location)	0.514
spaCy NER (narrow location)	0.000
spaCy NER (broad location)	0.006
Basic BiLSTM+CRF (Lample et al., 2016)	0.595
DM_NLP (toponym recognition) (Wang et al., 2019)	0.723
NeuroTPR	**0.821**

Ju2016 (Ju et al. 2016)

Conclusions

- EUPEG: an extensible and unified platform for evaluating geoparsers
- Are we there yet? Partly yes! Good performance of existing geoparsers on well-formatted text that mainly contain city names
- NeuroTPR: A Neuro-net ToPonym Recognition model for extracting locations from social media messages
- NeuroTPR can be further combined with a toponym resolution model to form a complete geoparser

Thank you!

Yingjie Hu

University at Buffalo, SUNY
Email: yhu42@buffalo.edu
GeoAI@UB: https://geoai.geog.buffalo.edu