How can voting mechanisms improve the robustness of individual toponym resolution approaches?

XUKE HU[*], YERAN SUN[**], JENS KERSTEN[*], KLAN FRIEDERIKE[*]

[*]Data Science Institute • German Aerospace Center (DLR) • Jena • Germany

(Xuke.Hu@dlr.de, Jens.kersten@dlr.de, Friederike.Klan@dlr.de)

[**]Department of Geography • University of Lincoln • Lincoln • United Kingdom

(yesun@lincoln.ac.uk)

Keywords: Geoparsing, toponym resolution, voting

Summary: *This paper investigates the feasibility of implementing a general and robust toponym resolution approach by ensembling multiple existing approaches through voting mechanisms. Experiments were conducted to compare two voting ensembles with nine individual approaches based on seven public datasets. The results show that the voting ensembles can achieve consistent measures of Accuracy@161km and mean error, outperforming the individual approaches.*

Introduction

Apart from sensor equipment, natural language texts (e.g., social media posts and news) are other important sources that can contain location information in the form of toponyms or location descriptions. Extracting location information from texts is called geoparsing, which includes two subtasks: toponym recognition, i.e., to recognize toponyms from texts, and toponym resolution, i.e., to identify their geospatial representations.

This paper focuses on toponym resolution, whose main task is to remove geo/geo-ambiguities. Geo/geo-ambiguities refer to the situation in which one toponym can refer to more than one geographical location. For instance, by searching 'Paris' in GeoNames, an open global gazetteer, 547 different geographical locations named 'Paris' are returned. Many novel approaches have been proposed. This paper investigates if the voting approaches that ensemble multiple individual approaches could further improve the robustness of individual approaches. To answer the question, we propose two voting ensembles and compare them with nine individual approaches with regard to correctness based on seven public datasets. The nine approaches are CLAVIN[1], TopoCluster (Delozier et al., 2015), Adaptive Learning (Lieberman et al., 2012), CamCoder (Gritta et al., 2018), CBH (Kamalloo et al., 2018), Edinburgh Geoparser (Grover et al., 2010), Mordecai (Halterman, 2017), DBpedia Spotlight[2], and Population-Heuristics (Speriosu et al., 2013).

The principle of the voting mechanism is that the minority is subordinate to the dominant (majority). The workflow of the voting approaches is as follows: (1) Cluster the coordinate estimation of individual approaches by using DBSCAN (Khan et al., 2014); (2) Treat the centroid of all the coordinate estimations of the largest cluster as the voting result; (3) When no clusters are generated, the estimation by CLAVIN is treated as the voting result. If the estimation by CLAVIN is invalid, a valid estimation by other approaches is randomly selected as the voting result. An example is given (shown in Fig. 1) to explain the principle of the voting mechanism.

---

[1] https://github.com/Novetta/CLAVIN

[2] https://www.dbpedia-spotlight.org

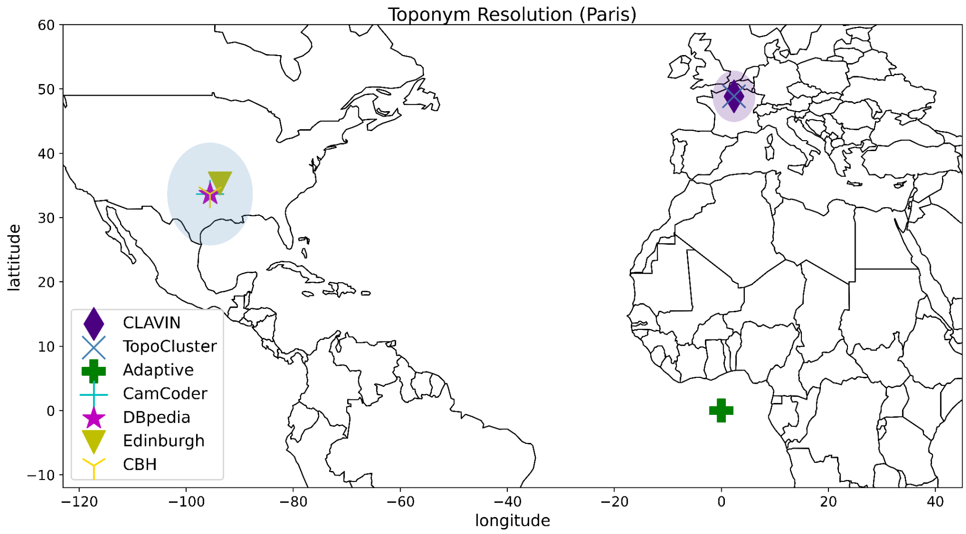

Fig. 1: An example to show how the voting mechanism works. The target toponym is 'Paris', whose true location is in the largest cluster (light blue circle), including the estimations of four approaches. The text is: 'Stepchildren: Dani Pipkins and husband, Jerome, Treva Battle, Tojquah Battle and Kelvin Battle all of **Paris**, Texas; brother, Thomas Smith and wife, Gwen of Detroit, Texas.'

Experiments

Table 1: Summary of seven public datasets. There are in total 29,074 toponyms (TPs).

	LGL	NEEL	TR-News	GeoWebNews	GeoCorpora	GeoVirus	WOTR
TP Count	5,088	481	1,319	5,121	3,100	2,170	11,795
Type	News	Tweet	History	News	Tweet	News	History

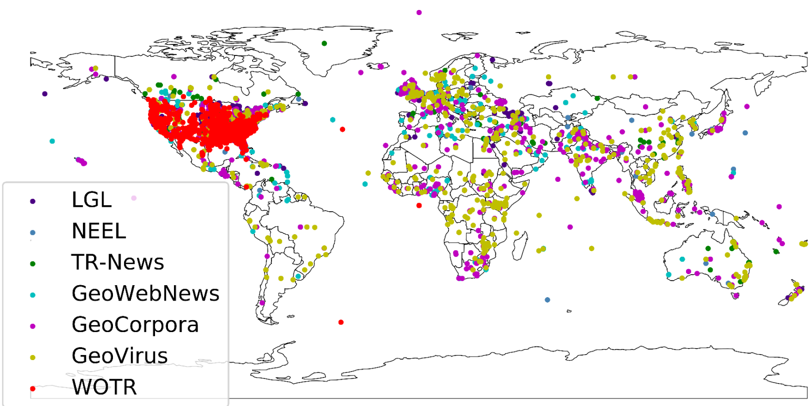

Fig. 2: Spatial distribution of toponyms in the seven datasets.

Table 1 summarizes the test datasets. They are LGL (Lieberman et al., 2010), NEEL (Rizzo et al., 2015), TR-News (Kamalloo et al., 2018), GeoWebNews (Gritta et al., 2020), GeoCorpora (Wallgrün et al., 2018), GeoVirus (Gritta et al., 2018), and WOTR (DeLozier et al., 2016). Fig. 2 shows the spatial distribution of the toponyms in the datasets, scattering across the globe. Notably, the nine individual approaches are globally applicable and thus can be evaluated based on the datasets. In the datasets, each toponym is assigned with a coordinate (latitude and longitude). From the standard metrics defined in Gritta et al. (2018), we adopt two metrics to evaluate toponym resolution approaches: mean distance error (*ME*) and percentage of estimations whose distance error is within 100 miles (*Accuracy@161km*). Given nine individual approaches, there are many possible ensembles. According to our preliminary test results, two optimal voting ensembles are selected, named Voting1 and Voting2, respectively. Voting1 ensembles CLAVIN, TopoCluster, Adaptive Learning, CamCoder, CBH, Edinburgh Geoparser, and DBpedia Spotlight, while Voting2 ensembles CLAVIN, Adaptive Learning, CamCoder, CBH, and Edinburgh Geoparser.

To achieve fair evaluation, we assume that all the toponyms can be correctly recognized by CLAVIN, TopoCluster, Adaptive Learning, CamCoder, CBH, and Population-Heuristics. We then compare the two voting ensembles with the six individual approaches on the full set of toponyms. The result is shown in Fig. 3. We can observe that when ignoring the voting ensembles, no individual approach performs best on all the datasets, while the voting ensembles perform the best (*Accuracy@161km* and ME) on all the datasets except the *Accuracy@161km* on LGL. We retrained the Adaptive Learning approach on LGL and it thus achieved the best *Accuracy@161km* on this dataset. On average, Voting1 improves the best *Accuracy@161km* by 9% and reduces the lowest ME by 25%, while Voting2 improves the best *Accuracy@161km* by 8% and reduces the lowest ME by 19%.

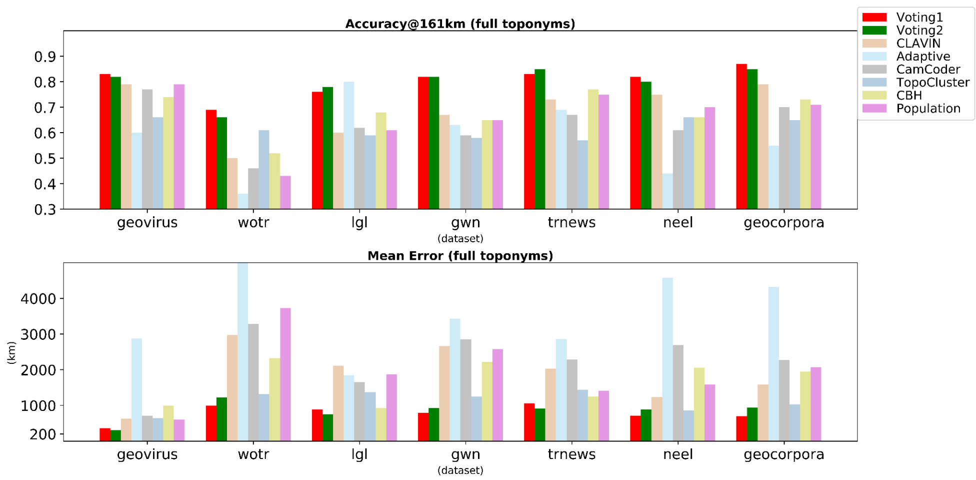

Fig. 3: Evaluation results on the full set of toponyms. *gwn* denotes the dataset GeoWebNews.

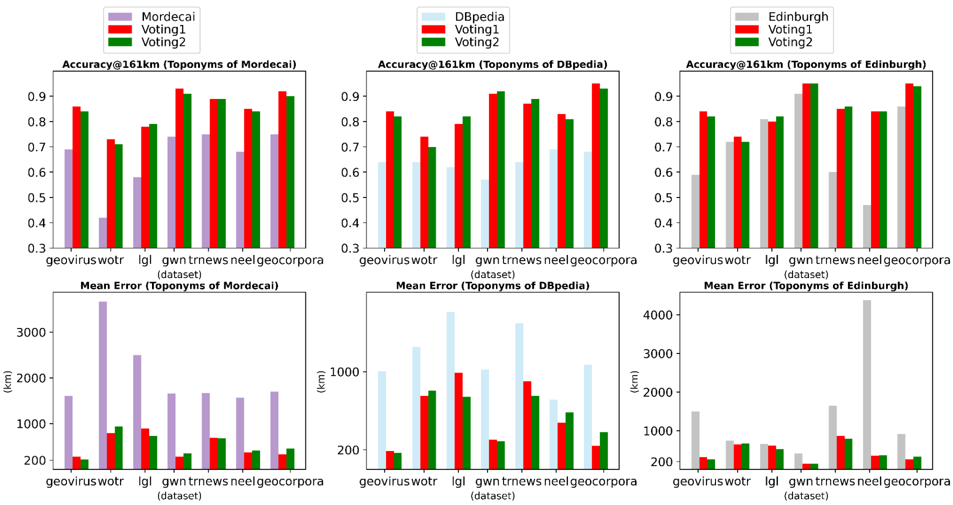

Fig. 4: Evaluation results on the correctly recognized toponyms by three approaches. *gwn* denotes the dataset GeoWebNews.

For DBPedia Spotlight, Edinburgh Geoparser, and Mordecai, we compare the two voting systems on the correctly recognized toponyms by the three approaches, respectively, since it is challenging or impossible to replace their toponym recognition modules. The result is shown in Fig. 4. The two voting systems outperform the three approaches on nearly all the datasets, except on the dataset LGL, on which Edinburgh Geoparser achieves slightly higher *Accuracy@161km* than Voting1.

Future work

To thoroughly evaluate the effectiveness of the voting mechanisms, in the future, we will include more individual approaches, such as TagMe (Ferragina et al., 2010) and use more test datasets, such as Corpus of Lake District Writing (Rayson et al., 2017) .

References

DELOZIER, G., BALDRIDGE, J. & LONDON, L., 2015: Gazetteer-independent toponym resolution using geographic word profiles. – In Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, Texas, USA, 2382-2388.

LIEBERMAN, M. D. & SAMET, H., 2012: Adaptive context features for toponym resolution in streaming news. – In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, Portland, Oregon, USA, 731-740.

GRITTA, M., PILEHVAR, M. & COLLIER, N., 2018: – Which melbourne? augmenting geocoding with maps. – In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 1285–1296.

KAMALLOO, E. & RAFIEI, D., 2018: A coherent unsupervised model for toponym resolution. – In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 1287-1296.

GROVER, C., TOBIN, R., BYRNE, K., WOOLLARD, M., REID, J., DUNN, S. & BALL, J., 2010: Use of the Edinburgh geoparser for georeferencing digitized historical collections. – Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 368(1925): 3875-3889.

HALTERMAN, A., 2017: Mordecai: Full Text Geoparsing and Event Geocoding. – J. Open Source Softw., 2(9): 91.

SPERIOSU, M. & BALDRIDGE, J., 2013: Text-driven toponym resolution using indirect supervision. – In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 1466-1476.

KHAN, K., REHMAN, S. U., AZIZ, K., FONG, S. & SARASVADY, S., 2014: DBSCAN: Past, present and future. – In The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014), Bengalura, India, 232-238.

LIEBERMAN, M. D., SAMET, H. & SANKARANARAYANAN, J., 2010: Geotagging with local lexicons to build indexes for textually-specified spatial data. – In 2010 IEEE 26th international conference on data engineering (ICDE 2010), Long Beach, CA, 201-212.

RIZZO, G., BASAVE, A. E. C., PEREIRA, B., VARGA, A., ROWE, M., STANKOVIC, M. & DADZIE, A., 2015: Making Sense of Microposts (# Microposts2015) Named Entity rEcognition and Linking (NEEL) Challenge. - In # MSM, 44-53.

GRITTA, M., PILEHVAR, M. T. & COLLIER, N., 2020: A pragmatic guide to geoparsing evaluation. – Language resources and evaluation 54(3): 683-712.

WALLGRÜN, J. O., KARIMZADEH, M., MACEACHREN, A. M. & PEZANOWSKI, S., 2018: GeoCorpora: building a corpus to test and train microblog geoparsers. – International Journal of Geographical Information Science 32(1):1-29.

DELOZIER, G., WING, B., BALDRIDGE, J. & NESBIT, S., 2016: Creating a novel geolocation corpus from historical texts. - In Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016), Berlin, Germany, 188–198.

FERRAGINA, P. & SCAIELLA, U., 2010: Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). - In Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, ON, Canada, 1625-1628.

RAYSON, P., REINHOLD, A., BUTLER, J., DONALDSON, C., GREGORY, I. & TAYLOR, J., 2017: A deeply annotated testbed for geographical text analysis: The corpus of lake district writing. - In Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities, Redondo Beach, CA, USA, 9-15.