International Journal of Applied Earth Observations and Geoinformation 117 (2023) 103191

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

How can voting mechanisms improve the robustness and generalizability of toponym disambiguation?

Xuke Hu [a,*], Yeran Sun [b], Jens Kersten [a], Zhiyong Zhou [c], Friederike Klan [a], Hongchao Fan [d]

[a] Institute of Data Science, German Aerospace Center, Germany
[b] Department of Geography, University of Lincoln, UK
[c] Department of Geography, University of Zurich, Switzerland
[d] Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Norway

ARTICLE INFO

Dataset link: https://github.com/uhuohuy/toponym-disambiguation-voting

Keywords:
Toponym disambiguation
Toponym resolution
Geocoding
Geoparsing
Entity linking
Voting
Evaluation

ABSTRACT

Natural language texts, such as tweets and news, contain a vast amount of geospatial information, which can be extracted by first recognizing toponyms in texts (toponym recognition) and then identifying their geospatial representations (toponym disambiguation). This paper focuses on toponym disambiguation, which can be approached by toponym resolution and entity linking. Recently, many novel approaches, especially deep learning-based, have been proposed, such as CamCoder, GENRE, and BLINK. However, these approaches were not compared on the same and large datasets. Moreover, there is still a need and space to improve their robustness and generalizability further. To mitigate the two research gaps, in this paper, we propose a spatial clustering-based voting approach combining several individual approaches and compare a voting ensemble with 20 latest and commonly-used approaches based on 12 public datasets, including several highly challenging datasets (e.g., WikToR). They are in six types: tweets, historical documents, news, web pages, scientific articles, and Wikipedia articles, containing 98,300 toponyms. Experimental results show that the voting ensemble performs the best on all the datasets, achieving an average *Accuracy@161km* of 0.86, proving its generalizability and robustness. It also drastically improves the performance of resolving fine-grained places, i.e., POIs, natural features, and traffic ways. The detailed evaluation results can inform future methodological developments and guide the selection of proper approaches based on application needs.

1. Introduction

Huge and ever-increasing amounts of semi- and unstructured text data, like news articles, historical archives, and social media posts, are available online and offline. These documents often contain valuable but hidden geospatial information in the form of toponyms, place names, or location references. The information is useful for many applications (Hu et al., 2022a), such as spatial humanities (Gregory et al., 2015), disaster management (Zhang et al., 2021), and disease surveillance (Scott et al., 2019). Extracting geospatial information from texts is also named geoparsing, which consists of two steps: toponym recognition, i.e., to recognize toponyms from texts, and toponym disambiguation, i.e., to determine their geo-coordinates. Specifically, toponym disambiguation handles the situation in which one toponym can refer to multiple geographical locations, as shown in Fig. 1. Toponym recognition has been extensively studied (Hu et al., 2022b; Qiu et al.; Wang et al., 2020; Hu et al., 2021), seeing Hu et al. (2022a) for an overview. This paper focuses on toponym disambiguation, which is still challenging.

Toponym disambiguation can be approached by entity linking and toponym resolution (Ardanuy and Sporleder, 2017). Entity linking aims to link an entity (e.g., *Person*, *Organization*, and *Location*) mentioned in texts to an entry of Knowledge Bases (KBs), such as Wikipedia (Wikipedia, 2004) and DBpedia (Auer et al., 2007). Recently, many deep learning-based entity linkers (ELs) emerged, such as GENRE (De Cao et al., 2021) and BLINK (Wu et al., 2020), showing superior performance (Sevgili et al., 2022). However, current KBs contain only a small proportion of places, lacking many small, unpopular, or fine-grained places (e.g., roads and shops). For instance, the largest KB, Wikipedia, contains about 1.5 million places, while over 23 million and 12 million places have been recorded in two open gazetteers, OpenStreetMap[1] and GeoNames[2] (Hu et al., 2022a). Toponym resolution aims to determine the coordinates of toponyms,

* Corresponding author.
E-mail address: xuke.hu@dlr.de (X. Hu).
[1] https://www.openstreetmap.org/
[2] http://www.geonames.org/

https://doi.org/10.1016/j.jag.2023.103191
Received 16 October 2022; Received in revised form 23 December 2022; Accepted 11 January 2023
Available online 31 January 2023

Fig. 1. Example of toponym ambiguity. *'Disneyland'* can refer to multiple different places, such as the park in Paris (France), California (US), Orlando (Canada), and other places named with *'Disneyland'*.

focusing only on location entities, typically leveraging gazetteers. Quite a few toponym resolution approaches have been proposed, such as Edinburgh Geoparser (Grover et al., 2010), CamCoder (Gritta et al., 2018a), and CHF (Kamalloo and Rafiei, 2018). However, according to Wang and Hu (2019), the performance of toponym resolution approaches varies by datasets, and no one can always perform the best.

We identify two research gaps in toponym disambiguation. First, although many approaches have been proposed, a thorough evaluation of these approaches on the same and large datasets is still missing. Previous comparative studies (Wang and Hu, 2019; Liu et al., 2022) ignore deep learning-based entity linkers and use a few datasets. Second, both entity linkers and toponym resolution approaches have their limitations, and there is still space and a need to improve their robustness and generalizability. To mitigate the two gaps, in this paper, we first propose a spatial clustering-based voting approach that combines several individual approaches to overcome the shortcomings of existing ones and further push state-of-the-art performance. We then thoroughly compare a voting ensemble with 20 latest and commonly-used toponym disambiguation approaches based on 12 public datasets regarding their correctness and computational efficiency.

Our contributions are twofold:

- We propose a more general and robust toponym disambiguation approach using voting mechanisms.
- It is the first time that many competing toponym disambiguation approaches (especially deep learning-based ELs) are thoroughly compared based on numerous datasets, which can help inform future methodological developments and guide the selection of proper approaches based on application needs.

2. Related works

Two main ways to disambiguate toponyms are entity linking and toponym resolution. Entity linking consists of two steps: named entity recognition and entity disambiguation. Recently, many surveys (Sevgili et al., 2022; Möller et al., 2022) on entity linking have been conducted. Therefore, we will review only toponym resolution approaches by dividing them into three groups: (1) Rules, (2) Learning and ranking, and (3) Learning and classification.

2.1. Rules

Given a toponym, rule-based approaches first search gazetteers to find all the candidates that match or partially match the toponym and then rank or score the candidates through manually defined IF-THEN rules, using heuristics or features like string similarity, the candidate's population and admin levels, spatial proximity, and one-sense-per-referent (Kamalloo and Rafiei, 2018; Karimzadeh et al., 2019). Representative approaches include Edinburgh Geoparser (Grover et al., 2010), CLAVIN,[3] GeoTxt (Karimzadeh et al., 2019), TAGGS (de Bruijn et al., 2018), and CHF, CBH, and SHS (Kamalloo and Rafiei, 2018). For a toponym in a PubMed article, Weissenbacher et al. (2015) select the candidate of the toponym, which is in the country mentioned in the GenBank record linked to the article. If an ambiguity cannot be resolved by the above heuristic, the candidate with the highest population will be chosen. Qi et al. (2019) select the candidate with the highest frequency in the training set or with the highest population when no candidate appears in the training set. Karimzadeh et al. (2019) calculate a candidate's score based on nine optional heuristics, such as population, the number of alternate names, and hierarchical and proximity relationship between two toponyms in the same tweet. Rule-based approaches are easy to implement and computationally efficient. However, manually defined rules are often fragile and not general enough.

2.2. Learning and ranking

The workflow of learning and ranking-based approaches (Lieberman and Samet, 2012; Santos et al., 2015; Ardanuy and Sporleder, 2017; Wang et al., 2019) is similar to the rule-based approaches. The only difference lies in the rules, which are not explicitly defined but learned from annotated examples. For example, Lieberman and Samet (2012) train a random forest model using context-free features, such as population and the distance of a candidate to a news's focus location, and adaptive context features, such as proximity relationships between the candidates of two toponyms in the same text. The input of the model is the features related to the pair of (*toponym*, *candidate*), and the output is 1 or 0, indicating if the toponym refers to the candidate. The classification confidence is regarded as the score of the candidate. Wang et al. (2019) train a LightGBM (Ke et al., 2017) model using features like population, name string similarity, spatial proximity, and context features that refer to the contextual similarity between the toponym and the candidate. The context of the candidate is obtained from its Wikipage. Apart from fully supervised approaches, weakly-supervised and unsupervised approaches (Speriosu and Baldridge, 2013; Ardanuy and Sporleder, 2017; Fize et al., 2021) have also been proposed to reduce the amount of annotated data required. For example, Ardanuy and Sporleder (2017) manually define a model to score candidates, using features like context similarity, spatial closeness to the focus location of documents, and spatial proximity between toponyms. The parameters of the model are learned from a small training dataset. Learning and ranking-based approaches can disambiguate toponyms without requiring as much expert knowledge as rule-based approaches. However, the trained models are often not general enough due to the lack of sufficient and high-quality training data, although unsupervised or weakly supervised techniques have been adopted.

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

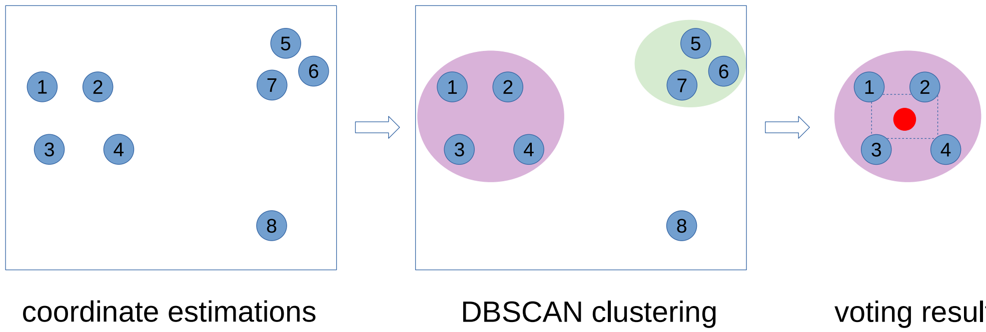

Fig. 2. Workflow of the voting approach.

2.3. Learning and classification

Learning and classification-based approaches (Gritta et al., 2018a; Kulkarni et al., 2020; Yan et al., 2021; Cardoso et al., 2021; DeLozier et al., 2015) divide the earth's surface into multiple cells and locate a toponym to a specific cell (class). For example, Gritta et al. (2018a) proposed a CNN-based model, named CamCoder, using features like the target toponym, the other toponyms in the same text, the contextual words removing the toponyms, and the prior probability of a candidate calculated based on its population. They train a model based on 1.4M examples generated from over 1M geographically annotated Wikipedia articles. Unlike CamCoder, which uses only local features (e.g., words and toponyms in texts), Yan et al. (2021) further utilizes global features, including topic and location embedding. Some studies also leverage language models, i.e., the spatial distribution of the words in texts (Speriosu and Baldridge, 2013; Wing and Baldridge, 2011; DeLozier et al., 2015), based on the assumption that apart from toponyms, common words, such as *'howdy'* and *'phillies'* can often be geographically indicative. For example, DeLozier et al. (2015) proposed TopoCluster, a gazetteer-independent approach. The spatial distribution of words is first learned based on 700,000 geographically annotated Wikipedia articles. The probability of a toponym in a specific cell is then calculated by merging the probability of the contextual words in the cell. Learning and classification-based approaches are typically trained on geographically annotated Wikipages currently containing around 1.5 million places. However, there are still many places that are not presented on Wikipedia.

3. Proposed approach

In this section, we introduce a voting approach, summarize 20 individual approaches used to form or to be compared with a voting ensemble, and illustrate the voting approach with four examples.

3.1. Voting approach

The idea of this study is inspired by Won et al. (2018), Hoang and Mothe (2018), which combine multiple existing toponym recognition approaches as a voting ensemble, achieving promising recognition performance. Each approach has its limitations, but combining multiple approaches can overcome these shortcomings. Different approaches normally return (or vote for) different locations (candidates) for a toponym in texts. We count the votes of the candidates and choose the one with the most votes. We set a higher weight to superior approaches by copying the coordinate estimation of the approaches multiple times. To realize the voting approach, we adopt DBSCAN (Khan et al., 2014), which groups together at least $minPts$ points that are below a distance measurement (denoted by eps) apart.

The workflow (as shown in Fig. 2) of the voting approach is described as follows:

Table 1
20 Representative approaches for toponym disambiguation. EL and TR denote entity linker and toponym resolution, respectively. ML and DL denote traditional machine learning algorithms based on feature engineering and deep learning algorithms, respectively.

Name	Method type
DBpedia Spotlight (Mendes et al., 2011)	EL
Entity-Fishing (ent, 2016–2022)	EL
MulRel-NEL (Le and Titov, 2018)	EL
DCA (Yang et al., 2019)	EL
BLINK (Wu et al., 2020)	EL
Bootleg (Orr et al., 2020)	EL
GENRE (De Cao et al., 2021)	EL
ExtEnD (Barba et al., 2022)	EL
LUKE (Yamada et al., 2022)	EL
Nominatim[a]	TR (rule)
Adaptive learning (Lieberman and Samet, 2012)	TR (ML+ranking)
Edinburgh Geoparser (Grover et al., 2010)	TR (rule)
Population-Heuristics (Speriosu and Baldridge, 2013)	TR (rule)
CLAVIN	TR (rule)
TopoCluster (DeLozier et al., 2015)	TR (ML+classification)
Mordecai (Halterman, 2017)	TR (rule)
CBH, SHS, CHF (Kamalloo and Rafiei, 2018)	TR (rule)
CamCoder (Gritta et al., 2018a)	TR (DL+classification)

[a] https://nominatim.org/.

(1) Group the coordinate estimation of the individual approaches combined in a voting ensemble with DBSCAN.
(2) If clusters are formed, select the largest cluster or randomly select one when multiple clusters of the same size exist. Treat the centroid of the coordinate estimations in the selected cluster as the voting result.
(3) If no clusters are formed, traverse the individual approaches of the ensemble and treat the first valid estimation as the voting result.

Invalid estimation refers to the situation where an approach fails to estimate the coordinates of a toponym, such as the one not included in gazetteers or linked to a non-location entity of KBs. The maximum error distance (20,039 km, half of the earth's circumference) is assigned to an invalid estimation (Gritta et al., 2020).

3.2. Individual approaches

Table 1 lists representative approaches, covering all types as discussed in Section 2. By default, we would modify their implementation to input gold toponyms to their entity (toponym) disambiguation step. We obtain the coordinates of DBpedia and Wikipedia entities from their geo-properties if they are geographically annotated. Otherwise, the coordinates of (0,0) are returned, denoting an invalid estimation. Details of the 20 approaches are as follows:

- **DBpedia Spotlight** is a popular EL. We use the provided HTTPS API[4] to annotate and link entities in texts.
- **Entity-Fishing** is a machine learning-based EL. We use its spaCy wrapper[5].
- **MulRel-NEL** is a neural EL. We use the provided API[6] of Radboud Entity Linker (REL) (van Hulst et al., 2020), which uses **mulrel-nel** for entity disambiguation.
- **DCA** is a neural EL. We retrain the model[7] based on the public and widely used AIDA CoNLL-YAGo dataset (Hoffart et al., 2011).
- **Bootleg** adopts a transformer architecture. We use the provided model[8] which is trained on weakly-labeled training data.
- **BLINK** is an EL based on fine-tuned BERT (Devlin et al., 2018). We use the provided model[9] that was pre-trained on nearly 9M examples generated from Wikipedia.
- **GENRE** uses a transformer-based architecture. We use the provided model[10] that was first pre-trained on nearly 9M examples generated from Wikipedia and then fine-tuned with the AIDA dataset.
- **LUKE** is an entity disambiguation model based on BERT, using both word-based and entity-based contextual information. We use the provided model[11] that was trained on a large entity-annotated corpus generated from Wikipedia.
- **ExtEnD** adopts Transformer-based architectures, which was first pretrained on the same Wikipedia dataset as BLINK and then fine-tuned on the AIDA dataset. We use the fine-tuned model[12] directly.
- **Nominatim** is a geocoder, built on OpenStreetMap, which is used as a baseline system. We request Nominatim with a toponym and keep the first (most popular) place it returns.
- **Population-Heuristics** uses the heuristic of the largest population, which is used as a baseline system. We implement the approach based on GeoNames.
- **Edinburgh Geoparser** is a rule-based geoparsing tool developed by the Language Technology Group (LTG) at Edinburgh University. We use its provided code and API[13].
- **CLAVIN** applies several heuristics for toponym resolution. We use its provided code.
- **Adaptive Learning** is a random forest-based toponym resolution approach. We use its implementation[14] to retrain a model based on the LGL dataset (Lieberman et al., 2010).
- **Mordecai** is a geoparser and we use its provided code[15].
- **CBH,SHS,CHF** are three rule-based approaches, and we use their implementation[16] to resolve toponyms.
- **TopoCluster** is a language model-based geoparser. We use the trained model and implementation[17].
- **CamCoder** is a CNN-based geoparser. We use the trained model and implementation[18].

3.3. Examples

We use four examples to illustrate the principle of the voting approach. We assume that the voting ensemble combines seven individual approaches, with each having one vote: **GENRE**, **BLINK**, **LUKE**, **CamCoder**, **Edinburgh Geoparser**, **SHS**, and **CBH**. Figs. 3, 4, 5, and 6 show the estimated location of *'SA'*, *'False River'*, *'Victoria Park'*, and *'Mount Sheridan'* by the seven individual approaches and formed clusters, respectively. In GeoNames, we can find 58 records of *'SA'*, 23 records of *'False River'*, 589 records of *'Victoria Park'*, and 25 records of *'Mount Sheridan'*, which actually refer to South Africa, a county in Louisiana, US, a park in London, UK, and a suburb of Cairns in the Cairns Region, Queensland, Australia, respectively. Their true locations are all in the largest cluster, denoted by the purple circle. Note that Edinburgh Geoparser cannot recognize *'Victoria Park'*, which thus cannot vote. From the four examples, we can see that no individual approach can correctly resolve all the toponyms, but the voting ensemble can.

4. Experiments

In this section, we first introduce the used test data and evaluation metrics. We then propose a voting ensemble and compare it with the 20 representative approaches regarding the correctness and computational efficiency. Finally, we conduct a sensitivity analysis of the voting approach.

4.1. Test data

We use 12 public datasets as test data, which are summarized in Table 2 and illustrated in Fig. 7. Note that ELs normally use Wikipedia as the target KB, while most datasets' toponyms are linked to GeoNames. However, the coordinates of some coarse-grained places (e.g., country) in Wikipedia and GeoNames are inconsistent. For instance, *'United States'* is geocoded to (40,-100) and (39.76, −98.5) and *'China'* is geocoded to (35, 103) and (35, 105) in Wikipedia and GeoNames, respectively. Such places frequently appear in the datasets, which can cause incorrect evaluation. From the datasets, we found 3,147 records of 28 frequent and misaligned places, including *'China', 'Chinese', 'Russia', 'Russian', 'Russians', 'Australia', 'Canada', 'Canadians', 'Canadian', 'United States', 'American', 'USA', 'America', 'U.S', 'United States of America', 'Americans', 'North America', 'South America', 'India', 'Algeria', 'Europe', 'European', 'Western Europe', 'Asia', 'Africa', 'West Africa', 'North Africa'*, and *'Middle East'*. They will be ignored during the evaluation.

Details of the 12 datasets are as follows:

- **LGL**[19] (Local-Global Lexicon) corpus was created by Lieberman et al. (2010), containing 588 human-annotated news articles published by 78 local newspapers.

[4] https://www.dbpedia-spotlight.org/api
[5] https://github.com/Lucaterre/spacyfishing
[6] https://github.com/informagi/REL
[7] https://github.com/YoungXiyuan/DCA
[8] https://github.com/HazyResearch/bootleg
[9] https://github.com/facebookresearch/BLINK
[10] https://github.com/facebookresearch/GENRE
[11] https://github.com/studio-ousia/luke
[12] https://github.com/SapienzaNLP/extend
[13] https://www.ltg.ed.ac.uk/software/geoparser/
[14] https://github.com/ehsk/CHF-TopoResolver
[15] https://github.com/openeventdata/mordecai
[16] https://github.com/ehsk/CHF-TopoResolver
[17] https://github.com/grantdelozier/TopoCluster
[18] https://github.com/milangritta/Geocoding-with-Map-Vector
[19] https://github.com/milangritta/Pragmatic-Guide-to-Geoparsing-Evaluation/blob/master/data/Corpora/lgl.xml

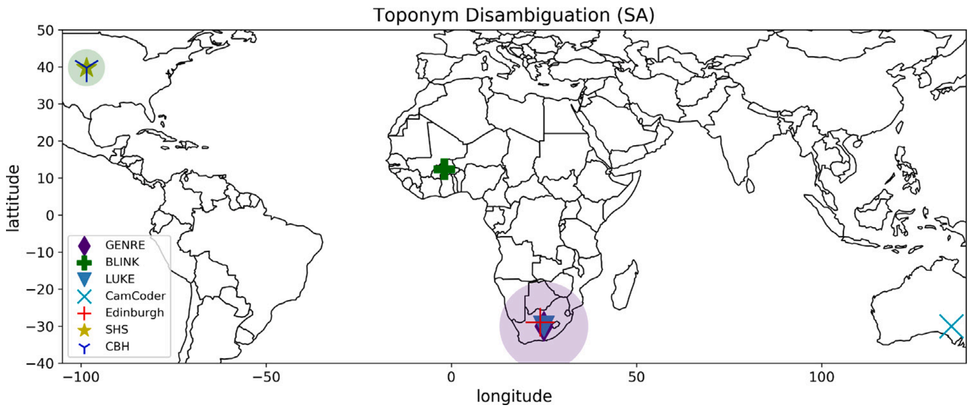

Fig. 3. An example to show how the voting approach works. The target toponym is *'SA'*, whose true location is in the largest cluster (purple circle). The context of the toponym is: *'Kgosi (chief) Nyalala Pilane of the Bakgatla-ba-Kgafela community — perhaps even more than any other chief in* ***SA*** *— has been the subject of a litany of maladministration and corruption allegations'*.

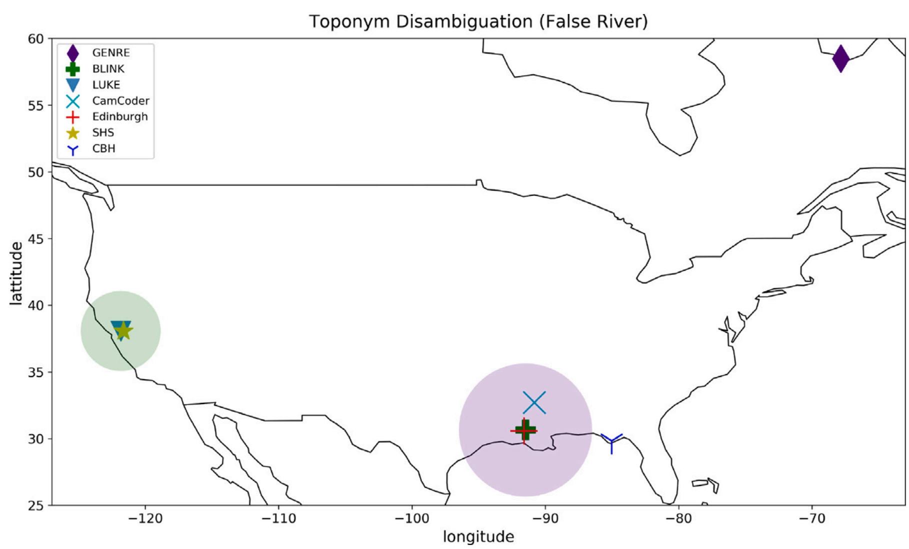

Fig. 4. An example to show how the voting approach works. The target toponym is *'False River'*, whose true location is in the largest cluster (purple circle). The context of the toponym is: '*The enemy have now left Waterloo, and that is of no importance, but the Rosedale country is of to visit, with the cavalry, and so also is the* ***False River*** *country. The cavalry must go to Rosedale and return by False River*'.

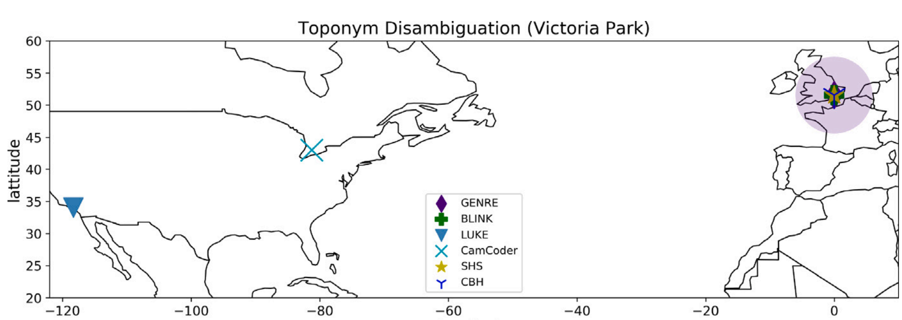

Fig. 5. An example to show how the voting approach works. The target toponym is *'Victoria Park'*, whose true location is in the largest cluster (purple circle). The context of the toponym is: '*The Clash - White Riot (Live 1978* ***Victoria Park****, London): via @YouTube Let's start our shift!*'.

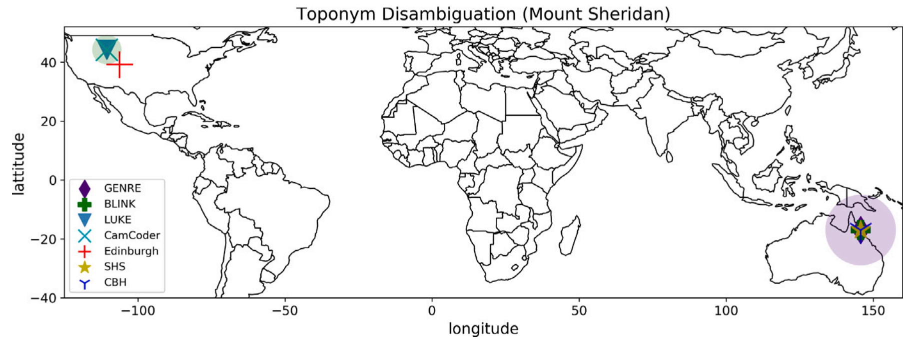

Fig. 6. An example to show how the voting approach works. The target toponym is *'Mount Sheridan'*, whose true location is in the largest cluster (purple circle). The context of the toponym is: *'Nine cases of the mosquito-borne illness have been confirmed in the Cairns suburbs of Edmonton,* ***Mount Sheridan****, Bentley Park, and Trinity Beach.'*.

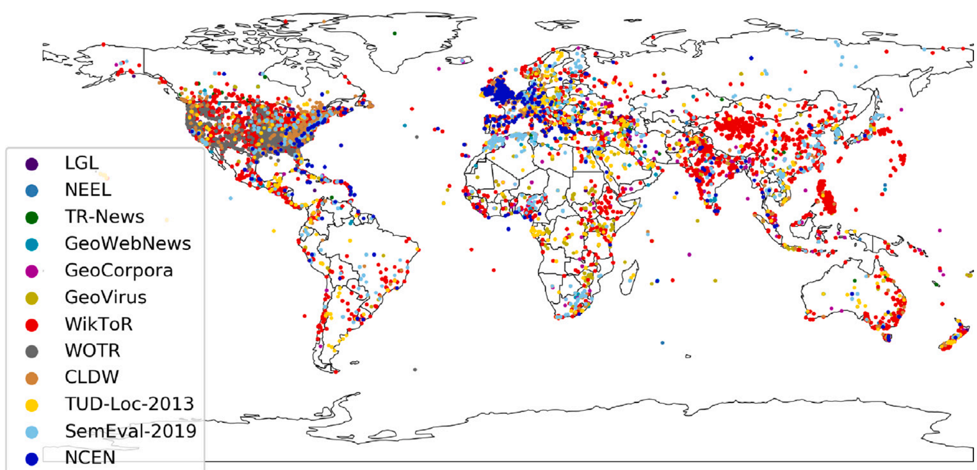

Fig. 7. Spatial distribution of the 98,300 toponyms in the 12 datasets.

Table 2
Summary of test datasets.

Dataset	Text count	Toponym count	Type	KB/Gazetteer
LGL	588	5088	News	GeoNames
NEEL	4078	481	Tweet	DBpedia
TR-News	118	1319	News	GeoNames
GeoWebNews	200	5121	News	GeoNames
GeoCorpora	6648	3100	Tweet	GeoNames
GeoVirus	229	2170	News	Wikipedia
WikToR	5000	31,500	Wikipedia article	Wikipedia
WOTR	1644	11,795	History	OpenStreetMap
CLDW	62	3814	History	Unlock & GeoNames
TUD-Loc-2013	152	3850	Web page	GeoNames
SemEval-2019	90	8360	Scientific article	GeoNames
NCEN	455	3364	History	Wikipedia

- **NEEL**[20] is the gold dataset of 2016 Named Entity rEcognition and Linking challenge, including tweets covering multiple noteworthy events from 2011 to 2013.
- **TR-News**[21] was created by Kamalloo and Rafiei (2018) from news articles of different sources.
- **GeoWebNews**[22] was created by Gritta et al. (2018a) from news articles collected from April 1st to 8th in 2018.
- **GeoCorpora**[23] was created by Wallgrün et al. (2018), containing tweets related to multiple events (e.g., ebola, flood, and rebel) that happened across the world in 2014 and 2015.
- **GeoVirus**[24] was created by Gritta et al. (2018a), containing news articles about epidemics, such as Ebola and Swine Flu.
- **WikToR**[25] was created by Gritta et al. (2018b) in an automatic manner, containing 5,000 Wikipedia articles with many ambiguous places, such as (*Santa Maria, California*), (*Santa Maria, Bulacan*), (*Santa Maria, Ilocos Sur*), and (*Santa Maria, Romblon*).
- **WOTR**[26] was created by DeLozier et al. (2016) based on a set of American Civil War archives, known as *Official Records of the War of the Rebellion*, using Unlock[27] and GeoNames.
- **CLDW**[28] (The Corpus of Lake District Writing) was created by Rayson et al. (2017) based on writing samples about the English Lake District between the early seventeenth and the beginning of the twentieth century.
- **TUD-Loc-2013**[29] was first utilized in Katz and Schill (2013), containing 152 texts from web pages.

20 http://microposts2016.seas.upenn.edu/challenge.html
21 https://github.com/milangritta/Pragmatic-Guide-to-Geoparsing-Evaluation/blob/master/data/Corpora/TR-News.xml
22 https://github.com/milangritta/Pragmatic-Guide-to-Geoparsing-Evaluation/tree/master/data
23 https://github.com/geovista/GeoCorpora
24 https://github.com/milangritta/Pragmatic-Guide-to-Geoparsing-Evaluation/blob/master/data/Corpora/GeoVirus.xml
25 https://github.com/milangritta/Pragmatic-Guide-to-Geoparsing-Evaluation/blob/master/data/Corpora/WikToR.xml
26 https://github.com/barbarainacioc/toponym-resolution/tree/master/corpora/WOTR
27 https://unlock.blogs.edina.ac.uk/
28 https://github.com/UCREL/LakeDistrictCorpus
29 https://bitbucket.org/palladian_pk/tud-loc-2013/src/master/

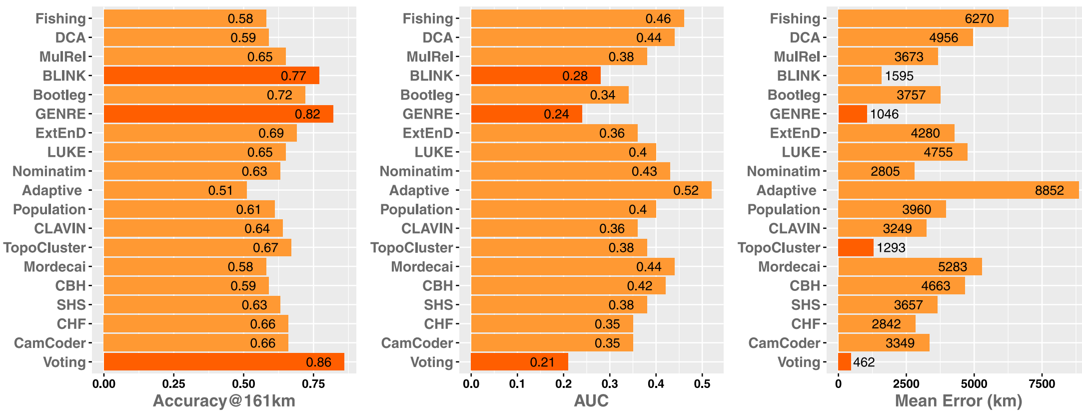

Fig. 8. Average *Accuracy@161km* (↑), *AUC* (↓), and *ME* (↓) on gold toponyms.

- **SemEval-2019-12**[30] is the gold dataset of Task 12 (Toponym Resolution in Scientific Papers) of the 13th International Workshop on Semantic Evaluation (SemEval) (Weissenbacher et al., 2019).
- **NCEN**[31] (The Nineteenth-Century English Newspapers) was created by Ardanuy et al. (2022), containing news articles published between 1780 and 1870 in England.

4.2. Evaluation metrics

To fairly evaluate toponym disambiguation approaches, we assume that all the toponyms in the datasets can be correctly recognized at the toponym recognition step. However, DBpedia Spotlight and Edinburgh Geoparser provide only an online API and deploy the toponym recognition module on servers. Therefore, when evaluating the correctness of the two approaches, we will compare them with the other approaches on the correctly recognized toponyms (a subset of gold toponyms) by them.

We adopt the three most important metrics from the standard metrics defined in Gritta et al. (2020). They are: (1) *Accuracy@161km*, which is the percentage of geocoding errors that are below 100 miles (161 km); (2) *Mean Error (ME)*, which is the mean distance error of toponyms; (3) *Area Under the Curve (AUC)*, which is calculated using Eq. (1), where x_i denotes the distance error of the ith toponym, N denotes the count of toponyms, and 20039 is the maximum possible error in km on earth.

$$AUC = \int_{i=1}^{N} \frac{ln(x_i + 1)}{ln(20039)} dx \tag{1}$$

4.3. Voting ensemble

We propose a voting ensemble that combines seven individual approaches with each assigned with a weight (count of votes), denoted by the number in the brackets. To determine the combination manner and weights, we first investigate the contribution of every single approach to a voting ensemble, which can be seen in Section 4.7. Based on the experimental results, we select from the approaches that positively contribute to *Accuracy@161km* and *ME* and assign an integer weight (from 1 to 4) for each approach to create a voting ensemble and then evaluate it on the test datasets. The optimal combination manner and weights are determined after several rounds. We set the DBSCAN parameters $minPts$ and eps to 2 and 10 (km), respectively. In Section 4.7, we investigate the impact of the two parameters on the disambiguation performance of the voting ensemble.

- **Voting ensemble**: GENRE (3), BLINK (2), LUKE (2), CamCoder (1), SHS (1), CBH (1), Edinburgh Geoparser (1).

4.4. Results

We average the metric of the 12 datasets. Fig. 8 shows the result of the 18 individual approaches (excluding Edinburgh Geoparser and DBpedia Spotlight) and the voting ensemble on the gold toponyms of the datasets. The voting ensemble achieves an *Accuracy@161km* of 0.86, *ME* of 462 (km), and *AUC* of 0.21, improving the best individual approach, GENRE, by 5%, 57%, and 13%, respectively. Figs. 9 and 10 show the average *Accuracy@161km*, *Mean Error*, and *AUC* of the approaches on the subset of gold toponyms, which are correctly recognized by DBpedia Spotlight and by Edinburgh Geoparser, respectively. The voting ensemble still performs the best. Among individual approaches, the state-of-the-art ELs, GENRE and BLINK, achieve promising results, outperforming the other ELs and toponym resolution approaches.

We provide the detailed result of each dataset in the supplementary data (seen in Appendix A), from which we can see that on the gold toponyms, the voting ensemble achieves the best result on 33/36 (3 metrics evaluated for 12 datasets) indicators. Besides, GENRE and BLINK are especially effective on highly challenging datasets, including WikToR, WOTR, and LDC, performing much better than the other individual approaches. These datasets contain many less-common or low-frequency places, such as *'Paris, Missouri'* and *'Lima, Oklahoma'*, on which the two baseline systems, Nominatim and Population-Heuristics adopting simple heuristics (i.e., popularity) thus perform poorly. Whether challenging or general datasets, combining several individual approaches, the voting ensemble can consistently achieve state-of-the-art performance, proving its generalizability and robustness.

4.5. Place category

We investigate the disambiguation performance of the existing approaches and our voting approach on four types of places: admin units (e.g., country, state, and county), POIs (e.g., park, church, and hospital), traffic ways (e.g., street, highway, and bridge), and natural features (e.g., river, beach, and hill). We determine in total 13,878 admin units (e.g., *'EU'*, *'Berlin'*, and *'Boone County'*), 820 POIs (e.g., *'Lambert-St. Louis International Airport'*, *'Sam Houston High School'*, and *'westboro baptist church'*), 1605 natural features (e.g., *'Pine Island Bayou'*, *'Skiddaw Mountain'*, and *'Little Pine Creek'*), and 336 traffic ways (e.g., *'High Street'*, *'Lynchburg Railroad bridge'*, and *'Highway 49'*) from the test datasets based on the GeoNames ID of places provided in some datasets as well as through manual annotation.

We then calculate *Accuracy@161km* on each place type, ruling out Edinburgh Geoparser and DBpedia Spotlight since their toponym

[30] https://github.com/TharinduDR/SemEval-2019-Task-12-Toponym-Resolution-in-Scientific-Papers

[31] https://bl.iro.bl.uk/concern/datasets/f3686eb9-4227-45cb-9acb-0453d35e6a03

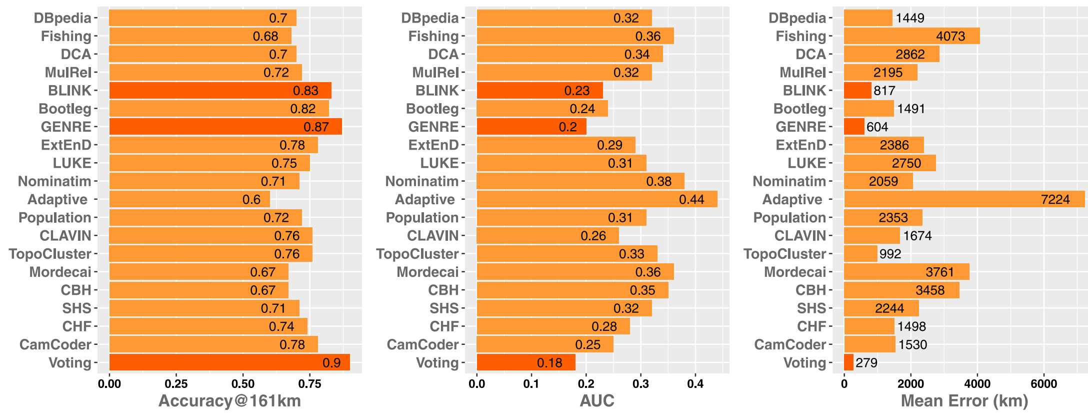

Fig. 9. Average *Accuracy@161km* (↑), *AUC* (↓), and *ME* (↓) on the subset (51%) of gold toponyms, which are correctly recognized by DBpedia Spotlight.

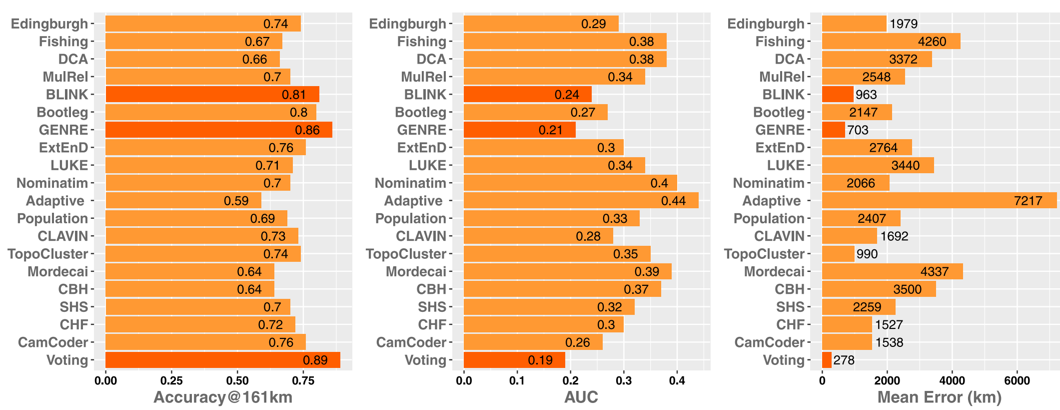

Fig. 10. Average *Accuracy@161km* (↑), *AUC* (↓), and *ME* (↓) on the subset (59%) of gold toponyms, which are correctly recognized by Edinburgh Geoparser.

recognition modules can only correctly recognize a small proportion of fine-grained toponyms. For example, they can only recognize 21/336 and 13/336 traffic ways, respectively. Fig. 11 shows that most of the approaches perform well in resolving coarse-grained places (i.e., admin units), with fifteen can correctly resolving over 70% of the admin units. However, they cannot resolve fine-grained places, with only four, three, and one correctly resolving over 60% of the POIs, natural features, and traffic ways, respectively. The voting ensemble performs the best, on average improving the best individual approach, GENRE, by 11% in resolving fine-grained places. However, there is still space for improving the performance of resolving fine-grained places.

4.6. Computational efficiency

We further investigate the computational efficiency of these approaches. We run each approach on the total datasets and record the consumed time without counting the training phase, as shown in Fig. 12. Note that we omit Edinburgh Geoparser, Nominatim, DBpedia Spotlight, and Entity-Fishing in the comparison since they are online services, and it is impossible to count their time consumption on the server. We run the toponym resolution approaches on a Dell laptop with an Intel Core i7-8650U CPU (1.90 GHz 8-Core) and a RAM of 16 GB, while we run the ELs on an NVIDIA Tesla V100 GPU of a cluster node since they normally require a GPU execution environment.

Generally, ELs take more time (from 3 h to 40 h) than toponym resolution approaches (from 2 min to 2.5 h) except TopoCluster and Mordecai since the former was normally built on large language models, such as BERT, and deals with more complex issues (disambiguating not only toponyms but also other types of entities) than the latter. TopoCluster is the slowest, taking nearly 191 h, while CLAVIN is the fastest, taking only 2 min. The time consumption for a voting ensemble equals the sum of the time of every individual approach that it combines. Therefore, the voting ensemble takes 72 h to resolve 98,300 toponyms. On average, resolving a toponym takes 2.6 s, which is acceptable for non-time-critical applications. For example, the voting approach can be used to assign geographic locations to news articles published online, which readers can then retrieve based on locations. There is a trade-off between correctness and speed.

4.7. Sensitivity analysis

4.7.1. Configuration

We first investigate how the removal of an individual approach would affect the performance of voting ensembles. A basic voting ensemble is first proposed, including all the 20 individual approaches, each having one vote. A degraded ensemble is then constructed by removing one approach from the basic ensemble. We then subtract the average *ME, Accuracy@161km*, and *AUC* achieved by the degraded ensemble from that of the basic ensemble. The result is shown in Fig. 13. Regarding *Accuracy@161km*, GENRE, BLINK, and SHS make the largest positive contribution, while CLAVIN and DBpedia Spotlight make the largest negative contribution. Regarding *ME*, GENRE and BLINK make the largest positive contribution, while CLAVIN and Population-Heuristics make the largest negative contribution. Regarding *AUC*, CamCoder, SHS, and Adaptive Learning make the largest positive contribution, while DCA and MulRel make the largest negative contribution. Generally, the disambiguation ability of every single approach determines their contribution to the voting ensemble, such as GENRE (with high disambiguation ability) and Population-Heuristics (with low disambiguation ability), which contribute positively and negatively, respectively.

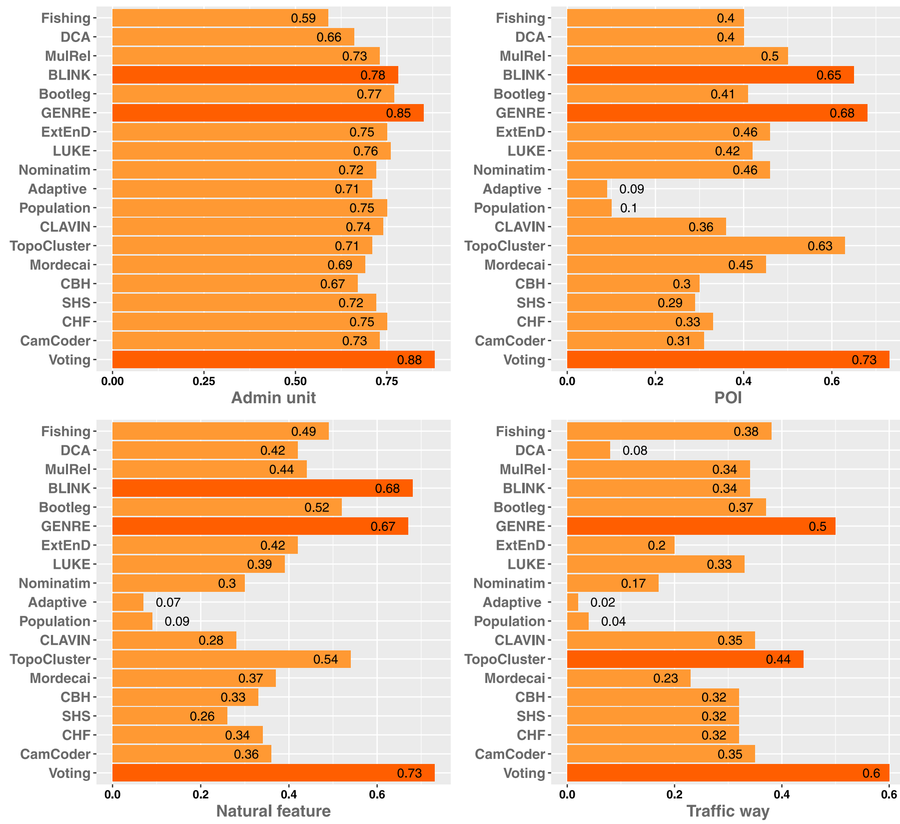

Fig. 11. *Accuracy@161km* on four place types with 13,878 admin units, 820 POIs, 1605 natural features, and 336 traffic ways.

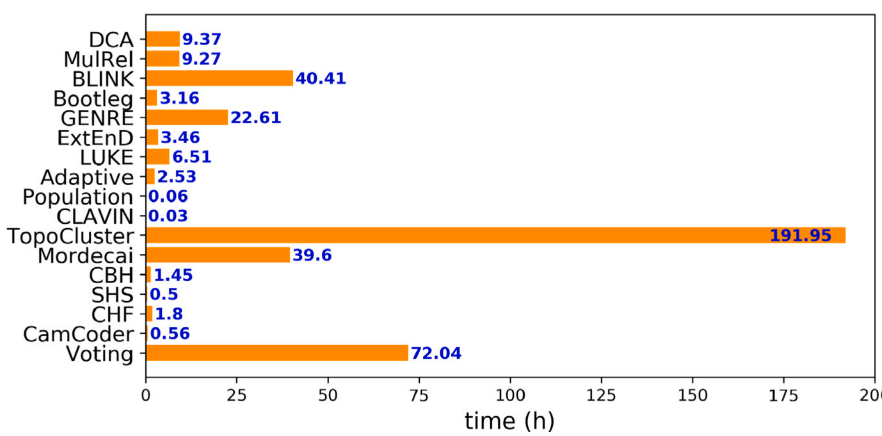

Fig. 12. Time consumption of the approaches running on the total test datasets.

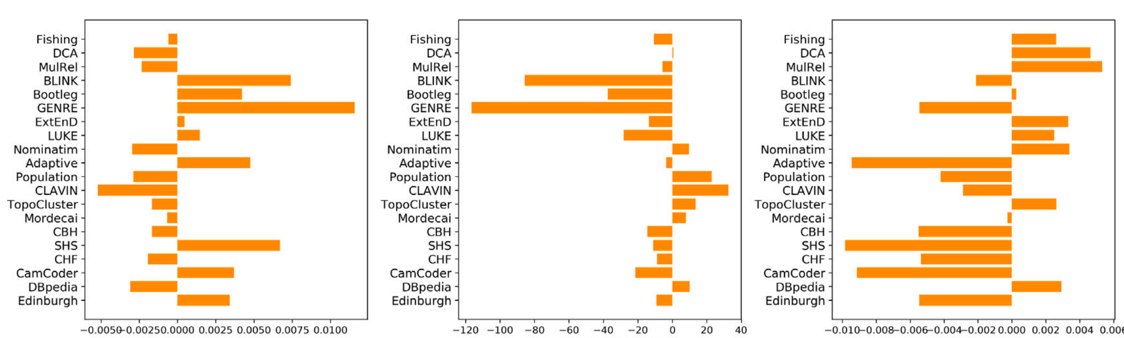

Fig. 13. Change of *Accuracy@161km*, *AUC*, and *ME* when adding an approach to a voting ensemble.

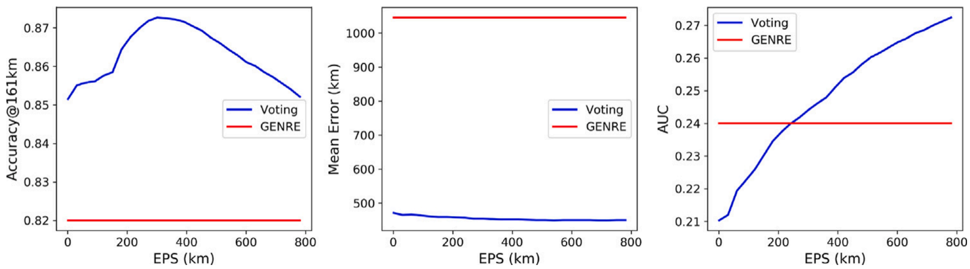

Fig. 14. Impact of eps on the performance of the voting ensemble.

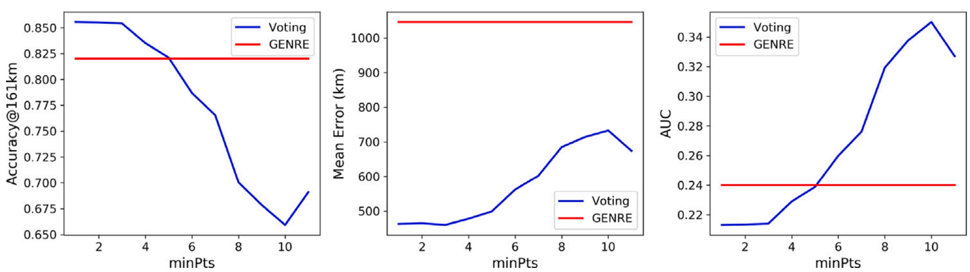

Fig. 15. Impact of $minPts$ on the performance of the voting ensemble.

4.7.2. Parameters

In the first experiment, the DBSCAN parameter eps was defined as $eps \in 1, \dots, 800$ at a step size of 30. Fig. 14 shows the performance of the voting ensemble as the change of eps. The red line denotes the performance of the best individual approach, GENRE. We can see that eps has a distinct impact on *ME, Accuracy@161km*, and *AUC*. The best *Accuracy@161km* is achieved when eps is set to 350 km, while as the increase of eps, *ME* decreases slightly from 470 km to 450 km, and *AUC* increases rapidly from 0.21 to 0.27.

In the second experiment, the DBSCAN parameter $minPts$ was defined as $minPts \in 1, \dots, 11$ at a step size of 1. Fig. 15 shows the result of the voting ensemble as the change of $minPts$. We can see $minPts$ has a large impact on the performance of the voting ensemble. The best performance is reached when $minPts$ is set to 1 and 2. The higher the $minPts$, the lower the performance.

5. Conclusion

In this paper, we thoroughly evaluate 20 toponym disambiguation approaches and investigate how voting ensembles that combine several individual approaches can further push the state-of-the-art performance. Experimental results on 12 public datasets of six types prove the generalizability and robustness of the voting approach. The deep learning-based ELs (i.e., GENRE and BLINK) that are pretrained on nearly 10 million Wikipedia entities show impressive disambiguation ability, performing much better than toponym resolution approaches. However, there is a trade-off between correctness and speed since the voting approach and the two ELs take much more time than most others. Moreover, there is still space for improving the performance of resolving fine-grained places, such as POIs, natural features, and traffic ways. This will be one of our future research tasks. Furthermore, we plan to provide HTTP APIs, with which users can easily utilize the voting approach and other approaches.

CRediT authorship contribution statement

Xuke Hu: Conceptualization, Methodology, Software, Data curation, Writing – original draft. **Yeran Sun:** Methodology, Data curation, Writing – review & editing. **Jens Kersten:** Conceptualization, Writing – review & editing. **Zhiyong Zhou:** Data curation. **Friederike Klan:** Writing – review & editing. **Hongchao Fan:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data and code availability

The code and data supporting this study's findings are available on GitHub with the link https://github.com/uhuohuy/toponym-disambiguation-voting.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jag.2023.103191.

References

2016–2022. entity-fishing. https://github.com/kermitt2/entity-fishing.

Ardanuy, Mariona Coll, Beavan, David, Beelen, Kaspar, Hosseini, Kasra, Lawrence, Jon, McDonough, Katherine, Nanni, Federico, van Strien, Daniel, Wilson, Daniel CS, 2022. A dataset for toponym resolution in nineteenth-century english newspapers. J. Open Humanit. Data 8.

Ardanuy, Mariona Coll, Sporleder, Caroline, 2017. Toponym disambiguation in historical documents using semantic and geographic features. In: Proceedings of the 2nd International Conference on Digital Access To Textual Cultural Heritage. pp. 175–180.

Auer, Sören, Bizer, Christian, Kobilarov, Georgi, Lehmann, Jens, Cyganiak, Richard, Ives, Zachary, 2007. Dbpedia: A nucleus for a web of open data. In: The Semantic Web. Springer, pp. 722–735.

Barba, Edoardo, Procopio, Luigi, Navigli, Roberto, 2022. ExtEnD: Extractive entity disambiguation. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online and Dublin, Ireland.

de Bruijn, Jens A., de Moel, Hans, Jongman, Brenden, Wagemaker, Jurjen, Aerts, Jeroen C.J.H., 2018. TAGGS: Grouping tweets to improve global geoparsing for disaster response. J. Geovisualization Spatial Anal. 2 (1), 2.

Cardoso, Ana Bárbara, Martins, Bruno, Estima, Jacinto, 2021. A novel deep learning approach using contextual embeddings for toponym resolution. ISPRS Int. J. Geo-Inf. 11 (1), 28.

De Cao, Nicola, Izacard, Gautier, Riedel, Sebastian, Petroni, Fabio, 2021. Autoregressive entity retrieval. In: International Conference on Learning Representations. URL https://openreview.net/forum?id=5k8F6UU39V.

DeLozier, Grant, Baldridge, Jason, London, Loretta, 2015. Gazetteer-independent toponym resolution using geographic word profiles. In: Twenty-Ninth AAAI Conference on Artificial Intelligence.
DeLozier, Grant, Wing, Benjamin, Baldridge, Jason, Nesbit, Scott, 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). pp. 188–198.
Devlin, Jacob, Chang, Ming-Wei, Lee, Kenton, Toutanova, Kristina, 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
Fize, Jacques, Moncla, Ludovic, Martins, Bruno, 2021. Deep learning for toponym resolution: Geocoding based on pairs of toponyms. ISPRS Int. J. Geo-Inf. 10 (12), 818.
Gregory, Ian, Donaldson, Christopher, Murrieta-Flores, Patricia, Rayson, Paul, 2015. Geoparsing, GIS, and textual analysis: current developments in spatial humanities research. Int. J. Humanities Arts Comput. 9 (1), 1–14.
Gritta, Milan, Pilehvar, Mohammad, Collier, Nigel, 2018a. Which melbourne? Augmenting geocoding with maps.
Gritta, Milan, Pilehvar, Mohammad Taher, Collier, Nigel, 2020. A pragmatic guide to geoparsing evaluation. Lang. Resour. Eval. 54 (3), 683–712.
Gritta, Milan, Pilehvar, Mohammad Taher, Limsopatham, Nut, Collier, Nigel, 2018b. What's missing in geographical parsing? Lang. Resour. Eval. 52 (2), 603–623.
Grover, Claire, Tobin, Richard, Byrne, Kate, Woollard, Matthew, Reid, James, Dunn, Stuart, Ball, Julian, 2010. Use of the Edinburgh geoparser for georeferencing digitized historical collections. Phil. Trans. R. Soc. A 368 (1925), 3875–3889.
Halterman, Andrew, 2017. Mordecai: Full text geoparsing and event geocoding. J. Open Source Softw. 2 (9), 91.
Hoang, Thi Bich Ngoc, Mothe, Josiane, 2018. Location extraction from tweets. Inf. Process. Manage. 54 (2), 129–144.
Hoffart, Johannes, Yosef, Mohamed Amir, Bordino, Ilaria, Fürstenau, Hagen, Pinkal, Manfred, Spaniol, Marc, Taneva, Bilyana, Thater, Stefan, Weikum, Gerhard, 2011. Robust disambiguation of named entities in text. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. pp. 782–792.
Hu, Xuke, Al-Olimat, Hussein, Kersten, Jens, Wiegmann, Matti, Klan, Friederike, Sun, Yeran, Fan, Hongchao, 2021. GazPNE: Annotation-free deep learning for place name extraction from microblogs leveraging gazetteer and synthetic data by rules. Int. J. Geogr. Inf. Sci. 1–28.
Hu, Xuke, Zhou, Zhiyong, Li, Hao, Hu, Yingjie, Gu, Fuqiang, Kersten, Jens, Fan, Hongchao, Klan, Friederike, 2022a. Location reference recognition from texts: A survey and comparison. arXiv preprint arXiv:2207.01683.
Hu, Xuke, Zhou, Zhiyong, Sun, Yeran, Kersten, Jens, Klan, Friederike, Fan, Hongchao, Wiegmann, Matti, 2022b. GazPNE2: A general place name extractor for microblogs fusing gazetteers and pretrained transformer models. IEEE Internet Things J. 1.
van Hulst, Johannes M., Hasibi, Faegheh, Dercksen, Koen, Balog, Krisztian, de Vries, Arjen P., 2020. REL: An entity linker standing on the shoulders of giants. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '20, ACM.
Kamalloo, Ehsan, Rafiei, Davood, 2018. A coherent unsupervised model for toponym resolution. In: Proceedings of the 2018 World Wide Web Conference. pp. 1287–1296.
Karimzadeh, Morteza, Pezanowski, Scott, MacEachren, Alan M., Wallgrün, Jan O., 2019. GeoTxt: A scalable geoparsing system for unstructured text geolocation. Trans. GIS 23 (1), 118–136.
Katz, Philipp, Schill, Alexander, 2013. To learn or to rule: two approaches for extracting geographical information from unstructured text. In: Data Mining and Analytics 2013 (AusDM'13). vol. 117.
Ke, Guolin, Meng, Qi, Finley, Thomas, Wang, Taifeng, Chen, Wei, Ma, Weidong, Ye, Qiwei, Liu, Tie-Yan, 2017. Lightgbm: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30.
Khan, Kamran, Rehman, Saif Ur, Aziz, Kamran, Fong, Simon, Sarasvady, Sababady, 2014. DBSCAN: Past, present and future. In: The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014). IEEE, pp. 232–238.
Kulkarni, Sayali, Jain, Shailee, Hosseini, Mohammad Javad, Baldridge, Jason, Ie, Eugene, Zhang, Li, 2020. Spatial language representation with multi-level geocoding. arXiv preprint arXiv:2008.09236.
Le, Phong, Titov, Ivan, 2018. Improving entity linking by modeling latent relations between mentions. arXiv preprint arXiv:1804.10637.
Lieberman, Michael D., Samet, Hanan, 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. pp. 731–740.
Lieberman, Michael D., Samet, Hanan, Sankaranarayanan, Jagan, 2010. Geotagging with local lexicons to build indexes for textually-specified spatial data. In: 2010 IEEE 26th International Conference on Data Engineering (ICDE 2010). IEEE, pp. 201–212.
Liu, Zilong, Janowicz, Krzysztof, Cai, Ling, Zhu, Rui, Mai, Gengchen, Shi, Meilin, 2022. Geoparsing: Solved or biased? An evaluation of geographic biases in geoparsing. AGILE: GISci. Ser. 3, 1–13.
Mendes, Pablo N., Jakob, Max, García-Silva, Andrés, Bizer, Christian, 2011. DBpedia spotlight: shedding light on the web of documents. In: Proceedings of the 7th International Conference on Semantic Systems. pp. 1–8.
Möller, Cedric, Lehmann, Jens, Usbeck, Ricardo, 2022. Survey on english entity linking on wikidata: Datasets and approaches. Semantic Web (Preprint), 1–42.
Orr, Laurel, Leszczynski, Megan, Arora, Simran, Wu, Sen, Guha, Neel, Ling, Xiao, Re, Christopher, 2020. Bootleg: Chasing the tail with self-supervised named entity disambiguation. arXiv preprint arXiv:2010.10363.
Qi, Tao, Ge, Suyu, Wu, Chuhan, Chen, Yubo, Huang, Yongfeng, 2019. THU_NGN at SemEval-2019 task 12: Toponym detection and disambiguation on scientific papers. In: Proceedings of the 13th International Workshop on Semantic Evaluation. pp. 1302–1307.
Qiu, Qinjun, Xie, Zhong, Wang, Shu, Zhu, Yunqiang, Lv, Hairong, Sun, Kai, ChineseTR: A weakly supervised toponym recognition architecture based on automatic training data generator and deep neural network. Trans. GIS.
Rayson, Paul, Reinhold, Alex, Butler, James, Donaldson, Chris, Gregory, Ian, Taylor, Joanna, 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. pp. 9–15.
Santos, João, Anastácio, Ivo, Martins, Bruno, 2015. Using machine learning methods for disambiguating place references in textual documents. GeoJournal 80 (3), 375–392.
Scott, Peter, Bader, Martin K-F, Burgess, Treena, Hardy, Giles, Williams, Nari, 2019. Global biogeography and invasion risk of the plant pathogen genus phytophthora. Environ. Sci. Policy 101, 175–182.
Sevgili, Özge, Shelmanov, Artem, Arkhipov, Mikhail, Panchenko, Alexander, Biemann, Chris, 2022. Neural entity linking: A survey of models based on deep learning. Semantic Web (Preprint), 1–44.
Speriosu, Michael, Baldridge, Jason, 2013. Text-driven toponym resolution using indirect supervision. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1466–1476.
Wallgrün, Jan Oliver, Karimzadeh, Morteza, MacEachren, Alan M., Pezanowski, Scott, 2018. GeoCorpora: building a corpus to test and train microblog geoparsers. Int. J. Geogr. Inf. Sci. 32 (1), 1–29.
Wang, Jimin, Hu, Yingjie, 2019. Are we there yet? Evaluating state-of-the-art neural network based geoparsers using EUPEG as a benchmarking platform. In: Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Geospatial Humanities. pp. 1–6.
Wang, Jimin, Hu, Yingjie, Joseph, Kenneth, 2020. NeuroTPR: A neuro-net toponym recognition model for extracting locations from social media messages. Trans. GIS.
Wang, Xiaobin, Ma, Chunping, Zheng, Huafei, Liu, Chu, Xie, Pengjun, Li, Linlin, Si, Luo, 2019. Dm_nlp at semeval-2018 task 12: A pipeline system for toponym resolution. In: Proceedings of the 13th International Workshop on Semantic Evaluation. pp. 917–923.
Weissenbacher, Davy, Magge, Arjun, O'Connor, Karen, Scotch, Matthew, Gonzalez, Graciela, 2019. Semeval-2019 task 12: Toponym resolution in scientific papers. In: Proceedings of the 13th International Workshop on Semantic Evaluation. pp. 907–916.
Weissenbacher, Davy, Tahsin, Tasnia, Beard, Rachel, Figaro, Mari, Rivera, Robert, Scotch, Matthew, Gonzalez, Graciela, 2015. Knowledge-driven geospatial location resolution for phylogeographic models of virus migration. Bioinformatics 31 (12), i348–i356.
Wikipedia, 2004. Wikipedia. PediaPress.
Wing, Benjamin, Baldridge, Jason, 2011. Simple supervised document geolocation with geodesic grids. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. pp. 955–964.
Won, Miguel, Murrieta-Flores, Patricia, Martins, Bruno, 2018. Ensemble named entity recognition (ner): evaluating ner tools in the identification of place names in historical corpora. Front. Digital Humanities 5, 2.
Wu, Ledell, Petroni, Fabio, Josifoski, Martin, Riedel, Sebastian, Zettlemoyer, Luke, 2020. Zero-shot entity linking with dense entity retrieval. In: EMNLP.
Yamada, Ikuya, Washio, Koki, Shindo, Hiroyuki, Matsumoto, Yuji, 2022. Global entity disambiguation with BERT. In: NAACL. Association for Computational Linguistics.
Yan, Zheren, Yang, Can, Hu, Lei, Zhao, Jing, Jiang, Liangcun, Gong, Jianya, 2021. The integration of linguistic and geospatial features using global context embedding for automated text geocoding. ISPRS Int. J. Geo-Inf. 10 (9), 572.
Yang, Xiyuan, Gu, Xiaotao, Lin, Sheng, Tang, Siliang, Zhuang, Yueting, Wu, Fei, Chen, Zhigang, Hu, Guoping, Ren, Xiang, 2019. Learning dynamic context augmentation for global entity linking. arXiv preprint arXiv:1909.02117.
Zhang, Yan, Chen, Zeqiang, Zheng, Xiang, Chen, Nengcheng, Wang, Yongqiang, 2021. Extracting the location of flooding events in urban systems and analyzing the semantic risk using social sensing data. J. Hydrol. 603, 127053.