

cases	doc_1		doc_2		decision	id
	authors	<ul style="list-style-type: none">Fan, HongchaoHu, XukeKersten, JensKlan, FriederikeSun, YeranZhou, Zhiyong	authors	<ul style="list-style-type: none">Fan, HongchaoHu, XukeKersten, JensKlan, FriederikeSun, YeranZhou, Zhiyong	NOT DUPLICATES	391
	title	How can voting mechanisms improve the robustness and generalizability of toponym disambiguation?	title	How can voting mechanisms improve the robustness and generalizability of toponym disambiguation?		
	publication_date	2022-09-17 00:00:00	publication_date	2022-01-01 00:00:00		
	source	SupportedSources.CORE	source	SupportedSources.CORE		
	journal		journal			
	volume		volume			
	doi	None	doi	None		
	urls	<ul style="list-style-type: none">http://arxiv.org/abs/2209.08286	urls	<ul style="list-style-type: none">https://core.ac.uk/download/542541244.pdf		
	id	id4859396598458892206	id	id901040983229638843		
	abstract	A vast amount of geographic information exists in natural language texts, such as tweets and news. Extracting geographic information from texts is called Geoparsing, which includes two subtasks: toponym recognition and toponym disambiguation, i.e., to identify the geospatial representations of toponyms. This paper focuses on toponym disambiguation, which is usually approached by toponym resolution and entity linking. Recently, many novel approaches have been proposed, especially deep learning-based approaches, such as CamCoder, GENRE, and BLINK. In this paper, a spatial clustering-based voting approach that combines several individual approaches is proposed to improve SOTA performance in terms of robustness and generalizability. Experiments are conducted to compare a voting ensemble with 20 latest and commonly-used approaches based on 12 public datasets, including several highly ambiguous and challenging datasets (e.g., WikToR and CLDW). The datasets are of six types: tweets, historical documents, news, web pages, scientific articles, and Wikipedia articles, containing in total 98,300 places across the world. The results show that the voting ensemble performs the best on all the datasets, achieving an average Accuracy@161km of 0.86, proving the generalizability and robustness of the voting approach. Also, the voting ensemble drastically improves the performance of resolving fine-grained places, i.e., POIs, natural features, and traffic ways.Comment: 32 pages, 15 figure	abstract	A vast amount of geospatial information exists in natural language texts, such as tweets and news. Extracting geospatial information from texts is called Geoparsing, which includes two subtasks: toponym recognition and toponym disambiguation, i.e., to identify the geospatial representations of toponyms. This paper focuses on toponym disambiguation, which is approached by toponym resolution and entity linking. Recently, many novel approaches have been proposed, especially deep learning-based, such as CamCoder, GENRE, and BLINK. In this paper, a spatial clustering-based voting approach combining several individual approaches is proposed to improve SOTA performance regarding robustness and generalizability. Experiments are conducted to 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 places across the world. 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. Besides, it drastically improves the performance of resolving fine-grained places, i.e., POIs, natural features, and traffic ways		
	versions		versions			