cases	doc_1		doc_2		decision	id
	authors	• Mozetic, Igor • Reba, Kristjan • Robnik-Sikonja, Marko	**authors**	• Marko Robnik-Sikonja • Kristjan Reba • Igor Mozetic	DUPLICATES	248
	title	Cross-lingual Transfer of Sentiment Classifiers	**title**	Cross-lingual Transfer of Sentiment Classifiers		
	publication_date	2021-03-24 00:00:00	**publication_date**	2020-05-15 10:15:27+00:00		
	source	SupportedSources.CORE	**source**	SupportedSources.ARXIV		
	journal	SlovenÅ¡Äina 2 0 empirical applied and interdisciplinary research	**journal**	None		
	volume		**volume**			
	doi	10.4312/slo2.0.2021.1.1-25	**doi**			
	urls	• https://core.ac.uk/download/475064809.pdf	**urls**	• http://arxiv.org/pdf/2005.07456v3 • http://arxiv.org/abs/2005.07456v3 • http://arxiv.org/pdf/2005.07456v3		
	id	id3664549880125155474	**id**	id3695174780573452741		
	abstract	Word embeddings represent words in a numeric space so that semantic relations between words are represented as distances and directions in the vector space. Cross-lingual word embeddings transform vector spaces of different languages so that similar words are aligned. This is done by constructing a mapping between vector spaces of two languages or learning a joint vector space for multiple languages. Cross-lingual embeddings can be used to transfer machine learning models between languages, thereby compensating for insufficient data in less-resourced languages. We use cross-lingual word embeddings to transfer machine learning prediction models for Twitter sentiment between 13 languages. We focus on two transfer mechanisms that recently show superior transfer performance. The first mechanism uses the trained models whose input is the joint numerical space for many languages as implemented in the LASER library. The second mechanism uses large pretrained multilingual BERT language models. Our experiments show that the transfer of models between similar languages is sensible, even with no target language data. The performance of cross-lingual models obtained with the multilingual BERT and LASER library is comparable, and the differences are language-dependent. The transfer with CroSloEngual BERT, pretrained on only three languages, is superior on these and some closely related languages.Comment: 18 pages, 8 table	**abstract**	Word embeddings represent words in a numeric space so that semantic relations between words are represented as distances and directions in the vector space. Cross-lingual word embeddings transform vector spaces of different languages so that similar words are aligned. This is done by constructing a mapping between vector spaces of two languages or learning a joint vector space for multiple languages. Cross-lingual embeddings can be used to transfer machine learning models between languages, thereby compensating for insufficient data in less-resourced languages. We use cross-lingual word embeddings to transfer machine learning prediction models for Twitter sentiment between 13 languages. We focus on two transfer mechanisms that recently show superior transfer performance. The first mechanism uses the trained models whose input is the joint numerical space for many languages as implemented in the LASER library. The second mechanism uses large pretrained multilingual BERT language models. Our experiments show that the transfer of models between similar languages is sensible, even with no target language data. The performance of cross-lingual models obtained with the multilingual BERT and LASER library is comparable, and the differences are language-dependent. The transfer with CroSloEngual BERT, pretrained on only three languages, is superior on these and some closely related languages.		
	versions		**versions**			