Using Multilingual Neural Re-ranking Models for Low Resource Target Languages in Cross-lingual Document Detection

Caitlin Westerfield[1]

[1]Language Information and Learning Lab, Yale University, New Haven, CT

LILY Lab

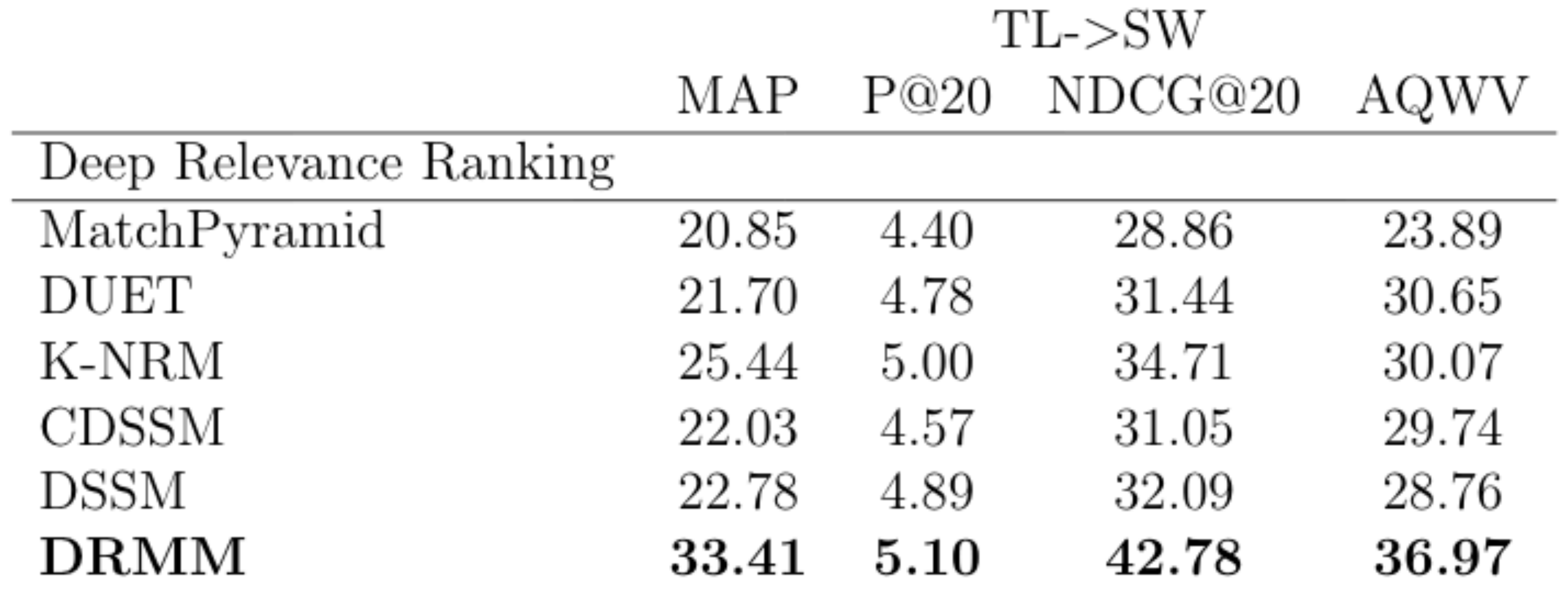
Abstract

Low-resource target languages introduce many challenges for cross-lingual document detection (CLDD) and re-ranking. First, while CLDD can be reduced to monolingual information retrieval by document translation using machine translation (MT) systems, such MT systems suffer from the lack of parallel data for low-resource target languages. Second, recent neural retrieval models that outperform traditional language modeling approaches suffer from the scarcity of relevance judgments in low-resource target languages. Due to these constraints, it is necessary to find ways to optimize the document retrieval process in ways that are not bound by the restraints on training data. This project focuses on exploring a variety of existing monolingual neural re-ranking models and applying them to the task of multilingual document retrieval.

Materials and Methods

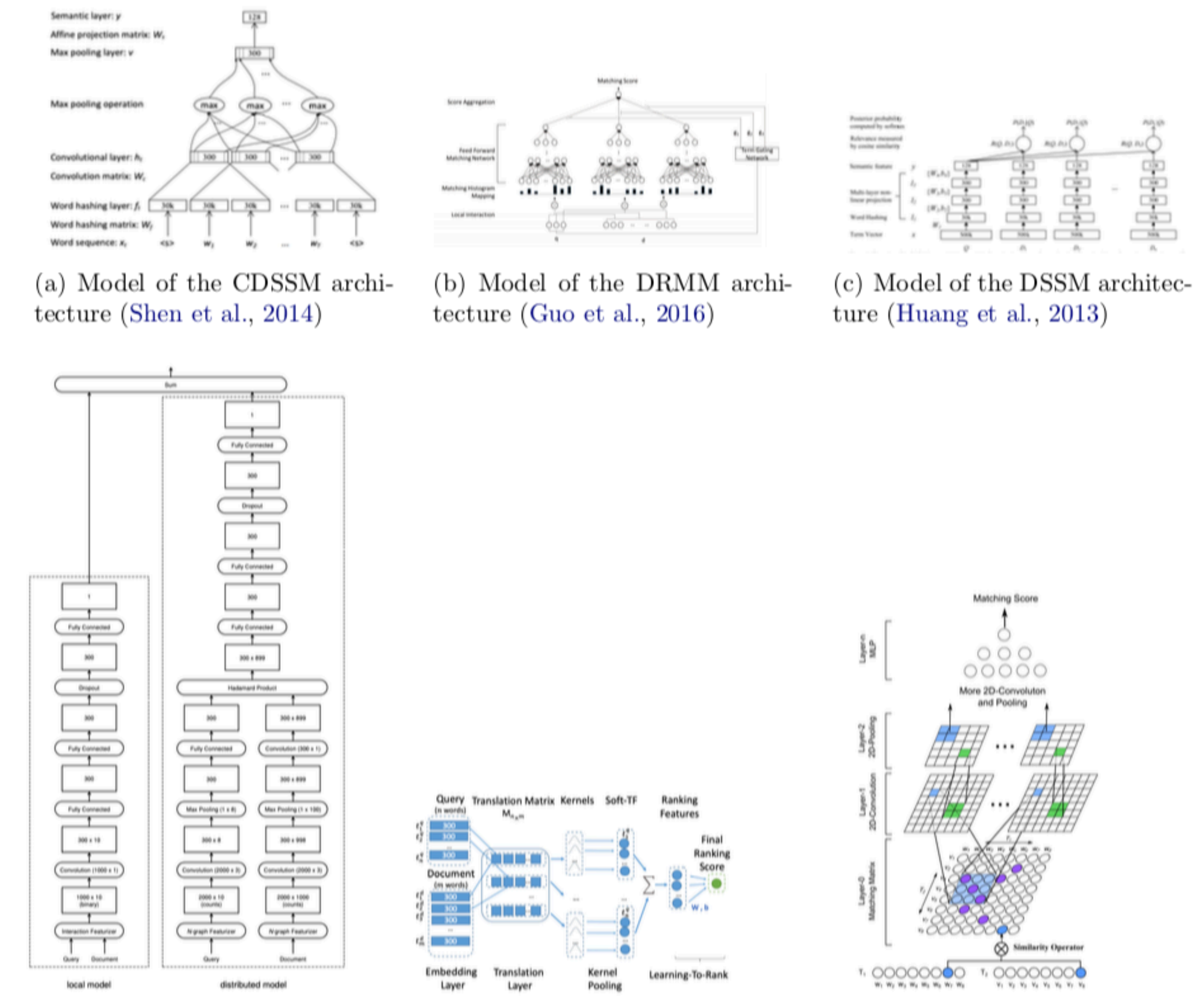
To go about determining the best neural re-ranking system to use on multilingual systems, I relied heavily on the MatchZoo pre-existing implementations of the CDSSM, DRMM, DSSM, DUET, K-NRM, and MatchPyramid models (Fan et al., 2017). While these models were already structured for monolingual systems, I altered the implementation to use English queries to train on Swahili documents and test on Tagalog documents, as well as train on Tagalog documents and test on Swahili documents. I then calculated necessary evaluation metrics using scripts found in MatchZoo as well as in the TREC dataset that is associated with MATERIAL 1. The output of the evaluation provided the necessary information to compare the re-ranking system and MT system pairs.

The re-ranking systems were trained on the filtered outputs of the machine translation systems. This was beneficial because the systems such as DBQT, PSQ, SMT, and NMT filter out any documents that they deem extraordinarily irrelevant which means that the training data was not primarily negatively labeled documents.



	TL->SW			
	MAP	P@20	NDCG@20	AQWV
Deep Relevance Ranking				
MatchPyramid	20.85	4.40	28.86	23.89
DUET	21.70	4.78	31.44	30.65
K-NRM	25.44	5.00	34.71	30.07
CDSSM	22.03	4.57	31.05	29.74
DSSM	22.78	4.89	32.09	28.76
DRMM	**33.41**	**5.10**	**42.78**	**36.97**

Table 1. MATERIAL Results by Re-Ranking System



(a) Model of the CDSSM architecture (Shen et al., 2014)

(b) Model of the DRMM architecture (Guo et al., 2016)

(c) Model of the DSSM architecture (Huang et al., 2013)

(d) Model of the DUET architecture (Mitra et al., 2017)

(e) Model of the K-NRM architecture (Xiong et al., 2017)

(f) Model of the MatchPyramid architecture (Pang et al., 2016)

Figure 1. Pre-Existing Monolingual Neural Re-ranking Models

Results

The results displayed in Table 1 show that the implementation of DRMM drastically outperforms other methods in the MAP, NDCG@20, and AQWV metrics. These results indicate that training on the original pre-translated document text for information retrieval tasks is highly beneficial because it assigns relevance scores that more closely match the accurate relevance scores and thus makes cutoff learning more beneficial.

It is clear that overall DRMM produces the net best outcome for multilingual systems, which indicates that the features that are critical to DRMM also are most important for multilingual tasks. Since DRMM does not rely on location of terms within the document, this indicates that for multilingual tasks term location is not critical for predicting relevance to a query. Additionally, DRMM's signal matching approach using histograms is something to explore further as a potentially good structure for multilingual CLIR tasks.

Conclusion

This work proved that using neural re-ranking models significantly alter performance of CLIR systems for English Queries on Swahili Documents. Additionally, DRMM is the pre- existing monolingual re-ranking system that performs the best in a multilingual setting. The systems that work for re-ranking are also representative of what representation techniques could work on other CLIR tasks. Thus, due to the success of DRMM, it is likely that other tasks could benefit from focusing on signal-based matching using histograms and the discarding of term location information.

Acknowledgement

Thank you to Dragomir Radev, Rui Zhang, and others on the MATERIAL project for their guidance and support on this project.