|  |  |
| --- | --- |
| Title | Deep Learning for Entity Matching: A Design Space Exploration |
| Author | Sidharth Mudgal,Han Li,Theodoros Rekatsinas,AnHai Doan,Youngchoon Park  ,Ganesh Krishnan,Rohit Deep,Esteban Arcaute,Vijay Raghavendra |
| URL | http://pages.cs.wisc.edu/~anhai/papers1/deepmatcher-sigmod18.pdf |
| Content | 本文尝试将深度学习应用在实体匹配问题上  3 kinds of EM problem: structured data instances, textual instances, dirty instances  4 kinds of DL methods: SIF, RNN, Attention, Hybrid  The results show that DL does not outperform current solutions on structured EM, but it can significantly outperform them on textual and dirty EM.  Experiment:  11 EM tasks for structured instances, 6 EM tasks for dirty instances, 6 tasks for dirty instances with the number of labeled instances ranging from 450 to 250k  It compare the four DL solutions( SIF, RNN, Attention, Hybrid) with Magellan( a state-of-the-art open-source learning-based EM solution  Result:  Our results show that DL solutions are competitive with Magel-  lan on structured instances (87.9% vs 88.8% average F 1 ), but require far longer training time (5.4h vs 1.5m on average). Thus, it is not clear to what extent DL can help structured EM (compared to just using today learning-based EM solutions). |
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|
|