

ThValRec: Threshold Value Recommendation Approach for Ontology Matching*

Kumar Vidhani^[0000-0002-2412-6391], Gurpriya Bhatia^[0000-0002-7511-8543],
Mangesh Gharote^[0000-0002-4942-2429], and Sachin Lodha^[0000-0001-5771-4977]

54B, TRDDC, Tata Consultancy Services Ltd., Hadapsar, Pune, Maharashtra
-411013 {kumar.vidhani, gurpriya.bhatia, mangesh.g, sachin.lodha}@tcs.com

Abstract. The determination of threshold is a complex and a time consuming task. Existing threshold value recommendation approaches are either not generalizable or requires further improvement in accuracy. In this paper, we propose an approach that computes two properties namely, symmetric and transitive, on the confidence values computed by an ontology matching algorithm in order to recommend the threshold. We demonstrate the effectiveness of our solution through experiments by comparing our solution with the hierarchical agglomerative clustering.

Keywords: Threshold Value Recommendation · Symmetric and Transitive Properties · Machine Set · Ontology Matching.

1 Introduction

Martinez-Gil and Aldana-Montes have highlighted the determination of threshold as a complex and time consuming task [1]. After producing an ontology alignment, a threshold value is specified to produce final alignment. In this paper, we propose a *Threshold Value Recommendation* (ThValRec) approach that defines two properties namely, symmetric and transitive on the confidence values computed by an ontology matching algorithm. Through these properties, ThValRec captures whether ontology matching algorithm computes a confidence value for a pair of concepts consistently or not and hence only use consistent pairs to compute final threshold.

2 Approach

As shown in the figure 1, ThValRec consists of the following steps.

Run the ontology matching algorithm on a pair of ontologies and generate a set of correspondences.

Convert the set of correspondences which is in many-to-many form into one-to-one form using the linear optimization.

Select the correspondences (of step 2) and filter them with respect to symmetric

* Copyright © for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

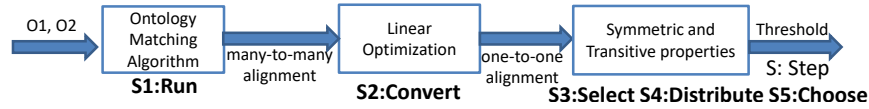


Fig. 1. ThValRec approach to compute threshold value

Table 1. Comparison between ThValRec (Thvr) approach and HAC. $\delta = 0.1$

OntologyPair	Threshold Value						F-measure					
	fastText		WuPalmer		NGram		fastText		WuPalmer		NGram	
	T_{Thvr}	T_{Hac}	T_{Thvr}	T_{Hac}	T_{Thvr}	T_{Hac}	F_{Thvr}	F_{Hac}	F_{Thvr}	F_{Hac}	F_{Thvr}	F_{Hac}
cmt_Conference	0.915	0.9	1	0.3	1	0.8	0.417	0.4	0.462	0.136	0.435	0.429
cmt_confOf	1	0.9	1	0.3	1	0.8	0.417	0.417	0.5	0.175	0.417	0.417
cmt_edas	1	0.9	0.941	0.3	1	0.8	0.609	0.609	0.615	0.174	0.667	0.667
cmt_ekaw	1	0.9	1	0.3	1	0.8	0.556	0.556	0.5	0.192	0.526	0.5
cmt_iasted	1	0.9	1	0.2	1	0.8	0.889	0.889	0.6	0.082	0.889	0.727
cmt_sigkdd	1	0.9	1	0.2	1	0.8	0.727	0.782	0.667	0.235	0.696	0.696
Conference_confOf	1	0.9	1	0.2	1	0.8	0.667	0.667	0.519	0.227	0.667	0.643
Conference_edas	1	0.9	1	0.1	1	0.8	0.581	0.581	0.5	0.159	0.581	0.514
Conference_ekaw	1	0.9	0.938	0.2	0.917	0.8	0.41	0.41	0.375	0.274	0.439	0.444
Conference_iasted	1	0.9	0.933	0.2	1	0.8	0.4	0.4	0.333	0.088	0.4	0.4
Conference_sigkdd	1	0.9	1	0.2	1	0.8	0.583	0.56	0.538	0.205	0.56	0.519
confOf_edas	1	0.9	0.952	0.1	1	0.8	0.564	0.564	0.524	0.283	0.564	0.55
confOf_ekaw	1	0.9	1	0.2	1	0.8	0.606	0.606	0.629	0.374	0.606	0.611
confOf_iasted	1	0.9	1	0.2	1	0.5	0.615	0.714	0.471	0.148	0.615	0.363
confOf_sigkdd	1	0.9	1	0.2	1	0.5	0.727	0.727	0.667	0.111	0.667	0.444
edas_ekaw	1	0.9	0.929	0.3	1	0.8	0.474	0.474	0.4	0.124	0.462	0.537
edas_iasted	1	0.9	0.933	0.3	1	0.8	0.519	0.519	0.457	0.1	0.519	0.551
edas_sigkdd	1	0.9	0.967	0.2	1	0.8	0.609	0.609	0.56	0.228	0.583	0.56
ekaw_iasted	1	0.9	0.967	0.2	1	0.8	0.706	0.706	0.476	0.104	0.706	0.632
ekaw_sigkdd	1	0.9	1	0.2	1	0.8	0.667	0.667	0.7	0.214	0.632	0.6
iasted_sigkdd	1	0.9	1	0.2	1	0.8	0.733	0.774	0.595	0.273	0.71	0.765

and transitive properties.

Distribute the filtered correspondences (of step3) into a set of δ -length intervals. $\delta \in [0, 1]$ is a value chosen by a user.

Choose the top interval's correspondences to determine a threshold value.

3 Experiments

We have conducted experiments on the OAEI 2019 conference dataset to compare threshold values recommended by ThValRec with the hierarchical agglomerative clustering (HAC) [2] viz-a-viz three ontology matching algorithms: fastText (v0.9.1), WuPalmer (nltk v3.4.5) and NGram (strsim v0.0.3 of python).

As shown in the table 1, HAC mostly recommends three threshold values, 0.5, 0.8 and 0.9, for the fastText and NGram algorithms across all ontology pairs. In case of WuPalmer, HAC recommends low threshold values viz-a-viz fastText and NGram, and, performs very poorly in comparison to ThValRec approaches. This demonstrates that HAC may not recommend consistent values for different ontology matching algorithms.

References

1. Martinez-Gil, J., Aldana-Montes, J.F.: An overview of current ontology meta-matching solutions. The Knowledge Engineering Review **27**(4), 393–412 (2012)
2. dos Santos, J.B., Heuser, C.A., Moreira, V.P., Wives, L.K.: Automatic threshold estimation for data matching applications. Information Sciences **181**(13), 2685–2699 (2011)