

# Aggregation of Ontology Matchers in Lieu of a Reference Ontology

(Full/Regular Research Papers)

Mohammad Al Boni

*Student Member, IEEE*

Center for Advanced Vehicular Systems  
Mississippi State University, MS 39762 USA  
Email: mboni@cavs.msstate.edu

Derek T. Anderson

*Member, IEEE*

Electrical and Computer Engineering Department  
Mississippi State University, MS 39762 USA  
Email: anderson@ece.msstate.edu

**Abstract**—Ontologies are widely used to represent knowledge in different domains. As a result, numerous methods have been put forth to match ontologies. No technique has been shown to be robust across all domains. Furthermore, ontology matchers typically make use of a reference ontology. However, this is not guaranteed to exist. In this article, the fuzzy integral is used to aggregate multiple ontology matchers in lieu of a reference ontology. Specifically, we present a way to derive the fuzzy measure based on ideas from crowd sourcing when the worth of individuals is not known. Preliminary results are presented to show the robustness of our approach across different domains.

**Keywords**—Ontology matching, fuzzy measure, fuzzy integral, crowd sourcing, measure of agreement

## I. INTRODUCTION

An ontology is a knowledge representation language. Specifically, it is a set of concepts and relationships between those concepts to describe a certain knowledge domain. Ontologies have been used in many disciplines, especially biology (e.g., the gene ontology [4]). Different researchers have designed distinct ontologies with respect to various domains. This often causes redundancy as many ontologies may partially, or fully, overlap. Ontology matching techniques have been proposed to help address and mitigate this challenge. In the last decade, many matchers have been put forth. A recent survey is presented in [2]. In order to improve stability, several aggregation methods have been used to combine results from different matchers [1, 14, 21]. Many datasets have been deployed to evaluate the performance of ontology matchers and their aggregation such as the OAEI [12] and I<sup>3</sup>CON [13]. Metrics such as Precision, Recall and Fmeasure are used to evaluate the performance of different matchers and their aggregation. These metrics rely on using a reference ontology. A reference ontology (R) is an alignment, designed by experts, contains exact matching between two ontologies. However, real world problems are typically not provided a reference ontology. Otherwise, there is no need for the help of matchers in the first place. Experiments demonstrate that matchers behave distinctively when used on different domains. Therefore, without a reference ontology, there is no way to find the best matcher with respect to a certain domain.

The goal of this article is to create a technique to fuse multiple matchers using the fuzzy integral (FI) based on ideas from crowd sourcing. The main contribution of this article is

the idea of extracting how to aggregate the matchers based on agreement, versus a prior non-flexible assignment of worth to the different matchers (which is often not known). In this respect, aggregation is flexible and can change with the problem and domain (versus a global solution which does not likely exist). Namely, we propose a fuzzy measure (FM) of agreement that we previously applied to crowd sourcing.

This paper is organized as follows. Section II presents related work in ontology matcher aggregation. The process of ontology matching is then described in section III. Foundations of the FM and FI are provided in section IV and a FM of agreement is defined in section IV. Experiments, preliminary results and evaluation are presented in section V. Section VI concludes the paper and it discusses the future work.

## II. RELATED WORK

A number of works have been put forth in the literature regarding combining several ontology matching techniques. In [14, 21], the authors used a genetic algorithm (GA) to learn the weight of each matching algorithm for an operator like or is the weighted sum. In [1], we put forth a more powerful non-linear aggregation based on the FI. Although these works improve the reliability of matching by aggregating multiple methods; these works used mechanisms (fitness functions) that depend on a reference ontology. Therefore, none of these methods are applicable when the reference ontology is not available. Without the reference ontology, all existing aggregation methods fail to capture the importance of each individual matcher and likely result in less than desirable performance. Herein, a unique way to measure agreement can address this gap. The claim is, if we cannot determine the worth of each matcher with respect to a certain domain, then we look towards the agreement to help improve the robustness of an approach. The question is how to do this for ontology matching?

## III. ONTOLOGY MATCHING AND EVALUATION

### A. Ontology Matching

Ontology matching is a process that tries to find, for each concept within one ontology (the source), the closest matched concept in the second ontology (the destination). There are several techniques to tackle this problem which can be based on: statistical analysis, string similarity, linguistic methods,

taxonomy analysis, data analysis, graph-mapping, data-type comparison, and inheritance analysis [14]. The output of the matching process is called alignment (A). The alignment consists of a set of cells, and each cell contains two terms, relation type and measure value. The alignment and the reference ontology are used to evaluate matching algorithms.

### B. Ontology matcher evaluation

Several metrics, originally from the information retrieval field [19], have been adopted in the evaluation process such as semantic Precision (P) and Recall (Re) [7]. The precision reflects the accuracy of the matching algorithm, i.e., it answers the question how many cells in the alignment are also found in the reference? The precision is:

$$P = \frac{|A \cap R|}{|A|}. \quad (1)$$

Recall determines the completeness of the matching algorithm, i.e., it answers the question how many correct matches are found by a matcher? The Recall is:

$$Re = \frac{|A \cap R|}{|R|}. \quad (2)$$

Finally, Fmeasure (F) is a compound metric that reflects both correctness and completeness,

$$F = \frac{2 \cdot P \cdot Re}{P + Re}. \quad (3)$$

We notice that all the previous metrics depend on R, which is always missing in real world applications.

## IV. FUZZY MEASURE OF AGREEMENT

### A. Fuzzy measure and integral foundations

The FI, introduced by Sugeno [17], is a powerful tool for data fusion and aggregation. The FM encodes the importance of various subsets of input sources and the FI is defined with respect to the FM. The FI has been widely used in different engineering fields such as multi-criteria decision making [8], image processing [18], and even robotics [16].

For a non-empty finite set of  $N$  input sources,  $X = \{x_1, x_2, \dots, x_N\}$ , the FM is a (set-valued) function  $g : 2^X \rightarrow [0, 1]$ . The FM has three properties [9]:

- (1)  $g(\emptyset) = 0$ ,  $g(X) = 1$  (boundary conditions),
- (2) For  $A, B \subseteq X$ , such that  $A \subseteq B$ ,  $g(A) \leq g(B)$  (the monotonicity constraint).
- (3) If  $A_n \in X$  and  $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ , then  $\lim_{n \rightarrow \infty} g(A_n) = g(\lim_{n \rightarrow \infty} A_n)$ . However, this property is not applicable in the case of finite sets..

The FM can be discussed in terms of its corresponding lattice, which is induced by the monotonicity constraints. The value in each node in the lattice is the measure for a certain subset of sources, i.e., the worth of a particular group. Figure 1 is an illustration of the FM for  $N = 3$ .

There are multiple ways to build the FM. In some applications, experts can manually define the measure. However, it might be very difficult, if possible at all, to determine all

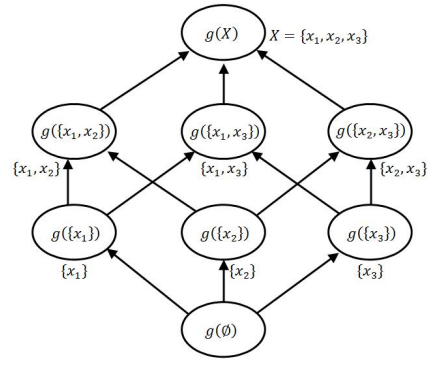


Fig. 1: Lattice representation of the FM for  $N = 3$ .

values other than the singletons. In other settings, data-driven learning methods are used to learn the measure. One method is to have a quadratic program (QP) to learn the measure based on a given data set [8]. However, the QP's complexity is relatively high and it does not scale well [15]. In order to reduce the complexity, several optimization techniques have been considered to learn the measure such as GAs [1, 3]. Furthermore, when we have a relatively small number of input sources and the values of the singletons are known, measure building techniques are employed, e.g., a S-Decomposable measure such as the Sugeno  $\lambda$ -fuzzy measure [9]. The Sugeno  $\lambda$ -measure is one of the well-known and widely used FMs. In the  $\lambda$ -measure, several sources can be combined using the rule

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B), \quad (4)$$

where  $\lambda$  is calculated using the formula

$$\lambda + 1 = \prod_{i=1}^N (1 + \lambda g^i), \quad \lambda > -1. \quad (5)$$

Data is aggregated using the FI based on the FM. The FI can take many forms, two of which we focus on here. For a finite set  $X = \{x_1, x_2, \dots, x_N\}$ , the discrete Sugeno and Choquet FIs are

$$\int_{Sugeno} h \circ g = \bigvee_{i=1}^N (h(x_{\pi(i)}) \wedge G(x_{\pi(i)})), \quad (6)$$

$$\int_{Choquet} h \circ g = \sum_{i=1}^N h(x_{\pi(i)}) [G(x_{\pi(i)}) - G(x_{\pi(i-1)})], \quad (7)$$

where:

- $h$  is the partial support function,  $h : X \rightarrow [0, 1]$ .
- $h(x_{\pi(i)})$  is the evidence provided by source  $\pi(i)$ .
- $\pi$  is a re-permutation function such that  $h(x_{\pi(1)}) \geq h(x_{\pi(2)}) \geq \dots \geq h(x_{\pi(N)})$ .
- The value  $G(x_{\pi(i)}) = g(\{x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(i)}\})$  is the measure of a set of information sources.
- $G(x_{\pi(0)}) = 0$ .

Note, both the  $\int_{Sugeno} h \circ g$  and the  $\int_{Choquet} h \circ g$  are bounded between  $\left[ \bigwedge_{i=1}^N h(x_i), \bigvee_{i=1}^N h(x_i) \right]$  and its method of aggregation is often anecdotally explained as “the best pessimistic agreement”.

## B. Measure of Agreement

In [20] we proposed a FM of agreement between  $N$  interval-valued inputs. That work was extended in [10] to overcome a natural bias with respect to the length of intervals. Herein, we consider the extension of this FM,  $g^{AG}$ , for ontology matchers. The result of an ontology matcher is a matrix of similarity values. Specifically, we put forth a similarity calculation based on the population standard deviation. It computes agreement between multiple ontology matchers in the context of  $g^{AG}$ .

A number of works have been put forth in the literature regarding distance and similarity indices. A recent survey is presented in [5]. However, all these measures compute the similarity between two objects only. In our case, we have a set of  $N$  input sources. Each one represents a  $(l * h)$  matrix which resulted from a matching algorithm where  $l$  ( $h$  respectively) is the number of terms in the source (destination respectively) ontology. Herein, we need to calculate the similarity between all combinations of sources, i.e., between multiple matrices.

The population standard deviation is a statistical measure that shows the variation in the values of elements in the population, compared to the population mean value. For a population of  $M$  elements,  $Y = \{y_1, y_2, \dots, y_M\}$ , the population standard deviation is

$$\sigma = \sqrt{\frac{1}{M} \sum_{i=1}^M (y_i - \mu)^2}, \quad (8)$$

$$\mu = \frac{\sum_{i=1}^M y_i}{M}. \quad (9)$$

Herein,  $\sigma$  is used to measure distance between two or more matrices. Herein,  $g^{AG}$  is defined as

$$\begin{aligned} g^{AG}(A_0) &= g^{AG}(A_1) = 0, \\ g^{AG}(A_i) &= \left[ \bigvee_{k_1=1}^{i-1} \bigvee_{k_2=k_1+1}^i d_\sigma(C_{\pi(k_1)}, C_{\pi(k_2)}) \right] z_2 + \\ &\left[ \bigvee_{k_1=1}^{i-2} \bigvee_{k_2=k_1+1}^{i-1} \bigvee_{k_3=k_2+1}^i d_\sigma(C_{\pi(k_1)}, C_{\pi(k_2)}, C_{\pi(k_3)}) \right] z_3 + \\ &\dots + d_\sigma(C_{\pi(1)}, C_{\pi(2)}, \dots, C_{\pi(i)}) z_i, \quad i = [2 : N], \end{aligned}$$

where

$$d_\sigma(C_{\pi(1)}, C_{\pi(2)}, \dots, C_{\pi(i)}) = \frac{\sum_{q=1}^i \sum_{j=1}^h (1 - \sigma(C_{\pi(1)}(q,j), C_{\pi(2)}(q,j), \dots, C_{\pi(i)}(q,j)))}{l * h},$$

such that  $A_0 = \emptyset, A_i = \{C_{\pi(1)}, C_{\pi(2)}, \dots, C_{\pi(i)}\}$  is the permuted set of matrices such that  $C_i$  is matrix resulted from input source (matching algorithm)  $x_i, z_i$  is the weight of each term such that  $z_2 \leq z_3 \leq \dots \leq z_N$  ( $z_i$  can take any value, such as  $\frac{i}{N}$ , which puts more importance on larger sets of sources), and  $max$  is used as a t-conorm ( $\bigvee$ ) operator. Note, since the values are each matrix are in  $[0,1]$ ,  $\sigma$  is also bound in  $[0,1]$ . Therefore,  $1 - \sigma \in [0,1]$ .

In [10] we proved that  $g^{AG}$  is monotonic and non-decreasing. Also, in order to guarantee the boundary conditions

( $g(\emptyset) = 0$  and  $g(A_N) = 1$ ),  $g^{AG}$  is normalized by

$$\tilde{g}^{AG}(A_i) = \frac{g^{AG}(A_i)}{g^{AG}(A_N)}. \quad (10)$$

The following two descriptions help to describe the inner workings of  $\tilde{g}^{AG}$  :

- (1) The worth of an individual is defined to be zero. The proposed measure of agreement is based on groups of individuals and their consensus.
- (2) The agreement of a set of matchers includes the worth of all sub-combinations to better characterize and account for all sub-agreement in the smaller groups.

## V. EXPERIMENTS AND RESULTS

In this section, several experiments were performed to show how the proposed technique behaves on different knowledge domains. We prefer the use of the I<sup>3</sup>CON [13] dataset over OAEI [12] for two reasons. First, I<sup>3</sup>CON provides a set of ontologies from diverse knowledge domains while OAEI focuses only on the bibliographic references domain. Having a wide range and real life related domains helps to support our claim that different ontology matchers have different performance characteristics in different domains. Also, all tests in OAEI (except one) are systematically built, and researchers take that into consideration when designing their matchers. Therefore, matchers might not cover all cases found in real life problems.

Several tests were performed, and we compared the performance of our aggregation relative to the individual matchers. In Figure 2, we show the matching matrices returned by each individual algorithm and our aggregated result. The first (second respectively) dimension of the matrix corresponds to concepts in the source (destination respectively) ontology. In Figure 3, we compared the individual and fused alignments. Cells detected by both algorithms are reported as blue. We are able to see how the fuzzy agreement measure captured the agreement between the different matchers.

In Figures 4 through 9, we compare the performance of the aggregated result with the individual matchers. The results show the inconsistent behavior of the existing matchers in different domains. For example, FALCON ([11]) has a very good Precision in the ‘‘Animals’’ ontology (Figure 4a), whereas in the ‘‘Pets’’ ontology (Figure 5a) it is a poor performer (1 and 0.5 respectively). The inconsistent behavior is also manifested in the big difference in Recall between the ‘‘Animals’’ and the ‘‘Sport Events’’ ontologies (0.083 and 0.893 respectively), see Figures 4b and 6b. Another good example is the Structure matcher, which got a Recall of 0.903 in the ‘‘Pets’’ ontology and it scored just 0.007 for ‘‘Sport Event’’ (Figures 5b and 6b). Even FOAM ([6]), which scored high in most ontologies (Figures 4c, 5c and 7c), scored low in the ‘‘Vehicles’’ and ‘‘Tourism’’ ontologies (Figures 8c and 9c). No matcher is expected to universally outright solve this problem. On the other hand, the performance of our tool was relatively stable throughout all domains (Figures 4 through 9). We were able to achieve high Precision (top 2) in all domains —as expected— because when many matchers agree on an alignment cell, then it will most likely be valid. And for the same reason, our tool did not score very high at the Recall factor. The crowd sourcing model is cautious enough to wait for many

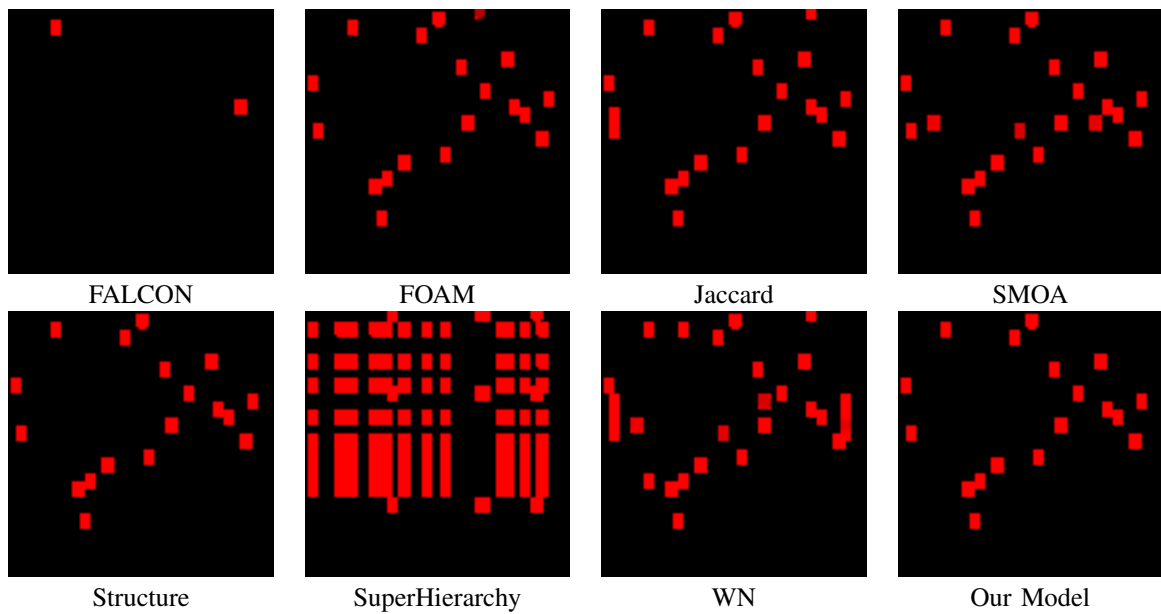


Fig. 2: Visualization of the result of different matchers and our aggregated result for the Animal ontology. Note, the rows (columns respectively) represent terms in the source ontology (destination respectively) and the red cells are values in range [0,1] that stand for the matching result which typically tend to be close to 1.

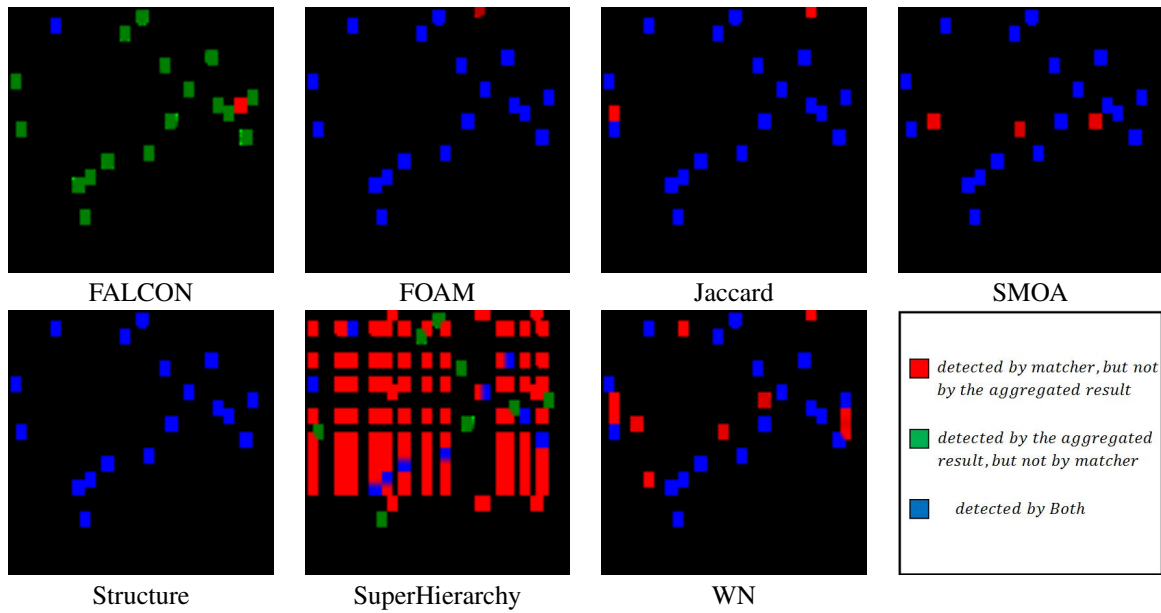


Fig. 3: Visualization of the differences between the individual matchers in figure 2. and our aggregated result.

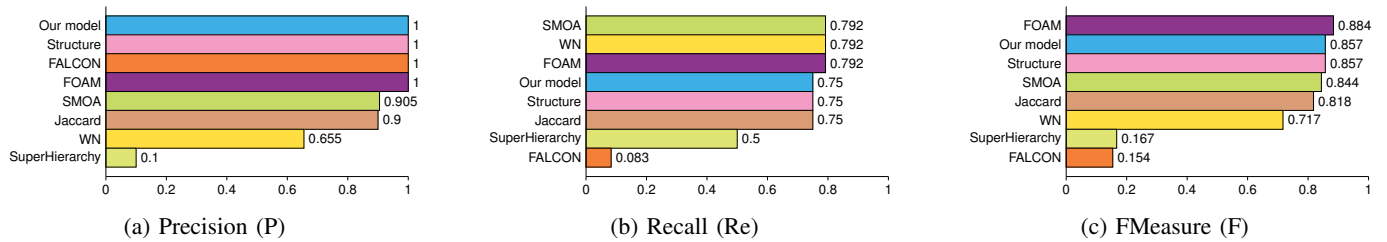


Fig. 4: P, Re, and F for the Animal ontology.

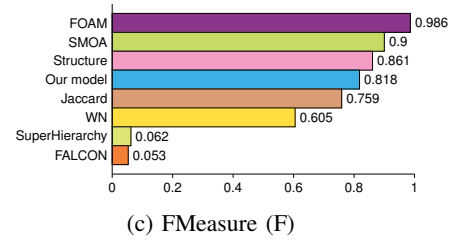
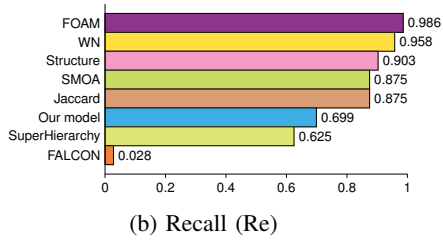
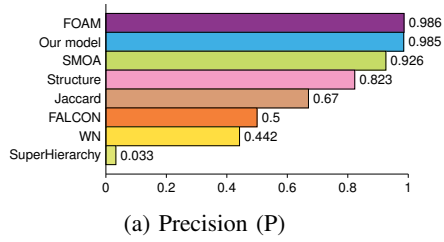


Fig. 5: P, Re, and F for the Pets ontology.

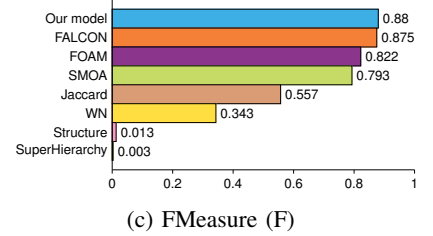
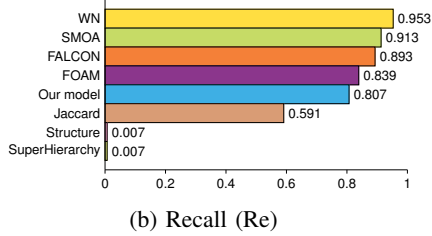
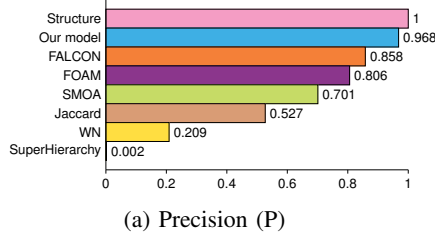


Fig. 6: P, Re, and F for the SportEvent ontology.

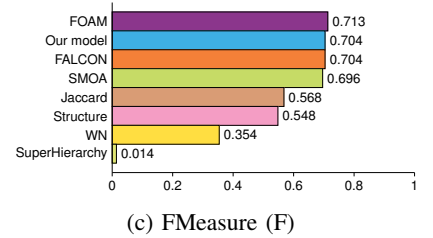
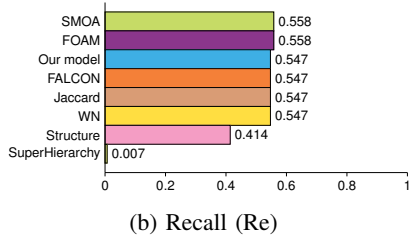
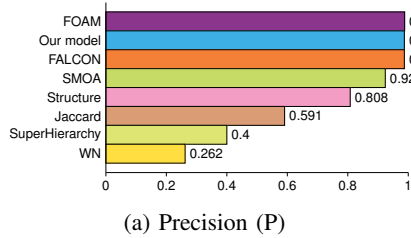


Fig. 7: P, Re, and F for the Russia ontology.

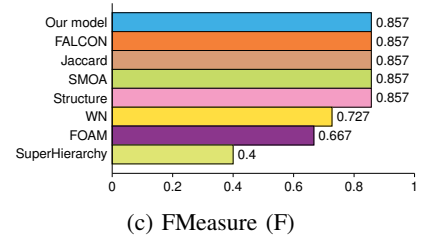
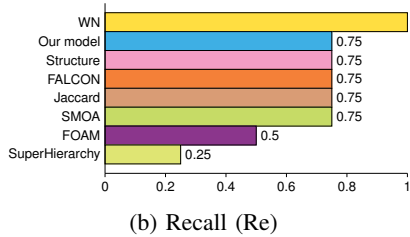
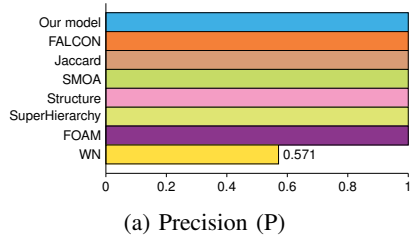


Fig. 8: P, Re, and F for the Vehicles ontology.

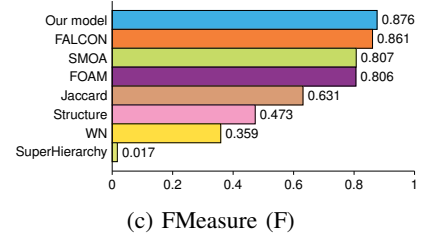
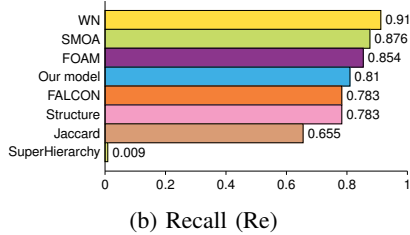
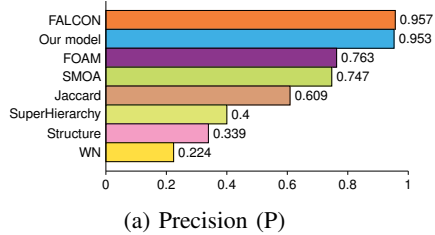


Fig. 9: P, Re, and F for the Tourism ontology.

votes from the matchers before deciding whether any cell should be considered in the output alignment or not. Given no prior knowledge on how good the individual matchers are and without the need for any reference ontology, our model did provide remarkably robust results when aggregating several matchers (top 3 with respect to Fmeasure).

## VI. CONCLUSION AND FUTURE WORK

In this paper, we presented a novel way to aggregate ontology matchers without relying on the use of a reference ontology. Specifically, we employed the FI with respect to a FM of agreement to fuse different ontology matching algorithms. That is, when nothing is known about the quality of the individuals, we can look to identify agreement among the inputs and use that measure back on the same data to guide its aggregation. In addition, we proposed the use of population standard deviation to measure the distance between two or more matrices. We showed that our model can give satisfactory and robust results in different study domains. However, few important questions need to be addressed for future work, e.g., how to increase the Recall? Should more matchers be added to increase the number of voters in the crowd? If yes, this will considerably increase the number of free parameters in the FM ( $2^N - 2$  for  $N$  sources) and thus increase complexity. How to deal with the case of having a large number of matchers? Can we use the k-additive measure? This means that we will be listening to votes of certain number of matchers (k-matchers). The use of k-additive measure might be considered risky for a small k because the measure of agreement might lose its significance. These questions and some others need to be answered so we may better understand and improve the performance of heterogeneous crowds.

## ACKNOWLEDGMENT

The first author would like to acknowledge and thank the Fulbright scholarship program for its financial support of the work performed herein.

## REFERENCES

- [1] M. Al Boni, D. T. Anderson, and R. L. King. Constraints preserving genetic algorithm for learning fuzzy measures with an application to ontology matching. In *2013 Third Annual World Conference on Soft Computing (WCSC 2013)*, San Antonio, USA, Dec. 2013.
- [2] S. Amrouch and S. Mostefai. Survey on the literature of ontology mapping, alignment and merging. In *Information Technology and e-Services (ICITeS), 2012 International Conference on*, pages 1–5. IEEE, 2012.
- [3] D. T. Anderson, J. M. Keller, and T. C. Havens. Learning fuzzy-valued fuzzy measures for the fuzzy-valued sugeno fuzzy integral. In *Computational Intelligence for Knowledge-Based Systems Design*, pages 502–511. Springer, 2010.
- [4] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000.

- [5] S.-H. Cha. Comprehensive survey on distance/similarity measures between probability density functions. *City*, 1(2):1, 2007.
- [6] M. Ehrig and Y. Sure. Foam-framework for ontology alignment and mapping-results of the ontology alignment evaluation initiative. In *Workshop on integrating ontologies*, volume 156, pages 72–76, 2005.
- [7] J. Euzenat. Semantic precision and recall for ontology alignment evaluation. In *Proc. 20th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 348–353, 2007.
- [8] M. Grabisch. The application of fuzzy integrals in multicriteria decision making. *European journal of operational research*, 89(3):445–456, 1996.
- [9] M. Grabisch, M. Sugeno, and T. Murofushi. *Fuzzy measures and integrals: theory and applications*. Springer-Verlag New York, Inc., 2000.
- [10] T. C. Havens, D. T. Anderson, C. Wagner, H. Deilamsalehy, and D. Wonnacott. Fuzzy integrals of crowd-sourced intervals using a measure of generalized accord. In *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, pages 1–8, 2013.
- [11] W. Hu and Y. Qu. Falcon-ao: A practical ontology matching system. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(3):237–239, 2008.
- [12] O. A. E. Initiative. Ontology Alignment Evaluation Dataset. <http://oaei.ontologymatching.org/tests/>, 2006. [accessed 11-October-2013].
- [13] I. Interpretation and I. Conference. Ontology Alignment Evaluation Dataset. <http://www.atl.lmco.com/projects/ontology/i3con.html>, 2004. [accessed 11-October-2013].
- [14] J. Martinez-Gil, E. Alba, and J. F. Aldana-Montes. Optimizing ontology alignments by using genetic algorithms. In *Proceedings of the workshop on nature based reasoning for the semantic Web. Karlsruhe, Germany*, 2008.
- [15] A. Mendez-Vazquez and P. Gader. Sparsity promotion models for the choquet integral. In *Foundations of Computational Intelligence, 2007. FOCI 2007. IEEE Symposium on*, pages 454–459. IEEE, 2007.
- [16] G.-Y. Pan, W.-Y. Wang, C.-P. Tsai, W.-Y. Wang, and C.-R. Tsai. Fuzzy measure based mobile robot controller for autonomous movement control. In *System Science and Engineering (ICSSE), 2011 International Conference on*, pages 649–653, 2011.
- [17] M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.
- [18] H. Tahani and J. M. Keller. Information fusion in computer vision using the fuzzy integral. *Systems, Man and Cybernetics, IEEE Trans on*, 20(3):733–741, 1990.
- [19] C. J. K. van Rijsbergen. *Information retrieval*. Butterworths, London (UK), 1975.
- [20] C. Wagner and D. Anderson. Extracting meta-measures from data for fuzzy aggregation of crowd sourced information. In *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, pages 1–8, 2012.
- [21] J. Wang, Z. Ding, and C. Jiang. Gaom: genetic algorithm based ontology matching. In *Services Computing, 2006*, pages 617–620. IEEE, 2006.