Example 5
Let us reconsider the sources from Example 4. If we apply F^* *by using the objective function given in Equation* (1)*, then we find that* $\mathrm{F}_1^*(M) = \{b, c, c, a\}$*, which is cardinality preserving, but not with an* $\mathcal{X}$*-majority. If we add the constraint of cardinality preservation with* $\mathcal{X}$*-majority to the definition of* F^**, we find that* $\mathrm{F}_2^*(M) = \{b, c, c\}$*. Moreover, if we calculate the values of the objective function, we find that:*

$$\begin{aligned} \mathcal{O}\big(\mathrm{F}_1^*(M), M\big) &= 0.51 \\ \mathcal{O}\big(\mathrm{F}_2^*(M), M\big) &= 0.47. \end{aligned}$$

This is a rather small difference in average similarity, compensated by the fact that F_2^* *offers us a more intuitive solution w.r.t. its cardinality.*

6. Conclusion

In this paper, the concept of θ-preservation of second order fusion functions (i.e., fusion functions that operate on multi-valued sources) has been introduced. Whereas traditional preservative functions choose one of the sources (i.e., one of the sources is preserved), θ-preservative fusion functions preserve the feature θ of one of the sources. Three such features have been proposed: the cardinality, the k-cut(s) and the multiplicity distribution. It has been shown how θ-preservation implies a weaker notion of idempotence. More specifically, it has been shown how θ-preservative fusion functions are idempotent in the feature space $\mathcal{X}$. Next, it was pointed out that θ-preservative fusion functions can be designed by defining fusion as an optimization problem. In that setting, θ-preservation can be added as a boundary constraint. Finally, the principle of majority voting was introduced in the framework of θ-preservation.

References

[1] Howard Newcombe, James Kennedy, S Axford, and A James. Automatic linkage of vital records. *Science*, 130(3381):954–959, 1959.

[2] Ivan Fellegi and Alan Sunter. A theory for record linkage. *American Statistical Association Journal*, 64(328):1183–1210, 1969.

[3] Matthew Jaro. Advances in record linking methodology as applied to the 1985 census of tampa florida. *Journal of the American Statistical Society*, 84(406):414–420, 1989.

[4] Antoon Bronselaer, Axel Hallez, and Guy De Tré. Extensions of fuzzy measures and the sugeno integral for possibilistic truth values. *International Journal of Intelligent Systems*, 24(2):97–117, 2009.

[5] Isabelle Bloch. Information combination operators for data fusion: A comparative review with classification. *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans*, 26(1):52–67, 1996.

[6] Alexander Bilke, Jens Bleiholder, Christoph Böhm, Karsten Draba, Felix Naumann, and Melanie Weis. Automatic data fusion with hummer. In *VLDB*, pages 1251–1254, 2005.

[7] Felix Naumann, Alexander Bilke, Jens Bleiholder, and Melanie Weis. Data fusion in three steps: Resolving inconsistencies at schema-, tuple-, and value-level. In *Bulletin of the technical committee on data engineering*, pages 21–31, 2006.

[8] Amihai Motro and Philipp Anokhin. Fusionplex: resolution of data inconsistencies in the integration of heterogeneous information sources. *Information Fusion*, 7(2):176–196, 2006.

[9] Jens Bleiholder and Felix Naumann. Data fusion. *ACM Computing Surveys*, 41(1), 2008.

[10] Antoon Bronselaer and Guy De Tré. Aspects of object merging. In *Proceedings of the NAFIPS Conference*, pages 27–32, Toronto, Canada, 07/2010 2010.

[11] Chitta Baral, Sarit Kraus, Jack Minker, and V Subrahmanian. Combining knowledge bases consisting of first-order theories. *Computational Intelligence*, 8(1):45–71, 1992.

[12] Jinxin Lin and Alberto Mendelzon. Merging databases under constraints. *International Journal of Cooperative Information Systems*, 7(1):55–76, 1998.

[13] Sébastien Konieczny and Ramon Pérez. Merging information under constraints: a logical framework. *Journal of Logic and Computation*, 12(1):111–120, 2002.

[14] Jens Bleiholder, Melanie Herschel, and Felix Naumann. Eliminating nulls with subsumption and complementation. *IEEE Data Engineering Bulletin*, 34(3):18–25, 2011.

[15] Omar Benjelloun, Hector Garcia-Molina, David Menestrina, Qi Su, Steven Euijong Whang, and Jennifer Widom. Swoosh: a generic approach to entity resolution. *VLDB Journal*, 18(1):255–276, 2009.

[16] Nicholas Rescher and Ruth Manor. On inferences from inconsistent premises. *Theory and Decision*, 1:179–217, 1970.

[17] Antoon Bronselaer, Daan Van Britsom, and Guy De Tré. A framework for multiset merging. *Fuzzy Sets and Systems*, 191:1–20, 2012.

[18] Ronald Yager. On the theory of bags. *International Journal of General Systems*, 13(1):23–27, 1986.