Landmark Explanation: a Tool for Entity Matching

(Discussion Paper)

Andrea Baraldi[1], Francesco Del Buono[1], Matteo Paganelli[1] and Francesco Guerra[1]

[1]*DIEF - University of Modena and Reggio Emilia, Modena, Italy*

Abstract

We introduce Landmark Explanation, a framework that extends the capabilities of a post-hoc perturbation-based explainer to the EM scenario. Landmark Explanation leverages on the specific schema typically adopted by the EM datasets, representing pairs of entity descriptions, for generating word-based explanations that effectively describe the matching model.

Keywords

Entity Matching, Post-hoc Explanation, Perturbation of EM datasets

1. Introduction

Machine Learning (ML) and Deep Learning (DL) models have been successfully applied to the Entity Matching (EM) problem as the state-of-the-art approaches demonstrate (e.g., DeepER [1], DeepMatcher [2], DITTO [3], AutoML [4] and others [5, 6, 7]). Nevertheless, they are black-box models: the difficulty to evaluate [8] and to interpret their behaviors [9] hampers their adoption in business scenarios.

Although many explanation systems have already been proposed in the literature (e.g., LIME [10], Shapley [11], Anchor [12], and Skater[1]), their application to EM tasks is not straightforward and only few approaches have partially addressed it [13, 14, 15, 16]. EM is conceived as a binary classification problem, where the classes show if the pairs of entities described in the dataset records are or are not matching. The structure of the datasets is then "unusual" for ML and DL techniques used to manage single evidence records and generic techniques for explaining ML and DL models cannot be straightforwardly applied.

In this paper, we present Landmark Explanation a post-hoc perturbation-based local explainer for EM approaches. Post-hoc perturbation-based explainers build a surrogate linear model that approximates the model locally to the instance to explain. The surrogate linear model is trained with synthetic data. The dataset is generated by creating a number of alterations of the record to explain (in the so-called perturbation phase) and predicting their class by applying them the original model (in the so-called reconstruction phase). The explanation is directly obtained from

SEBD 2022: The 30th Italian Symposium on Advanced Database Systems, June 19-22, 2022, Tirrenia (PI), Italy

andrea.baraldi96@unimore.it (A. Baraldi); francesco.delbuono@unimore.it (F. D. Buono); matteo.paganelli@unimore.it (M. Paganelli); francesco.guerra@unimore.it (F. Guerra)
http://morespace.unimore.it/francescoguerra/ (F. Guerra)
0000-0001-8119-895X (M. Paganelli); 0000-0001-6864-568X (F. Guerra)

[1] https://github.com/oracle/Skater

left_description	left_name	right_description	right_name
sony white cybershot t series digital camera jacket case with stylus lcjthcw for 2007 cybershot t series camera stylus include...	sony white cybershot t series digital camera jacket case with stylus lcjthcw	top loading leather black	sony lcs-csl cyber-shot camera case

Table 1
Pairs of non-matching entity descriptions.

the surrogate model. The importance of a feature in the decision is computed by multiplying its value in the record with the linear coefficient of the surrogate model. In textual databases, as the ones considered in this paper, the features of the model are typically the words used in the entity descriptions.

Example 1. *Table 1 shows an example of non-matching descriptions. Both the entities refer to camera cases produced by the same brand, but since their product code is different they are not be considered as the same entity. An explanation of for this record consists of a values associated to each word in the description. Words are extracted from the descriptions via a tokenization process (we evaluated the application of stemming techniques and the deletion of stop words). For this reason the terms “token" and “word" are used as synonym in this paper.*

Landmark Explanation leverages the specificity of the EM dataset by introducing two main innovations. The first is the generation of two explanations per dataset entry, one for each entity described in the record. The second is a mechanism for computing meaningful explanations, especially for records belonging to non-matching classes. The descriptions of a non-matching entity are composed of different words, and selecting the ones that mostly contributed in the decision is a complex task even for humans. To address the problem, we inject additional words extracted from one entity into the second entity before the perturbation. The result is that the number of different words in non-matching entities decreases, while the similarity increases, thus enabling the approach to select the most relevant elements for the decision.

We implemented Landmark Explanation as an add-on component of the LIME system. The results of the experiments show that the explanations generated for EM datasets outperform the ones of the competing approaches in accuracy and “interest" for the users. This paper summarizes the Landmark Explanation presentations in [17, 18].

2. **The** Landmark Explanation **approach**

2.1. Landmark Explanation **principles**

Landmark Explanation adapts a local post-hoc explanation technique to the EM scenario. Indeed, the direct application of a perturbation mechanism based on token removals is not effective for EM datasets. The reason is that removing random tokens is likely to affect both the entities represented by the two descriptions. The generated synthetic records may then contain ***null or non coherent perturbations*** where the same tokens referring to the different entities are removed. These inconsistent perturbations lead to biased explanations. Moreover, post-hoc

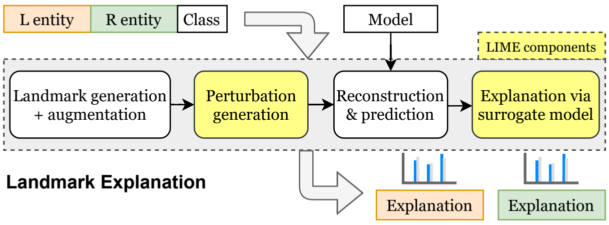


Figure 1: Landmark Explanation workflow

explanation systems adopt techniques for generating perturbations based on token removal. The resulting explanations for non-matching entity descriptions (the greatest parts of the records generally in EM datasets) are not useful as we will describe later on. Landmark Explanation addresses these issues by introducing the following two main innovations.

Double explanation. The first innovation consists of the generation of two explanations for each dataset entry. When we compute an explanation, we perturb a description (the *varying entity*) and keep unchanged its paired description (the *landmark entity*). The explanation assigns an impact to each token of the perturbed description. We repeat the computation by exchanging varying and landmark entities. Each result explains the model decision from the perspective of one of the two entities described in the record.

Injection of features. The second is a mechanism is for contrasting the asymmetric nature of the EM problem: an explanation of a matching pair is always composed of “interesting" tokens since they express the reasons why the entities have been considered as matching. The same does not happen for non-matching entities that have many reasons to be different. We address this issue by injecting additional tokens extracted from the landmark entity into the varying entity before the perturbation. Therefore, such a dataset contains entities close to the landmark, and the surrogate model trained with these entities will be able to highlight the distinctive tokens, that mainly contribute to the decision. Without the injection, descriptions of non-matching entities would have a large number of tokens that would uniformly contribute to the decision with the same low impact.

2.2. Landmark Explanation **explanations**

Let r be a record in an EM dataset representing a pair of entity descriptions (e_x, e_y), each one composed of a collection of tokens $\{t_{i1}, ..., t_{in^i}\}$, where $i \in \{x, y\}$, and n^i is the number of tokens belonging to the description of the entity i. The application of an EM binary classification model to r returns $\{0, 1\}$ when r is composed of non-matching or matching entity descriptions, respectively. An explanation is composed of a score for each description token $E_i = \{s_{i1}, ..., s_{in^i}\}$, where $i \in \{x, y\}$, $s \in \mathbb{R}$, s_{ij} is the score of token t_{ij}. S_x is the explanation generated by selecting e_y as the landmark and, vice-versa, S_y by selecting e_x as the landmark. Positive scores push the decision towards the class of matching entities, negative towards non-matching. The highest the absolute value of the score, the highest the importance of the token associated with the score. An explanation with augmented features assumes the form of $E_i = \{s_{x1}, ..., s_{xn^x}, s_{y1}, ..., s_{yn^y}\}$, where for the explanation E_x, the scores s_{yj} are the ones of the injected features from the entity description e_y (and vice-versa for the explanation E_y).

2.3. Landmark Explanation **workflow**

Figure 1 shows the description of the end-to-end workflow implemented by Landmark Explanation. The yellow boxes are the ones provided by a generic explanation system. The white boxes are provided by Landmark Explanation.

Landmark generation and entity augmentation. The descriptions of the entities are tokenized, and a prefix is added to each token to mark the provenance attribute. We set as landmark the set of tokens of the first entity, the other set of tokens will be perturbed. In the case of non-matching predictions, tokens are injected in the varying entity as described in Section 2.1. The process is repeated exchanging the landmark and the varying entities.

Perturbation generation. A representation of the neighborhood for *varying entities* is generated by perturbing its tokens in multiple ways. We used LIME which generates a series of textual phrases containing many combinations of the tokens of the varying description.

Reconstruction and prediction. We reconstruct the schema of the synthetic textual records obtained in the last step. We concatenate each of these new records with the original *landmark entity*. The produced pairs of entities are finally provided as input to the original EM model in order to obtain the relative prediction scores.

Explanation via surrogate model. Finally, a surrogate linear model (one for each workflow, one for the left and right entities, respectively) is trained on the perturbed dataset to learn an approximation of the behavior of the original model in those localities. The surrogate model takes in input the bag of words representation of the perturbed tokens and is trained to learn the relation between the input and the prediction score produced by the model under explanation. The coefficients learned during training represent the impact of each token in the prediction, and are used to generate the explanations of the original EM model for each EM record. In our implementation we adopt LIME to perform this task, but our approach is transparent to the explanation tool selected.

2.4. Explaining ER Models

Studies applying interpretation techniques in the entity matching area [16, 14], and tools, like Mojito [15] and Explainer [13], have been proposed. ExplainER provides a unified interface for applying well-known interpretation techniques (e.g., LIME, Shapley, Anchor, and Skater) in the EM scenario. Mojito adapts LIME for the explanation of single EM predictions and represents the work closer to our approach. It extends LIME in two ways: 1) it exploits the subdivision of EM data into attributes, 2) it introduces a new form of data perturbation, called LIME-COPY[2], which allows generating match elements starting from non-match elements. Differently Landmark Explanation, Mojito treats attributes atomically, distributing its impact equally to its constituent tokens. Furthermore, Landmark Explanation analyzes the diversified impact that the same token can generate depending on the entity considered as a landmark for the explanation.

[2] In Section 3 we refer to this technique as Mojito Copy since it is part of the Mojito tool.

3. Experimental evaluation

We evaluated the explanations generated by Landmark Explanation according to two main perspectives: the fidelity in representing the EM Model (in Section 3.1) and the "quality" of the explanation. For this last evaluation, we introduce a measure for assessing the interest of the explanations (in Section 3.2) and we propose an example of explanation for non-matching entity descriptions (in Section 3.3). This shows the importance of the token injection mechanism.

Dataset and Model. We perform an experimental evaluation against the datasets provided by the Magellan library[3] which is considered as a standard benchmark for the evaluation of EM tasks. The datasets are divided into structured (iTunes-Amazon S-IA, DBLP-ACM S-DA, DBLP-GoogleScholar S-DG, Walmart-Amazon S-WA), textual (Abt-Buy T-AB) and Dirty (iTunes-Amazon D-IA, DBLP-ACM D-DA, DBLP-GoogleScholar D-DG, Walmart-Amazon D-WA). The records in all datasets represent pairs of entities described with the same attributes. A label is provided to express if the record represents a matching / non-matching pair of entities. A simple logistic regression model is experimented as matcher, where the features are the similarities of the paired attributes in the descriptions. We compute the similarity by applying the jaccard measure on the trigrams of the attribute values. The experiments are performed by sampling 100 records per label (all records in datasets with smaller cardinality) and computing their explanations. We generate **base explanations**, by using the tokens from an entity description and **augmented explanations**, by generating explanations with the tokens of entity description with the ones injected from the second entity description.

3.1. Fidelity of the explanations

To evaluate the fidelity of the explanations, i.e., if the weights assigned by Landmark Explanation to the tokens generate a surrogate model that is consistent with the EM model, we randomly remove 25% tokens from the record to explain, defining a new item. We then compared the probability score obtained passing the new item to the EM model with the one of the original records, where we have subtracted the sum of the coefficients associated with the removed tokens. If the explanation model correctly represents the EM model these two values should be close. The experiment is repeated 100 times per class, and the performance measured by means of two metrics: the mean absolute error (MAE) between the explanation and the EM Model and the accuracy that measures the percentage of times that the probability score of the new item changes consistently with to the sum of the impacts of the tokens removed. Table 2 shows the results of the experiment. The column LIME shows the results obtained with LIME with the same setting. Non-matching settings also include a comparison with the Mojito Copy technique.

Discussion. The experiments show that the surrogate model built by Landmark Explanation with the base perturbation provides an accurate representation of the EM model for records representing matching pairs of entities. At the same time, the model built with the augmented perturbation is an accurate representation of the EM model for record representing non-matching pairs of entities. In particular, Table 2a shows that Landmark Explanation, applied to records

[3] https://github.com/anhaidgroup/deepmatcher/blob/master/Datasets.md

	Base		**Augmented**		**LIME**	
	Acc.	MAE	Acc.	MAE	Acc.	MAE
S-IA	**0.940**	**0.226**	0.793	0.251	0.847	0.240
S-DA	0.887	0.171	**0.894**	**0.164**	0.573	0.337
S-DG	**0.836**	**0.196**	0.823	**0.196**	0.757	0.200
S-WA	**0.954**	**0.071**	0.928	0.115	0.659	0.228
T-AB	**0.908**	**0.066**	0.854	0.146	0.758	0.118
D-IA	0.899	**0.090**	**0.975**	0.112	0.780	0.156
D-DA	0.942	**0.030**	**0.979**	0.041	0.940	**0.025**
D-D	0.929	**0.107**	**0.963**	0.152	0.891	0.115
D-WA	**0.916**	**0.045**	0.901	0.090	0.813	0.074

(a) Matching label.

Base		**Augmented**		**LIME**		**Mojito Copy**	
Acc.	MAE	Acc.	MAE	Acc.	MAE	Acc.	MAE
0.669	0.248	**0.736**	**0.127**	0.624	0.267	0.022	0.569
0.975	**0.021**	0.590	0.287	**0.985**	0.066	0.005	0.574
0.895	**0.086**	0.660	0.306	**0.935**	0.107	0.005	0.504
0.990	**0.028**	0.955	0.217	0.890	0.352	0.000	0.746
0.860	0.076	0.680	**0.047**	0.795	0.092	0.045	0.328
0.874	**0.019**	0.291	0.070	0.390	0.129	0.242	0.191
0.615	0.071	0.300	**0.027**	**0.690**	0.036	0.010	0.173
0.540	0.305	0.375	**0.118**	**0.640**	0.235	0.040	0.437
0.500	0.184	**0.785**	**0.078**	0.500	0.192	0.005	0.380

(b) Non-matching label.

Table 2
Evaluation of the fidelity of the explanations.

labeled as matching entity, performs better than LIME in the datasets when the perturbation is generated with the base technique (it obtains better accuracy in all datasets and low MAE in 8/9 datasets). The augmented generation technique performs slightly worse: in 8/9 it obtains better accuracy and in 5/9 lower MAE). Note that this can be motivated also by the increased number of tokens in the augmented explanations. Nevertheless, the scores, when worst, are very close to LIME. Table 2b shows the accuracy and the MAE obtained analyzing records referring to non-matching labels. In this scenario, the augmented entity perturbation obtains the best scores with an accuracy better than LIME in 3/9 datasets and a lower MAE in 7/9 datasets. Finally, the copying technique introduced by Mojito to manage records associated with non-matching labels does not show high performance. The reason is that Mojito generates a perturbation by duplicating entire attributes. The result of this operation is that the tokens of the replaced attribute have the same weights, and decrease the performance.

3.2. Quality of the explanations

Since there are many reasons to be dissimilar for two entities, the explanations of non-matching entity descriptions are typically “slightly polarized" having negative values distributed in a range close to zero and no value dominating the others. For the user, this means not being able to grasp a strong motivation for the non-matching decision. To evaluate if we are able to generate “interesting explanations", we introduced a heuristic according to which an explanation for non-matching entities is interesting if it contains tokens that, if injected into the second entity, would make the record classified as matching. These are the elements that make the explanation interesting for the users. To evaluate if the explanations generated by Landmark Explanation satisfy this property, we perform the same experiment described in Section 3.1, but selecting the tokens to remove: negative tokens are removed when the label represents a non-matching record (all tokens that contribute to the decision). Positive tokens are removed in case of matching records. In Table 3 we measure the *interest*, which is the percentage of records where the removal of the tokens was able to generate a change in the label.

Discussion. Landmark Explanation generates interesting explanations, and the perturbation generated with the augmented technique effectively increases “the interest" of non-matching

	Base	**Augmented**	**LIME**
S-IA	0.652	0.404	**0.702**
S-DA	**1.000**	0.940	0.965
S-DG	0.660	0.610	**0.925**
S-WA	**1.000**	0.785	0.870
T-AB	0.985	0.575	**0.995**
D-IA	**0.561**	0.278	0.311
D-DA	0.695	0.715	**0.800**
D-DG	0.635	0.530	**0.735**
D-WA	**0.915**	0.545	0.880

(a) Matching label.

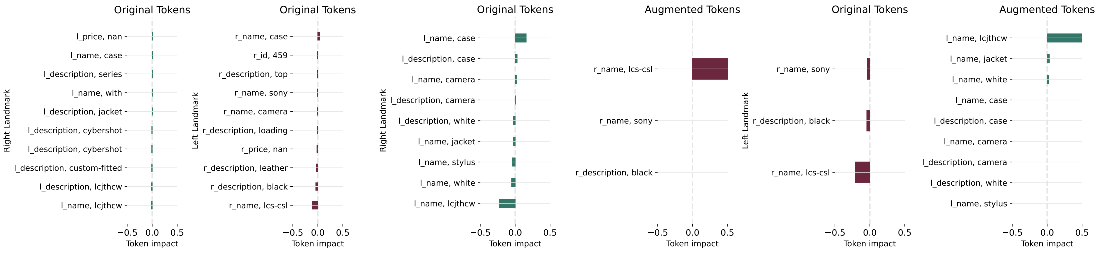
	Base	**Augmented**	**LIME**	**Mojito Copy**
S-IA	0.545	**0.736**	0.393	0.000
S-DA	0.000	0.030	0.000	0.005
S-DG	0.020	**0.545**	0.020	0.000
S-WA	0.015	**0.955**	0.000	0.000
T-AB	0.305	**0.680**	0.340	0.045
D-IA	**0.670**	0.291	0.379	0.027
D-DA	0.205	**0.300**	0.125	0.000
D-DG	0.200	**0.375**	0.160	0.030
D-WA	0.190	**0.785**	0.130	0.005

(b) Non-matching label.

Table 3
Evaluation of the interest associated to the computed explanations.

record explanations. In particular, Table 3a shows that Landmark Explanation is good but slightly worse than LIME in terms of interest, when the records are labeled as matching class. This happens even if the surrogate model is accurate (the MAE score is the lowest for all experiments with the single-entity configuration). The problem is that in most of the cases, even removing all tokens, the explanation created by Landmark Explanation belongs to the same class as before the token removal. Note that if we set a decision threshold to 0.4, our approach has the best results in all datasets. Table 3b shows that the augmented explanations of non-matching entities generated by Landmark Explanation outperform LIME and Mojito Copy.

3.3. Showing the explanations



(a) The base technique.

(b) The augmented technique.

Figure 2: Visualizing an explanation. Red (green) bars are associated to the right (left) entity description.

Figure 2a shows the explanations computed with the base technique for the entity descriptions in Table 1 1. We recall that positive impacts push towards the match decision, negative towards a non-match decision. Landmark Explanation generates two explanations per record and we can see that no token assumes a particular importance. The resulting explanation is therefore not interesting (and useful) for the user. Figure 2b shows the explanation obtained by the injection of the tokens from the landmark. The first explanation (where the right entity is the landmark) clearly shows that the token `case` pushes towards the match decision (both the entities refer to camera cases) and the code `lcjthcw` towards the non-match decision (it is different from the code in the second description). The augmented tokens show that the code `lcs-csl` pushes towards a match decision. This means that if that code had been part of the description for the left entity, it would have pushed the model towards a match decision. Similar considerations can be done by observing the second explanation obtained setting the left entity as landmark.

4. Conclusion

This paper introduces Landmark Explanation a tool that makes a post-hoc perturbation-based explainer able to deal with ML and DL models describing EM datasets. The approach has been experimented coupled with the LIME explainer on a simple EM model based on logistic regression. The results show that the explanations generated by Landmark Explanation outperform the ones generated by the competing approaches.

References

[1] M. Ebraheem, S. Thirumuruganathan, S. R. Joty, M. Ouzzani, N. Tang, Distributed representations of tuples for entity resolution, Proc. VLDB Endow. 11 (2018) 1454–1467.

[2] S. Mudgal, H. Li, T. Rekatsinas, A. Doan, Y. Park, G. Krishnan, R. Deep, E. Arcaute, V. Raghavendra, Deep learning for entity matching: A design space exploration, in: SIGMOD Conference, ACM, 2018, pp. 19–34.

[3] Y. Li, J. Li, Y. Suhara, A. Doan, W.-C. Tan, Deep entity matching with pre-trained language models, Proc. VLDB Endow. 14 (2020) 50–60. URL: https://doi.org/10.14778/3421424.3421431. doi:10.14778/3421424.3421431.

[4] M. Paganelli, F. D. Buono, M. Pevarello, F. Guerra, M. Vincini, Automated machine learning for entity matching tasks, in: EDBT, OpenProceedings.org, 2021, pp. 325–330.

[5] L. Gagliardelli, S. Zhu, G. Simonini, S. Bergamaschi, BigDedup: A Big Data Integration Toolkit for Duplicate Detection in Industrial Scenarios, in: TE, volume 7 of *Advances in Transdisciplinary Engineering*, IOS Press, 2018, pp. 1015–1023.

[6] R. Cappuzzo, P. Papotti, S. Thirumuruganathan, Creating embeddings of heterogeneous relational datasets for data integration tasks, in: SIGMOD Conference, ACM, 2020, pp. 1335–1349.

[7] U. Brunner, K. Stockinger, Entity matching with transformer architectures - A step forward in data integration, in: EDBT, OpenProceedings.org, 2020, pp. 463–473.

[8] M. Paganelli, F. D. Buono, F. Guerra, N. Ferro, Evaluating the integration of datasets, in: Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing, SAC '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 347–356. URL: https://doi.org/10.1145/3477314.3507688. doi:10.1145/3477314.3507688.

[9] M. Du, N. Liu, X. Hu, Techniques for interpretable machine learning, Commun. ACM 63 (2020) 68–77.

[10] M. T. Ribeiro, S. Singh, C. Guestrin, " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD, 2016, pp. 1135–1144.

[11] A. Ghorbani, J. Y. Zou, Data shapley: Equitable valuation of data for machine learning, in: ICML, volume 97 of *Proceedings of Machine Learning Research*, PMLR, 2019, pp. 2242–2251.

[12] M. T. Ribeiro, S. Singh, C. Guestrin, Anchors: High-precision model-agnostic explanations, in: AAAI, AAAI Press, 2018, pp. 1527–1535.

[13] A. Ebaid, S. Thirumuruganathan, W. G. Aref, A. Elmagarmid, M. Ouzzani, Explainer: Entity resolution explanations, in: 2019 IEEE 35th Int. Conf. on Data Engineering (ICDE), IEEE, 2019, pp. 2000–2003.

[14] S. Thirumuruganathan, M. Ouzzani, N. Tang, Explaining entity resolution predictions: Where are we and what needs to be done?, in: Proceedings of the Workshop on Human-In-the-Loop Data Analytics, 2019, pp. 1–6.

[15] V. D. Cicco, D. Firmani, N. Koudas, P. Merialdo, D. Srivastava, Interpreting deep learning models for entity resolution: an experience report using LIME, in: aiDM@SIGMOD, ACM, 2019, pp. 8:1–8:4.

[16] X. W. L. H. A. Meliou, Explaining data integration, Data Engineering (2018) 47.

[17] A. Baraldi, F. D. Buono, M. Paganelli, F. Guerra, Landmark explanation: An explainer for entity matching models, in: CIKM, ACM, 2021, pp. 4680–4684.

[18] A. Baraldi, F. D. Buono, M. Paganelli, F. Guerra, Using landmarks for explaining entity matching models, in: EDBT, OpenProceedings.org, 2021, pp. 451–456.