Similarity Measure Learning for Analogical Transfer

Chunyang Fan[1,2,*]

[1]*Sorbonne Université, CNRS, LIP6, F-75005 Paris, France*

[2]*Université Sorbonne Paris Nord, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé – LIMICS, INSERM, UMR 1142, F-93000, Bobigny, France*

Abstract

Analogical transfer infers unknown information by leveraging known similarities between situations. This doctoral research investigates similarity measure learning tailored for analogical transfer, focusing on classification. A unified theoretical framework, called Similarity Measure Learning Architecture (SiMeLAr), is introduced to optimize systematically similarity measures for analogical tasks. Beside this generic question, we focus on the specific case of the method called CoAT (Complexity-based Analogical Transfer). To address its non-continuity, we propose a continuous variant which enables gradient-based optimization. Future work will explore the application of the variant to real-world domains, such as culinary and medical use cases.

Keywords

Analogical Transfer, Similarity Measure Learning, Metric Learning, Case-Based Reasoning, Machine Learning

1. Introduction

Analogical transfer infers information from known similar cases by assuming that their similarity in certain components implies similarity in others [1]. Case-based prediction (CBP) methods (see e.g. [2, 3] for surveys) apply this general principle to solve supervised machine learning tasks, classification or regression: they consider a data instance as a case with two components, respectively called situation and outcome, that correspond to the instance features and its label. They then apply the general analogical transfer principle that, in that context, is expressed as follows: if two situations are similar, their outcomes should also be similar. Therefore, they predict the outcome of a new situation leveraging its similarities with situations from the case base.

Similarity measures thus obviously play a central role to analogical transfer, that can be seen as essentially transferring similarity knowledge [3]. This role raises questions related to the topic of metric learning [4, 5].

This thesis aims at studying the interactions between analogical transfer and metric learning and at developing a methodology to learn optimized similarity measures for case-based prediction. Validation will occur in two domains: culinary recipe transfer [6] and breast cancer management decision-making [7].

1.1. State of the Art

This research lies at the intersection of similarity measure learning and analogical transfer, necessitating an overview of both fields and their interactions.

Similarity Measure Learning Similarity measures quantify how alike two objects are, often derived from distances transformed by decreasing functions [8, 9]. Metric learning optimizes these measures using data-driven constraints [4, 5]. Numerous approaches have been proposed, they for instance include linear [10], nonlinear [11], local metrics [12], and histogram-based methods [13, 14].

More recently, Deep metric learning (DML) employs neural networks projecting data into embedding spaces, leading to approaches that can be categorized into pair-based methods using contrastive

loss [15], triplet-based methods [16], similarity cloud-based method [17] and clustering-based methods capturing global data structures [18]. Additional approaches integrate attribute-specific measures, e.g., k-Prototypes [19], or unsupervised hyperbolic methods for hierarchical data [20].

In the context of case-based reasoning, some works further propose data-driven approaches to learning or tuning similarity measures that rely less on expert-driven designs [21, 22].

Analogical Transfer   Analogical transfer infers new situations' outcomes from similar known instances by mapping similarities (see e.g. [1]). The analogical transfer principle can be formulated as: **if two situations are similar with respect to some criteria, then it is plausible that they are also similar with respect to other criteria** (see e.g. [23]). Computational analogy formalizes this principle through various methods that can for instance be categorized into [24]: instance-label alignment [25, 26], negative constraints [27], label support measures [28], domain knowledge integration [29], and Complexity-based Analogical Transfer (CoAT) [30].

Case based prediction (CBP) considers a case base CB, which is a set of cases where each case c is a pair $(s, r) \in \mathcal{S} \times \mathcal{R}$, where $\mathcal{S}$ and $\mathcal{R}$ respectively denote the situation and the outcome spaces. They are respectively equipped with two similarity measures $\sigma_{\mathcal{S}} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}^+$ and $\sigma_{\mathcal{R}} : \mathcal{R} \times \mathcal{R} \to \mathbb{R}^+$. We describe in more details the CoAT [30] approach below. It constitutes a recent model that has shown promising results in analogical transfer tasks.

CoAT relies on a global indicator, computed on the whole case base instead of evaluating locally similarity between source-target pairs: applying the analogical transfer to CBP, according to which similar situations imply similar outcomes, the incompatibility indicator computes the number of triplets (c_i, c_j, c_k) in CB that violate this principle, i.e. $\sigma_{\mathcal{S}}(s_i, s_j) \geq \sigma_{\mathcal{S}}(s_i, s_k)$ but $\sigma_{\mathcal{R}}(r_i, r_j) < \sigma_{\mathcal{R}}(r_i, r_k)$:

$$\Gamma(\sigma_{\mathcal{S}}, \sigma_{\mathcal{R}}, \text{CB}) := \sum_{(t_i, t_j, t_k) \in \text{CB}^3} \text{Ind}_\theta(t_i, t_j, t_k) \tag{1}$$

where $\theta = (\sigma_{\mathcal{S}}, \sigma_{\mathcal{R}}, \text{CB})$, and $\text{Ind}_\theta(t_i, t_j, t_k) := \mathbb{1}\left\{\sigma_{\mathcal{S}}(s_i, s_j) \geq \sigma_{\mathcal{S}}(s_i, s_k)\right\} \mathbb{1}\left\{\sigma_{\mathcal{R}}(r_i, r_j) < \sigma_{\mathcal{R}}(r_i, r_k)\right\}$.

As a classifier, for a new situation s_{new}, when leveraging the situation similarity to predict its outcome, CoAT then identifies the most plausible outcome, defined as the one minimizing the incompatibility of the case base augmented with the candidate new case (s_{new}, r)

$$\hat{r}_{\text{new}} := \arg\min_{r \in R} E_\theta^{\text{CoAT}}(s_{\text{new}}, r) \quad \text{where} \quad E_\theta^{\text{CoAT}}(s_{\text{new}}, r) := \Gamma\left(\sigma_{\mathcal{S}}, \sigma_{\mathcal{R}}, \text{CB} \cup \{(s_{\text{new}}, r)\}\right)$$

1.2. Research Questions

Similarity measure learning and analogical transfer are interdependent: effective analogical transfer requires suitably learned measures, and optimal measures can be guided by analogical prediction tasks. Taking CoAT as an example, the incompatibility indicator Γ enables metric learning beyond classification. In fact, defining parameters θ and incompatibility E_θ, Badra et al. [31] identify CoAT as an energy-based model (EBM). Many loss functions have been proposed to optimize the performance for EBM, such as hinge loss $\ell_\theta(s, r) := \max\left(0, E_\theta(s, r) - \min_{r' \neq r} E_\theta(s, r') + \lambda\right)$. Such approaches open the way to optimise the similarity measure used in CoAT.

This opens the following questions: Do similarity measures optimized for a specific algorithm, for example CoAT, generalize across analogical classifiers? Can we tailor measures specifically for analogical classifiers or vice versa?

This research investigates how metric learning algorithms influence analogical transfer classifier performance, e.g. measured by their accuracy, and considering in particular the case of CoAT. Challenges include studying classifier-metric interactions, defining mathematical frameworks for analogical similarity learning, and taking into account computational complexity constraints.

2. Research Direction and Methods

2.1. Objectives

The objectives of this research include both theoretical and practical aspects. The former aims to study the relationship between similarity measure learning and analogical transfer, with a focus on the classification task. This involves (1) establishing a theoretical framework that unifies similarity measure learning and analogical transfer, and (2) studying the optimization of similarity measures using a specific analogical transfer model. On the practical side, the analogical transfer model will be applied to different domains, including culinary recipe cases and medical use cases.

2.2. Methodology

In order to achieve the first objective in the theoretical part, I propose a unified framework that combines similarity measure learning and analogical transfer, by abstracting the common processes, as measure, interaction and optimization. This work includes the following targets. Firstly, it aims to provide a clear mathematical framework. By doing so, my aim is to identify parts of existing metric learning that can be considered as using components of analogical transfer like E_θ. After that, I would like to provide a base of theoretical proofs for the research questions introduced in Section 1.2, for example, whether guarantee can be provided about the prediction accuracy, i.e., studying the probability $\mathbb{P}\left(\arg\min_{r \in R} E_{\hat{\theta}}^{\mathrm{CoAT}}(s_{\mathrm{new}}, r)\}) = r_{\mathrm{new}}\right)$. Eventually, this work could help to guide the development of new models. This part will be evaluated by a systematic literature survey and mathematical demonstrations: we will investigate models that fit into this framework and characterize formally their expression within this framework. In addition, we will prove the theorems under this framework using mathematical methods.

After establishing the theoretical framework, the next step is to study the optimization of similarity measures for the analogical transfer model CoAT [30] presented in Section 1.1. Since its incompatibility indicator Γ (see Eq.1) is a sum of binary indicators, and therefore a step function, it is non-continuous at a finite number of points, and its gradient is zero at all other points. As a consequence, traditional optimisation methods based on differentiation cannot be applied to find the optimal parameter values θ, in particular the similarity measure σ_S.

I would like to explore the use of metric learning methods to optimize σ_S for CoAT. This involves a more detailed theoretical analysis the limitations of the original CoAT model, such as its discrete and non-differentiable nature limiting its optimization capabilities, and proposing a method to overcome these limitations. The proposed method will be evaluated on real datasets to assess its effectiveness in improving the performance of CoAT, using classical supervised learning quality metrics, such as accuracy and algorithmic complexity.

3. Progress Summary

At present, the theoretical part of my study comprises a framework that integrates similarity measure learning with analogical transfer (Section 3.1), and an investigation into optimizing similarity measures for the analogical transfer model CoAT (Section 3.2).

3.1. Identification of Common Architecture in Metric Learning

In order to explore the relationship between metric learning and analogical transfer, I first propose a theoretical framework based on the literature in the state of the art (especially inspired by CoAT [30]), I propose the following common structure of models in metric learning called **Si**milarity **Me**asure **L**earning **Ar**chitecture (SiMeLAr), which consists of three core components:

1. The measure component defines the similarity measures (σ_S and σ_R) independently within the situation space (S) and the outcome space (R), establishing intra-space relationships.

2. The interaction component introduces an interaction function linking metrics from the situation and outcome spaces, establishing inter-space relationships. This function ensures consistency, so that similar input features correspond to similar outputs.

3. The optimization component employs a parameterized loss function and an aggregated total loss function to optimize the model parameters. It guides the training process, aligning the interaction of metrics toward effective learning.

Based on this framework I develop several theorems for discussing when the loss function can be guaranteed to be differentiable and convex, and study the relationship between the loss function and the accuracy of the model.

3.2. Continuous CoAT

Beyond the general case, I look at the particular case of CoAT [30]. The original CoAT model is based on the incompatibility function Γ given in Eq. (1), which is defined as the sum of binary indicators Ind_θ. However, it does not allow to apply gradient-based optimization methods as discussed in Section 2.2. Besides, it is not sensitive to small changes in the input data or parameters, which can lead to poor performance in some cases.

To overcome these limitations, I propose a new continuous energy function, which measures the extent of violation of the analogical transfer principle rather than merely counting violations. Specifically, in Ind_θ, I propose to replace the first term of the product with $\max\left(0, \eta - \sigma_\mathcal{X}(s_i, s_j) + \sigma_\mathcal{X}(s_i, s_k)\right)$, where η is a margin parameter. It quantifies the degree of violation. By doing so, the proposed method enables gradient-based optimization due to its continuity and differentiability. Additionally, I propose to modify the number of triplets (t_i, t_j, t_k) involved in the computation of Γ (see Eq. 1), to reduce the algorithmic complexity, and to introduce a normalization term in order to deal with multi-class classification scenarios.

Experimental Study Ongoing works study the proposed function for metric leaning, combining it with several common loss functions in energy-based models, such as MCE loss, hinge loss and direct loss [32]. We evaluate the proposed model Continuous CoAT (C-CoAT) against the original CoAT model on classical datasets, employing various similarity measures (Polynomial and Mahalanobis) and optimization methods (Adam and SGD).

Results show that the new continuous energy-based method outperforms the original CoAT in 4 out of the 6 considered datasets, particularly when the similarity measures are optimized. Notably, datasets with purely categorical attributes present challenges, that we are currently investigating in more details.

4. Conclusion and Future Work

My research advances the integration of similarity measure learning with analogical transfer for classification tasks. A unified theoretical framework (SiMeLAr) is proposed to clarify structural elements common to metric learning and analogical transfer. The proposed Continuous Complexity-based Analogical Transfer (C-CoAT) resolves discontinuity and zero-gradient limitations of the original CoAT, and opens the way to improving the classification performance via gradient-based optimization of the similarity measure it relies on, validated through empirical evaluations.

Future work will further explore whether the similarity measures optimized for C-CoAT are also effective for other classifiers, and extend to real-world applications such as cooking and medical use cases.

Acknowledgments

Declaration on Generative AI

The authors did not use any generative AI during the preparation of this work.

References

[1] H. Gust, U. Krumnack, K.-U. Kühnberger, A. Schwering, Analogical Reasoning: A Core of Cognition., KI 22 (2008) 8–12.

[2] I. Gilboa, D. Schmeidler, Case based Predictions: Introduction, Introduction to Case-Based Prediction. World Scientific Publishers (2010).

[3] F. Badra, M.-J. Lesot, Case-based prediction – A survey, Int. Journal of Approximate Reasoning 158 (2023) 108920.

[4] A. Bellet, A. Habrard, M. Sebban, Metric Learning, Synthesis Lectures on Artificial Intelligence and Machine Learning, Springer, 2015.

[5] J. L. Suárez, S. García, F. Herrera, A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges, Neurocomputing 425 (2021) 300–322.

[6] F. Badra, R. Bendaoud, R. Bentebibel, P.-A. Champin, J. Cojan, A. Cordier, S. Després, S. Jean-Daubias, J. Lieber, T. Meilender, A. Mille, E. Nauer, A. Napoli, Y. Toussaint, TAAABLE: Text Mining, Ontology Engineering, and Hierarchical Classification for Textual Case-Based Cooking, in: M. Schaaf (Ed.), 9th European Conf. on Case-Based Reasoning - ECCBR 2008, Workshop Proc., 2008, pp. 219–228.

[7] A. Redjdal, J. Bouaud, G. Guézennec, J. Gligorov, B. Seroussi, Reusing Decisions Made with One Decision Support System to Assess a Second Decision Support System: Introducing the Notion of Complex Cases, Studies in Health Technology and Informatics 281 (2021) 649–653.

[8] S. Santini, R. Jain, Similarity measures, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (1999) 871–883.

[9] M.-J. Lesot, M. Rifqi, H. Benhadda, Similarity measures for binary and numerical data: a survey, Int. Journal of Knowledge Engineering and Soft Data Paradigms 1 (2009) 63–84.

[10] E. Xing, M. Jordan, S. J. Russell, A. Ng, Distance Metric Learning with Application to Clustering with Side-Information, in: Advances in Neural Information Processing Systems, volume 15, MIT Press, 2002.

[11] J. V. Davis, B. Kulis, P. Jain, S. Sra, I. S. Dhillon, Information-theoretic metric learning, in: Proc. of the 24th Int. Conf. on Machine Learning, ICML '07, ACM, 2007, pp. 209–216.

[12] J. Wang, A. Kalousis, A. Woznica, Parametric Local Metric Learning for Nearest Neighbor Classification, in: Advances in Neural Information Processing Systems, volume 25, 2012.

[13] D. Kedem, S. Tyree, F. Sha, G. Lanckriet, K. Q. Weinberger, Non-linear Metric Learning, in: Advances in Neural Information Processing Systems, volume 25, 2012.

[14] F. Wang, L. J. Guibas, Supervised Earth Mover's Distance Learning and Its Computer Vision Applications, in: A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid (Eds.), Computer Vision – ECCV 2012, Springer, 2012, pp. 442–455.

[15] A. Hermans, L. Beyer, B. Leibe, In Defense of the Triplet Loss for Person Re-Identification, ArXiv (2017).

[16] E. Hoffer, N. Ailon, Deep Metric Learning Using Triplet Network, in: A. Feragen, M. Pelillo, M. Loog (Eds.), Similarity-Based Pattern Recognition, Springer, 2015, pp. 84–92.

[17] T. Gabel, E. Godehardt, Top-Down Induction of Similarity Measures Using Similarity Clouds, in: E. Hüllermeier, M. Minor (Eds.), Case-Based Reasoning Research and Development, Springer, 2015, pp. 149–164.

[18] H. O. Song, Y. Xiang, S. Jegelka, S. Savarese, Deep Metric Learning via Lifted Structured Feature Embedding, 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2016) 4004–4012.

[19] Z. Huang, Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values, Data Mining and Knowledge Discovery 2 (1998) 283–304.

[20] J. Yan, L. Luo, C. Deng, H. Huang, Unsupervised Hyperbolic Metric Learning, in: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 12465–12474.

[21] A. Jaiswal, K. Bach, A Data-Driven Approach for Determining Weights in Global Similarity Functions, in: K. Bach, C. Marling (Eds.), Case-Based Reasoning Research and Development, volume 11680, Springer, 2019, pp. 125–139.

[22] B. M. Mathisen, A. Aamodt, K. Bach, H. Langseth, Learning similarity measures from data, Progress in Artificial Intelligence 9 (2020) 129–143.

[23] T. R. Davies, S. J. Russell, A Logical Approach to Reasoning by Analogy, in: J. P. McDermott (Ed.), Proc. of the 10th Int. Joint Conf. on Artificial Intelligence (IJCAI'87), Morgan Kaufmann Publishers, 1987, pp. 264–270.

[24] F. Badra, M.-J. Lesot, A. Barakat, C. Marsala, Theoretical and Experimental Study of a Complexity Measure for Analogical Transfer, in: M. T. Keane, N. Wiratunga (Eds.), Case-Based Reasoning Research and Development, volume 13405, Springer, 2022, pp. 175–189.

[25] F. Badra, K. Sedki, A. Ugon, On the Role of Similarity in Analogical Transfer, in: M. T. Cox, P. Funk, S. Begum (Eds.), Case-Based Reasoning Research and Development, volume 11156, Springer, 2018, pp. 499–514.

[26] L. Miclet, S. Bayoudh, A. Delhay, Analogical Dissimilarity: Definition, Algorithms and Two Experiments in Machine Learning, Journal of Artificial Intelligence Research 32 (2008) 793–824.

[27] E. Hüllermeier, Case-Based Approximate Reasoning, number 44 in Theory and Decision Library, Springer, 2007.

[28] E. Hüllermeier, Possibilistic instance-based learning, Artificial Intelligence 148 (2003) 335–383.

[29] J. Lieber, E. Nauer, H. Prade, When Revision-Based Case Adaptation Meets Analogical Extrapolation, in: 29th Int. Conf. on Case-Based Reasoning (ICCBR 2021), volume 12877 of *Lecture Notes in Computer Science book series (LNCS)*, 2021, pp. 156–170.

[30] F. Badra, A Dataset Complexity Measure for Analogical Transfer, in: Int. Joint Conf. on Artificial Intelligence, volume 2, 2020, pp. 1601–1607.

[31] F. Badra, M.-J. Lesot, E. Marquer, M. Couceiro, Some Perspectives on Similarity Learning for Case-Based Reasoning and Analogical Transfer, in: Workshop on the Interactions between Analogical Reasoning and Machine Learning, IARML@IJCAI'2023:, volume 3492, CEUR-WS.org, 2023, pp. 16–29.

[32] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, F. Huang, A tutorial on energy-based learning, in: G. Bakir, T. Hofman, B. Scholkopf, A. Smola, B. Taskar (Eds.), Predicting structured data, MIT Press, 2006.