

Learning in Parallel Universes

Michael R. Berthold and Bernd Wiswedel

University of Konstanz
Department of Computer and Information Science
Germany
{Michael.Berthold,Bernd.Wiswedel}@uni-konstanz.de

Abstract. This abstract summarizes a brief, preliminary formalization of learning in parallel universes. It also attempts to highlight a few neighboring learning paradigms to illustrate how parallel learning fits into the greater picture.

Keywords. parallel universes

1 Introduction

Parallel Universes offer a uniform framework for learning approaches that build global models from a set of connected universes, holding different object representations. In each universe potentially only a part of the model is created and during model construction the different object representations but also the intermediate models in each universe aid model construction in other universes.

There are, of course, many overlaps to existing techniques and terminologies. Most are somewhat borderline Learning in Parallel Universes approaches – the most prominent examples are (among others):

- Multi View Learning: focuses on finding global models in (or across) all universes [1,2],
- Feature Selection: finds one, global universe based on the projection of an existing, larger one [3],
- ensemble methods: find a mix of models which are built in universes without interaction [4].
- Subspace clustering: finds a set of potentially overlapping clusters operating using an individual subset of features.

Figure 1 illustrates how these approaches fit in with respect to Parallel Universes.

2 Patterns in Parallel Universes

Assumption of Existence:

- a set of objects Ω denoting all possible instances, following some underlying distribution P_Ω . Objects can be e.g. molecules, pieces of music, or images.

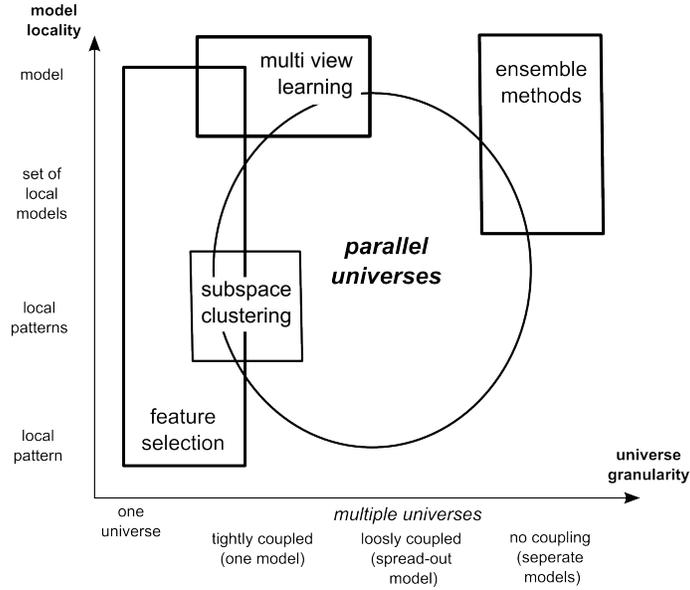


Fig. 1. Parallel Universes shown against universe granularity and model locality together with related approaches. Note that the crisp boundaries are only for illustrative purposes – in reality the borders are rather fuzzy.

- a target concept $c : \Omega \mapsto T$ (which can also represent an underlying true clustering).

Input:

- for each space U_l (called *universe*, $l = 1, \dots, u$), a representation for a set of objects $O_l \subset \Omega$ in the corresponding universe, that is $\forall o_i \in O_l : \exists^1 \mathbf{x}_{l,i} \in U_l$ (i. e. the data in universe l). Representations can be image features, descriptors such as molecular weight and others. (Sometimes also only a distance metric $d^l(i, j)$ between two objects o_i and o_j will be available.)

Output:

- partial models m_l (e. g. collections of local patterns) in each universe l , $m_l : U'_l \mapsto T$, whereby $U'_l \subseteq U_l$. These models assign a label or another output of the target domain to all or a subset of all object representations. If the model in a universe decides itself which objects it can assign a target value to, one could also see this as a function $m_l : U_l \mapsto T \cup \{?\}$, assigning either a target value or *unknown*. The important part here is that models in universes do not need to classify, cluster, or otherwise produce a response for all objects in U_l .

Remarks:

- Note that a model can be empty in one or more universes ($U'_l = \emptyset$ or $\forall \mathbf{x}_{l,i} \in U_l : m_l(\mathbf{x}_{l,i}) = ?$), that is not all universes are used.

- In some cases $U'_l = U_l$ and it is not desirable that m_l assigns an output to each object but the family of models m_l stems from, can not itself distinguish between known and unknown objects. In this case an explicit assignment of objects to models in these universes can be provided, for instance using a membership function (see [5] for an example).
 - If all patterns belong to a model in all universes this approach is similar to multi view learning.
 - If each local model uses a subset of only one universe as its own universe (e.g. only a subset of the available features) this could be regarded as an example of subspace clustering [6].
- a model joiner $M := f(m_1, \dots, m_u)$ denoting an overall model based on partial models in the universes.

Remarks:

- The output of joiner M can depend on the learning of the individual models m_l , that is M is build throughout learning of all or some m_l .
- If M is a simple combination operator (e.g. weighted average) of all individually learned models, this would resemble ensemble learning. Learning in Parallel Universes benefits from “mixing” during learning.

Goal:

- The joined model M outperforms each individual partial model m_l and, in general, produces responses for more objects than individual models.
- during learning, information was shared among universes. (In contrast to a simply merging of finalized models after training has finished in all universes – similar to ensembles.)

Acknowledgments

The participants of Dagstuhl Seminar 07182 “Local Patterns and Parallel Universes” really created this abstract.

References

1. Rüping, S., Scheffer, T., eds.: Proceedings of the ICML 2005 Workshop on Learning with Multiple Views. (2005) <http://www-ai.cs.uni-dortmund.de/MULTIVIEW2005/MultipleViews.pdf>.
2. Bickel, S., Scheffer, T.: Multi-view clustering. In: Proceedings of the Fourth IEEE International Conference on Data Mining (ICDM'04). (2004) 19–26
3. Blum, A., Langley, P.: Selection of relevant features and examples in machine learning. *Artificial Intelligence* (1-2) (December 1997) 245–271
4. Dietterich, T.G.: Ensemble methods in machine learning. *Lecture Notes in Computer Science* **1857** (2000) 1–15
5. Wiswedel, B.: Fuzzy clustering in parallel universes. In Berthold, M.R., Morik, K., Siebes, A., eds.: Local Patterns and Parallel Universes. Number 07181 in Dagstuhl Seminar Proceedings, Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany (2007)
6. Parsons, L., Haque, E., Liu, H.: Subspace clustering for high dimensional data: a review. *SIGKDD Explor. Newsl.* **6**(1) (2004) 90–105