Homomorphisms Between Transfer, Multi-Task, and Meta-Learning Systems

Tyler Cody

National Security Institute
Virginia Tech
Arlington, VA, USA
tcody@vt.edu

Abstract. Transfer learning, multi-task learning, and meta-learning are well-studied topics concerned with the generalization of knowledge across learning tasks and are closely related to general intelligence. But, the formal, general systems differences between them are underexplored in the literature. This lack of systems-level formalism leads to difficulties in coordinating related, inter-disciplinary engineering efforts. This manuscript formalizes transfer learning, multi-task learning, and meta-learning as abstract learning systems, consistent with the formal-minimalist abstract systems theory of Mesarovic and Takahara. Moreover, it uses the presented formalism to relate the three concepts of learning in terms of composition, hierarchy, and structural homomorphism. Findings are readily depicted in terms of input-output systems, highlighting the ease of delineating formal, general systems differences between transfer, multi-task, and meta-learning.

Keywords: Abstract Learning Systems · Transfer Learning · Multi-Task Learning · Meta-Learning · Abstract Systems Theory

1 Introduction

Transfer learning, multi-task learning, and meta-learning are three different concepts of learning that aim to generalize knowledge across learning tasks. As such, they are common topics in artificial general intelligence [1,7]. They are informally described as similar in their respective, prominent surveys [6, 9, 11]. Formally, however, the general systems character of this similarity is left undiscussed. Likely, this is because the formalism of their respective learning algorithms quickly represents their differences. While this gap may seem inconsequential to algorithm designers, who typically work very closely to solution methods, to systems engineers, this gap muddles basic questions about composition and hierarchy.

In this manuscript, a recently proposed abstract systems theory (AST) model of learning [2,4] is used to formally relate transfer, multi-task, and meta-learning. Each concept of learning is modeled as an abstract system [5], i.e., as a relation on component sets, and their structural homomorphism is studied. The presented results extend previous work that synthesizes AST with statistical learning theory [2] and transfer learning [3,4] with novel definitions of multi-task and meta-learning as abstract systems, and with an investigation of their structural similarities.

This manuscript is structured as follows. First, preliminaries on abstract learning systems and transfer learning systems are given in Sections 2 and 3. Subsequently, multi-task learning and meta-learning are formalized as systems from their informal descriptions in Section 4, and the homomorphism between transfer, multi-task, and meta-learning is investigated in Section 5. The manuscript concludes with a synopsis and remarks on the pitfalls of the existing informal taxonomy in light of the presented material.

2 Abstract Learning Systems

Abstract systems S are relations on (non-empty) abstract sets

$$S \subset \times\{V_i | i = 1, ..., I\},$$

where $\times$ is the Cartesian product, V_i are (component) sets, and $\overline{S} = \{V_i | i = 1, ..., I\}$ [5]. Input-output systems are (elementary) systems

$$S \subset \times\{\mathcal{X}, \mathcal{Y}\},$$

where $\mathcal{X} \cap \mathcal{Y} = \varnothing$, $\mathcal{X} \cup \mathcal{Y} = \overline{S}$, and $\varnothing$ is the empty set. The set $\mathcal{X}$ is termed the input and the set $\mathcal{Y}$ is termed the output. Functional systems are input-output systems of the form $S : \mathcal{X} \to \mathcal{Y}$. AST is primarily concerned with input-output systems, with their composition, and with categories of systems [5].

Recent work presented a stratified model of abstract learning systems as a cascade connection of learning algorithms $A : D \to \Theta$ and hypotheses $H : \Theta \times \mathcal{X} \to \mathcal{Y}$ where D are data and Θ are parameters [2]. This follows the treatment of learning as function approximation [10]. Learning systems are defined as follows.

Definition 1 (Learning Systems.).
A learning system S is a relation

$$S \subset \times\{A, D, \Theta, H, \mathcal{X}, \mathcal{Y}\}$$

such that

$$D \subset \mathcal{X} \times \mathcal{Y}, A : D \to \Theta, H : \Theta \times \mathcal{X} \to \mathcal{Y}$$
$$(d, x, y) \in \mathcal{P}(S) \leftrightarrow (\exists\theta)[(\theta, x, y) \in H \wedge (d, \theta) \in A]$$

where

$$x \in \mathcal{X}, y \in \mathcal{Y}, d \in D, \theta \in \Theta.$$

The algorithm A, data D, parameters Θ, hypotheses H, input $\mathcal{X}$, and output $\mathcal{Y}$ are the component sets of S, $\mathcal{P}$ is the power set, and learning is specified in the relation among them.

This AST model of learning is depicted in Figure 1 at the elementary (input-output) and cascade levels of abstraction (as presented in [2]).

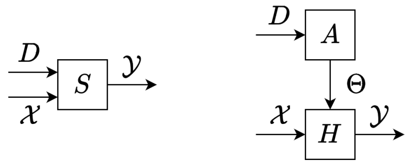

Fig. 1. Learning systems at the elementary (left) and cascade (right) levels of abstraction [2].

3 Transfer Learning Systems

The concept of learning tasks is widely used in artificial intelligence [8]. Transfer learning is conventionally defined in terms of domains $\mathcal{D} = \{\mathcal{X}, P(X)\}$ and tasks $\mathcal{T} = \{\mathcal{Y}, P(Y|X)\}$, where P denotes a probability measure. Given a source domain $\mathcal{D}_S$ and learning task $\mathcal{T}_S$, a target domain $\mathcal{D}_T$ and learning task $\mathcal{T}_T$, Pan and Yang define transfer learning as a learning paradigm that [6],

> "aims to help improve the learning of the target predictive function f_T[1] in $\mathcal{D}_T$ using the knowledge in $\mathcal{D}_S$ and $\mathcal{D}_T$, where $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$."

Alternatively, previous work describes transfer learning as [4],

> "...a relation on the source and target (learning) systems that combines knowledge from the source with data from the target and uses the result to select a hypothesis that estimates the target learning task."

Transfer learning systems are defined as follows.

Definition 2 (Transfer Learning System.).
Given source and target learning systems S_S and S_T

$$S_S \subset \times\{A_S, D_S, \Theta_S, H_S, X_S, Y_S\}$$
$$S_T \subset \times\{A_T, D_T, \Theta_T, H_T, X_T, Y_T\}$$

a transfer learning system S_{Tr} is a relation on the component sets of the source and target systems $S_{Tr} \subset \overline{S_S} \times \overline{S_T}$ such that

$$K_S \subset D_S \times \Theta_S, D \subset D_T \times K_S$$

and

$$A_{Tr} : D \to \Theta_{Tr}, H_{Tr} : \Theta_{Tr} \times X_T \to Y_T$$
$$(d, x_T, y_T) \in \mathcal{P}(S_{Tr}) \leftrightarrow$$
$$(\exists \theta_{Tr})[(\theta_{Tr}, x_T, y_T) \in H_{Tr} \wedge (d, \theta_{Tr}) \in A_{Tr}]$$

where

$$x_T \in X_T, y_T \in Y_T, d \in D, \theta_{Tr} \in \Theta_{Tr}.$$

[1] $f_T \sim P(Y_T|X_T)$

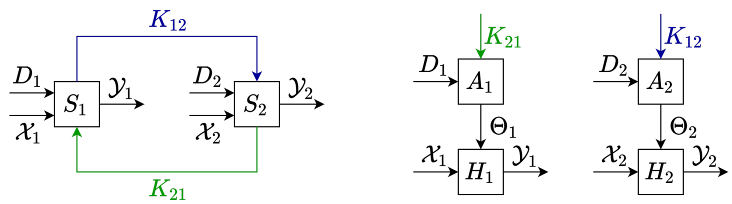

Fig. 2. Two learning systems transferring knowledge to each other depicted at the elementary (left) and cascade (middle, right) levels of abstraction [4].

The nature of source knowledge K_S[2]*, the transfer learning algorithm* A_{Tr}*, hypotheses* H_{Tr}*, and parameters* Θ_{Tr} *specify transfer learning as a relation on* $\overline{S_S}$ *and* $\overline{S_T}$*.*

This AST model of transfer learning is depicted in Figure 2. Previous work extensively elaborates on and beyond Definition 2 [4].

4 Multi-Task and Meta-Learning Systems

4.1 Multi-Task Learning

Zhang and Yang define multi-task learning as [11],

> "a learning paradigm in machine learning and its aim is to leverage useful information contained in multiple related tasks to help improve the generalization performance of all the tasks."

Multi-task learning systems are defined herein as follows.

Definition 3 (Multi-Task Learning Systems.).
Given N *learning systems* $S_1, ..., S_N$*, a multi-task learning system is a learning system* $S \subset \times\{A, D, \Theta, H, \mathcal{X}, \mathcal{Y}\}$ *where,*

$$D = (D_1, ..., D_N), H = (H_1, ..., H_N),$$
$$\Theta = (\Theta_1, ..., \Theta_N), \mathcal{X} = (\mathcal{X}_1, ..., \mathcal{X}_N),$$
$$\mathcal{Y} = (\mathcal{Y}_1, ..., \mathcal{Y}_N),$$

i.e., $A : (D_1, ..., D_N) \to (\Theta_1, ..., \Theta_N)$.

Multi-task learning systems are simply learning systems that jointly learn multiple, distinct hypotheses. Multi-task learning systems are depicted in Figure 3A. A trivial multi-task learning system can be defined as follows.

[2] Here transferred knowledge K_S is defined as D_S and Θ_S, the source data and parameters, following convention [6].

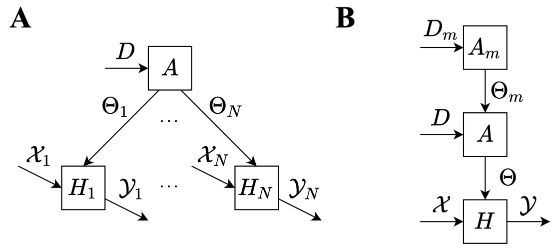

Fig. 3. A multi-task learning system (A) and a meta-learning system (B).

Definition 4 (Trivial Multi-Task Learning Systems.).
Given N learning systems $S_1, ..., S_N$, a trivial multi-task learning system is a multi-task learning system $S \subset \times\{A, D, H, \Theta, \mathcal{X}, \mathcal{Y}\}$ defined over $S_1, ..., S_N$ where $A = (A_1, ..., A_N)$.

In other words, the trivial case of multi-task learning is a superficial grouping of algorithms $(A_1, ..., A_N)$ where A simply uses $D_1, ..., D_N$ as input to each respective algorithm A_n for $n \in N$. A non-trivial multi-task learning system can be defined as follows.

Definition 5 (Non-Trivial Multi-Task Learning Systems.).
Given N learning systems $S_1, ..., S_N$, a non-trivial multi-task learning system is a multi-task learning system $S \subset \times\{A, D, H, \Theta, \mathcal{X}, \mathcal{Y}\}$ defined over $S_1, ..., S_N$ where $A \neq (A_1, ..., A_N)$.

4.2 Meta-Learning

Vanschoren defines meta-learning as [9]

> "the science of systematically observing how different machine learning approaches perform on a wide range of learning tasks, and then learning from this experience, or meta-data, to learn new tasks much faster than otherwise possible."

Meta-learning systems are defined herein as follows.

Definition 6 (Meta-Learning System.).
Meta-learning systems are learning systems $S \subset \times\{A_m, \Theta_m, D_m, H_m, \mathcal{X}_m, \mathcal{Y}_m\}$ with hypotheses H_m that are algorithms A, inputs $\mathcal{X}_m$ that are data D, outputs $\mathcal{Y}_m$ that are parameters Θ for hypotheses $H : \Theta \times \mathcal{X} \to \mathcal{Y}$, and where $S \subset \times\{A, D, \Theta, H, \mathcal{X}, \mathcal{Y}\}$ is a learning system.

Meta-learning systems are learning systems whose hypotheses are learning algorithms. Meta-learning systems are depicted in Figure 3B.

5 Homomorphisms Between Learning Systems

Similarity of systems is a fundamental notion. Structural similarity describes the *homomorphism* between two systems' structures. In accord with category theory, a map from one system to another is termed a morphism. Homomorphism specifies the morphism to be onto. Homomorphism is formally defined as follows.

Definition 7. Homomorphism.
An input-output system $S_1 \subset \times\{\mathcal{X}_1 \times \mathcal{Y}_1\}$ is homomorphic to $S_2 \subset \times\{\mathcal{X}_2, \mathcal{Y}_2\}$ if there exists a pair of maps,

$$\varrho : \mathcal{X}_1 \to \mathcal{X}_2, \vartheta : \mathcal{Y}_1 \to \mathcal{Y}_2$$

such that for all $x_1 \in \mathcal{X}_1$, $x_2 \in \mathcal{X}_2$, and $y_1 \in \mathcal{Y}_1$, $y_2 \in \mathcal{Y}_2$, $\varrho(x_1) = x_2$ and $\vartheta(y_1) = y_2$.

Let a two-way transfer learning system be a pair of transfer learning systems that both transfer knowledge to each other. In the following, it is proven that two transfer learning systems sharing knowledge with each other are homomorphic to a non-trivial multi-task learning system, as depicted in Figure 4.

Theorem 1. *Two-way transfer learning systems are homomorphic to a non-trivial multi-task learning system.*

Proof. Consider two learning systems S'_1 and S'_2,

$$S'_1 \subset \times\{A'_1, D'_1, \Theta'_1, H'_1, \mathcal{X}_1, \mathcal{Y}_1\},$$
$$S'_2 \subset \times\{A'_2, D'_2, \Theta'_2, H'_2, \mathcal{X}_2, \mathcal{Y}_2\}.$$

Let transfer learning be used to transfer knowledge $K_{12} \subset D'_1$ from S'_1 to S'_2 and knowledge $K_{21} \subset D'_2$ from S'_2 to S'_1. This creates two transfer learning systems, termed S_1 and S_2, respectively,

$$S_1 \subset \times\{A_1, D_1, \Theta_1, H_1, \mathcal{X}_1, \mathcal{Y}_1\},$$
$$S_2 \subset \times\{A_2, D_2, \Theta_2, H_2, \mathcal{X}_2, \mathcal{Y}_2\}.$$

where $D_1 \subset D'_1 \times D'_2$ and $D_2 \subset D'_1 \times D'_2$, as in Figure 2. Consider a multi-task learning system

$$S \subset \times\{A, D, \Theta, H, \mathcal{X}, \mathcal{Y}\}$$

such that $A = (A_1, A_2)$, $D = (D_1, D_2)$, $\Theta = (\Theta_1, \Theta_2)$, $H = (H_1, H_2)$, $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2)$, and $\mathcal{Y} = (\mathcal{Y}_1, \mathcal{Y}_2)$. Clearly, by the identity, A, D, Θ, H, $\mathcal{X}$, and $\mathcal{Y}$ are homomorphic to (A_1, A_2), (D_1, D_2), (Θ_1, Θ_2), (H_1, H_2), $(\mathcal{X}_1, \mathcal{X}_2)$, and $(\mathcal{Y}_1, \mathcal{Y}_2)$, respectively. Thus, there exists a set of onto maps $\{\varrho_A, \varrho_D, \varrho_\Theta, \varrho_H, \varrho_\mathcal{X}, \varrho_\mathcal{Y}\}$ from $(S_1, S_2) \to S$. Thus, the two-way transfer learning system (S_1, S_2) is homomorphic to the multi-task learning system S. Since $A = (A_1, A_2)$ and since $A'_1 \neq A_1$ and $A'_2 \neq A_2$ (they have different supports), S is therefore necessarily a non-trivial multi-task learning system with respect to S'_1 and S'_2. $\square$

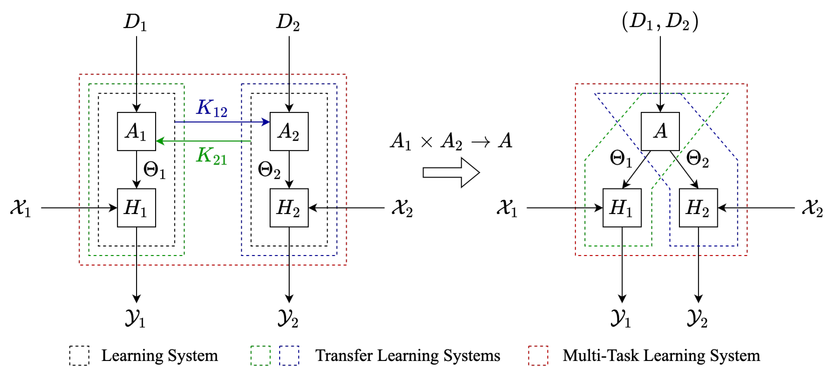

Fig. 4. A set of learning systems all transferring knowledge to each other implicitly forms a multi-task learning system.

The above hints that multi-task learning systems are related to parallel connections of learning systems. To see this, first define a parallel connection as follows.

Definition 8 (Parallel Connections).
A parallel connection of systems $S_1 : \mathcal{X}_1 \to \mathcal{Y}_1$ *and* $S_2 : \mathcal{X}_2 \to \mathcal{Y}_2$ *is an operator* $\| : \overline{S_1} \times \overline{S_2} \to \overline{S_2}$ *such that* $S_2 : (\mathcal{X}_1 \times \mathcal{X}_2) \to (\mathcal{Y}_1 \times \mathcal{Y}_2)$ *and*

$$((x_1, x_2), (y_1, y_2)) \in S_2 \leftrightarrow$$
$$((x_1, y_1)) \in S_1 \wedge ((x_2, y_2) \in S_2).$$

Theorem 2. *Trivial multi-task systems are a parallel connection of learning systems.*

Proof. Consider a set of N learning systems S_1, ..., S_N. Let $S_{1-2} = S_1 \| S_2$. Let $S_{1-n} = S_{1-(n-1)} \| S_n$. Thus, S_{1-N}, at the elementary (input-output) level of abstraction is,

$$S_{1-N} = (D_1 \times ... \times D_N) \times (\mathcal{X}_1 \times ... \times \mathcal{X}_N) \to (\mathcal{Y}_1 \times ... \times \mathcal{Y}_N),$$

which simplifies to,

$$S_{1-N} : (D_1, ..., D_N) \times (\mathcal{X}_1, ..., \mathcal{X}_N) \to (\mathcal{Y}_1, ..., \mathcal{Y}_N).$$

Let $D = (D_1, ..., D_N)$, $\mathcal{X} = (\mathcal{X}_1, ..., \mathcal{X}_N)$, and $\mathcal{Y} = (\mathcal{Y}_1, ..., \mathcal{Y}_N)$. Let A, Θ, and H be defined similarly. By definition, the system

$$S \subset \times \{A, D, \Theta, H, \mathcal{X}, \mathcal{Y}\}$$

is a trivial multi-task learning system. □

When triviality does not hold, multi-task learning is a *shallow* parallel connector in the sense that it is always a parallel connector in (elementary-level) terms of D, $\mathcal{X}$, and $\mathcal{Y}$, but not always a parallel connector in the (cascade-level) terms of the relation on them given by A, Θ, and H since $\exists A \neq (A_1, ..., A_N)$.

In contrast to the parallel connections of multi-task learning, meta-learning is related to cascade connections and hierarchy. To see this, first define a cascade connection as follows.

Definition 9 (Cascade Connections).
Let $\circ : \overline{S} \times \overline{S} \to \overline{S}$ *be such that* $S_1 \circ S_2 = S_3$*, where,*

$$S_1 \subset X_1 \times (Y_1 \times (Z_1)), S_2 \subset (X_2 \times Z_2) \times Y_2$$
$$S_3 \subset (X_1 \times X_2) \times (Y_1 \times Y_2), Z_1 = Z_2 = Z$$

and,

$$((x_1, x_2), (y_1, y_2)) \in S_3 \leftrightarrow$$
$$(\exists z)((x_1, (y_1, z)) \in S_1 \wedge ((x_2, z), y_2) \in S_2)$$

$\circ$ *is termed the cascade (connecting) operator.*

Theorem 3. *Meta-learning systems are a cascade connection of a learning algorithm and a learning system.*

Proof. Consider a learning algorithm $A_2 : D_2 \to \Theta_2$ and a learning system S_1. Let $S_3 = A_2 \circ S_1$. Thus $S_3 = \subset \times\{A_2, D_2, \Theta_2, A_1, D_1, \Theta_1, H_1, \mathcal{X}_1, \mathcal{Y}_1\}$ where $A_1 : \Theta_2 \times D_1 \to \Theta_1$. Let $S_m \subset S_3$ such that $S_m \subset \times\{A', D', \Theta', A, D, \Theta\}$. S_m is a meta-learning system where A_1, D_1, and Θ_1 are the hypotheses, inputs, and outputs of S_m. $\square$

5.1 Discussion

The preceding provides the basic math needed to support the use of transfer, multi-task, and meta-learning as basic elements of modeling abstract learning systems. Because of the generality and homomorphism of these three concepts, a systems modeler clearly has many representations to choose from when modeling learning systems. Consider the learning system shown in Figure 5, which shows a series of homomorphisms on a learning system from Figure 5A to 5F.

Figure 5A shows two learning systems with meta-learning systems that transfer knowledge to each other. In Figure 5B, a parallel connection of A_1 and A_2 transforms the system in 5A into a multi-task learning system with a decomposed meta-learning system. In Figure 5C, the meta-learning system is recomposed into A_m using a parallel connection. Figure 5D shows the hypotheses of the multi-task learning system are composed into H by a parallel connection. Figure 5E shows the meta-learning hierarchy is collapsed into algorithm A' using a series connection. And lastly, in Figure 5F, sets are redefined to recover a general learning system as in Definition 1. All of these morphisms are homomorphisms—they are onto and thus structure-preserving.

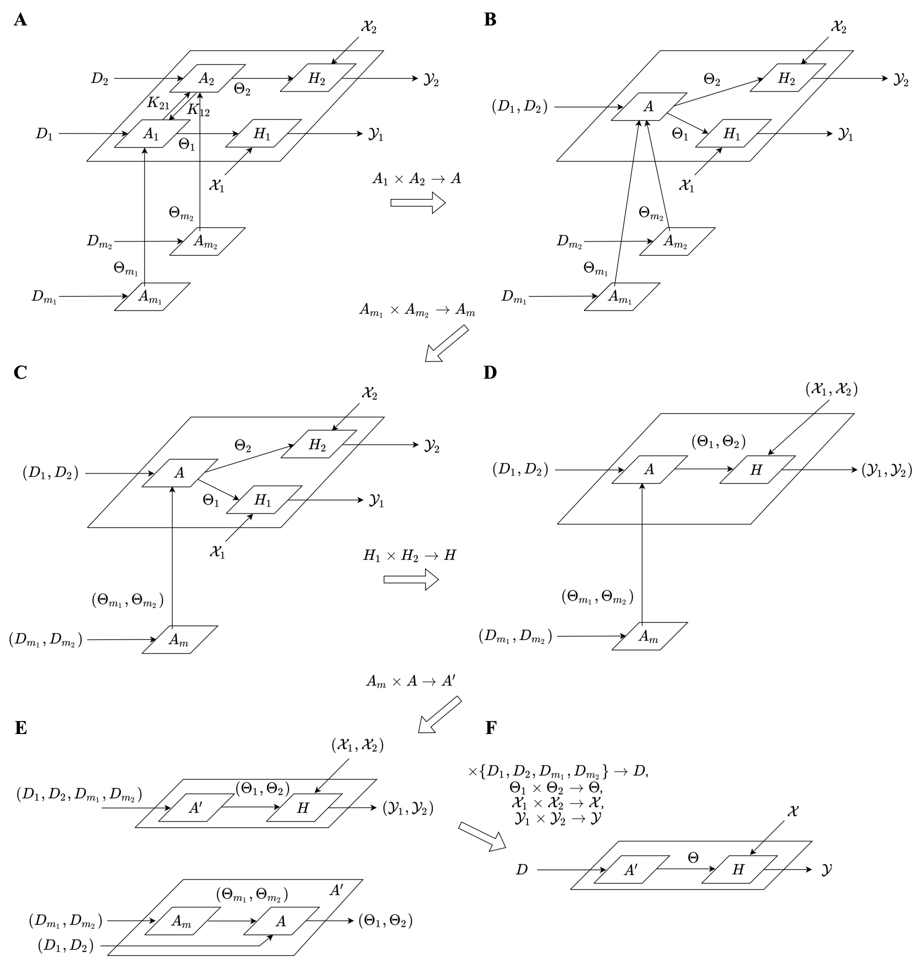

Fig. 5. A series of homomorphisms from a two-way transfer learning system with meta-learning in (A) to a generic learning system in (F), as described in detail in Section 5.1.

6 Conclusion

Appreciating the compositional and hierarchical relations between the three concepts of learning makes clear a simple point about learning systems. Transfer, multi-task, and meta-learning are basic compositions of learning systems that display a recursive self-similarity. A multi-task learning system can use transfer learning from other multi-task learning systems that all have meta-learning systems that use transfer learning, and so on.

What is a multi-task transfer learning system? What is a transfer multi-task learning system? What is a meta-transfer multi-task learning system? What is a meta-transfer multi-task transfer learning system? What is ... etc.? Clearly what is a useful taxonomy for organizing machine learning solution methods in the literature becomes burdensome and tedious when applied to systems modeling. Using the presented material, modelers have simple, general formalism for describing the compositions of learning systems possible by way of transfer, multi-task, and meta-learning.

References

1. Al-Shedivat, M., Bansal, T., Burda, Y., Sutskever, I., Mordatch, I., Abbeel, P.: Continuous adaptation via meta-learning in nonstationary and competitive environments. In: International Conference on Learning Representations (2018)
2. Cody, T.: Mesarovician abstract learning systems. In: International Conference on Artificial General Intelligence. pp. 55–64. Springer (2021)
3. Cody, T., Adams, S., Beling, P.A.: A systems theoretic perspective on transfer learning. In: 2019 IEEE International Systems Conference (SysCon). pp. 1–7. IEEE (2019)
4. Cody, T., Beling, P.A.: A systems theory of transfer learning. arXiv preprint arXiv:2107.01196 (2021)
5. Mesarovic, M.D., Takahara, Y.: Abstract systems theory. Springer (1989)
6. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering **22**(10), 1345–1359 (2009)
7. Sheikhlar, A., Thórisson, K.R., Eberding, L.M.: Autonomous cumulative transfer learning. In: International Conference on Artificial General Intelligence. pp. 306–316. Springer (2020)
8. Thórisson, K.R., Bieger, J., Thorarensen, T., Sigurðardóttir, J.S., Steunebrink, B.R.: Why artificial intelligence needs a task theory. In: International Conference on Artificial General Intelligence. pp. 118–128. Springer (2016)
9. Vanschoren, J.: Meta-learning: A survey. arXiv preprint arXiv:1810.03548 (2018)
10. Wang, P., Li, X.: Different conceptions of learning: function approximation vs. self-organization. In: International Conference on Artificial General Intelligence. pp. 140–149. Springer (2016)
11. Zhang, Y., Yang, Q.: A survey on multi-task learning. IEEE Transactions on Knowledge and Data Engineering (2021)