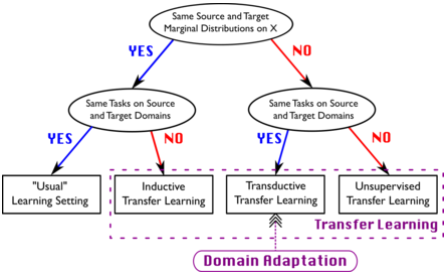


Domain adaptation

Domain adaptation^{[1][2][3]} is a field associated with machine learning and transfer learning. This scenario arises when we aim at learning from a source data distribution a well performing model on a different (but related) target data distribution. For instance, one of the tasks of the common spam filtering problem consists in adapting a model from one user (the source distribution) to a new user who receives significantly different emails (the target distribution). Domain adaptation has also been shown to be beneficial for learning unrelated sources.^[4] Note that, when more than one source distribution is available the problem is referred to as multi-source domain adaptation.^[5]



Distinction between usual machine learning setting and transfer learning, and positioning of domain adaptation.

Contents

Overview

Domain shift

Examples

Formalization

The different types of domain adaptation

Four algorithmic principles

Reweighting algorithms

Iterative algorithms

Search of a common representation space

Hierarchical Bayesian Model

References

Overview

Domain adaptation is the ability to apply an algorithm trained in one or more "source domains" to a different (but related) "target domain". Domain adaptation is a subcategory of transfer learning. In domain adaptation, the source and target domains all have the same feature space (but different distributions); in contrast, transfer learning includes cases where the target domain's feature space is different from the source feature space or spaces.^[6]

Domain shift

A **domain shift**,^[7] or **distributional shift**,^[8] is a change in the data distribution between an algorithm's training dataset, and a dataset it encounters when deployed. These domain shifts are common in practical applications of artificial intelligence. Conventional machine-learning

algorithms often adapt poorly to domain shifts. The modern machine-learning community has many different strategies to attempt to gain better domain adaptation.^[7]

Examples

- An algorithm trained on newswires might have to adapt to a new dataset of biomedical documents.^[9]
- A spam filter, trained on a certain group of email users during training, must adapt to a new target user when deployed.^[10]
- Applying AI diagnostic algorithms, trained on labeled data associated with previous diseases, to new unlabeled data associated with the COVID-19 pandemic.^[11]
- A sudden societal change, such as a pandemic outbreak, can constitute domain shift and cause machine learning algorithms trained on now-obsolete consumer data to fail and require intervention.^{[12][13]}

Other applications include wifi localization detection and many aspects of computer vision.^[6]

Formalization

Let \mathbf{X} be the input space (or description space) and let \mathbf{Y} be the output space (or label space). The objective of a machine learning algorithm is to learn a mathematical model (a hypothesis) $h : \mathbf{X} \rightarrow \mathbf{Y}$ able to attach a label from \mathbf{Y} to an example from \mathbf{X} . This model is learned from a learning sample $\mathcal{S} = \{(\mathbf{x}_i, \mathbf{y}_i) \in (\mathbf{X} \times \mathbf{Y})\}_{i=1}^m$.

Usually in supervised learning (without domain adaptation), we suppose that the examples $(\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{S}$ are drawn i.i.d. from a distribution $\mathbf{D}_{\mathcal{S}}$ of support $\mathbf{X} \times \mathbf{Y}$ (unknown and fixed). The objective is then to learn h (from \mathcal{S}) such that it commits the least error possible for labelling new examples coming from the distribution $\mathbf{D}_{\mathcal{S}}$.

The main difference between supervised learning and domain adaptation is that in the latter situation we study two different (but related) distributions $\mathbf{D}_{\mathcal{S}}$ and $\mathbf{D}_{\mathcal{T}}$ on $\mathbf{X} \times \mathbf{Y}$. The domain adaptation task then consists of the transfer of knowledge from the source domain $\mathbf{D}_{\mathcal{S}}$ to the target one $\mathbf{D}_{\mathcal{T}}$. The goal is then to learn h (from labeled or unlabelled samples coming from the two domains) such that it commits as little error as possible on the target domain $\mathbf{D}_{\mathcal{T}}$.

The major issue is the following: if a model is learned from a source domain, what is its capacity to correctly label data coming from the target domain?

The different types of domain adaptation

There are several contexts of domain adaptation. They differ in the information considered for the target task.

1. The **unsupervised domain adaptation**: the learning sample contains a set of labeled source examples, a set of unlabeled source examples and a set of unlabeled target examples.
2. The **semi-supervised domain adaptation**: in this situation, we also consider a "small" set of labeled target examples.

3. The **supervised domain adaptation**: all the examples considered are supposed to be labeled.

Four algorithmic principles

Reweighting algorithms

The objective is to reweight the source labeled sample such that it "looks like" the target sample (in terms of the error measure considered).^{[14][15]}

Iterative algorithms

A method for adapting consists in iteratively "auto-labeling" the target examples.^[16] The principle is simple:

1. a model h is learned from the labeled examples;
2. h automatically labels some target examples;
3. a new model is learned from the new labeled examples.

Note that there exist other iterative approaches, but they usually need target labeled examples.^{[17][18]}

Search of a common representation space

The goal is to find or construct a common representation space for the two domains. The objective is to obtain a space in which the domains are close to each other while keeping good performances on the source labeling task. This can be achieved through the use of Adversarial machine learning techniques where feature representations from samples in different domains are encouraged to be indistinguishable.^{[19][20]}

Hierarchical Bayesian Model

The goal is to construct a Bayesian hierarchical model $p(\mathbf{n})$, which is essentially a factorization model for counts \mathbf{n} , to derive domain-dependent latent representations allowing both domain-specific and globally shared latent factors.^[4]

References

1. Redko, Ievgen; Morvant, Emilie; Habrard, Amaury; Sebban, Marc; Bennani, Younès (2019). *Advances in Domain Adaptation Theory* (<https://www.elsevier.com/books/advances-in-domain-adaptation-theory/redko/978-1-78548-236-6>). ISTE Press – Elsevier. p. 187. ISBN 9781785482366.
2. Bridle, John S.; Cox, Stephen J (1990). "RecNorm: Simultaneous normalisation and classification applied to speech recognition" (<http://papers.nips.cc/paper/328-recnorm-simultaneous-normalisation-and-classification-applied-to-speech-recognition.pdf>) (PDF). *Conference on Neural Information Processing Systems (NIPS)*. pp. 234–240.

3. Ben-David, Shai; Blitzer, John; Crammer, Koby; Kulesza, Alex; Pereira, Fernando; Wortman Vaughan, Jennifer (2010). "A theory of learning from different domains" (<https://link.springer.com/content/pdf/10.1007/s10994-009-5152-4.pdf>) (PDF). *Machine Learning*. **79** (1–2): 151–175. doi:10.1007/s10994-009-5152-4 (<https://doi.org/10.1007%2Fs10994-009-5152-4>).
4. Hajiramezanali, Ehsan; Siamak Zamani Dadaneh; Karbalayghareh, Alireza; Zhou, Mingyuan; Qian, Xiaoning (2018). "Bayesian multi-domain learning for cancer subtype discovery from next-generation sequencing count data". arXiv:1810.09433 (<https://arxiv.org/abs/1810.09433>) [stat.ML (<https://arxiv.org/archive/stat.ML>)].
5. Crammer, Koby; Kearns, Michael; Wortman, Jeniifer (2008). "Learning from Multiple Sources" (<http://www.jmlr.org/papers/volume9/crammer08a/crammer08a.pdf>) (PDF). *Journal of Machine Learning Research*. **9**: 1757–1774.
6. Sun, Shiliang; Shi, Honglei; Wu, Yuanbin (July 2015). "A survey of multi-source domain adaptation". *Information Fusion*. **24**: 84–92. doi:10.1016/j.inffus.2014.12.003 (<https://doi.org/10.1016%2Fj.inffus.2014.12.003>).
7. Sun, Baochen, Jiashi Feng, and Kate Saenko. "Return of frustratingly easy domain adaptation." In Thirtieth AAAI Conference on Artificial Intelligence. 2016.
8. Amodei, Dario, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. "Concrete problems in AI safety." arXiv preprint arXiv:1606.06565 (2016).
9. Daumé III, Hal. "Frustratingly easy domain adaptation." arXiv preprint arXiv:0907.1815 (2009).
10. Ben-David, Shai, John Blitzer, Koby Crammer, and Fernando Pereira. "Analysis of representations for domain adaptation." In Advances in neural information processing systems, pp. 137–144. 2007.
11. Hu, Yipeng; Jacob, Joseph; Parker, Geoffrey J. M.; Hawkes, David J.; Hurst, John R.; Stoyanov, Danail (June 2020). "The challenges of deploying artificial intelligence models in a rapidly evolving pandemic" (<https://www.nature.com/articles/s42256-020-0185-2>). *Nature Machine Intelligence*. **2** (6): 298–300. arXiv:2005.12137 (<https://arxiv.org/abs/2005.12137>). doi:10.1038/s42256-020-0185-2 (<https://doi.org/10.1038%2Fs42256-020-0185-2>). ISSN 2522-5839 (<https://www.worldcat.org/issn/2522-5839>).
12. Matthews, Dylan (26 March 2019). "AI disaster won't look like the Terminator. It'll be creepier" (<https://www.vox.com/future-perfect/2019/3/26/18281297/ai-artificial-intelligence-safety-disaster-scenarios>). *Vox*. Retrieved 21 June 2020.
13. "Our weird behavior during the pandemic is messing with AI models" (<https://www.technologyreview.com/2020/05/11/1001563/covid-pandemic-broken-ai-machine-learning-amazon-retail-fraud-humans-in-the-loop/>). *MIT Technology Review*. 11 May 2020. Retrieved 21 June 2020.
14. Huang, Jiayuan; Smola, Alexander J.; Gretton, Arthur; Borgwardt, Karster M.; Schölkopf, Bernhard (2006). "Correcting Sample Selection Bias by Unlabeled Data" (<http://papers.nips.cc/paper/3075-correcting-sample-selection-bias-by-unlabeled-data.pdf>) (PDF). *Conference on Neural Information Processing Systems (NIPS)*. pp. 601–608.
15. Shimodaira, Hidetoshi (2000). "Improving predictive inference under covariate shift by weighting the log-likelihood function" (<https://www.researchgate.net/publication/230710850>). *Journal of Statistical Planning and Inference*. **90** (2): 227–244. doi:10.1016/S0378-3758(00)00115-4 (<https://doi.org/10.1016%2FS0378-3758%2800%2900115-4>).
16. Gallego, A.J.; Calvo-Zaragoza, J.; Fisher, R.B. (2020). "Incremental Unsupervised Domain-Adversarial Training of Neural Networks" (https://www.pure.ed.ac.uk/ws/files/17203566/0/Incremental_Unsupervised_GALLEGO_DOA18092020_AFV.pdf) (PDF). *IEEE Transactions on Neural Networks and Learning Systems*. **PP**: 1–15. doi:10.1109/TNNLS.2020.3025954 (<https://doi.org/10.1109%2FTNNLS.2020.3025954>). hdl:20.500.11820/72ba0443-8a7d-4cdd-8212-38682d4f0730 (<https://hdl.handle.net/20.500.11820%2F72ba0443-8a7d-4cdd-8212-38682d4f0730>). PMID 33027004 (<https://pubmed.ncbi.nlm.nih.gov/33027004>).

17. Arief–Ang, I.B.; Salim, F.D.; Hamilton, M. (2017–11–08). *DA–HOC: semi-supervised domain adaptation for room occupancy prediction using CO2 sensor data* (<https://dl.acm.org/citation.cfm?id=3137146>). 4th ACM International Conference on Systems for Energy–Efficient Built Environments (BuildSys). Delft, Netherlands. pp. 1–10. doi:10.1145/3137133.3137146 (<https://doi.org/10.1145%2F3137133.3137146>). ISBN 978–1–4503–5544–5.
18. Arief–Ang, I.B.; Hamilton, M.; Salim, F.D. (2018–12–01). "A Scalable Room Occupancy Prediction with Transferable Time Series Decomposition of CO2 Sensor Data". *ACM Transactions on Sensor Networks*. **14** (3–4): 21:1–21:28. doi:10.1145/3217214 (<https://doi.org/10.1145%2F3217214>).
19. Ganin, Yaroslav; Ustinova, Evgeniya; Ajakan, Hana; Germain, Pascal; Larochelle, Hugo; Laviolette, François; Marchand, Mario; Lempitsky, Victor (2016). "Domain–Adversarial Training of Neural Networks" (<http://jmlr.org/papers/volume17/15–239/15–239.pdf>) (PDF). *Journal of Machine Learning Research*. **17**: 1–35.
20. Hajiramezanali, Ehsan; Siamak Zamani Dadaneh; Karbalayghareh, Alireza; Zhou, Mingyuan; Qian, Xiaoning (2017). "Addressing Appearance Change in Outdoor Robotics with Adversarial Domain Adaptation". arXiv:1703.01461 (<https://arxiv.org/abs/1703.01461>) [cs.RO (<https://arxiv.org/archive/cs.RO>)].

Retrieved from "https://en.wikipedia.org/w/index.php?title=Domain_adaptation&oldid=1064795990"

This page was last edited on 10 January 2022, at 07:08 (UTC).

Text is available under the Creative Commons Attribution–ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non–profit organization.