

Transfer and transport: incorporating causal methods for improving predictive models

Predicting patient outcome is an important task in medical decision making, as clinician expectations of outcome drive testing and treatment decisions. Accurate models can assist clinicians by capitalizing on information from a broad spectrum of features to predict outcome. In an article in this journal, 'A study in transfer learning: leveraging data from multiple hospitals to enhance hospital-specific predictions,' Wiens, Guttag, and Horvitz¹ explore the use of transfer learning for improving a predictive model of *Clostridium difficile* infection (CDI). Their discussion focuses on the need to aggregate data for studying rare diseases but notes the failure of global models to predict accurately for specific institutions. Transfer learning attempts to rectify the generalizability problem by applying evidence from multiple sources on a related target task. Their work demonstrates how transfer learning can be utilized to create a 'source+target' model matching or outperforming models trained with source or target data alone. A number of important considerations when pooling data are raised by the authors. Here, we note the need for further discussion by revisiting two points raised in their paper affecting transfer: (1) feature similarity and (2) feature selection. We briefly discuss limitations of transfer learning tied to a lack of causal knowledge and posit that causal information can complement transfer learning to improve model generalization.

Wiens *et al* combined datasets by comparing the overlap of features in data collected across hospitals. However, it is

important to note that overlapping features do not guarantee feature similarity. This issue was explored in the section, 'Not all transfer is created equal'; hospital B was determined as the most different of the three, transferring evidence poorly to target tasks at hospitals A and C. When differences are minimal, the task of transfer learning is straightforward and data can be aggregated freely. But uninformed assumptions of similarity could have a detrimental effect on model accuracy through population and confounding effects. Population differences are either systematic collection differences or inherent differences in the given populations at each location. Confounding differences are tied to the causal interactions and subsequent correlations between chosen features in the model. Wiens *et al* discuss the apparent difference between the three hospitals examined in the work, yet no explanation is provided concerning why such differences exist or the probable source of differences (population/confounding). Consideration should be given to feature similarity early in the modeling task to appropriately constrain the model. Poor assumptions may lead to spurious associations or remove generalization.

To enhance feature similarity exploration, causal assumptions between variables can be drawn from expert knowledge and previous research. CDI risk, for example, is associated with increased age, duration of hospitalization, and exposure to antimicrobial agents in past scientific literature (eg, randomized clinical trials).² Age, hospitalization, and antimicrobial drug knowledge can be combined with other empirical evidence to define causal assumptions and construct a causal graph, providing a linked consideration of feature effects. Transportability theory, introduced by Pearl and Bareinboim,³ offers a basis for using causal graph relations to describe which variables' probability distributions 'transport' or are more likely to generalize, between populations. Transport encompasses transfer learning in attempting to use statistical evidence from a source on a target, but differs by incorporating causal assumptions derived from a combination of empirical source data and outside domain knowledge (table 1). The final product of transportability analysis is a formula dictating what information should be combined from the considered domains. Transport techniques allow information to be pulled from appropriate data resources: a larger source can strengthen a weaker target or a strong target can build on portions of a source. Outside domain knowledge can also define assumptions or

strengthen weak empirical findings. Depending on the causal network structure, features can be deemed 'non-transportable', indicating that source data will never accurately predict the target task despite overlap. When multiple source tasks are available, causal considerations for each source can be combined to yield a transport formula using the expanded rubrics from meta-transportability.⁴

Feature selection is another important consideration of the learning task that received limited discussion. Causal assumptions cannot be properly described until a modeler understands variables available to the model. Chosen predictive features should make sense as causal indicators for the medical task of raising a CDI alert. For example, physician and location features, as suggested in the paper, may be predictive of increased CDI risk, but these features are not independent: location strongly mediates the physician (ie, a physician is associated with a location). As such, the physician feature may be unstable.⁵ Feature selection is also unstable when source and target locations cannot supply enough data. Additional evidence must be obtained before transfer or transport techniques can proceed. Also large feature spaces can be beneficial to finding predictive correlations, but these large spaces are challenging for considering causal relationships and require larger datasets. Consider Wiens *et al*'s model with 256 features common to all sites and features specific to the target task. The final set of selected features and the predictive weights of individual variables are not described, making it difficult for a reader to ascertain which available features contribute to prediction. Applying a deeper understanding of available features avoids blind application of large data sets. Understanding guides feature selection and describes what associations may be influenced by other factors. Striking a balance between big data and causal feature selection methods will be important for developing future learning techniques.

Additionally, the transfer method demonstrated by Wiens *et al* removes all source-only features from the model during feature selection. However, indiscriminately removing source features may decrease predictive performance. Including a source-only feature can be advantageous when data are difficult or expensive to obtain in routine care. Genetic phenotypes, for example, include important information about disease and are increasingly collected by academic research hospitals. Rural clinics, however, lack the ability to measure these same data. By considering population

Table 1 Definitions of 'Transfer' and 'Transport' terminology used in this correspondence

Technique	Definition
'Transfer' learning	Leveraging evidential knowledge from a source domain(s) to improve a predictive function for a target task
Transportability or 'transport' learning	Leveraging causal assumptions from source evidence and domain knowledge (experts and literature) to improve a predictive function for a target task

and causal assumptions through transportability, source genetic features from a larger research hospital can be evaluated for transport to a target. Similarly, it may not be evident why a source-only variable is missing from target data. For instance, Wiens *et al*'s data lack home medication values for hospitals A and B. Medication features obtained at hospital C might contain information useful to these targets, but determining their utility is tied up in the context of why the features were not available.

Wiens *et al* illustrate how transfer learning is suited to handling target features that are distinct from features in a global model, improving model prediction. Yet it remains unclear if transfer learning alone can account for the influence of variable difference and confounding in models. The reporting of selected features and other model assumptions is lacking in much of the modeling literature. Reporting model characteristics through a shared representation, such as an expanded version of the Predictive Model Markup Language,⁶ would be beneficial to future publications and models. There is potential for using causal techniques derived from transportability theory to determine when source-only variables may be used and when features specifically require target information.⁷ Transportability theory provides an opportunity to develop a formal

process for evaluating causal assumptions in order to dictate in a principled fashion what data are transportable across models. Options for temporal exploration of causality have also been described by Kleinberg and Hripcsak.⁵ We believe it will be beneficial to examine the contributions of Wiens *et al* in transfer learning in light of causal transport and other techniques to determine where each can provide improvement for model generalization in medicine.

Kyle W Singleton,^{1,2} Alex A T Bui,^{1,2} William Hsu^{1,2}

¹Department of Bioengineering, University of California, Los Angeles, USA

²Department of Radiological Sciences, Medical Imaging Informatics, University of California, Los Angeles, USA

Correspondence to Kyle W Singleton, Medical Imaging Informatics, Department of Radiological Sciences, University of California, Los Angeles, 924 Westwood Blvd, Suite 420, Los Angeles, CA 90024, USA; kwsingleton@ucla.edu

Contributors All authors contributed to writing and approval of the final manuscript.

Funding This work was supported by National Cancer Institute grant R01 CA157553.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.



CrossMark

To cite Singleton KW, Bui AAT, Hsu W. *J Am Med Inform Assoc* 2014;**21**:e374–e375.

Received 6 May 2014

Revised 29 May 2014

Accepted 26 June 2014

Published Online First 9 July 2014

J Am Med Inform Assoc 2014;**21**:e374–e375.

doi:10.1136/amiajnl-2014-002968

REFERENCES

- 1 Wiens J, Guttat J, Horvitz E. A study in transfer learning: leveraging data from multiple hospitals to enhance hospital-specific predictions. *J Am Med Inform Assoc* 2014;**21**:699–706.
- 2 Cohen SH, Gerding DN, Johnson S, *et al*. Clinical practice guidelines for Clostridium difficile infection in adults: 2010 update by the Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA). *Infect Control Hosp Epidemiol* 2010;**31**:431–55.
- 3 Pearl J, Bareinboim E. Transportability of Causal and Statistical Relations: A Formal Approach. *Twenty-Fifth AAAI Conference on Artificial Intelligence*. 2011.
- 4 Bareinboim E, Pearl J. Meta-Transportability of Causal Effects: A Formal Approach. *JMLR* 2013;**31**: 135–43.
- 5 Kleinberg S, Hripcsak G. A review of causal inference for biomedical informatics. *J Biomed Inform* 2011;**44**:1102–12.
- 6 Data Mining Group—PMML version 4.2. <http://www.dmg.org/pmml-v4-2.html> (accessed May 2014).
- 7 Singleton KW, Speier W, Bui AAT, *et al*. Motivating the Additional Use of External Validity: Examining Transportability in a Model of Glioblastoma Multiforme. *AMIA Annu Symp Proc* 2014. In press.