Monitoring Air Quality using Domain Adaptation and GAN

Arunavo Dey and Dr. Tanzima Islam
Department of CSE, College of Science and Engineering, Texas State University

Objective & Motivation

Recent environmental pollution is most detrimental to air which is one of most necessary elements of nature for living beings. Recent excessive ozone pollution made the Texas Commission on Environmental Quality (TCEQ) to call for an Ozone Action Day for the Austin area for Saturday, March 19, 2022 [1] to prevent ozone pollution. But in most of the cases predicting air quality with the help of data collected from somewhere else doesn't help much. Moreover measuring directly the amount of pollution may not be feasible always. Collecting directly the elements to predict one specific amount also may not be feasible also. So in this work it has been tried to measure one element from other available elements without taking account their contribution to the specific element. Here measuring ozone from other data has been tried out. Also taking elements from one city air it has been experimented to measure pollution in a complete different city. By studying this real life scenario it has been tried to measure the performance among different domain adaptation methods including optimal DANN and optimal transport and subspace alignment to measure their effectiveness in this real life scenerio.

Background

Domain Adaptation has been studied well in visual domain by different methods including DNS(domain separation networks), optimal transport and subspace mapping. [2] Moreover most of the works addressed classification problems where very few addressed the regression problem. [3] In optimal transport [5] it has been tried in different ways to approximate the source and target distributions by creating a transport amp which includes pseudo labels and taking target labels into account. In this work for optimal transport a 5 shot version of this problem has been considered. For the sake of comparing domain adversarial networks, generative adversarial networks has been used to approximate the target domain from the source domain data and that has been used to train a predictor to later generalize on target domain.

Research Questions

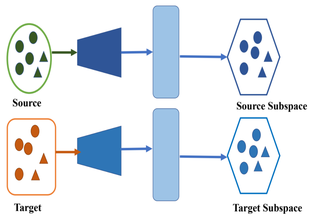
The specific research questions answered in this work are given below

How the Domain adaptation for real life regression problems work

How well domain can be adapted with subspace alignment approach compared to the DANN and Optimal Transport method

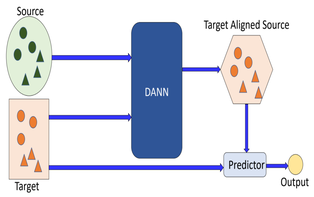
Here the domains sued for each experiment were containing homogeneous features in each case. The values though belonged from different distributions which makes this a real life distribution shift problem. Moreover the focus has been on not seeing any target labels which makes this a unsupervised adaptation problem.

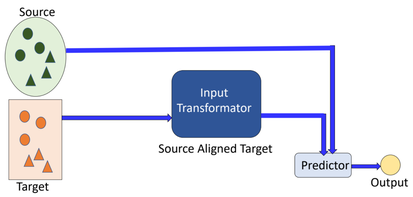
Methodology



Subspace alignment technique

DANN.

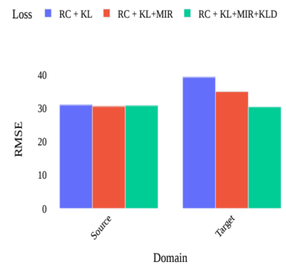




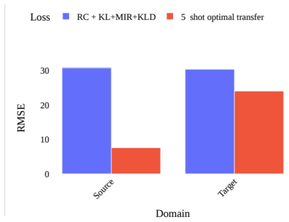
Optimal Transfer

Results

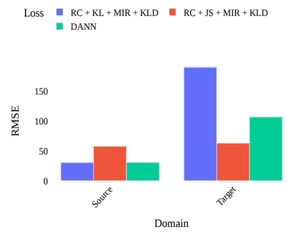
Here in the first study it can be seen that proposed loss function effectively smoothness the error with addition of each component. In the second study we try to experiment subspace alignment vs optimal transfer. For optimal transfer, 5 target labels were used which made it more privileged over subspace alignment which does not see any target labels. But the error is closely approximated in the target domain. The third study proves its effectiveness over DANN methods. All the experiments were hyper parameter tuned.



Ablation study of loss function.



Optimal Transport vs subspaace



DANN vs subspaace

Conclusion

Here different techniques have been studied with the subspace mapping but whether seeing target data improves the performance or not is still to be studied. Nevertheless it performs better without seeing any target label which makes it more applicable to use in real life scenario.

Social Implication

This real life problem is severe as we are facing the air pollution at a severe rate. We may not have the data of the intended location and the target location may have a distribution shift among features. Our study shows it may be fruitful to use subspace mapping for this domain of problems. And by utilizing this we may be able to see improvements in generalizing unseen real life domains with regression problems. Moreover this study paves the way to measure different domain adaptation methods for problems of various fields.

References

1. https://www.weatherusa.net/wxwarnings?id=116473761
2. A review of domain adaptation
without target labels Wouter M. Kouw, Marco Loog
3. Domain Adaptation in Regression
Corinna Cortes1 and Mehryar Mohri2,1
4. Unsupervised Supervised Learning I: Estimating
Classification and Regression Errors without Labels Pinor Gomez, Guy Lebanon
5. Optimal Transport for Domain Adaptation
Nicolas Courty, Rémi Flamary, Devis Tuia, Senior Member, IEEE, Alain Rakotomamonjy, Member, IEEE

TEXAS STATE UNIVERSITY
MEMBER THE TEXAS STATE UNIVERSITY SYSTEM