

CSC5520 – Project Update 3

Jonathan A. Gibson
Andrew P. Worley
Raunak Shakya

Motivation



Platform



Methodology

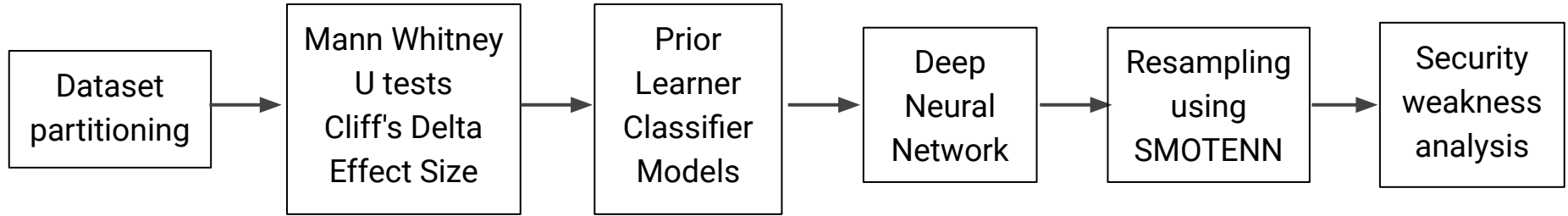
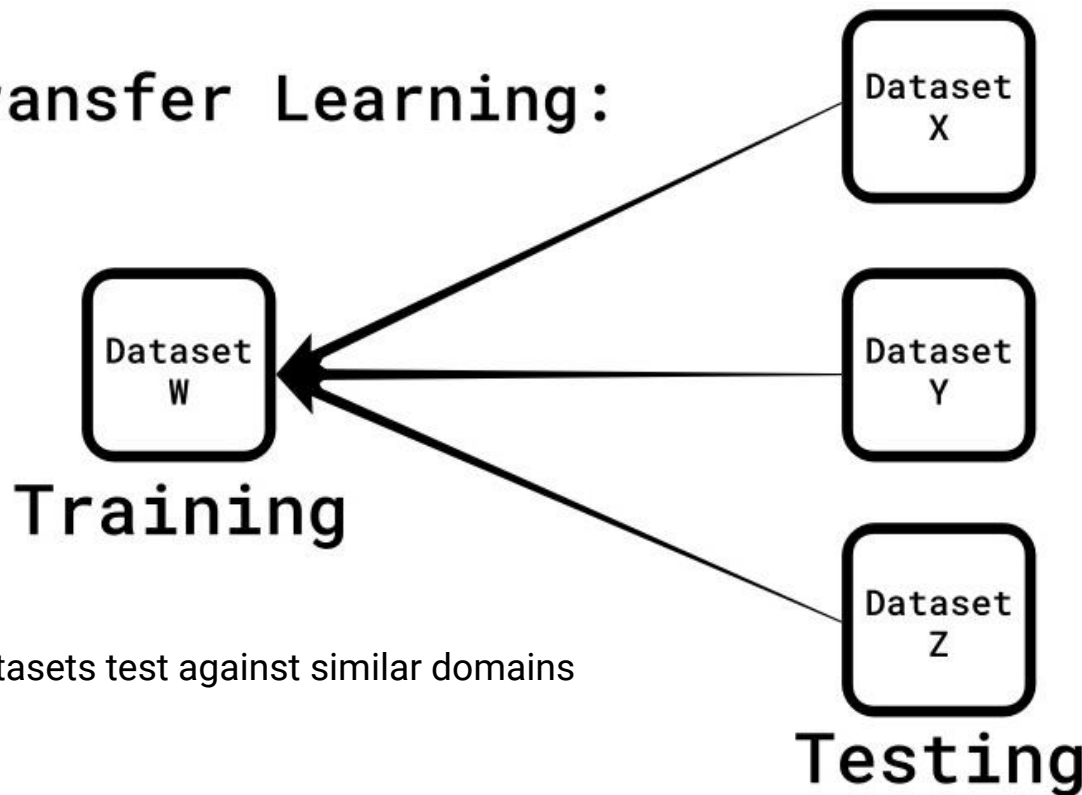


Figure 1. Different phases of Project update 3.

Methodology

Transfer Learning:



- See how well datasets test against similar domains

Methodology

- Sequential model with 3 fully connected layers using Keras
- epochs = 50 and batch size = 10

Table 1. Deep Neural Network model

Layer (type)	Output Shape	# Parameters
Input (Dense)	(None, 12)	96
Hidden (Dense)	(None, 7)	91
Output (Dense)	(None, 1)	8

Results

- Cliff's Delta ($0 \leq \partial \leq 1$)
- Addressing similarity between NEUTRAL and INSECURE feature values
- $\partial \rightarrow 0$: Features are similar
- $\partial \rightarrow 1$: Features are different

Table 2. Results from Cliff's Delta

	Cliff's Delta						
	ADD_LOC	DEL_LOC	TOT_LOC	DEV_EXP	DEV_RECENT	PRIOR_AGE	CHANGE_FILE_CNT
Chemistry	0.12	0.07	0.1	0.07	0.03	0.02	0.05
Astronomy	0.07	0.03	0.05	0.05	0.07	0.01	0.01
Biology	0.1	0.03	0.07	0.13	0.17	0.01	0.05
Others	0.17	0.12	0.13	0.13	0.01	0.11	0.07

Results

- Class imbalance is very common in real-world datasets
- Difficult to accurately classify minority class
- Resampling algorithms make the data distribution more balanced

Table 3. Results from SMOTEENN

	Original				
	Whole	Chemistry	Astronomy	Biology	Others
NEUTRAL	0.92	0.94	0.93	0.91	0.96
INSECURE	0.08	0.06	0.07	0.09	0.04
	Resampled				
		0.46	0.46	0.46	0.46
NEUTRAL		0.46	0.46	0.46	0.46
INSECURE		0.54	0.54	0.54	0.54

Table 4. Prior Learners Results for Transfer Learning Before Resampling

Results

- Very low I_F1 during transfer learning
 - Except with DTC and RFC tested with “Other” dataset
- Less consistent
- Dependent on the testing set

Using Original Datasets						
DTC	(Acc, N_F1, I_F1)		Testing			
	Training		Chemistry	Astronomy	Biology	Others
		Chemistry		(0.85, 0.92, 0.08)	(0.85, 0.92, 0.10)	(1.00, 1.00, 0.99)
		Astronomy	(0.87, 0.93, 0.07)		(0.85, 0.92, 0.10)	(1.00, 1.00, 0.99)
		Others	(0.87, 0.93, 0.07)	(0.85, 0.92, 0.08)	(0.85, 0.92, 0.10)	(1.00, 1.00, 0.99)
RFC	(Acc, N_F1, I_F1)		Testing			
	Training		Chemistry	Astronomy	Biology	Others
		Chemistry		(0.93, 0.96, 0.00)	(0.91, 0.95, 0.00)	(1.00, 1.00, 0.98)
		Astronomy	(0.93, 0.96, 0.03)		(0.91, 0.95, 0.00)	(1.00, 1.00, 0.98)
		Others	(0.93, 0.96, 0.03)	(0.93, 0.96, 0.00)	(0.91, 0.95, 0.00)	(1.00, 1.00, 0.98)
KNN	(Acc, N_F1, I_F1)		Testing			
	Training		Chemistry	Astronomy	Biology	Others
		Chemistry		(0.92, 0.96, 0.02)	(0.90, 0.95, 0.01)	(0.96, 0.98, 0.32)
		Astronomy	(0.92, 0.96, 0.06)		(0.90, 0.95, 0.01)	(0.96, 0.98, 0.32)
		Others	(0.92, 0.96, 0.06)	(0.92, 0.96, 0.02)	(0.90, 0.95, 0.01)	(0.96, 0.98, 0.32)
ANN	(Acc, N_F1, I_F1)		Testing			
	Training		Chemistry	Astronomy	Biology	Others
		Chemistry		(0.83, 0.91, 0.09)	(0.74, 0.85, 0.08)	(0.93, 0.97, 0.10)
		Astronomy	(0.76, 0.86, 0.09)		(0.74, 0.85, 0.08)	(0.93, 0.97, 0.10)
		Others	(0.76, 0.86, 0.09)	(0.83, 0.91, 0.09)	(0.74, 0.85, 0.08)	(0.93, 0.97, 0.10)
NBC	(Acc, N_F1, I_F1)		Testing			
	Training		Chemistry	Astronomy	Biology	Others
		Chemistry		(0.26, 0.36, 0.13)	(0.20, 0.24, 0.16)	(0.71, 0.83, 0.11)
		Astronomy	(0.28, 0.40, 0.10)		(0.20, 0.24, 0.16)	(0.71, 0.83, 0.11)
		Others	(0.28, 0.40, 0.10)	(0.26, 0.36, 0.13)	(0.20, 0.24, 0.16)	(0.71, 0.83, 0.11)

Results

- Results consistent across all datasets
- High accuracy but 0.00 for all other metrics
- DNN predicted the NEUTRAL class only

Table 5. Acc, Prec, Rec, F1 values with DNN Before Resampling

Using Original Datasets						
DNN	(Acc, Prec, Rec, F1)		Testing			
			Chemistry	Astronomy	Biology	Others
	Training	Chemistry		(0.93, 0.00, 0.00, 0.00)	(0.91, 0.00, 0.00, 0.00)	(0.96, 0.00, 0.00, 0.00)
		Astronomy	(0.94, 0.00, 0.00, 0.00)		(0.91, 0.00, 0.00, 0.00)	(0.96, 0.00, 0.00, 0.00)
		Biology	(0.94, 0.00, 0.00, 0.00)	(0.93, 0.00, 0.00, 0.00)		(0.96, 0.00, 0.00, 0.00)
		Others	(0.94, 0.00, 0.00, 0.00)	(0.93, 0.00, 0.00, 0.00)	(0.91, 0.00, 0.00, 0.00)	

Table 6. Prior Learners Results for Transfer Learning After Resampling

Results

- Drastic improvements for minority class, but majority class suffered
- More consistent
- Less dependent on testing dataset

Using Resampled Datasets						
DTC	(Acc, N_F1, I_F1)		Testing			
	Training		Chemistry	Astronomy	Biology	Others
		Chemistry		(0.64, 0.65, 0.62)	(0.60, 0.63, 0.56)	(1.00, 1.00, 1.00)
		Astronomy	(0.61, 0.64, 0.56)		(0.60, 0.63, 0.56)	(1.00, 1.00, 1.00)
		Biolog	(0.61, 0.64, 0.56)	(0.64, 0.65, 0.62)		(1.00, 1.00, 1.00)
		Others	(0.61, 0.64, 0.56)	(0.64, 0.65, 0.62)	(0.60, 0.63, 0.56)	
RFC	(Acc, N_F1, I_F1)		Testing			
	Training		Chemistry	Astronomy	Biology	Others
		Chemistry		(0.62, 0.66, 0.57)	(0.59, 0.66, 0.48)	(1.00, 1.00, 1.00)
		Astronomy	(0.61, 0.67, 0.54)		(0.59, 0.66, 0.48)	(1.00, 1.00, 1.00)
		Biolog	(0.61, 0.67, 0.54)	(0.62, 0.66, 0.57)		(1.00, 1.00, 1.00)
		Others	(0.61, 0.67, 0.54)	(0.62, 0.66, 0.57)	(0.59, 0.66, 0.48)	
KNN	(Acc, N_F1, I_F1)		Testing			
	Training		Chemistry	Astronomy	Biology	Others
		Chemistry		(0.50, 0.54, 0.46)	(0.47, 0.57, 0.33)	(0.99, 0.99, 0.99)
		Astronomy	(0.51, 0.58, 0.40)		(0.47, 0.57, 0.33)	(0.99, 0.99, 0.99)
		Biolog	(0.51, 0.58, 0.40)	(0.50, 0.54, 0.46)		(0.99, 0.99, 0.99)
		Others	(0.51, 0.58, 0.40)	(0.50, 0.54, 0.46)	(0.47, 0.57, 0.33)	
ANN	(Acc, N_F1, I_F1)		Testing			
	Training		Chemistry	Astronomy	Biology	Others
		Chemistry		(0.56, 0.53, 0.58)	(0.41, 0.45, 0.36)	(0.72, 0.63, 0.78)
		Astronomy	(0.53, 0.43, 0.60)		(0.41, 0.45, 0.36)	(0.72, 0.63, 0.78)
		Biolog	(0.53, 0.43, 0.60)	(0.56, 0.53, 0.58)		(0.72, 0.63, 0.78)
		Others	(0.53, 0.43, 0.60)	(0.56, 0.53, 0.58)	(0.41, 0.45, 0.36)	
NBC	(Acc, N_F1, I_F1)		Testing			
	Training		Chemistry	Astronomy	Biology	Others
		Chemistry		(0.54, 0.02, 0.70)	(0.54, 0.03, 0.70)	(0.55, 0.07, 0.70)
		Astronomy	(0.53, 0.07, 0.69)		(0.54, 0.03, 0.70)	(0.55, 0.07, 0.70)
		Biolog	(0.53, 0.07, 0.69)	(0.54, 0.02, 0.70)		(0.55, 0.07, 0.70)
		Others	(0.53, 0.07, 0.69)	(0.54, 0.02, 0.70)	(0.54, 0.03, 0.70)	

Results

- Results consistent across all datasets
- Much better DNN results after resampling

Table 7. Acc, Prec, Rec, F1 Values with DNN After Resampling

Using Resampled Datasets						
DNN	(Acc, Prec, Rec, F1)		Testing			
			Chemistry	Astronomy	Biology	Others
	Training	Chemistry		(0.55, 1.00, 0.55, 0.70)	(0.54, 1.00, 0.54, 0.70)	(0.54, 1.00, 0.54, 0.70)
		Astronomy	(0.54, 1.00, 0.54, 0.70)		(0.54, 1.00, 0.54, 0.70)	(0.54, 1.00, 0.54, 0.70)
		Biology	(0.54, 1.00, 0.54, 0.70)	(0.54, 1.00, 0.54, 0.70)		(0.54, 1.00, 0.54, 0.70)
		Others	(0.54, 1.00, 0.54, 0.70)	(0.54, 1.00, 0.54, 0.70)	(0.54, 1.00, 0.54, 0.70)	

Potential Changes/Optimizations

- We used mostly default parameters for the API calls
- We could optimize SMOTE parameters for better resampling
- Calculate loss based on the recall metric to maximize minority classification
 - High accuracy even when predictions were only for NEUTRAL class
- We can experiment with different DNN shapes/layers
 - Optimize DNN parameters
- We could compare features against each other
 - May allow us to omit features that are too similar

Conclusion

- DTC and RFC did really well before and after resampling the datasets
- Transfer learning tells us if a classification model is good at generalizing information between similar domains
- Resampling is a good way to produce better results for the minority class at the expense of the majority class results
- Prior learners is better before resampling
- DNN is better after resampling
- No security weaknesses reported with Bandit for any project update

Thank you!