

Vorhersage von Demenz anhand von Sprachmustern unter Verwendung von LLMs

Erste Ergebnisse

Lukas Probst | 20. August 2023

Inhaltsverzeichnis

1. Vorgehensweise

- Wegweisendes Paper
- Aufbau
- Trainieren von Modellen

2. Ergebnisse

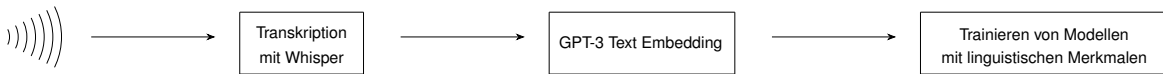
- Vergleich mit den Ergebnissen von Agbavor und Liang 2022 der Drexel Universität

3. Ausblick

- Mögliche Erweiterungen

Wegweisendes Paper von Agbavor und Liang 2022

- „Predicting dementia from spontaneous speech using large language models“ (Drexel Universität) liefert beeindruckende Ergebnisse für die Vorhersage von Demenz mit GPT-3 Text Embeddings sowie akustischen Merkmalen der Tonaufnahmen.
 - Text Embeddings können zuverlässig verwendet werden, um Personen mit Demenz von gesunden Kontrollpersonen zu unterscheiden und den kognitiven Testwert der Testperson (MMSE Score) abzuleiten, beides ausschließlich auf der Grundlage von Sprachdaten.
 - Die Ergebnisse zeigen, dass die Verwendung von GPT-3 Text Embeddings ein praktikabler Ansatz für die Bewertung von Demenz ist und das Potenzial hat, die Frühdiagnose von Demenz zu verbessern.



10-Fold Cross Validation

```
_n_splits = 10  
# Split the dataset into k equal partitions  
cv = KFold(n_splits=_n_splits, random_state=42, shuffle=True)
```

Hyperparameteroptimierung mit GridSearchCV: Parameter Grids

```
svc_param_grid = {
    'C': [0.1, 1, 10, 100],
    'gamma': [1, 0.1, 0.01, 0.001],
    'kernel': ['rbf', 'poly', 'sigmoid']
}
lr_param_grid = [
    {'penalty': ['l1', 'l2'],
     'C': np.logspace(-4, 4, 20),
     'solver': ['liblinear'],
     'max_iter': [100, 200, 500, 1000]},
    {'penalty': ['l2'],
     'C': np.logspace(-4, 4, 20),
     'solver': ['lbfgs'],
     'max_iter': [200, 500, 1000]},
]
rf_param_grid = {
    'n_estimators': [25, 50, 100, 150],
    'max_features': ['sqrt', 'log2', None],
    'max_depth': [3, 6, 9],
    'max_leaf_nodes': [3, 6, 9],
}
```

Hyperparameteroptimierung mit GridSearchCV: Ermitteln der besten Parameter

```
for model, name in zip(models, names):  
    # Tune hyperparameters with GridSearchCV  
    best_params = None  
    if name == 'SVC':  
        grid_search = GridSearchCV(estimator=model, param_grid=svc_param_grid, cv=cv, n_jobs=-1, error_score=0.0)  
        grid_search.fit(X, y)  
        best_params = grid_search.best_params_  
    elif name == 'LR':  
        grid_search = GridSearchCV(estimator=model, param_grid=lr_param_grid, cv=cv, n_jobs=-1, error_score=0.0)  
        grid_search.fit(X, y)  
        best_params = grid_search.best_params_  
    elif name == 'RF':  
        grid_search = GridSearchCV(estimator=model, param_grid=rf_param_grid, cv=cv, n_jobs=-1, error_score=0.0)  
        grid_search.fit(X, y)  
        best_params = grid_search.best_params_  
        model.set_params(**best_params)  
  
    # Perform cross-validation with custom scoring metrics and best params  
    results = cross_validation(name, model, X, y, cv)
```

Auswertung der trainierten Modelle

	Set	Model	Accuracy	Precision	Recall	F1
0	Train	SVC	0.801(0.133)	0.806 (0.131)	0.801 (0.133)	0.793 (0.135)
1	Train	LR	0.802(0.144)	0.809 (0.145)	0.802 (0.144)	0.797 (0.146)
2	Train	RF	0.731(0.134)	0.737 (0.139)	0.731 (0.134)	0.722 (0.134)
3	Test	SVC	0.801	0.806	0.801	0.793
4	Test	LR	0.802	0.809	0.802	0.797
5	Test	RF	0.731	0.737	0.731	0.722
6	Test	Dummy	0.5384615384615384			

Vergleich mit den Ergebnissen der Drexel Universität (Agbavor und Liang 2022)

	Embeddings	Model	Accuracy	Precision	Recall	F1
10-fold CV	Ada	SVC	0.788 (0.075)	0.798 (0.109)	0.819 (0.098)	0.799 (0.066)
		LR	0.796 (0.107)	0.798 (0.126)	0.835 (0.129)	0.808 (0.100)
		RF	0.734 (0.090)	0.738 (0.109)	0.763 (0.149)	0.743 (0.103)
	Babbage	SVC	0.802 (0.054)	0.823 (0.092)	0.804 (0.103)	0.806 (0.053)
		LR	0.809 (0.112)	0.843 (0.148)	0.811 (0.091)	0.818 (0.091)
		RF	0.760 (0.052)	0.780 (0.102)	0.781 (0.110)	0.770 (0.047)
Test Set	Ada	SVC	0.788	0.708	0.971	0.819
		LR	0.718	0.653	0.914	0.762
		RF	0.732	0.690	0.829	0.753
	Babbage	SVC	0.803	0.723	0.971	0.829
		LR	0.718	0.647	0.943	0.767
		RF	0.761	0.714	0.857	0.779

<https://doi.org/10.1371/journal.pdig.0000168.t002>

Wert konnte verbessert werden
 Wert ist etwas schlechter
 Nicht relevant

Vergleich mit den Ergebnissen der Drexel Universität (Agbavor und Liang 2022)

⇒ Vergleichbare Werte, möglicherweise sogar etwas besser nach Optimierung.

Vergleich mit Vorsicht zu genießen, denn...

- Es wurde eine andere Bibliothek für die Transkription verwendet und bisher nur das base-Modell davon verwendet.
 - Ergebnisse könnten durch größeres Whisper Modell verbessert werden.
- Es könnten andere Parameter für die Modelle verwendet worden sein, die die Ergebnisse beeinflussen.

Mögliche Erweiterungen

- Modelle mit akustischen Merkmalen (OpenSMILE) trainieren (WIP)
 - "With additional acoustic features, we observe only marginal improvement in the classification performance on the 10-fold CV. There is no clear difference in predicting the test set in terms of accuracy and F1 score when the acoustic features are combined with GPT-3 based text embeddings, but we instead observe higher precision at the expense of lower recall." – Agbavor und Liang 2022.
- Kombination aus Text Embedding & akustischen Merkmalen.
- Fine-Tuning eines GPT-3 Text Embedding Modells
 - "[...] the fine-tuned Babbage model underperforms the GPT-3 based text embeddings, a result in line with the recent findings that GPT-3 embedding model is even competitive with fine-tuned models." – Agbavor und Liang 2022.
- MMSE Prediction Task der ADReSSo-Challenge
 - Ausprägungsgrad der Demenz-Erkrankung ermitteln.