

Explainable model for time-series via ChatGPT

Xue Zhechang

CHAIR FOR PERVASIVE COMPUTING SYSTEMS, INSTITUTE OF TELEMATICS, DEPARTMENT OF COMPUTER SCIENCE



Outlines

- Summary of previous works
- Improvement
- Baseline
- New method
- Future works

Summary of previous works

Temporal Data Meets LLM

- Feature-based GPT classification

- GPT model: GPT-API(gpt-3.5-turbo)

- 5 classes, 15 samples, 10 features

- Highest accuracy: Gradient Boosting Tree: 54%

	Decision Tree	Gradient Boosting	Random Forest**
Average	51.9%	54%	40.2%
Highest	90% (5)	100% (1)	60%(2)
Lowest	20% (2)	25% (1)	20%(1)

Disadvantages of previous study

- Low accuracy
- 5 classes -> less accuracy
- Lack baseline
- Not few-shot

Improvement

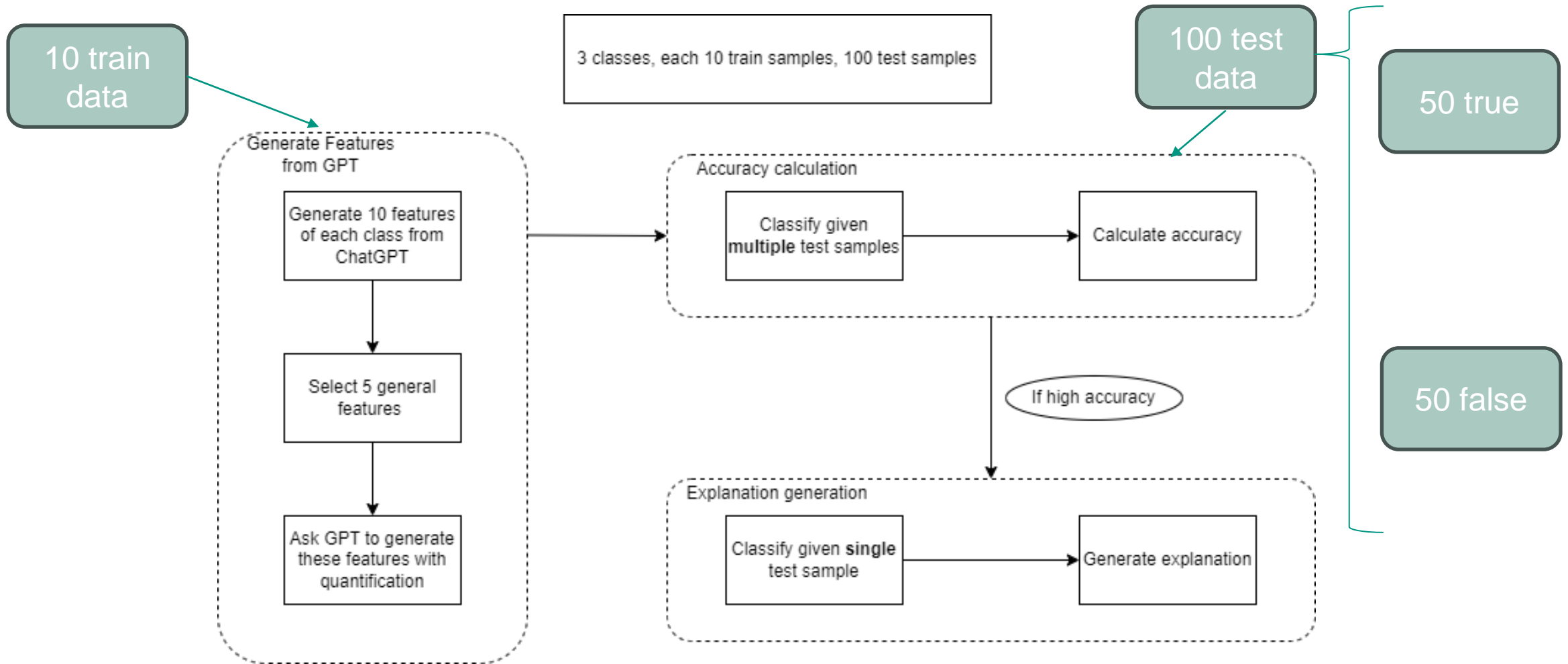
- Set baseline
- Decline class: Now 3 classes, let them do 2-class classification with each other
- A new method: GPT-generated Feature-based classification

Baseline

3 classes, 2 each other, train: 10, test: 40

- Feature-based GPT classification: 59.43%
- DummyClassifier from sklearn: 50%
- XGBoost*: 64.17%
- RandomForest: 74.17%
- DecisionTree: 63.33%
- Average: 62.22%

New method



Quantification

Problem

- GPT has bad quantification ability
- Features given by GPT are hard to redefine.

■ Solution:

- Standardize data (0-10)
 - Quantify by human
 - Rename features
- 'nA' means 0 or 1.
'nB' means 9 or 10.
'Situation 1' : When there's a 'nA' in the array, there should be a 'nB' after 'nA'.
'Situation 2' : When there's a 'nB' in the array, there should be a 'nA' after 'nB'.

- Prompt: Find situation 1 & 2 step by step

Prompt for classification

- Use **sharp drop** and **sharp increase** (more features decrease accuracy)
- Class2(shake hand): Have both sharp drop and sharp increase
- Class3(move to left): Have only sharp increase
- Class4(move to left): Have only sharp drop

Result

57.17%

- Class2(shake hand): True: 60%; False: 44%, average: 52%
- Class3(move to left): True: 84%; False: 40%: average: 62%
- Class4(move to right): True: 60%; False: 55%: average: 57.5 %

Future works

- Use ChatGPT or GPT-4 instead of GPT-3.5-turbo (86.67% for class2 true)
- Use 0-9 standardize instead
- Try to classify class 2 & 3, class 2 & 4, class 3 & 4