Appendix

Prompt for Data Preparation

The prompt for data preparation can be organized as follows:

Data Preparation Prompt

Research Objective: Analyzing causal relationships between text-derived features and stock price time series changes.
Key Requirements:

1. Identify elements from text data that form causal relationships with time series changes.
2. Output results in JSON format with fields: `time`, `original_text`, `features` (including subfields like `score`, `causal_factor`, `causal_impact`, etc.).
3. Include:
   - `causal_factor`: Elements influencing time series changes.
   - `affected_by_time_series`: Elements affected by time series changes.
   - Direction of impacts (`positive`/`negative`).
4. Do not include fictional dates; retain valid JSON structure.

JSON Format Example:

```
[
  {
    "time": "2018-01-02",
    "orginal_text": "",
    "features": {
      "keywords":
       ["CFI index",
        "strong start",
        "gap up"],
      "sentiment_score": 0.8,
      "policy_related": "false",
      "investment_strategy": "false",
      "sector_focus": ["finance"],
      "causal_factor": "true",
      "causal_impact": "positive",
      "affected_by_time_series": "
          true",
      "ts_effect_direction": "
          positive"
    }
  }
]
```

The concrete input content can compose the role of "System" and "User".

System: I'm researching stock price movements using multimodal data. Please analyze...Output the results in JSON format including...In addition to text content extraction, please list...

User: ["Forget Your iPhone X, Ignore The Samsung And Pixel 2 Choices, This Is Your Smartphone Of 2017.", "Apple's Executive Cash Bonus Plan.", "2 Warren Buffett Stocks to Consider Buying Now.", "Bitcoin or Stocks? Here's the One to Buy in 2018."]

Evaluation Metrics

The evaluation metric of this work is ACC and MCC, they can be calculated as follows:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

$$\mathrm{MCC} = \frac{\mathrm{TP} \cdot \mathrm{TN} - \mathrm{FP} \cdot \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}}$$

Information Related to Reproducing

The source code can be found in the zipfiles.