THE WEB CONFERENCE ACM

Short Paper Track

Automatic Construction of Expiration Time Expression Dataset from Retweets

Hirotaka Nagashima, Keishi Tajima
Kyoto University

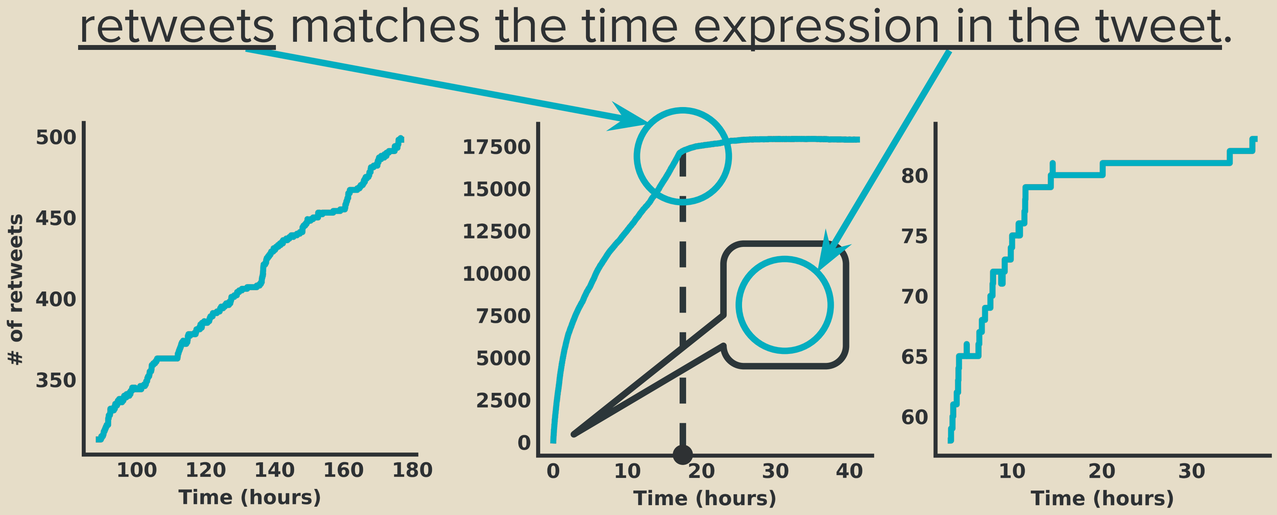
1. Introduction

- **Problem**:
  Distinguishing time expressions that specify expiration times.
  "The deadline is 23:00." ✔
  "Campaign began at 23:00." ✖
- **Our Contribution**:
  Developed a method for automatically generating a labeled dataset of expiration time expressions.
- **Expected Applications**:
  - Detection of approaching deadlines.
  - Filtering of obsolete information.

2. Methods

I. **Tracking**: Monitoring tweets with time expressions.

II. **Detection of Expiration Time Specifications**:
Determine if the inflection point of the number of retweets matches the time expression in the tweet.



Without expiration time | With expiration time | No clear inflection point

III. **Dataset Construction**: Label the matching tweets as positive and others as negative.

3. Experiment

- **Dataset**:
  Collected 25,046 tweets with time expressions, labeled them both manually and automatically.
- **Classifier Training and Evaluation**:
  - Trained classifiers using automated labels.
  - Compared performance with classifiers trained on manually assigned labels.

Achieved 70.7% accuracy.

Achieved 75.9% of Manual's improvement over Random.

Model	Precision	Recall	F1-score	Accuracy	MCC
Random	0.255	0.508	0.339	0.499	-0.001
Our Method	0.823	0.455	0.580	0.707	0.520
Manual Labeling	0.750	0.629	0.676	0.773	0.576
Gain Ratio	1.15	-0.438	0.715	0.759	0.903

4. Conclusions

- **Effectiveness**:
  - Demonstrated reliable dataset creation using proposed method.
  - Reduced manual labeling effort significantly.
- **Future Work**:
  - Refine accuracy.
  - Expand the dataset for broader applications.

PDF