We compare Jagger (proposal) with Need for speed. Because the amount of text has been increasing, whereas NLP models become inefficient (larger and slower)

Models focus on accuracy and become slower
• The efficient neural methods are only relatively efficient and much slower than classical methods

Classical methods are still used to process SNS posts in sociolinguistics and marketing

Need for speed-intensive approach to NLP

Jagger (proposal): A remarkably simple yet accurate pattern-based method for morphological analysis

• Let’s make the fastest method more accurate (instead of making the most accurate method slightly more efficient) for Japanese morphological analysis (word segmentation, POS tagging, and lemmatization; latter two as tagging)

Bypass expensive argmax operations via patterns
• Assume morphological analysis as single multiclass classification (where to segment and what to tag)
• Segment and tag words by greedily applying patterns inspired by longest-matching for word segmentation

Extract patterns as learning-based methods do
• Design a pattern template from feature templates of learning-based methods [Kudo+ 2004, Neubig+ 2011] as: posterior contexts + a previous tag
• Use the training data and a dictionary to extract patterns by frequency (offline argmax)

Issue: How to obtain reliable patterns?

Evaluation: Jagger can process 1,000,000 sents/s with accuracy comparable to learning-based baselines

We compare Jagger (proposal) with the learning-based baselines using two dictionaries on two common datasets
• Baselines: search-based method [Kudo+ 2004] (MeCab 0.996, Vibrato 0.5.0), classification-based method [Neubig+ 2011] (Vaporetto 0.6.2)
• Dictionaries: mecab-juman-5.1 (475,716 words) and mecab-juman-7.0 (702,358 words; augmented from Wikipedia words)

Jagger is 7-16x faster than baselines with 1/2 to 1/20 as much memory, while achieving comparable accuracy

Message to researchers
Because the accuracies are becoming saturated on NLP benchmark datasets with a larger foundation model, researchers may want to set diverse goals based on underrepresented metrics besides accuracy (e.g., efficiency)