

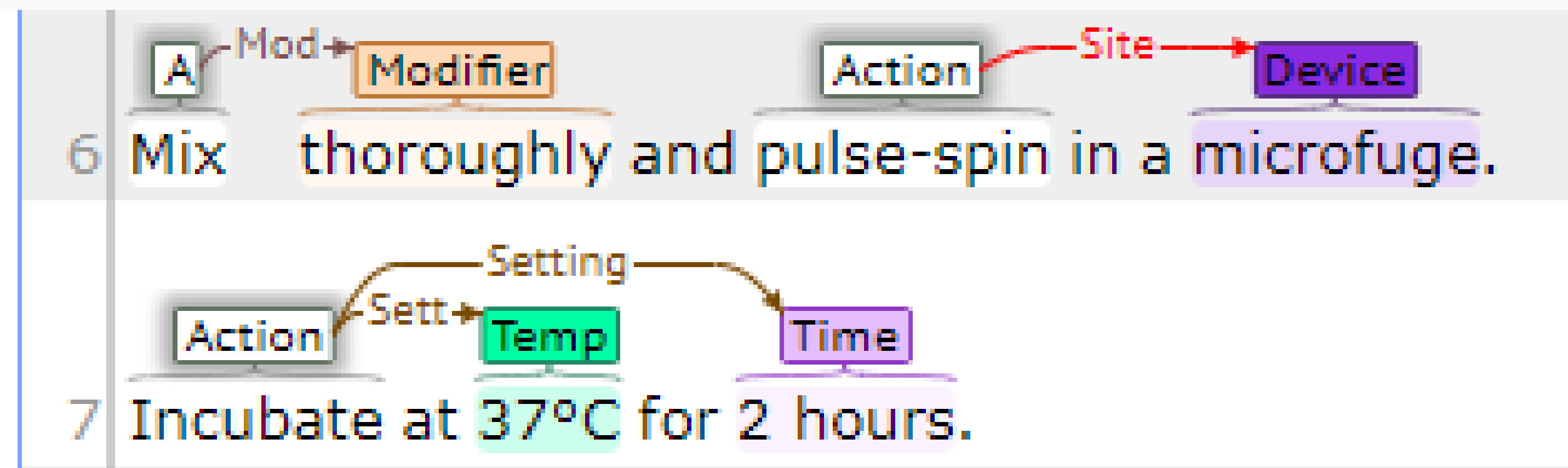
KAUSHIKACHARYA AT WNUT 2020 SHARED TASK-1: CRF BASED NER FOR WET LAB PROTOCOLS

Kaushik Acharya
Philips India Ltd.

The paper describes how a Linear-chain Conditional Random Field classifier model detects named entities in wet lab protocols. Wet lab protocols are sequence of instructions for conducting biology and chemistry experiments.

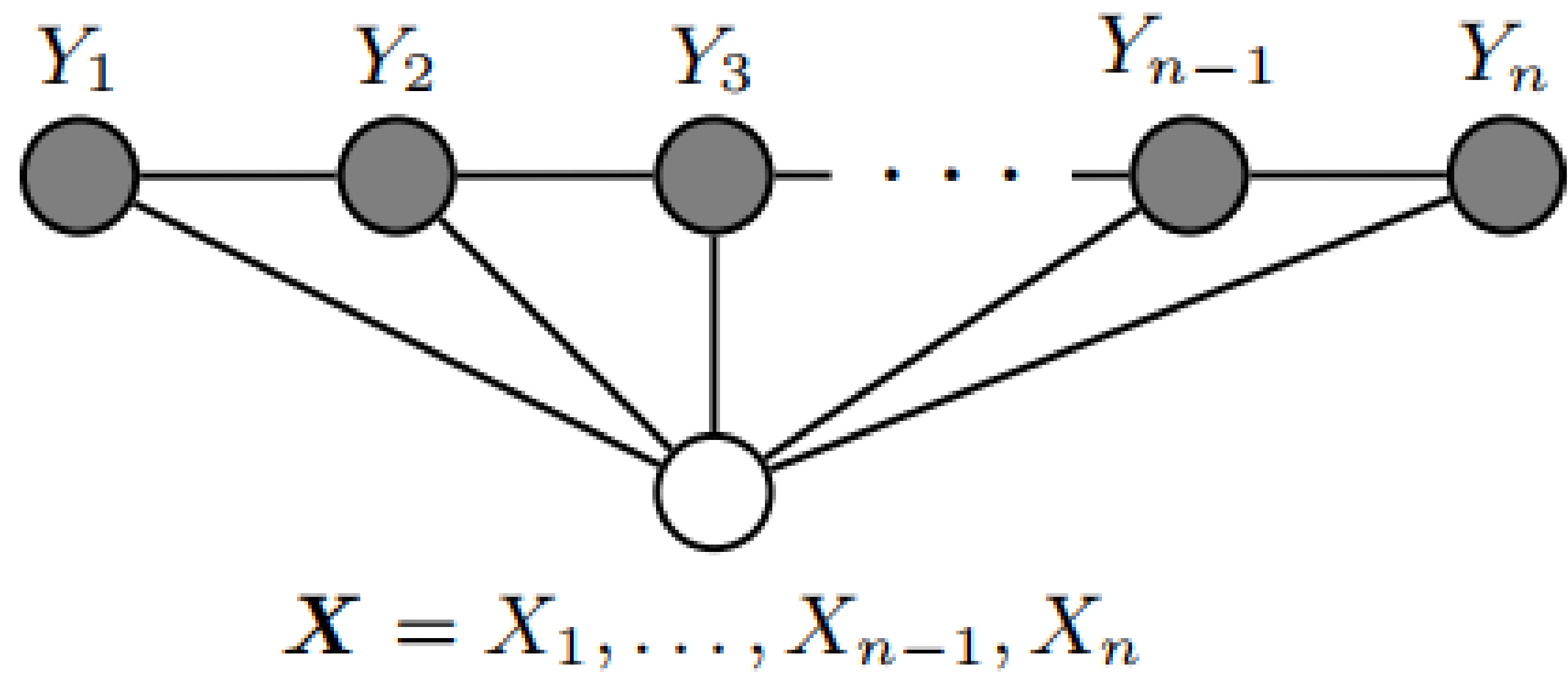
Problem Statement

Identify span of text corresponding to one of 18 fine-grained entities.



Example sentences from protocol with BRAT annotated entities.

Approach: Linear-chain Conditional Random Field



X: Random variable representing observation sequences.

Y: Random variable over corresponding label sequences.

Source: Conditional Random Fields: An Introduction by Hanna Wallach (2004)

Fine-grained entities with top frequent examples

Entity	Examples
ACTION	add, incubate, mix
Amount	50 l, equal volume
Concentration	1x
Device	filter, vacuum, microfuge
Generic-Measure	30-kd, several times, 100v
Location	tube, plate, well
Measure-Type	volume, concentration
Mention	It, them, this
Method	up and down, extraction
Modifier	each, gently, at least
Numerical	one, 3, several, several times
pH	ph 8.0, ph8.0
Reagent	cells, supernatant
Seal	Lid, cap, aluminum foil
Size	0.02 m, 12 x 75 mm
Speed	14,000xg, 10,000 rpm
Temperature	room temperature, overnight
Time	5 minutes

Features

- Lexical features
- Parts of speech (POS) features
- Dependency parse features

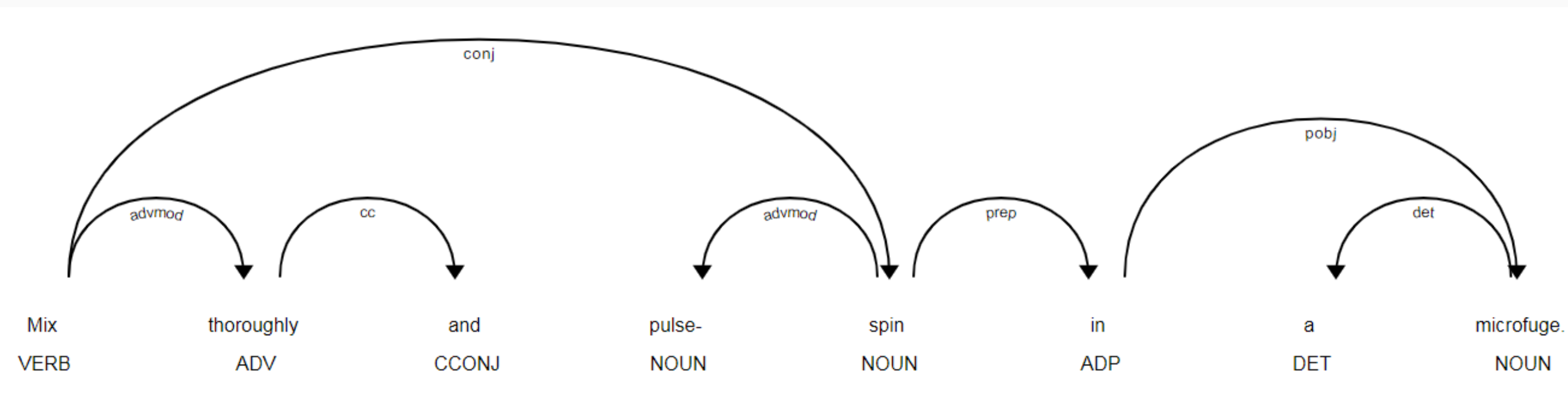


Figure shows POS and dependency parse tree.

Frequent confuser classes

Truth	Confusers
Generic-Measure	Concentration, Numerical
Method	Action, Reagent
Modifier	Reagent, Location, Action
Numerical	Amount, Generic-Measure
Size	Concentration, Location

Entity level metrics on test set

Average	P	R	F1
Micro	0.782	0.766	0.774
Macro	0.777	0.766	0.771

P: Precision, R:Recall, F1:F1 score

Conclusion

Error analysis shows need for better feature engineering. Future plan includes extraction of global structured information features from dependency tree.