COMP4650 / COMP6490
Document Analysis 2018

Introduction

Gabriela F

Australian
National
University

Sequence labeling II

Weakness of Markov approaches in that it limits the context from whic tracted

Anything outside the context windo pact on the decision being made...

* Acknowledgement: Some of the content originates from the Stanford NLP course at Coursera.org

Sequence labeling II

CRFs are indeed basically the sequential version of logistic regression

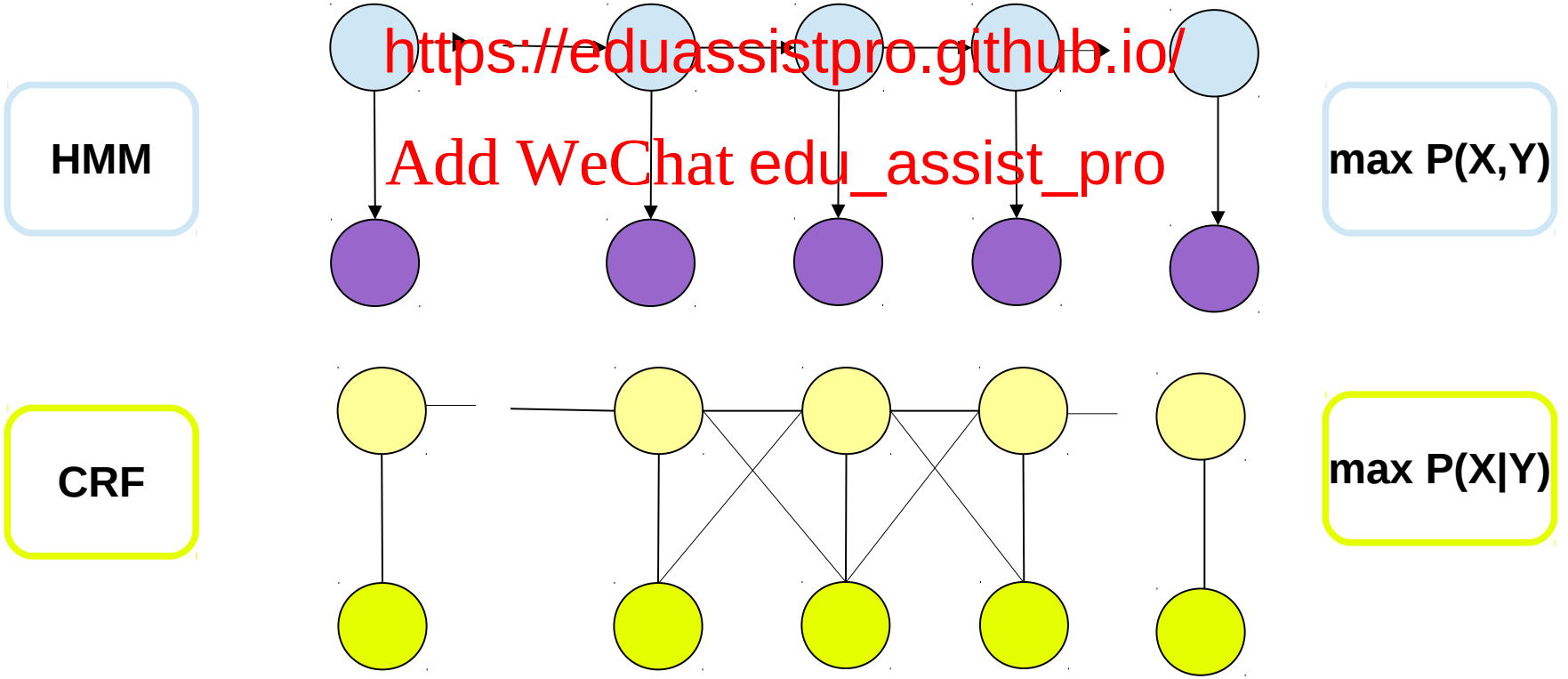
… whereas logistic regression is a l del for classification, **CRFs are a log-linea sequential labels**.

Sequence labeling II

CRFs can define a much larger set of features

HMMs are necessarily local in nature because they're constrained to binary transition and emission feature functions

which force each word to depend only on the current label and each label to depend only on the previous label



Sequence labeling II

How to label a sentence using CRF?

The naive way is to calculate $p(labelsequence)$ for every possible labeling l, and then choose the label that maximizes this probability.

However, this is intractable...

A better way is to realize that (linear-chain) CRFs satisfy an optimal substructure property that allows us to use a dynamic programming algorithm to find the optimal label, e.g., the Viterbi algorithm for HMMs.

Information Extraction

BIO enconding

(B) beginning

(I) inside

(O) other

2n + 1 tags, where *n* is the number of entity types

BIO enconding

Without the B tag IO tagging is unable to distinguish between two entities of the same type that are right next to each other.

Since this situation do ly there is at least some punctuation or other d

+ IO tagging may be sufficient

+ advantage of using only n + 1 tags

Word-by-word feature encoding

Named Entity Recognition as sequence labeling

The features available to the classifier during training and classification are those in the boxed area

Neuronal algorithm for NER

- Use a CRF layer on top of the bi-LSTM
- Use Viterbi for decoding for selecting the most likely tag sequence

Information Extraction

- Extracting time and dates
  - Question answering
  - Calendar assistance
  - Personal as
  - …

Needs normalization!

So we can reason about them...

Extracting time and dates

- Absolute → map to calendar dates
- Relative → r time throught some other r
  - A week from last Tuesday.
- Duration → spans of time with differente granularities
  - seconds, minutes, days, weeks, centuries, etc

Extracting time and dates

Extracting time and dates

Lexical triggers

A fare increase initiated <TIMEX3>last w by UALCorp's United Airlines was matched by competitors over <TIMEX3>the weekend</TIMEX3>, marking the second successful fare increase in<TIMEX3>two weeks</TIMEX3>.

(Pustejovskyet al. 2005, Ferro et al. 2005)

Extracting time and dates

Sequence labeling with BIO encoding

Extracting time and dates

Normalization: *VALUE attribute*

from the ISO 8601 standard for encoding temporal values

(ISO8601, 2004)

Events Extraction

- Identify mentions of events in texts

  events can be assigned to point (or interval) in time

- sequence labeling

- BIO encoding

- usually applied s ning methods

Feature	Explanation
Character affixes	Character-level prefixes and suffixes of target word
Nominalization suffix	Character level suffixes for nominalizations (e.g., *-tion*)
Part of speech	Part of speech of the target word
Light verb	Binary feature indicating that the target is governed by a light verb
Subject syntactic category	Syntactic category of the subject of the sentence
Morphological stem	Stemmed version of the target word
Verb root	Root form of the verb basis for a nominalization
WordNet hypernyms	Hypernym set for the target

Events Extraction

Events + temporal expressions → Temporal ordering of the events

- Timeline

Classify events according to temporal relations

- Similar to Relation ns between entities, relations are betwe
- Finite set of temporal relations (Allen, 1984)

Useful for Q&A and summarization

TimeBank corpus

Allen

relations

between

temporal

events

Benchmarking

How do you know your method is working?

How good it is in respect to other methods?

What is a baseline?

- Information that is used as a starting point by which to comp ion
- Benchmark
- Something you want to beat

How a baseline looks like?

- Random assignment
- Mayoririty class voting
- Simple heuristics
- Simple Machine Learning techniques
- Simple feature sets
- The system/method you want to beat!

Summary

- Named entities: `who and who's class (type)`
- Relation extraction: `who is doing what`
- Temporal ex `facilitate reasoning`
- Events: `facts`