A Multi-Grained Self-Interpretable Symbolic-Neural Model For Single/Multi-Labeled Text Classification

Xiang Hu, Xinyu Kong, Kewei Tu

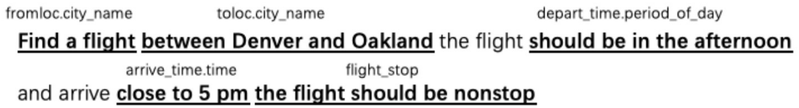
00 / Motivation & What we achieved

- Human rationales

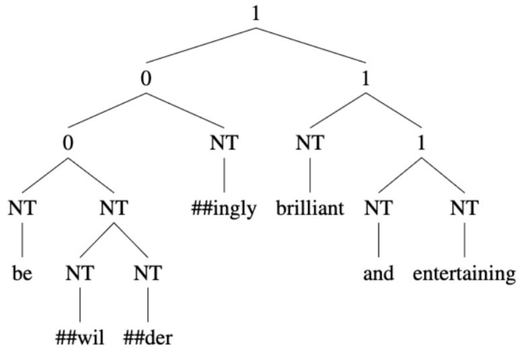
  Hierarchical, Multi-grained for language understanding
  e.g. For multi-intent understanding, we are able to recognize the relationships between intents and text spans.
  For semantic analysis, we are able to identify the sentiment and polarity reversal at the levels of words, phrases and clauses.

- Model rationales of ours

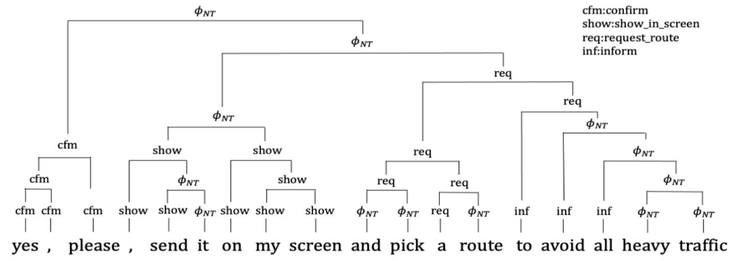
NER given labels only



Classification tasks with sentence labels only

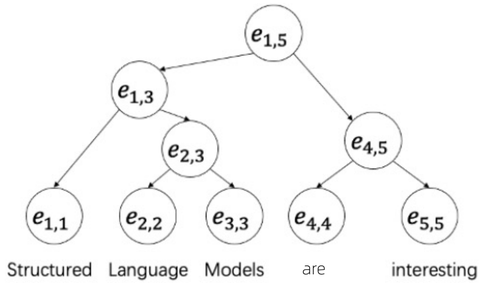


Bewilderingly brilliant and entertaining

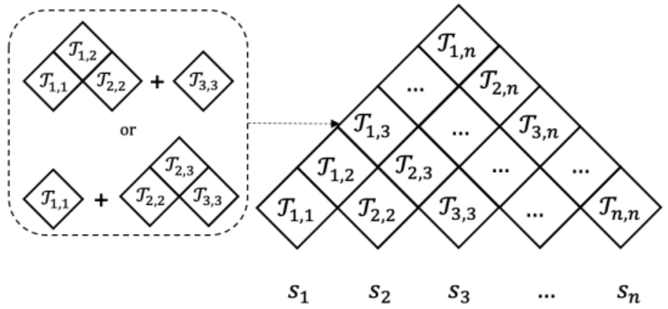


01 | Preliminary & Assumption

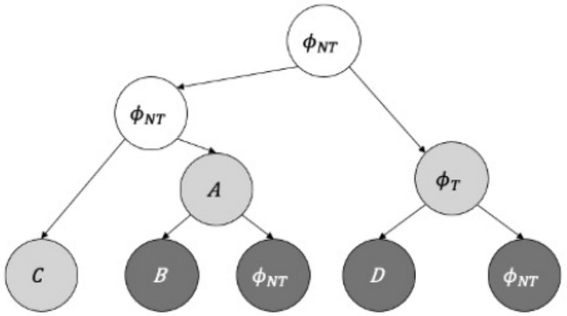
- Structured Language models



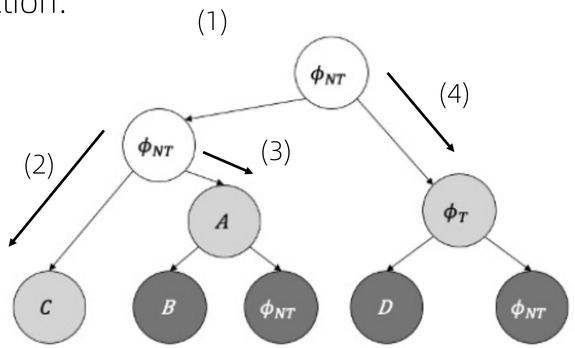
Fast-R2D2



- Our Assumption



Yield function:

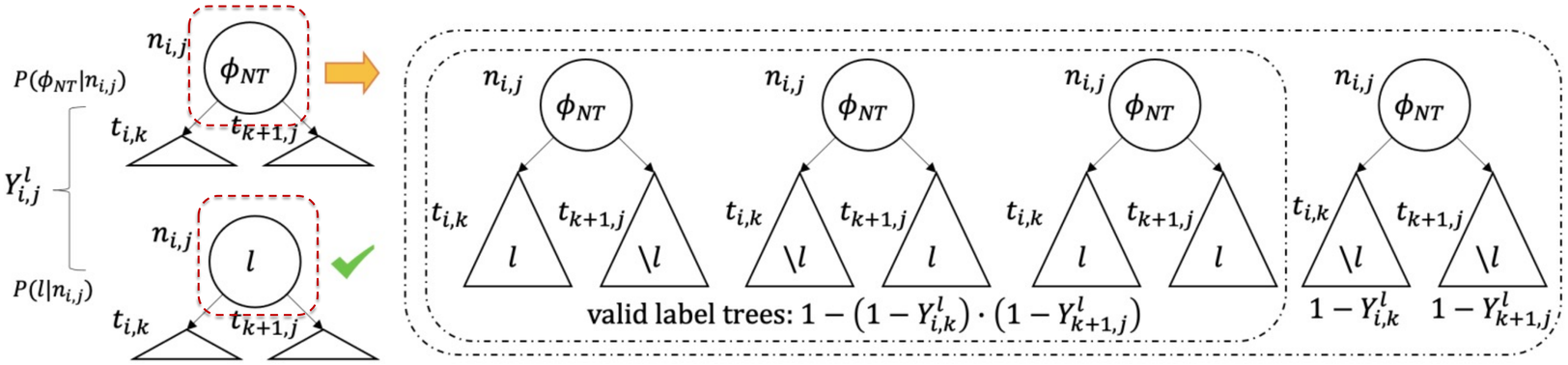


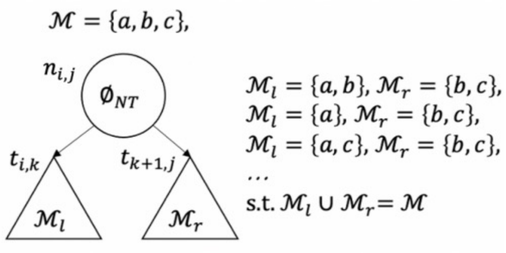
ϕ_{NT} continue traversing (NT stands for non-terminal)
ϕ_T stop traversing (T stands for terminal)

Final labels: {A, C}

02 | Approaches

What's the probability that yield result of a given tree contains a given label l ?





$$\mathcal{L}^t_{cls}(\Psi) = -\log P_\Psi(\hat{t}^{[\mathcal{Y}(\hat{t})=\mathcal{T}]}|t) = -\sum_{l\in\mathcal{T}} \log Y^l_{1,|\mathbf{S}|} - \log(1-Y^{\mathcal{O}}_{1,|\mathbf{S}|})$$

$\mathcal{T}$:target labels, $\mathcal{O}$: other labels

Thanks