

Dataset Mention Extraction and Classification

Animesh Prasad, Chenglei Si and Min-Yen Kan



Task (CI Richtext)

Publication:

....Source: **Monitoring the Future: National Survey on Drug Use, 1975-2009** Section 2 provides a brief summary of trends in adolescent drinking and smoking, using data for the US from the annual **Monitoring the Future survey** Trends in Adolescent Drinking and Smoking: **Monitoring the Future** Systematic annual data on the prevalence of underage drinking and smoking in the US are collected and tracked by several organizations. This section relies on data from the **Monitoring the Future (MTF)**

Datasets (Present): [...

56: Monitoring the Future: A Continuing Study of the Lifestyles and Values of Youth, 1984;

101: Monitoring the Future: A Continuing Study of the Lifestyles and Values of Youth, 1989; ...]

Datasets (Not Present): [...

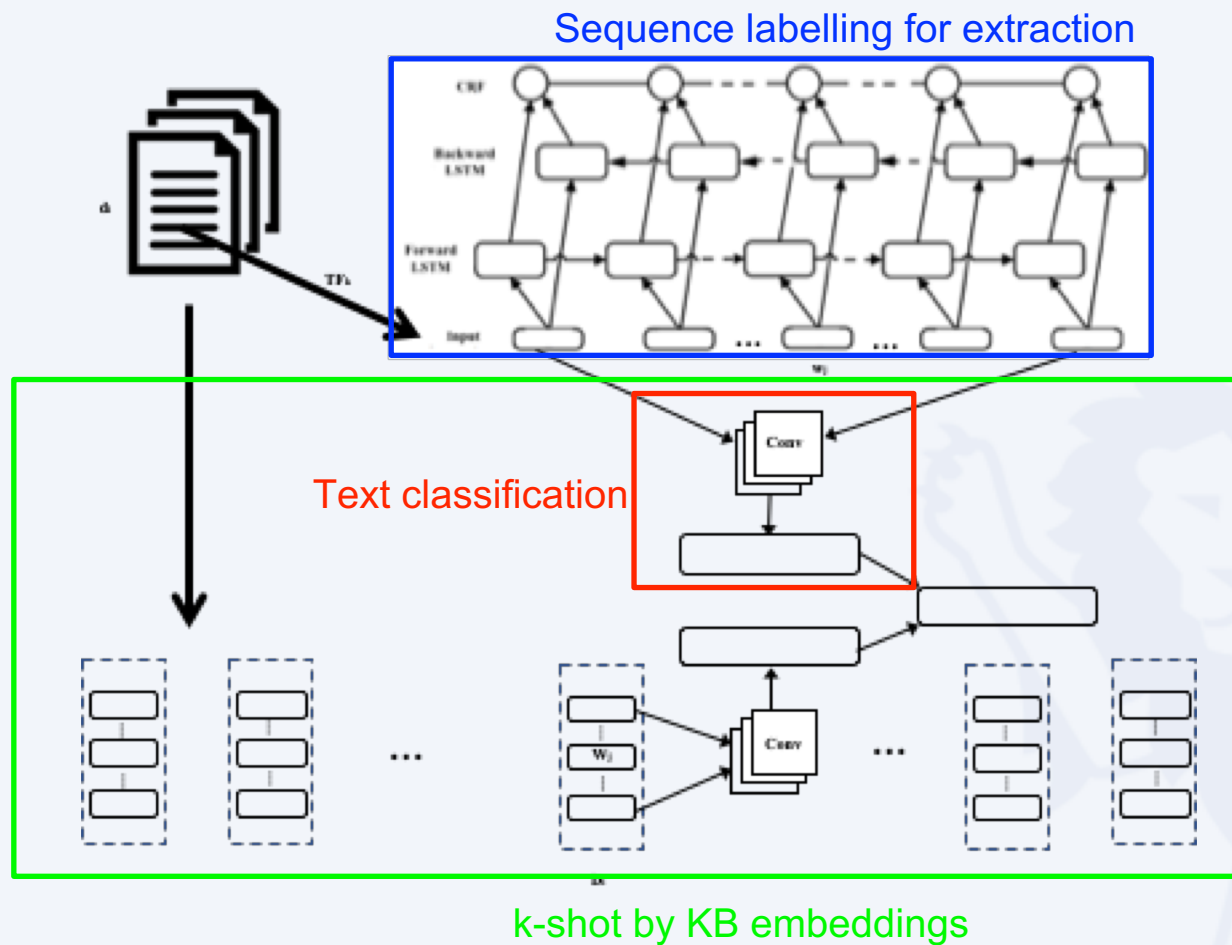
100: Monitoring the Future: A Continuing Study of American Youth (12th-Grade Survey), 1996;

108: Current Population Survey, May 1973; ...]

Subtask A: Extraction of the mentions

Subtask B: Classification (attachment) of the mention to a dataset

Model



Subtask A: Mention Extraction

	Partial			Exact		
Model	P	R	F1	P	R	F1
BiLSTM	29.4	32.1	30.7	11.2	12.8	12.0
CNN-BiLSTM	49.8	44.7	47.1	28.6	31.2	29.8
CNN-BiLSTM-CRF	54.1	44.8	48.9	35.6	33.8	34.7
CNN-BiLSTM-Attn-CRF	58.0	50.0	53.7	34.8	38.0	36.4
SL E-C	40.3	43.1	41.7	27.1	28.4	27.7

Subtask B: Classification

Model	P	R	F1
BiLSTM	27.5	47.4	34.8
CNN	42.8	46.5	44.6
SL E-C	31.8	49.3	49.3

Conclusion

Reasonable performance is easy to achieve, however human performance is difficult due to sparsity.

