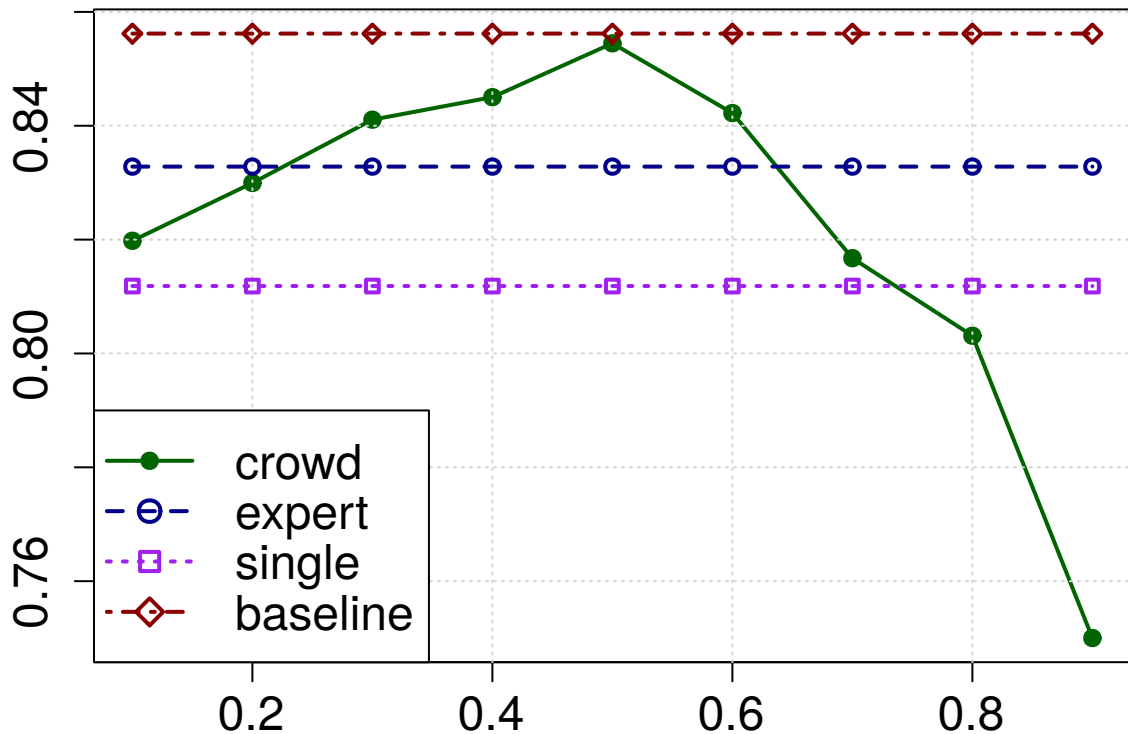*treat* relation

model evaluation F1 vs. neg/pos sentence-relation score threshold

Legend:
- crowd
- expert
- single
- baseline