Approximating Learning Curves for Active-Learning-Driven Annotation

Katrin Tomanek and Udo Hahn

Jena University Language and Information Engineering (JULIE) Lab
Agenda

- Introduction to Active Learning
- Stopping Conditions
- Experiments & Results
Passive versus Active Selection

passive annotation scenario (aka Random Sampling)

- Corpus selection
- Annotation
- Unlabeled
- Labeled
Passive versus Active Selection

passive annotation scenario (aka Random Sampling)

active annotation scenario (aka Active Learning)

"intelligent" example selection

annotation
Committee-based AL Framework

1. train on subsets (bagging)
2. predict labels
3. calculate disagreement
4. select by disagreement
5. annotate labeled examples

committee: $C_1, C_2, \ldots, C_n$

AL pool: unlabeled examples

$D_k(P_1, P_2, \ldots, P_n) \rightarrow u_k: P_1 P_2 \ldots P_n$

$D_1(P_1, P_2, \ldots, P_n) \rightarrow u_1: P_1 P_2 \ldots P_n$

$u_k: P_1 P_2 \ldots P_n \rightarrow u_k: P_1 P_2 \ldots P_n$
Reduction of Annotation Effort

learning curves

- t-score
- tokens

- AL selection
- random selection
Reduction of Annotation Effort

> 50% reduction

60K tokens | 130K tokens

F=0.83
When to Stop the Annotation?

Learning curves

![Graph showing learning curves with two lines representing active learning selection and random selection. The curve for active learning selection is above the curve for random selection, indicating better performance.](image-url)
When to Stop the Annotation?

Could gain a lot by further annotation
When to Stop the Annotation?

Could gain a lot by further annotation

Further annotation will not increase classifier performance significantly
When to Stop the Annotation?

Could gain a lot by further annotation

Further annotation will not increase classifier performance significantly

![Graph showing f-score over tokens with red and black lines indicating al selection and random selection respectively. The graph shows a comparison of f-score improvement with increasing tokens, highlighting the points where the annotation quality changes from bad to better and back to bad.](image)
Stopping Condition based on Learning Curve?

- **Pro**: stopping condition directly based on classifier performance
- **Contra**: requires *labeled* gold standard
  - not applicable in practice as gold standard not available

**Goal:**
- Estimate the (progression of) learning curve without need for gold standard
Approximating the Learning Curve

• Approach:
  – Based on agreement among committee members
  – Does not require extra labeling effort
  – Agreement curve approximates *progression* of learning curve

→ We can tell relative position in annotation process from it:
  • *relative* trade-off between annotation effort and gain in classifier performance from it
    – Steep slope ?
    – Convergence ?
Approximating the Learning Curve

• Intuition:
  – Agreement among committee:
    • Low in early AL iterations
    • High in later ones
  ➔ When agreement among committee members converges, also learning curve does
Approximating the Learning Curve

- Where to calculate the agreement:
  - On separate validation set
    - Not be involved in AL selection process itself
    - Agreement values comparable over different AL iteration
  - Otherwise agreement curve often not reliable approximation due to "simulation dilemma"
    - When e.g. agreement calculated on examples selected in each AL iteration:
      - Approximation of learning curve usually works well in simulation scenarios, because...
        » few hard cases left in later AL iterations (perfect agreement)
      - But fails in real-world annotation scenarios, because...
        » in practice AL will always find tricky cases...
Experiments

- For annotation of Named Entity mentions
- Whole sentences selected (20 each round)
- Simulation on CoNLL-2003 corpus
  - News-paper, MUC entities (PERS, LOC, ORG)
  - AL pool: ~ 14,000 sentences
  - Gold Standard: ~ 3,500 sentences
    - For learning curve
    - For agreement curve (labels ignored)
Results

learning curves

agreement curve
Summary & Conclusions

- AL has high potential to reduce annotation effort
- Proper stopping point necessary to profit from savings
  - Method to monitor progress of annotation needed
- Agreement curve
  - Works well: good approximation of learning curve
  - No extra annotation effort: does not require labeled gold standard
Approximating Learning Curves for Active-Learning-Driven Annotation

Thanks. Questions?

http://www.julielab.de/