

# Ideas for Approaches

## Natural Language Processing, Summer Semester 2021

Max Henze    Hanh Luu    Jan Heinrich Reimer

Martin Luther University Halle-Wittenberg

May 25, 2021



# Language Model Approach

- ▶ Using BERT we can create the following input structure
  - ▶ *[CLS]* argument *[SEP]* Key Point
- ▶ Hence learning contextual relations between words

---

pro	con
often used as baseline	computationally intensive
often good results but not too good	might need to use DistilBERT

---

# Language Model Approach cont.

- ▶ [StabMSRG2018] showed that integrating topic information "has a strong impact on argument prediction"
- ▶ Therefore maybe use BICLSTM and integrate topic information only on i and c gates

# Logistic Regression and Ensemble Voting

## All approaches with Stemming Snowball

name	params	mAP (relaxed)
regression-bow	C=14, CountVectorizer	0.70
regression-bow-pos	C=14, CountVectorizer, POS	0.66
regression-tfidf	C=1, TfidfVectorizer	0.55
ensemble-bow	LG, SVC, 0.55, 0.45	0.635
ensemble-bow	LG, SVC, 0.45, 0.55	0.647
ensemble-bow-pos	LG, SVC, 0.55, 0.45	0.696
ensemble-bow-pos	LG, SVC, 0.45, 0.55	0.713
svc-bow-pos	SVC, CountVectorizer, POS	0.74
svc-bow	SVC, CountVectorizer	0.70

# Logistic Regression and Ensemble Voting (cont.)

## Further

- ▶ features reducing: cutting words by certain length, ex: homeschooling - homesch (get max 6 characters)
- ▶ instead of stemming, using lemmatization
- ▶ other solvers for logistic regression
- ▶ using doc2vec embeddings for features

# References