Active Learning for text classification using rationales
Midwest Machine Learning Symposium (MMLS 2019)

Active Learning and Query Strategies

Active Learning is an iterative process that trained machine learning algorithms while minimizing the need for labeled data.

Random query strategy uniformly samples one unlabeled data point for human annotator to label.

Least certain query strategy returns the least certain unlabeled sample when being scored by the supervised learning process.

Simulation Results

Simulation =

IMDB movie review dataset +
gathered rationales from Mechanical Turk +
active learning process with Logistic Regression

See Results in Figure 1.

Document Features with Rationale

Human Annotator

Rationale $d_r$

Rationale-highlighted text why review is positive

Universal Sentence Encoder (USE) $\varphi_{\theta}$

$\varphi_{\theta}(d)$, $\varphi_{\varphi}(d_r)$

Least certain query strategy returns the least certain unlabeled sample when being scored by the supervised learning process.

Fig. 1. this figure, the least certain (lc) query strategy using the document features with rationale is converging faster, compared to the least certain and random query strategies with only document features.

Next Steps

Other query strategies for example Batch-Mode Active Learning (RAML) strategy.

Synthetic Rationales generated by Language Attention Models.

Online Learning updates the machine learning model incrementally with new training samples.

Please Contact:
Glenn Fung gfung@amfam.com,
Qian You qyou@amfam.com,
Rick Lentz rientz@amfam.com,
Teja Kanchinadam tkanchin@amfam.com

Simulation Results

Simulation =

IMDB movie review dataset +
gathered rationales from Mechanical Turk +
active learning process with Logistic Regression

See Results in Figure 1.

Active Learning and Human Annotation System

RoCKET

Robust Concept Knowledge Extraction From Text

Positive movie reviews?

Satisfied?

Movie Reviews

Yes/No and Rationale

Movie Review Classifier

Universal Sentence Encoder (USE) 4