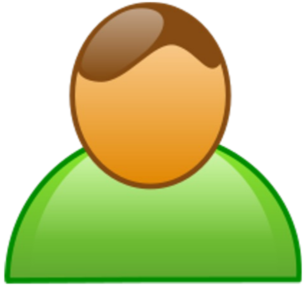
Efficiently Acquiring Human Feedback with Bayesian Deep Learning

Haishuo Fang, Jeet Gor, **Edwin Simpson**

University of Bristol

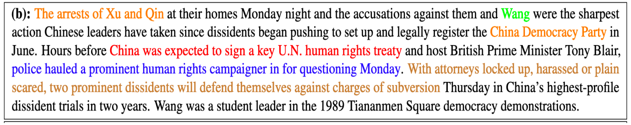
Learning from Human Feedback

- Vital to instruction tuning of LLMs + adaptation to specialised tasks
- Preferences are cheaper to acquire than gold-standard summaries, translations, paraphrases, etc.
- But which candidate outputs should we ask the human to compare?
  - Low training budget
  - End user preferences



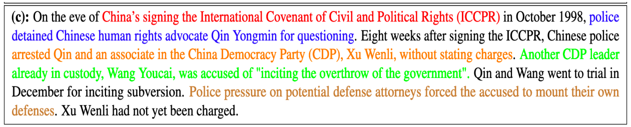
Which summary do you prefer?

□



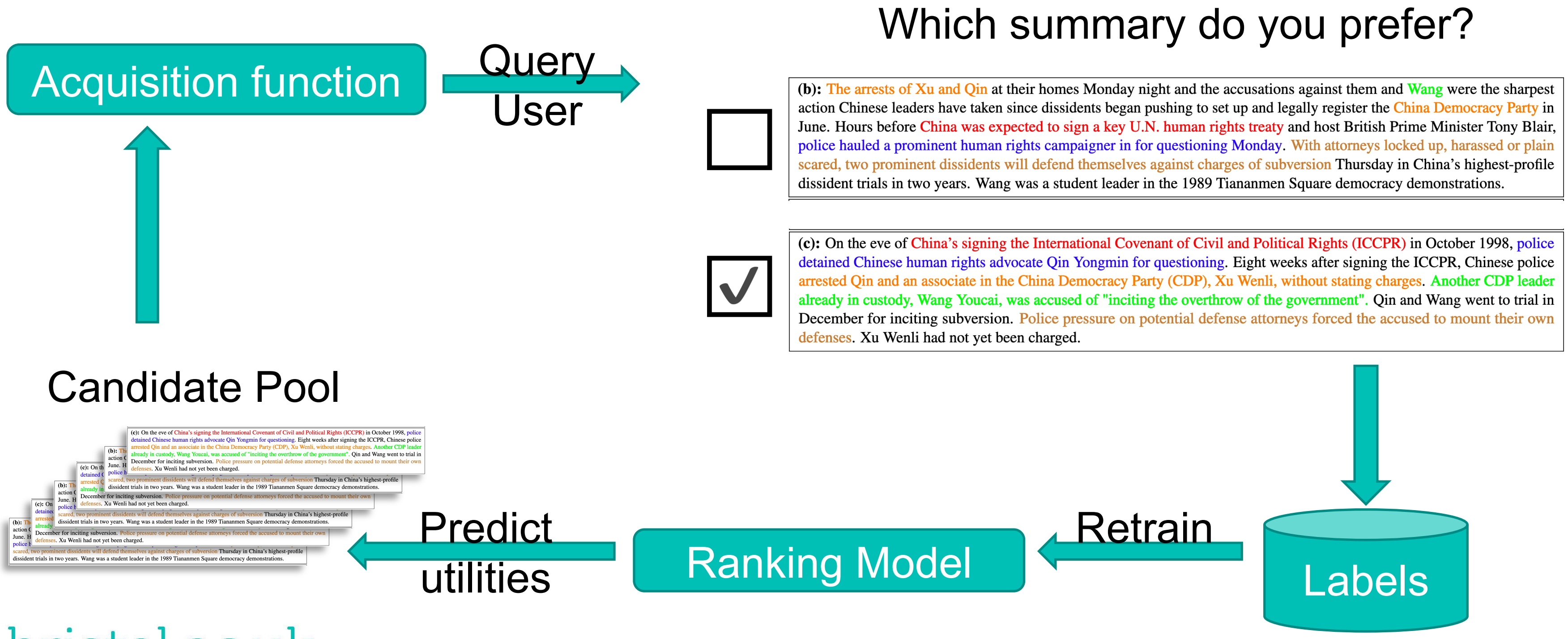
(b): The arrests of Xu and Qin at their homes Monday night and the accusations against them and Wang were the sharpest action Chinese leaders have taken since dissidents began pushing to set up and legally register the China Democracy Party in June. Hours before China was expected to sign a key U.N. human rights treaty and host British Prime Minister Tony Blair, police hauled a prominent human rights campaigner in for questioning Monday. With attorneys locked up, harassed or plain scared, two prominent dissidents will defend themselves against charges of subversion Thursday in China's highest-profile dissident trials in two years. Wang was a student leader in the 1989 Tiananmen Square democracy demonstrations.

☑



(c): On the eve of China's signing the International Covenant of Civil and Political Rights (ICCPR) in October 1998, police detained Chinese human rights advocate Qin Yongmin for questioning. Eight weeks after signing the ICCPR, Chinese police arrested Qin and an associate in the China Democracy Party (CDP), Xu Wenli, without stating charges. Another CDP leader already in custody, Wang Youcai, was accused of "inciting the overthrow of the government". Qin and Wang went to trial in December for inciting subversion. Police pressure on potential defense attorneys forced the accused to mount their own defenses. Xu Wenli had not yet been charged.

Interactive Text Ranking

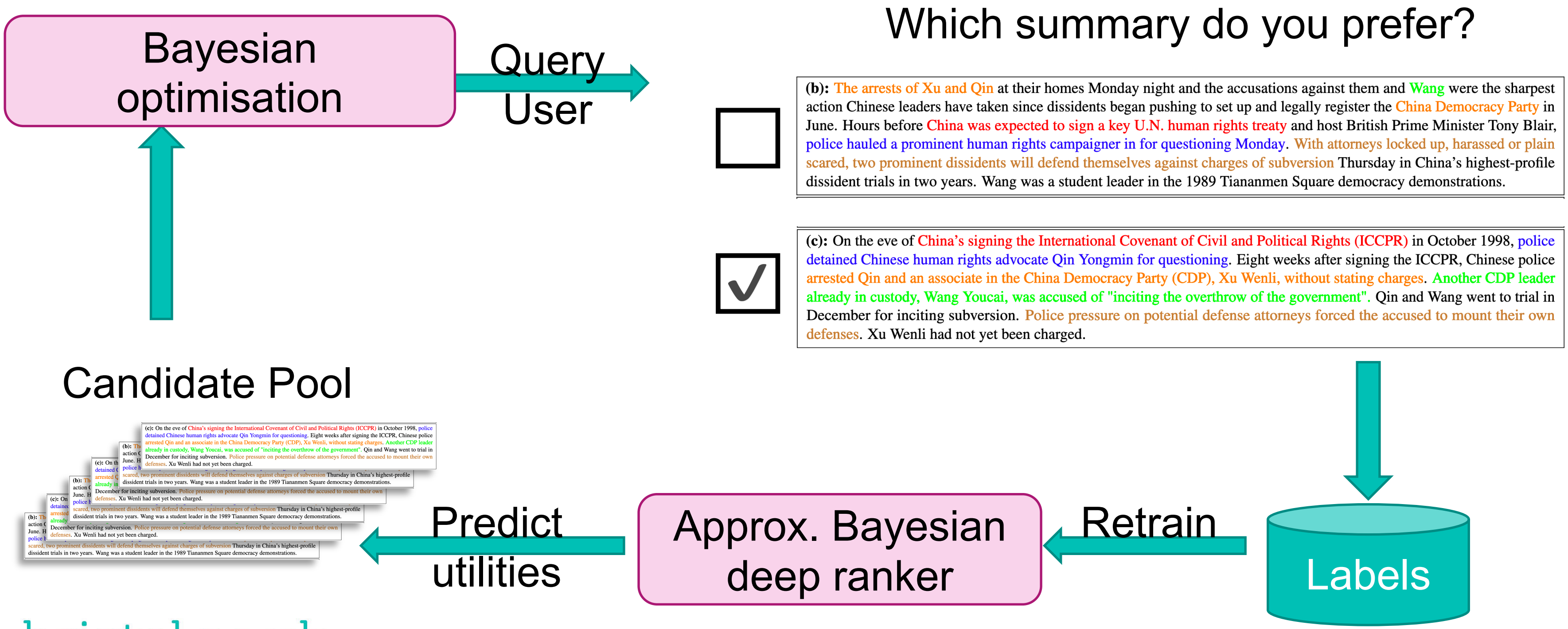


Which summary do you prefer?

□ **(b):** The arrests of Xu and Qin at their homes Monday night and the accusations against them and Wang were the sharpest action Chinese leaders have taken since dissidents began pushing to set up and legally register the China Democracy Party in June. Hours before China was expected to sign a key U.N. human rights treaty and host British Prime Minister Tony Blair, police hauled a prominent human rights campaigner in for questioning Monday. With attorneys locked up, harassed or plain scared, two prominent dissidents will defend themselves against charges of subversion Thursday in China's highest-profile dissident trials in two years. Wang was a student leader in the 1989 Tiananmen Square democracy demonstrations.

☑ **(c):** On the eve of China's signing the International Covenant of Civil and Political Rights (ICCPR) in October 1998, police detained Chinese human rights advocate Qin Yongmin for questioning. Eight weeks after signing the ICCPR, Chinese police arrested Qin and an associate in the China Democracy Party (CDP), Xu Wenli, without stating charges. Another CDP leader already in custody, Wang Youcai, was accused of "inciting the overthrow of the government". Qin and Wang went to trial in December for inciting subversion. Police pressure on potential defense attorneys forced the accused to mount their own defenses. Xu Wenli had not yet been charged.

Interactive Text Ranking

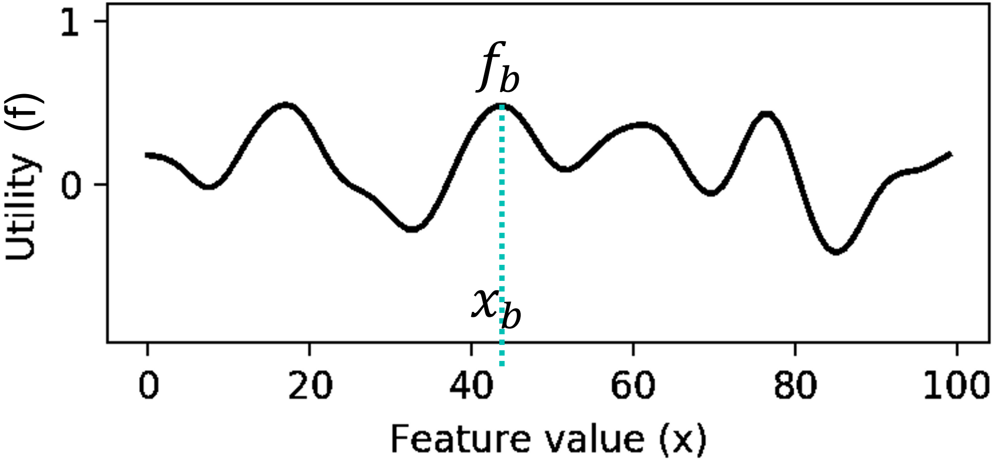


□ **(b):** The arrests of Xu and Qin at their homes Monday night and the accusations against them and Wang were the sharpest action Chinese leaders have taken since dissidents began pushing to set up and legally register the China Democracy Party in June. Hours before China was expected to sign a key U.N. human rights treaty and host British Prime Minister Tony Blair, police hauled a prominent human rights campaigner in for questioning Monday. With attorneys locked up, harassed or plain scared, two prominent dissidents will defend themselves against charges of subversion Thursday in China's highest-profile dissident trials in two years. Wang was a student leader in the 1989 Tiananmen Square democracy demonstrations.

☑ **(c):** On the eve of China's signing the International Covenant of Civil and Political Rights (ICCPR) in October 1998, police detained Chinese human rights advocate Qin Yongmin for questioning. Eight weeks after signing the ICCPR, Chinese police arrested Qin and an associate in the China Democracy Party (CDP), Xu Wenli, without stating charges. Another CDP leader already in custody, Wang Youcai, was accused of "inciting the overthrow of the government". Qin and Wang went to trial in December for inciting subversion. Police pressure on potential defense attorneys forced the accused to mount their own defenses. Xu Wenli had not yet been charged.

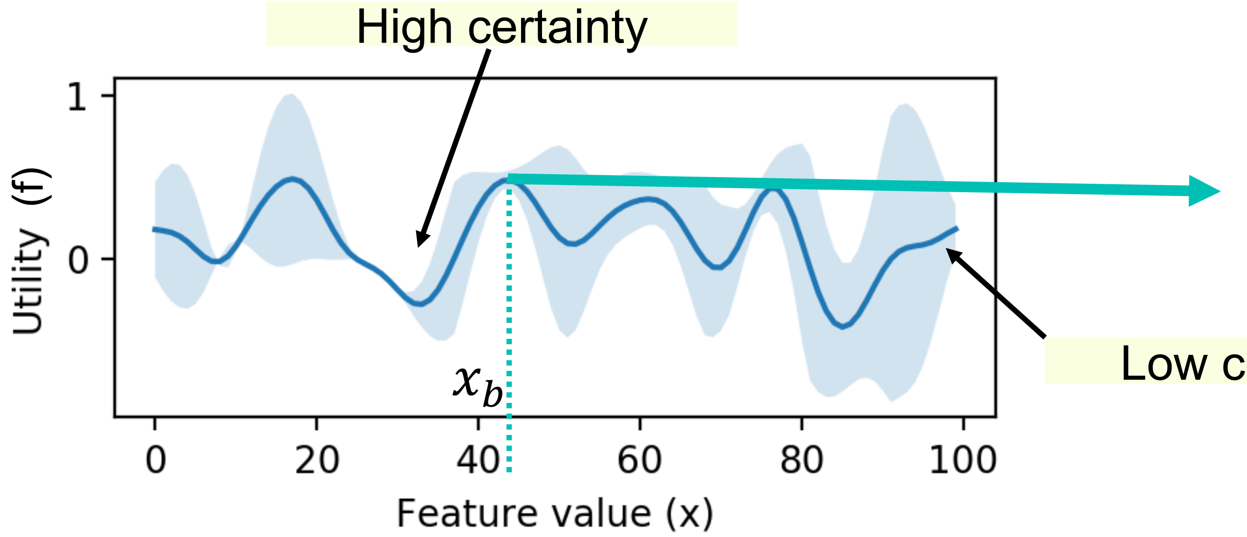
Ranking Model

- A ranking function maps candidates to a “utility”



Bayesian Ranker

- Estimate a probability distribution over ranking functions
- Compute the mean and variance of a set of sample predictions



High certainty

Low confidence

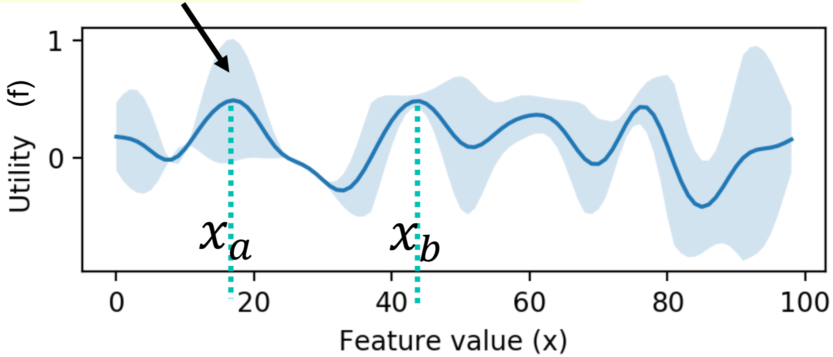
$$p(f_b) = N(\widehat{f_b}, v_b)$$

Bayesian Optimisation

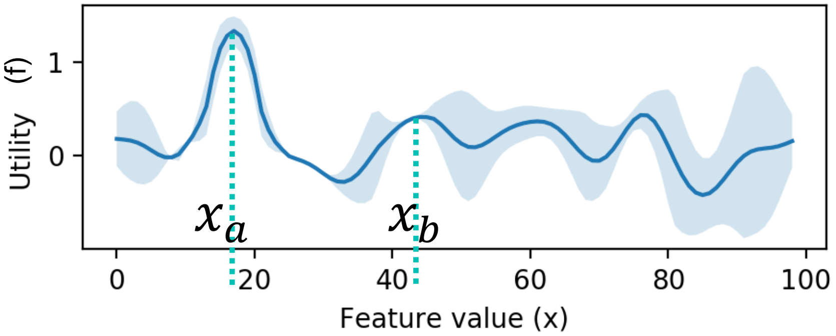
- Goal: find the best output, don’t learn the whole function!
- To select a pair, take current best, *b,* and the candidate *a* that maximises expected improvement:

$$\mathbb{E}[\max\{f_a - f_b, 0\}]$$

High uncertainty and mean

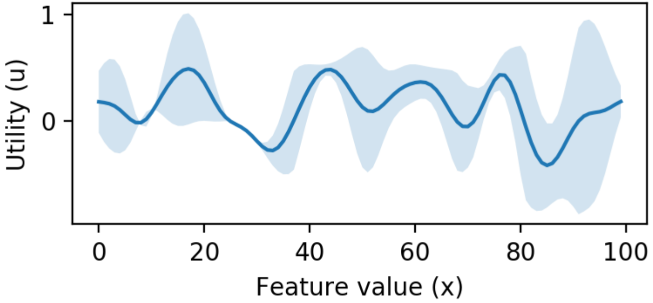


After new label $a \succ b$



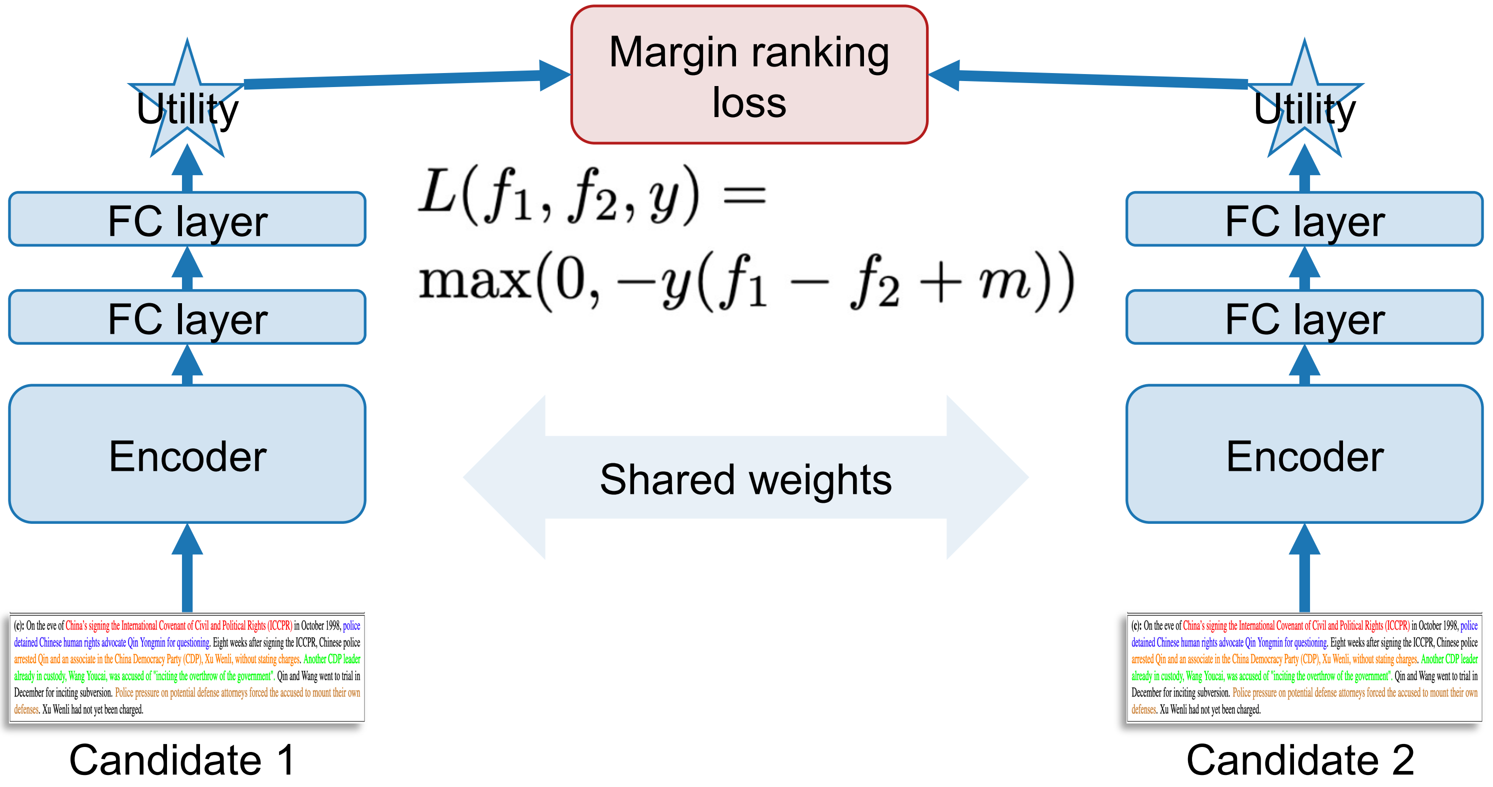
Bayesian Ranker

- Prior work used Gaussian processes
  - Frozen feature representation
  - Poor in high-dimensional feature space?
  - Uncertainty in the embeddings and prior?
  - Strong approximations for tractable inference
- Can we do better with approximate Bayesian deep learning?
  - Tune the whole model from preferences
  - Pretrained encoder (e.g., BERT)



Simpson et al., 2020, Interactive Text Ranking with Bayesian Optimisation: A Case Study on Community QA and Summarisation, TACL.

Deep Ranker

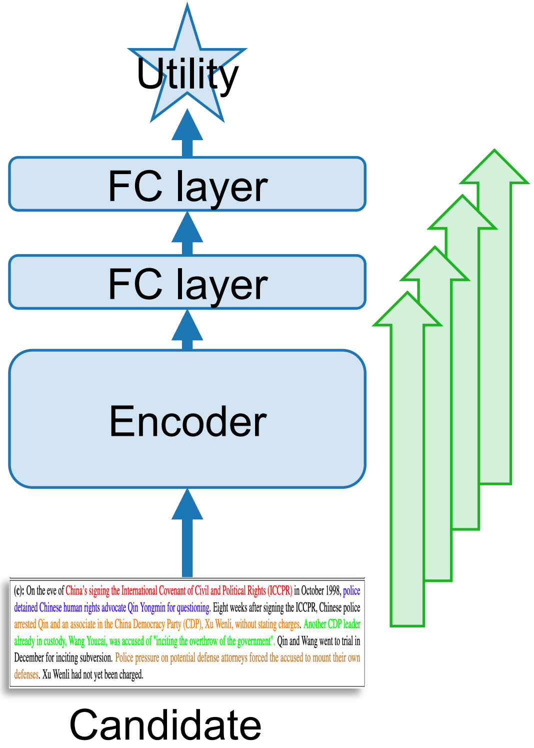


$$L(f_1, f_2, y) = \max(0, -y(f_1 - f_2 + m))$$

(c): On the eve of China's signing the International Covenant of Civil and Political Rights (ICCPR) in October 1998, police detained Chinese human rights advocate Qin Yongmin for questioning. Eight weeks after signing the ICCPR, Chinese police arrested Qin and an associate in the China Democracy Party (CDP), Xu Wenli, without stating charges. Another CDP leader already in custody, Wang Youcai, was accused of "inciting the overthrow of the government". Qin and Wang went to trial in December for inciting subversion. Police pressure on potential defense attorneys forced the accused to mount their own defenses. Xu Wenli had not yet been charged.

(c): On the eve of China's signing the International Covenant of Civil and Political Rights (ICCPR) in October 1998, police detained Chinese human rights advocate Qin Yongmin for questioning. Eight weeks after signing the ICCPR, Chinese police arrested Qin and an associate in the China Democracy Party (CDP), Xu Wenli, without stating charges. Another CDP leader already in custody, Wang Youcai, was accused of "inciting the overthrow of the government". Qin and Wang went to trial in December for inciting subversion. Police pressure on potential defense attorneys forced the accused to mount their own defenses. Xu Wenli had not yet been charged.

Approx. Bayesian Deep Ranker



(c): On the eve of China's signing the International Covenant of Civil and Political Rights (ICCPR) in October 1998, police detained Chinese human rights advocate Qin Yongmin for questioning. Eight weeks after signing the ICCPR, Chinese police arrested Qin and an associate in the China Democracy Party (CDP), Xu Wenli, without stating charges. Another CDP leader already in custody, Wang Youcai, was accused of "inciting the overthrow of the government". Qin and Wang went to trial in December for inciting subversion. Police pressure on potential defense attorneys forced the accused to mount their own defenses. Xu Wenli had not yet been charged.

MCD: dropout at inference time, collect 20 samples

SWAG: estimate distribution over weights using model checkpoints as samples

Community Question Answering

- Choose the best answer on StackExchange:
  - Select the correct answer from 100 candidates
  - StackExchange topics: Apple, Cooking, Travel
- Train a distilRoBERTa ranker on training questions
- Interactively tune the classifier for a specific question by collecting four simulated user preferences

cQA Results

Model	Strategy	accuracy
GPPL	BO expected improvement	0.580

cQA Results

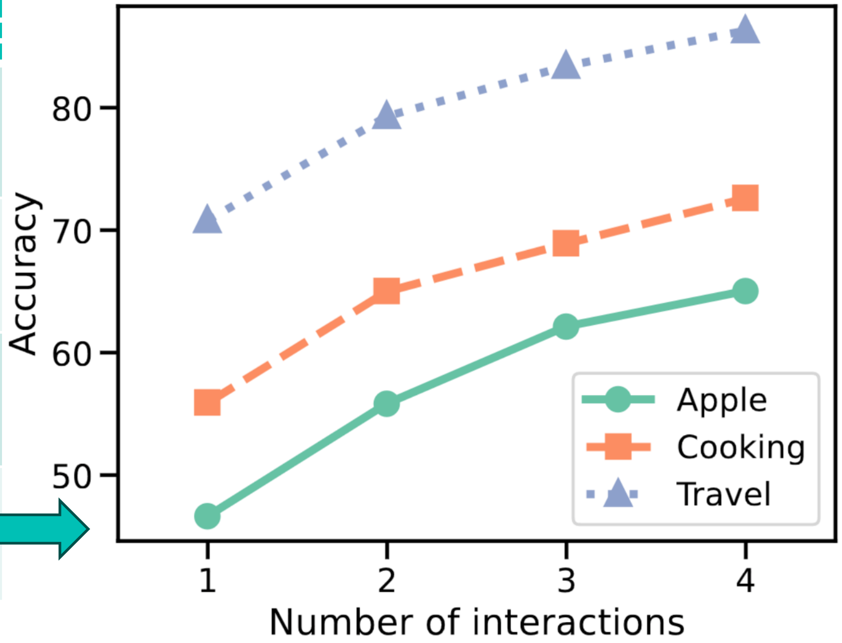
Model	Strategy	accuracy
GPPL	BO expected improvement	0.580
Deep Ranker	Uncertainty sampling	0.596

cQA Results

Model	Strategy	accuracy
GPPL	BO expected improvement	0.580
Deep Ranker	Uncertainty sampling	0.596
+SWAG	BO expected improvement	0.648
+MCD	**BO expected improvement**	**0.733**

cQA Results

Model	Strategy	accurac
GPPL	BO expected improvement	0.580
Deep Ranker	Uncertainty sampling	0.596
+SWAG	BO expected improvement	0.648
+MCD	**BO expected improvement**	**0.733**



bristol.ac.uk

Multi-Document Extractive Summarisation

- Extract a summary for a news topic
  - DUC 2001, 2002, 2004 newswire datasets
  - Rank 10,000 randomly-generated candidate summaries
- Train a ranker on training topics with SUPERT as encoder
- Interactively tune the topic summary for a particular simulated user
  - Each topic has three reference summaries written by different people
  - Collect six preferences per summary

Summarisation Results

Model	Strategy	Number of interactions	NDCG @1%
GPPL	BO expected improvement	20	0.636
Deep ranker	Uncertainty sampling	6	0.551
+ MCD	**BO expected improvement**	**6**	**0.660**

Conclusions

- Small amounts of carefully chosen human feedback can quickly identify the best model outputs
- Simple approximations to Bayesian deep learning provide effective uncertainty estimates for selecting feedback
- Limitations: users are simulated
- Future:
  - Can we apply BO to to tune LLMs or prompts with end-user feedback?
  - Do alternative approaches such as Bayesian layers (Tran et al, NeurIPS 2019) outperform MCD?