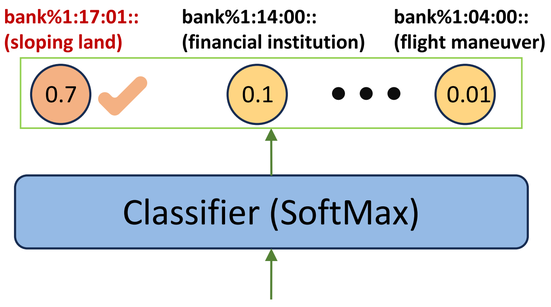
Ambiguity Meets Uncertainty: Investigating Uncertainty Estimation for Word Sense Disambiguation

Zhu Liu, Ying Liu
liuzhu22@mails.tsinghua.edu.cn

Motivation



- Word Sense Disambiguation (WSD) as classification
- Modern neural networks are poorly calibrated
- WSD needs uncertainty estimation

Model Uncertainty

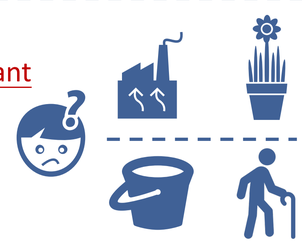
Imbalanced sense distribution

Domain shift

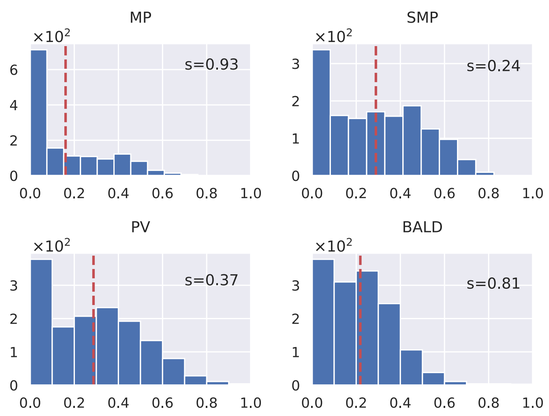
Data Uncertainty

The energy comes from a plant

Kick the bucket



Evaluation: Two Uncertainties



- Traditional SoftMax output (MP) is over-confident
- MC Dropout Sampling-based score (SMP) is better than MP

UE Score	Senseval-2		Senseval-3		SemEval-07		SemEval-13		SemEval-15	
	RCC ↓	RPP ↓	RCC ↓	RPP ↓	RCC ↓	RPP ↓	RCC ↓	RPP ↓	RCC ↓	RPP ↓
MP	**5.69**	9.50	7.11	10.37	**8.68**	11.40	5.78	8.02	**5.02**	**11.07**
SMP	5.78	**9.14**	**7.10**	**9.83**	8.81	**10.83**	**5.59**	**7.88**	5.34	11.16
PV	6.11	11.47	7.50	12.40	9.93	16.00	5.97	10.22	5.62	13.11
BALD	6.00	11.09	7.46	11.99	9.36	14.73	5.83	10.02	5.48	12.77

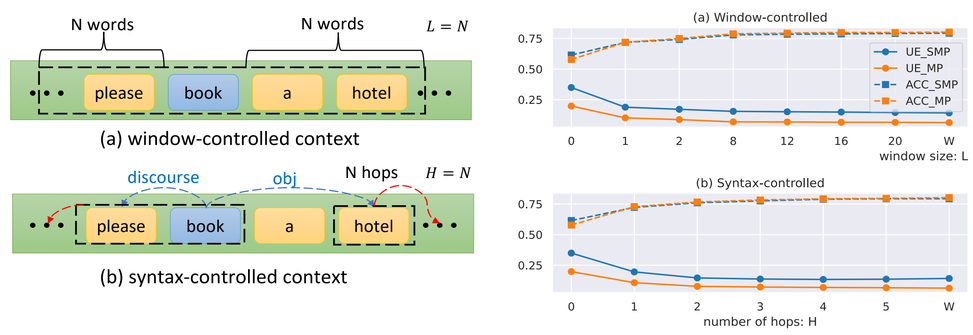
Table 1: UE score comparisons on five standard WSD datasets.

UE Score	NOUN		VERB		ADJ		ADV		ALL	
	RCC ↓	RPP ↓	RCC ↓	RPP ↓	RCC ↓	RPP ↓	RCC ↓	RPP ↓	RCC ↓	RPP ↓
MP	6.06	**7.47**	14.08	18.20	5.15	**8.25**	3.70	4.89	6.13	9.78
SMP	**4.94**	7.66	**13.76**	**17.45**	**4.39**	8.35	**2.65**	**4.85**	**6.11**	**9.44**
PV	6.25	9.17	15.38	22.02	4.97	9.37	3.20	5.33	6.48	11.91
BALD	5.18	9.39	14.42	20.96	4.59	9.80	2.66	5.56	6.36	11.52

Table 2: UE score comparisons on all the datasets with different kinds of POS.

Evaluation: Uncertainty Scores

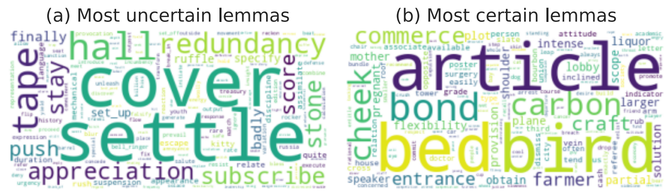
Two test scenarios for **data uncertainty**



- The model captures data uncertainty well
- Model uncertainty is underestimated

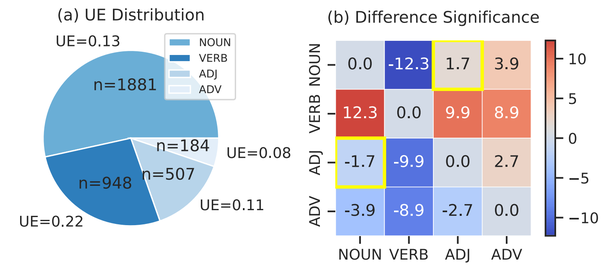
Out-of-distribution test for **model uncertainty**

Analysis: Effects



Significant difference among different levels in terms of

- Syntactic categories
- Morphology (nMorph)
- Sense granularity (nGT, nPD)
- Semantic relation (hyponym)



Effect	Condition	Agg.	Uncertainty Estimation			Difference Significance		
			L1	L2	L3	L1 ↔ L2	L1 ↔ L3	L2 ↔ L3
nMorph	nGT=1, POS=NOUN	L	0.13	0.11	0.07	**1.44e-2**	**1.35e-8**	**5e-4**
	nGT=1, POS=VERB		0.22	0.19	0.13	7.61e-2	**6.04e-4**	6.6e-2
	nGT=1, POS=ADJ		0.11	0.08	0.10	**3.6e-2**	4.21e-1	4.40e-1
	nGT=1, POS=ADV		0.11	0.06	0.02	7.6e-2	**6.04e-4**	6.60e-2
nGT	-	I	0.12	0.22	-	**1.61e-22**	-	-
nPD	nGT=1	L	0.04	0.16	0.22	**6.22e-96**	**3.42e-135**	**5.01e-10**
dHypo	nGT=1, POS=NOUN	L	0.14	0.12	0.09	**1.43e-2**	**1.91e-6**	**6e-3**
dSyno	nGT=1	S	0.14	0.14	0.14	5.55	5.38	5.67

Paper and Code:
https://github.com/RyanLiut/WSD-UE

61 ACL 2023

Tsinghua University