wsdm2019
12th ACM International Conference on Web Search and Data Mining
Melbourne Australia | 11-15 February 2019

Microsoft

KLM

DIR 19

UNIVERSITY OF AMSTERDAM

Domain Adaptation for Commitment Detection in Email

Hosein Azarbonyad[(1)], Robert Sim[(2)], and Ryen White[(2)]

(1) University of Amsterdam and KLM

(2) Microsoft AI, Seattle

WHAT IS COMMITMENT?

- Any sentence in email
  - where the sender is promising to do an action which can potentially be added to his/her TO-DO list (eg. sending a document)
  - can be worthy of a reminder (e.g. meeting a colleague)

```
From: sender
To: recipient
Subject: Opportunity for Enron

Chad, thank
you for your email. I will forward on to Dan Reck
who is responsible for our new Enron Freight Markets
business. I am sure you will be hearing from him.

Thanks,
m
```

WHY COMMITMENT DETECTION IS IMPORTANT?

➤ People use emails not only as a communication tool, but also as a means to create and manage tasks

➤ Automatic task management systems can assist users manage their tasks more efficiently

➤ Commitments are often hidden in emails and users struggle to recall and complete them in a timely manner

COMMITMENT DETECTION

- Commitment detection is a challenging task

  Challenge1: There is no publicly available large-enough dataset for this task

  Challenge2: There is a domain bias associated with email datasets

DATASETS

➤ We crowd-source a set of samples from Enron and Avocado and collect commitment labels

The statistics of commitment datasets

	Enron	Avocado
# samples	65,398	13,021
# positive samples	3,337	4484
avg. sentence length	12.1	14.5
median sentence length	10	13

The most informative Enron features regarding the positive class

"i will", "i", "will", "i'll", "let you know", "let you", "call you", "i shall", "we will", "will call"

CAN COMMITMENTS BE RELIABLY DETECTED?

- In-domain performance of a logistic regression classifier
- Task
  - Binary classification
  - Classify if the sample constitutes any commitment
  - Features: word n-grams

Dataset	Precision	Recall	F1
Avocado	0.82	0.81	0.81
Enron	0.80	0.77	0.78

The commitment model achieves a reasonable performance

COMMITMENT DETECTION

- Commitment detection is a challenging task

Challenge1: There is no publicly available large-enough dataset for this task

Challenge2: There is a domain bias associated with email datasets

PERFORMANCE OF COMMITMENT MODELS ACROSS DOMAINS

Train	Test	Precision	Recall	F1	AUC
Avocado	Avocado	0.82	0.81	0.81	0.86
	Enron	0.77	0.69	0.73	0.67
Enron	Enron	0.80	0.77	0.78	0.88
	Avocado	0.74	0.78	0.76	0.58

- Performance of commitment models degrade significantly when moving across domains
- *We cannot reliably train a model in one domain and use it to detect commitments on a different domain*

DOMAIN BIAS IN EMAIL DATASETS AND MODELS

- Most email-based models are derived from public datasets, which are skewed in a variety of ways
  - different organizations with very different and specific focus areas
  - being old and adding an element of obsolescence
  - different named entities and technical jargon

Our goal: Using transfer learning for transferring knowledge learned in one domain to other domains and achieve more robust and generalizable models for commitment extraction

DETECTING COMMITMENTS ACROSS DOMAINS

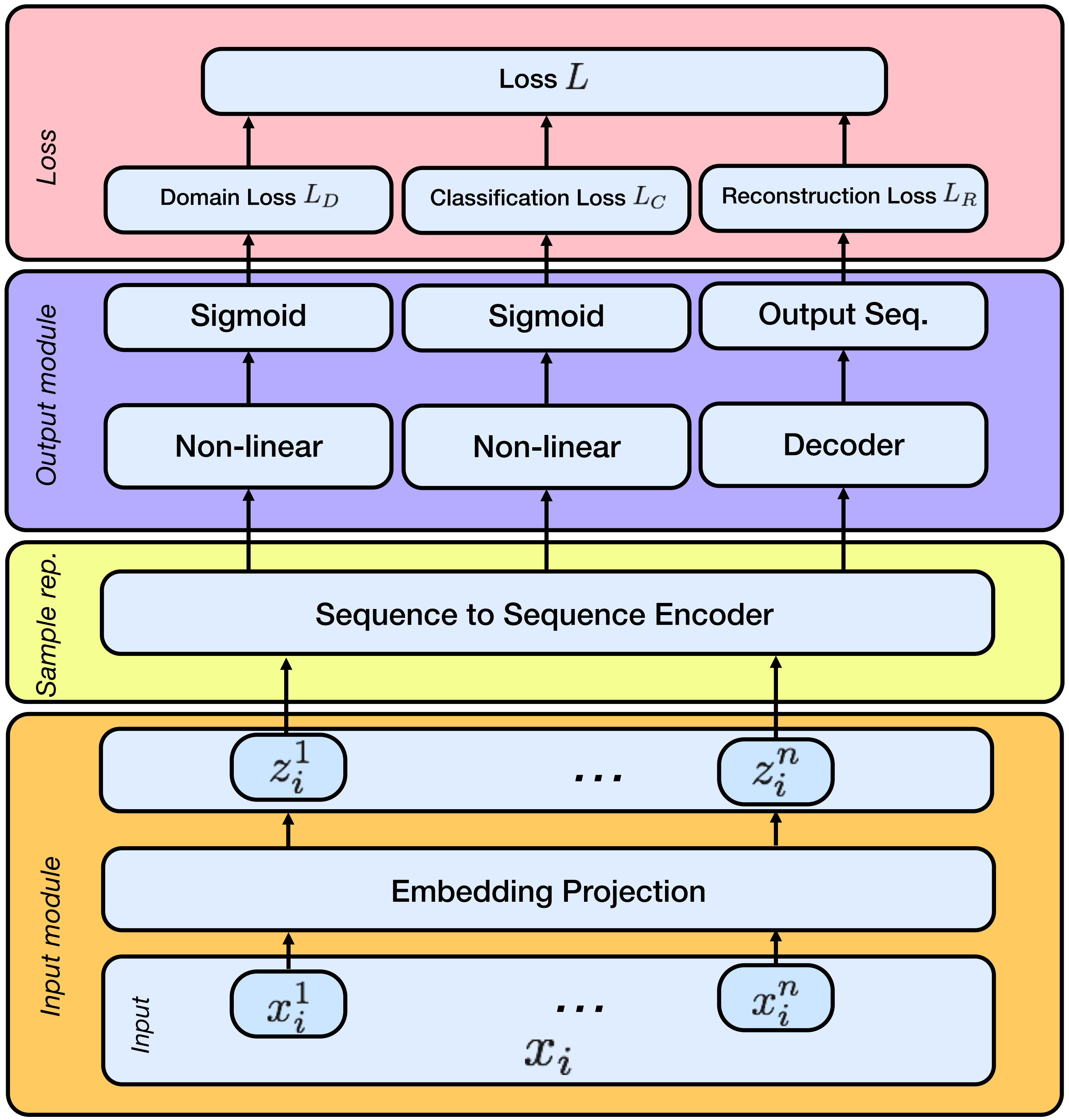
- Feature-level transfer learning
  - Feature selection
  - Feature mapping
- Sample-level transfer learning
  - Importance sampling
- **Deep autoencoder**

DEEP AUTOENCODERS: OBJECTIVE

- **Goal**: to achieve a domain independent representation for samples optimized for the commitment detection task

- Objectives
  - Achieve a good representation for samples: the representation should capture the core and essential parts of the input sample
    - **Conventional reconstruction loss**
  - Achieve a good performance in commitment detection task
    - **Commitment classification loss**
  - Remove domain bias
    - **Domain loss**

DEEP AUTOENCODERS: ARCHITECTURE OVERVIEW



AUTOENCODER RESULTS

Train	Test	Method	Precision	Recall	F1	AUC
Avocado	Enron	IS	0.81	0.75	0.77	0.74
		LR	0.80	0.79	0.79	0.78
		AE_R	0.80	0.78▲	0.79	0.76
		AE_{R+D}	0.81	0.79▲	0.80▲	0.77▲
		AE_{All}	0.82	0.81▲	0.81▲	0.79▲
Enron	Avocado	IS	0.78	0.85	0.81	0.71
		LR	0.80	0.82	0.81	0.70
		AE_R	0.77	0.82	0.79	0.68
		AE_{R+D}	0.77	0.84	0.80	0.69
		AE_{All}	0.79▲	0.87▲	0.83▲	0.72

- Proposed AE outperforms IS method significantly over all datasets
- All loss functions contribute to the performance of the AE method

CONCLUSIONS

- Commitments can be reliably detected in emails when models are trained and tested on a same domain (dataset).

- However, their performance degrades when moving across domains

- Domain bias can have a big impact on the performance of commitment models and email models in general

- We can detect and characterize this bias from email datasets

- This characterization can be used for training reliable and generalizable commitment models

Thank You!

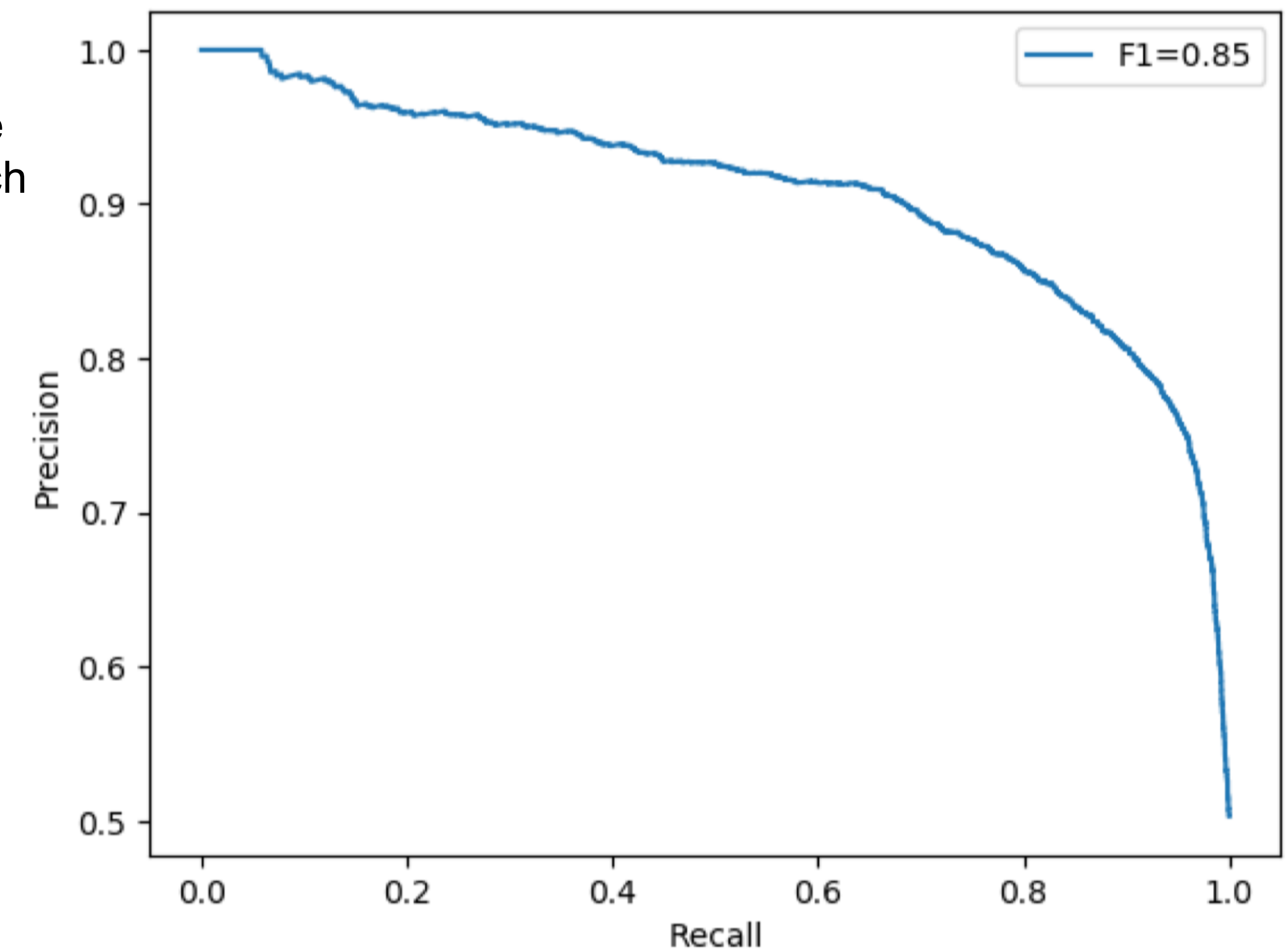
Hosein Azarbonyad

@HAzarbonyad

hosein.azarbonyad@klm.com

CHARACTERIZING DIFFERENCES BETWEEN DOMAINS

The Precision-Recall curve of the domain classifier (predicting which domain the samples come from)



The most informative unigram features indicating the Enron domain

"enron", "gas", "ena", "houston", "ferc", "eol", "energy", "ees", "counterparty"

CHARACTERIZING DIFFERENCES BETWEEN DOMAINS

- Can we use the characterization between domains to train domain-independent commitment models?

Train	Test	Method	Precision	Recall	F1	AUC
Avocado	Enron	LR	0.77	0.69	0.73	0.67
		IS	0.81▲	0.75▲	0.77▲	0.74▲
		LM	0.83▲	0.76▲	0.79▲	0.75▲
		FS	0.80▲	0.73▲	0.76▲	0.73▲
Enron	Avocado	LR	0.74	0.78	0.76	0.58
		IS	0.78▲	0.85▲	0.81▲	0.71▲
		LM	0.75	0.81▲	0.77	0.64▲
		FS	0.74	0.80▲	0.76	0.62

- All transfer learning approaches improve the performance of LR model
- More improvements for Enron->Avocado
  - Enron samples are more biased and domain specific

AUTOENCODER RESULTS

➤ How much data does AE need in the target domain to achieve a good performance?