Few-Shot Learning for Opinion Summarization

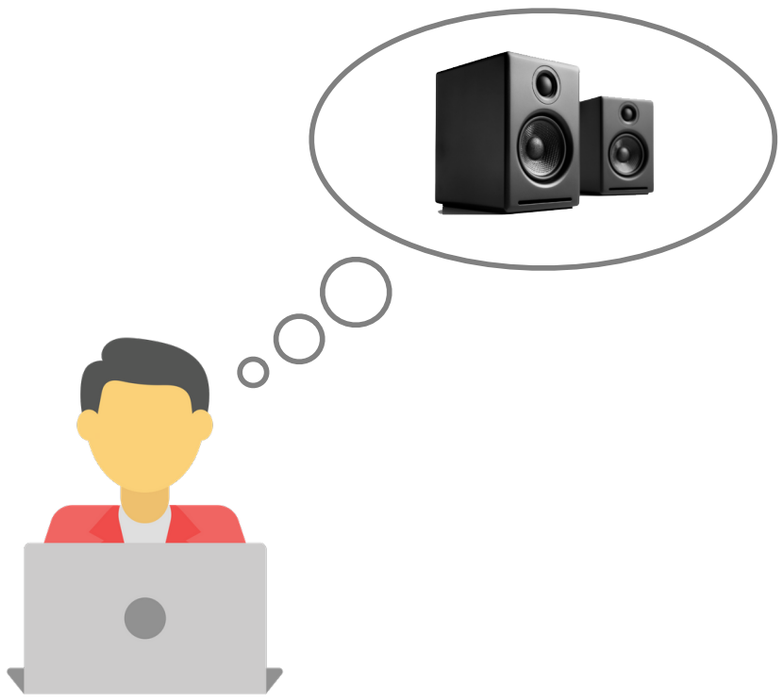
Arthur Bražinskas, Mirella Lapata, Ivan Titov

The University of Edinburgh, Scotland

EMNLP 2020

Opinion Summarization

James

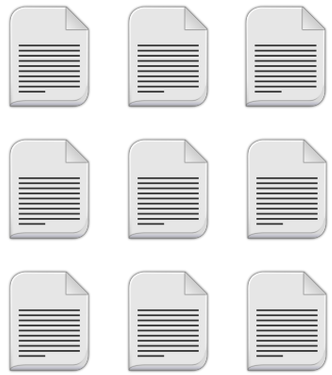


James

James

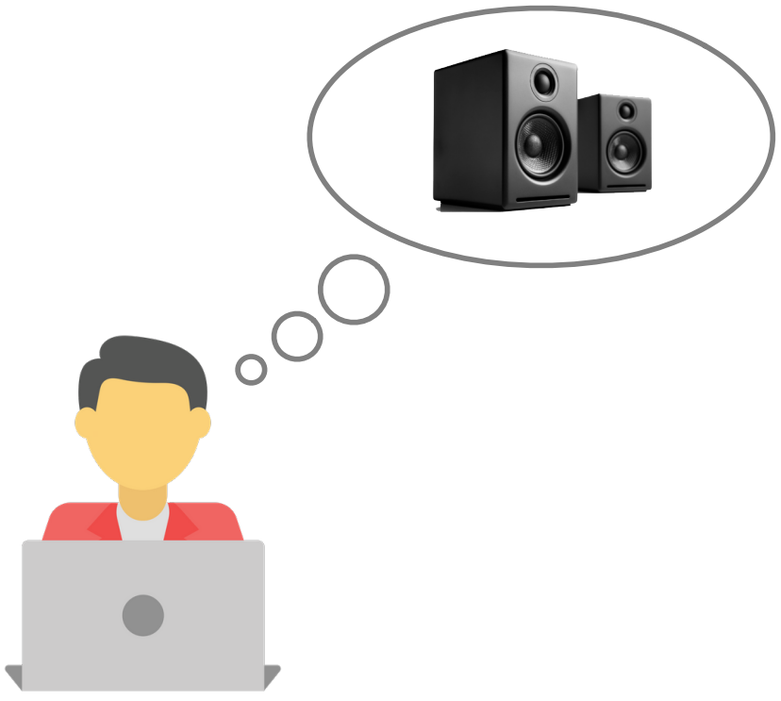
Online store

James

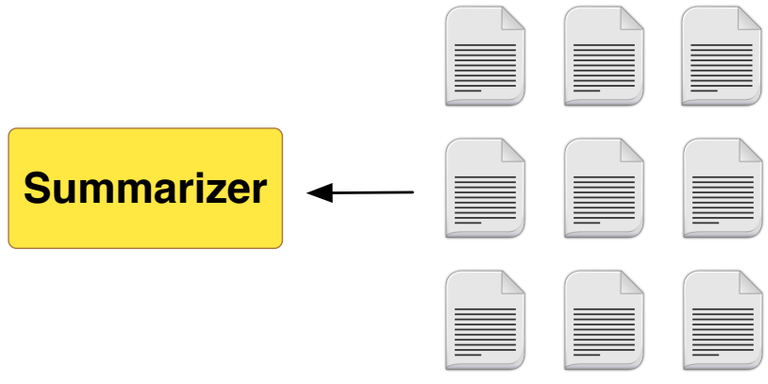


Reviews

Online store

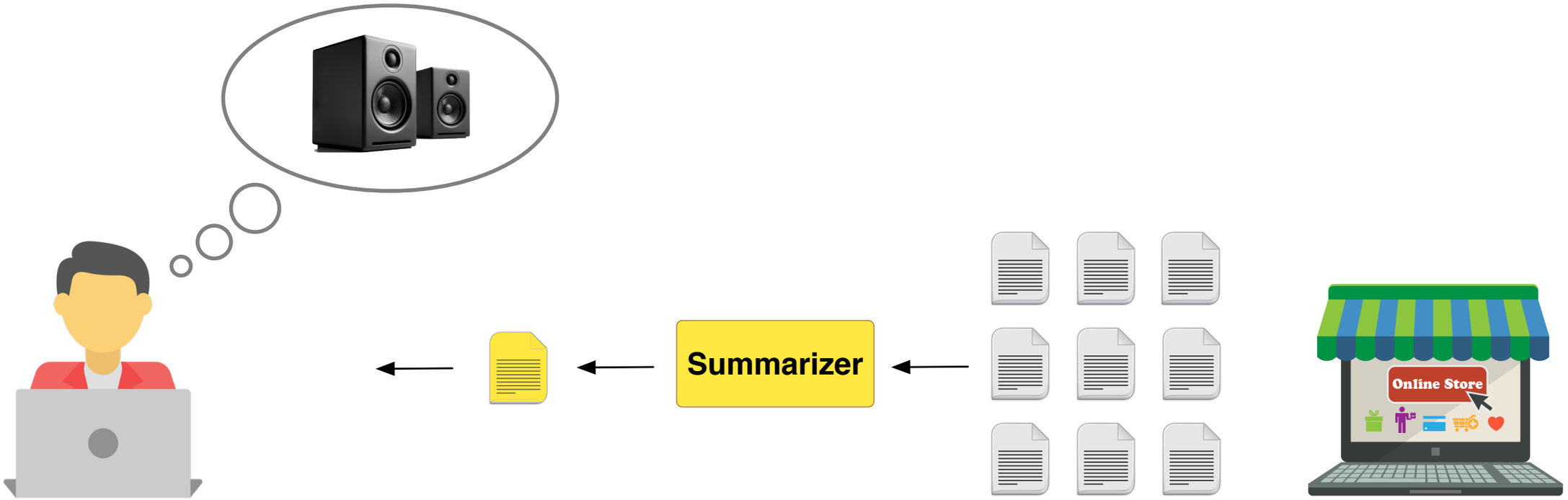


James



Reviews

Online store



James

Summary

Reviews

Online store

Opinion and news summarization

	News	Opinion
Setup	Single-document	Multi-document
Task	Objective facts	Subjective opinions
Annotated abstractive data	1M+ NewsRoom (Grusky et. al. 2018)	100 MeanSum (Chu and Liu, 2019)

Opinion summarization (unannotated data)

233 million reviews

8 million reviews

Extractive summarizers

- Are commonly used for the task (Ganesa et. al 2010; Angelidis and Lapata, 2018; Isonuma et al. 2019)
- Mostly **unsupervised** or **weakly-supervised**
- Select **summarizing input fragments**
- Concatenate to form a summary
- Can be **incoherent** and **unimportant details**

Abstractive summarizers

- Generate text (Paulus et. al. 2017; See et. al. 2017; Liu et al., 2018)
- Can use a **richer vocabulary** of words
- Can **rephrase, condense**, and **abstract**
- Can deal with **conflicting information**
- Require **large annotated datasets** for training

Unsupervised abstractive methods

- A few abstractive unsupervised summarizers:
  - **MeanSum** (Chu and Liu, 2019)
  - **Copycat** (Bražinskas et. al. 2020)
- Can induce **common opinions** to some extent
- Often the **output summaries** are:
  - written as **reviews**
  - some content is **unimportant**
- Reason: **never exposed** to **human-written summaries**

MeanSum

The shirt is very soft and comfortable. I bought a size larger than I normally wear and it fits fine. I'm 5 '4 and the top is a bit short. I guess I just got a good deal.

MeanSum

problem: superficial, unimportant details

The shirt is very soft and comfortable. ***I bought a size larger than I normally wear and it fits fine.*** *I'm 5 '4 and the top is a bit short. I guess I just got a good deal.*

MeanSum

problem: writing style

*The shirt is very soft and comfortable. **I** bought a size larger than **I** normally wear and it fits fine. **I'm** 5 '4 and the top is a bit short. **I** guess **I** just got a good deal.*

In this work

- Propose a **few-shot learning framework**
- Utilises a **handful** of **human-written summaries**
- Effectively switch an **unsupervised model** to a **summarizer**
- Summaries:
  - written in the **formal style**
  - have more **informative** content

Approach

Conditional language model

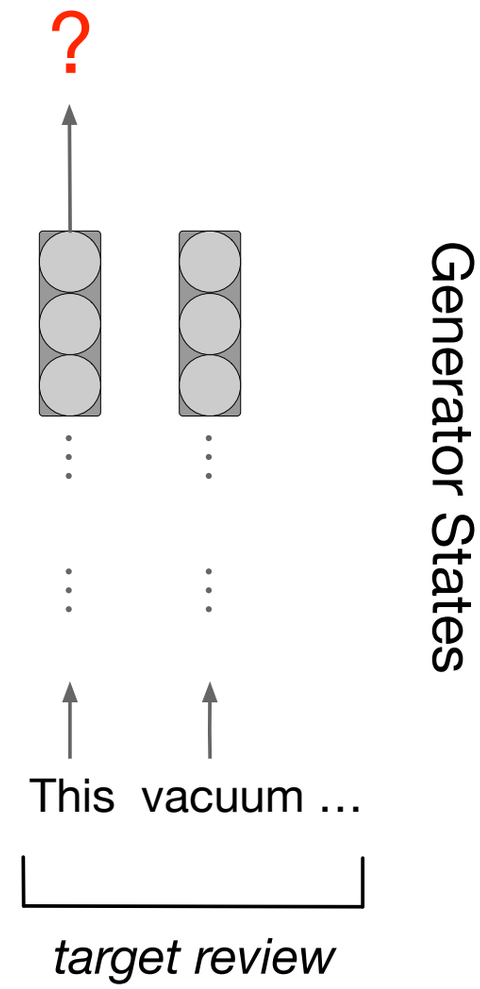
- Conditional language model (CLM)
- Encoder-generator architecture
- Unsupervised training
- On a large collection of customer reviews
- Using the **leave-one-out objective**

Leave-one-out

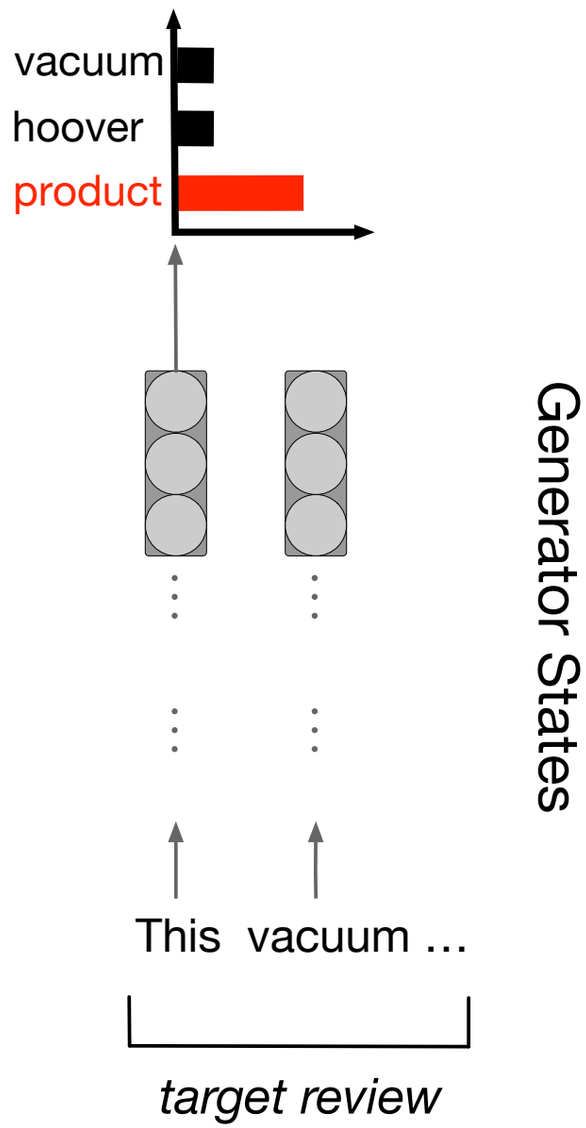
Great Italian restaurant with authentic food and great service! Recommend!	We ordered pasta, and it was very tasty. Would recommend this place to anyone.	This Italian place has the best spaghetti in the world! Strongly recommend!	We visited this place last week. The waiters were friendly, and the food was great!
review 1	review 2	review 3	review 4

Leave-one-out

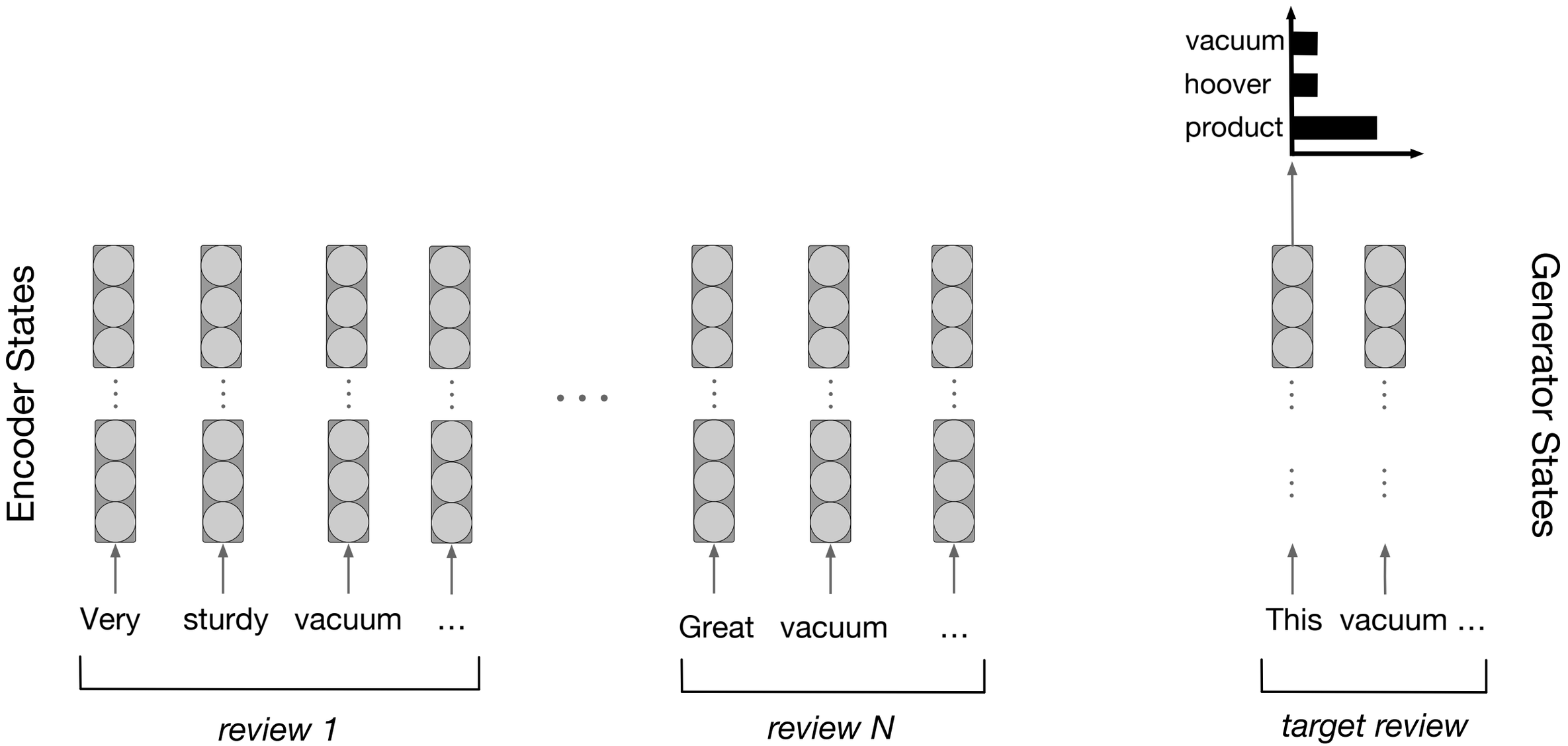
Leave-one-out



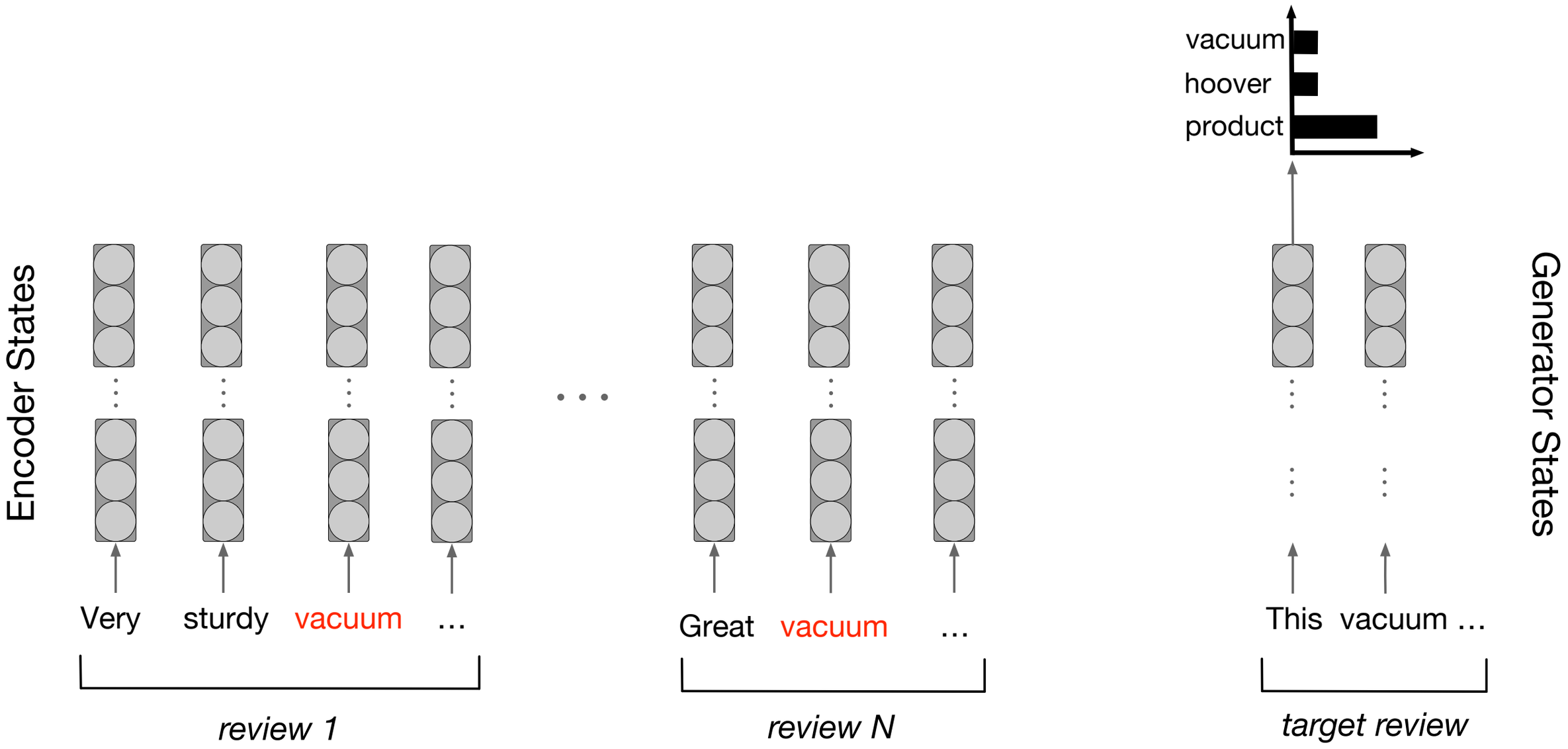
Leave-one-out



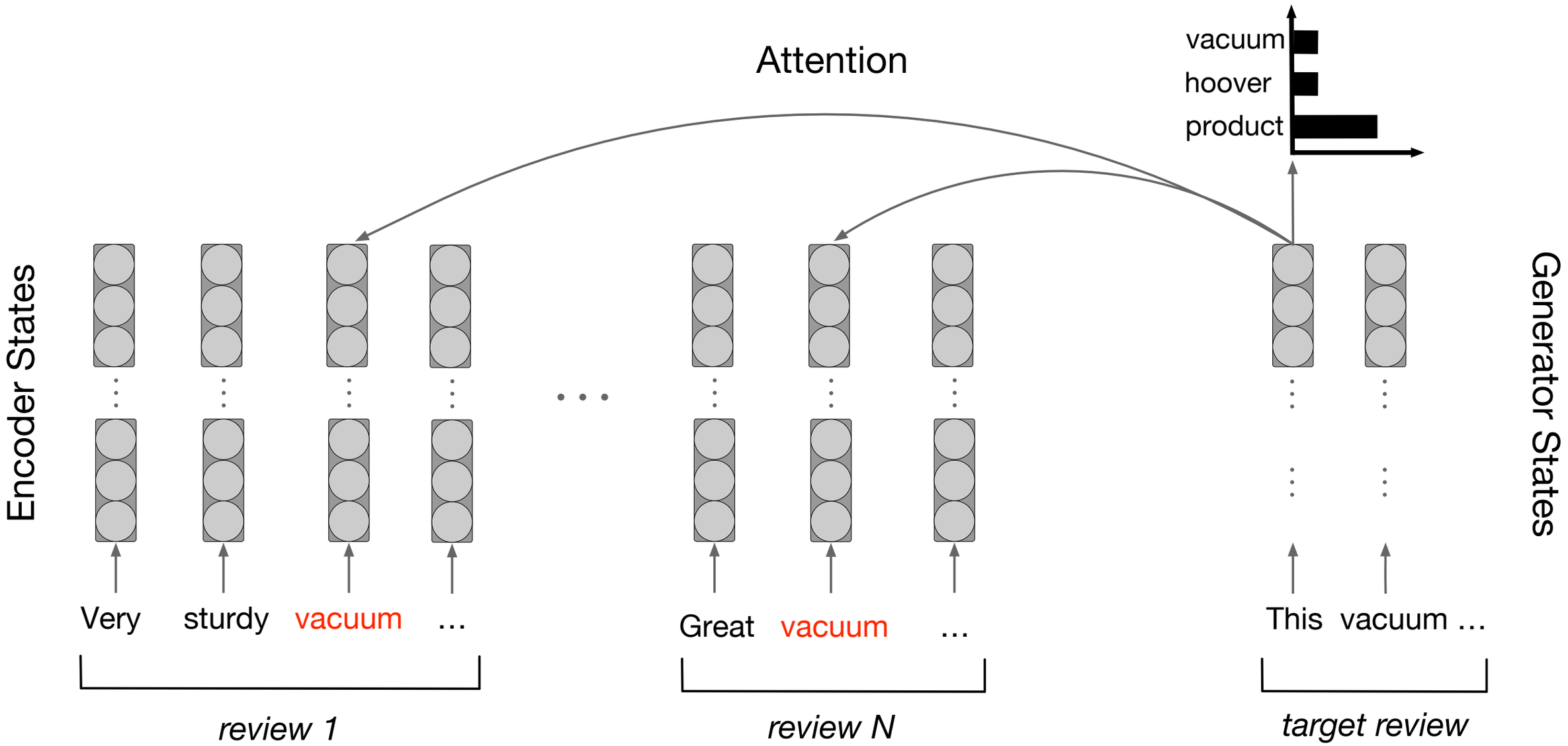
Leave-one-out



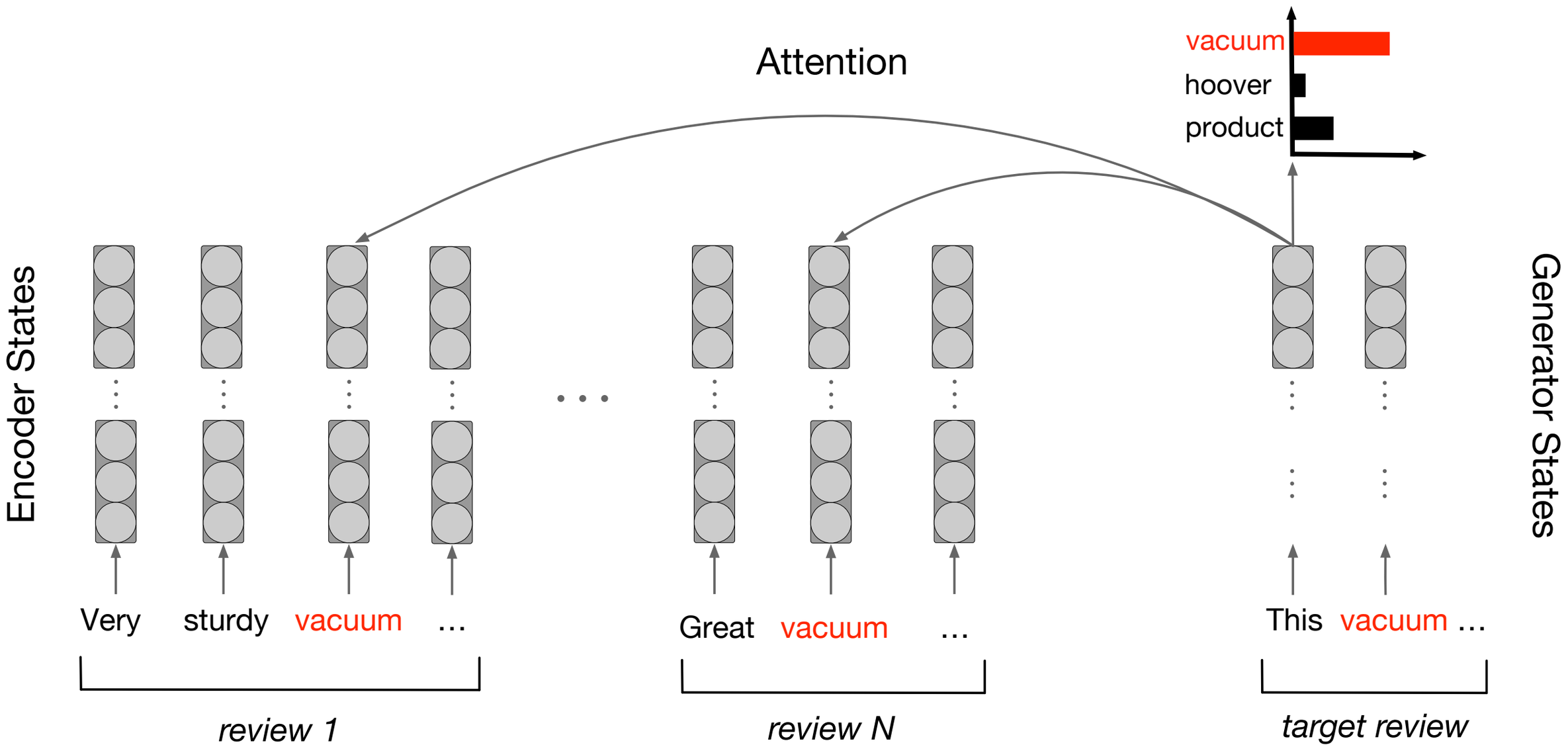
Leave-one-out



Leave-one-out



Leave-one-out



Review properties

- We also condition on **properties**
- Observation:
  - Some reviews are more like summaries
  - Some are less

Review 1

Varys

★★★☆☆

When I first got diabetes I got this. It has a lot of what we need. But later I have switched to another brand.

Unimportant content

Varys

★★★☆☆

When I first got diabetes I got this. It has a lot of what we need. But later I have switched to another brand.

Unimportant content

Varys

★★★☆☆

When I first got diabetes I got this. It has a lot of what we need. But later I have switched to another brand.

Informal writing style

Varys

When I first got diabetes I got this. It has a lot of what we need. But later I have switched to another brand.

Review 2

Jon Snow

★★★★★

These capsules are a natural alternative to other over-the-counter medications. They are easy to swallow and have a great taste. Overall, great value for money.

Introduction

Jon Snow

★★★★★

These capsules are a natural alternative to other over-the-counter medications. They are easy to swallow and have a great taste. Overall, great value for money.

Bottom line

Jon Snow

★★★★★

These capsules are a natural alternative to other over-the-counter medications. They are easy to swallow and have a great taste. Overall, great value for money.

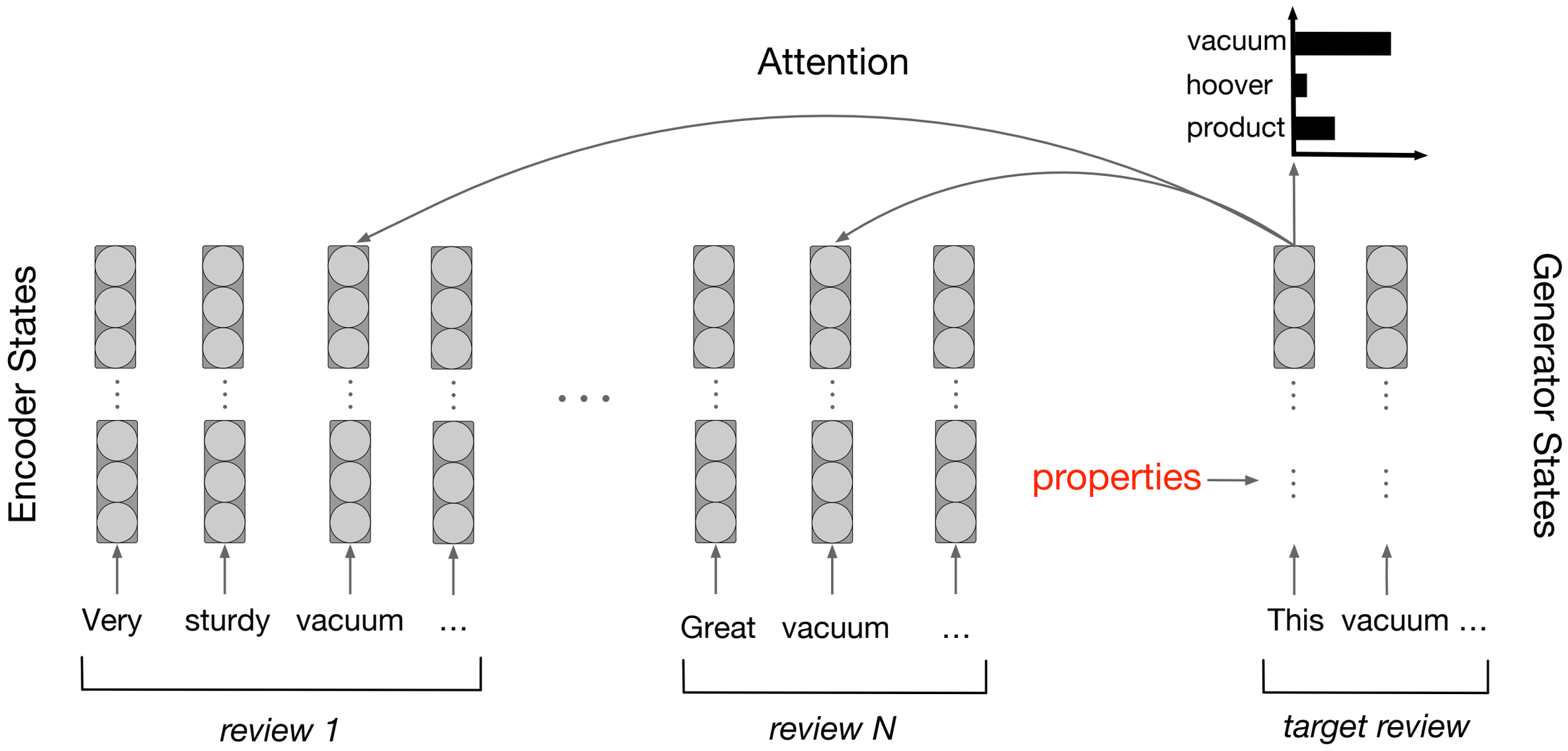
Formal writing style

Jon Snow

★★★★★

These capsules are a natural alternative to other over-the-counter medications. They are easy to swallow and have a great taste. Overall, great value for money.

Properties


Attention
vacuum
hoover
product
Encoder States
Very
sturdy
vacuum
…
review 1
Great
vacuum
…
review N
properties
This
vacuum
…
target review
Generator States

Properties

Property	Reviews	Summaries	Implementation
Information coverage	Uncommon	Common	ROUGE scores
Writing style	Informal	Formal	Pronoun counts
…	…	…	…

Oracle

- **Oracle** is used to compute **property values** based on:
  - **target review**
  - source reviews
- Has access to what needs to be predicted (target review)

Plug-in network

- At test time, want to generate **summaries**
- Have access only to **source reviews**
- Can’t use the **oracle**
- The model exposed to **a range of property values**
- Might **not know** what **property values** are needed
- Replace the **oracle** by a **trainable neural network**

Plug-in network

- Using a **handful** of summaries (~30 data-points)
- Can train the **plug-in network**
- Learns what property values lead to **generation of summaries**

Recap

- **Pre-train phase**
  - Large corpus of **unannotated reviews**
  - **CLM** with the **leave-one-out** objective
  - **Oracle** that computes **property values**
- **Fine-tune phase**
  - Replace the oracle by the **plug-in network**
  - Fine-tune it on a **handful** of **human-written summaries**

Example summary

Gold	These shoes run true to size, do a good job supporting the arch of the foot and are well-suited for exercise. They're good looking, comfortable, and the sole feels soft and cushioned. Overall they are a nice, light-weight pair of shoes and come in a variety of stylish colors.
FewSum	These running shoes are great! They fit true to size and are very comfortable to run around in. They are light weight and have great support. They run a little on the narrow side, so make sure to order a half size larger than normal.

Experiments

Setup

- Results on the **Amazon dataset** (He and McAuley, 2016)

- **Yelp** results can be found in the **paper**

Automatic evaluation

	ROUGE-1	ROUGE-2	ROUGE-L

Automatic evaluation

	ROUGE-1	ROUGE-2	ROUGE-L
Lead	27.00	4.92	14.95

Automatic evaluation

	ROUGE-1	ROUGE-2	ROUGE-L
MeanSum	26.63	4.89	17.11
Lead	27.00	4.92	14.95

Automatic evaluation

	ROUGE-1	ROUGE-2	ROUGE-L
Copycat	27.85	4.77	18.86
MeanSum	26.63	4.89	17.11
Lead	27.00	4.92	14.95

Automatic evaluation

	ROUGE-1	ROUGE-2	ROUGE-L
FewSum	**33.56**	**7.16**	**21.49**
Copycat	27.85	4.77	18.86
MeanSum	26.63	4.89	17.11
Lead	27.00	4.92	14.95

Best-worst selection

- Used Amazon Mechanical Turk (AMT)
- Workers selected the **best** and **worst** summary per **criterion**
- Range from:
  - -1 (unanimously the worst)
  - +1 (unanimously the best)

Best-worst selection

FewSum
Copycat
LexRank

Best-worst selection

	Flue.
FewSum	**0.1000**
Copycat	-0.1765
LexRank	-0.4848

Best-worst selection

	Flue.	Coher.
FewSum	**0.1000**	**0.1429**
Copycat	-0.1765	-0.5333
LexRank	-0.4848	-0.5161

Best-worst selection

	Flue.	Coher.	Non-Red.
FewSum	**0.1000**	**0.1429**	**0.1250**
Copycat	-0.1765	-0.5333	-0.2727
LexRank	-0.4848	-0.5161	-0.5862

Best-worst selection

	Flue.	Coher.	Non-Red.	Inform.
FewSum	**0.1000**	**0.1429**	**0.1250**	**0.2000**
Copycat	-0.1765	-0.5333	-0.2727	-0.7455
LexRank	-0.4848	-0.5161	-0.5862	-0.3488

Best-worst selection

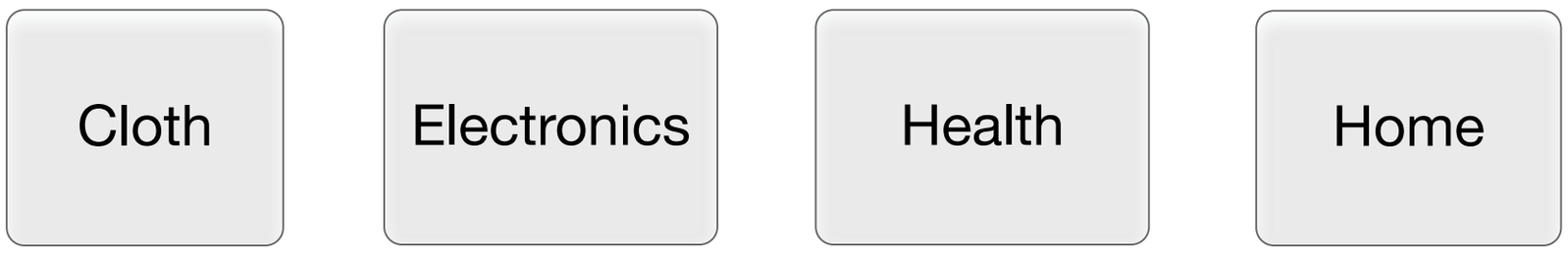
	Flue.	Coher.	Non-Red.	Inform.	Senti.
FewSum	**0.1000**	**0.1429**	**0.1250**	**0.2000**	**0.3061**
Copycat	-0.1765	-0.5333	-0.2727	-0.7455	-0.7143
LexRank	-0.4848	-0.5161	-0.5862	-0.3488	-0.0909

Cross-domain

1. We hypothesized that **fine-tuning** can be performed on **remotely related domains**

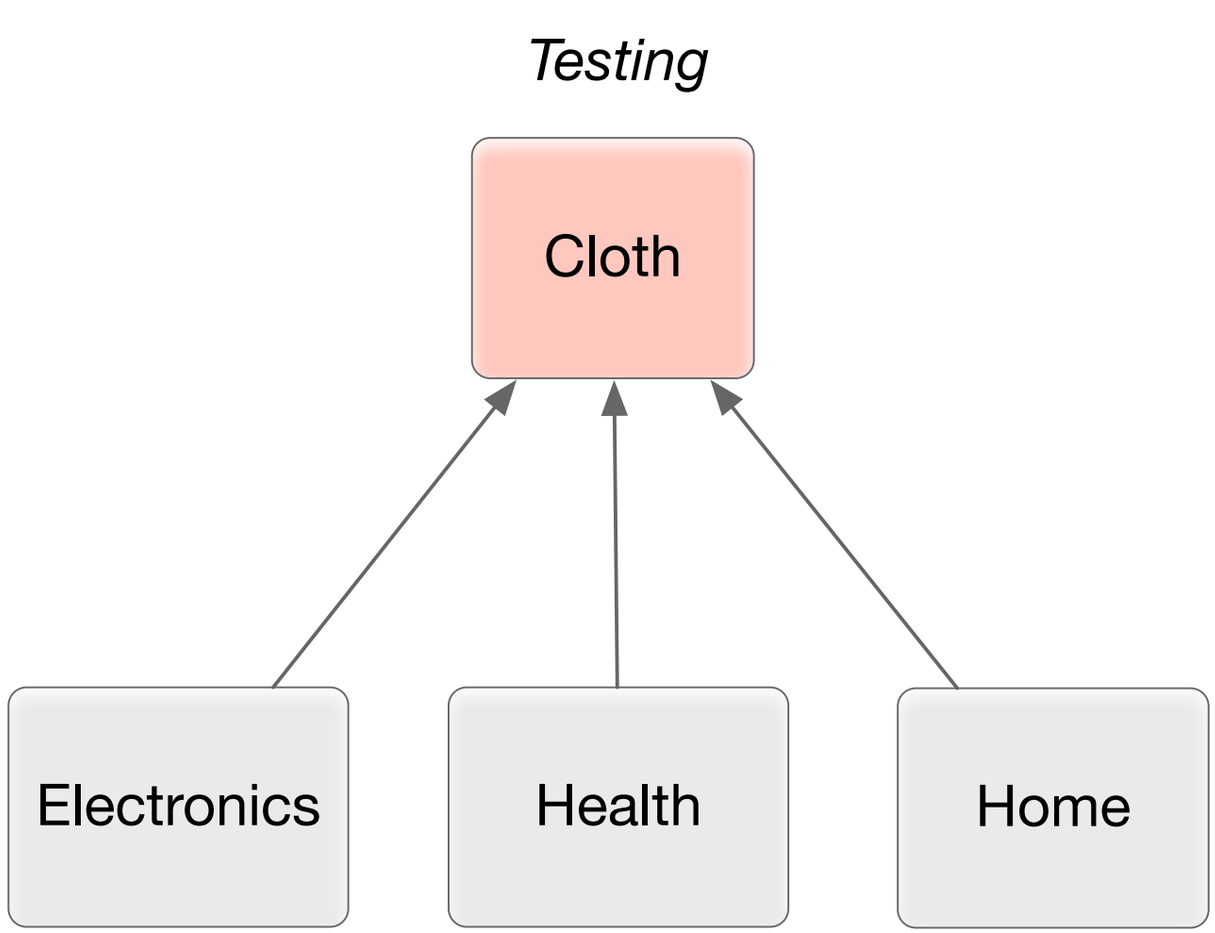
2. Plug-in network will **generalize**

Cross-domain

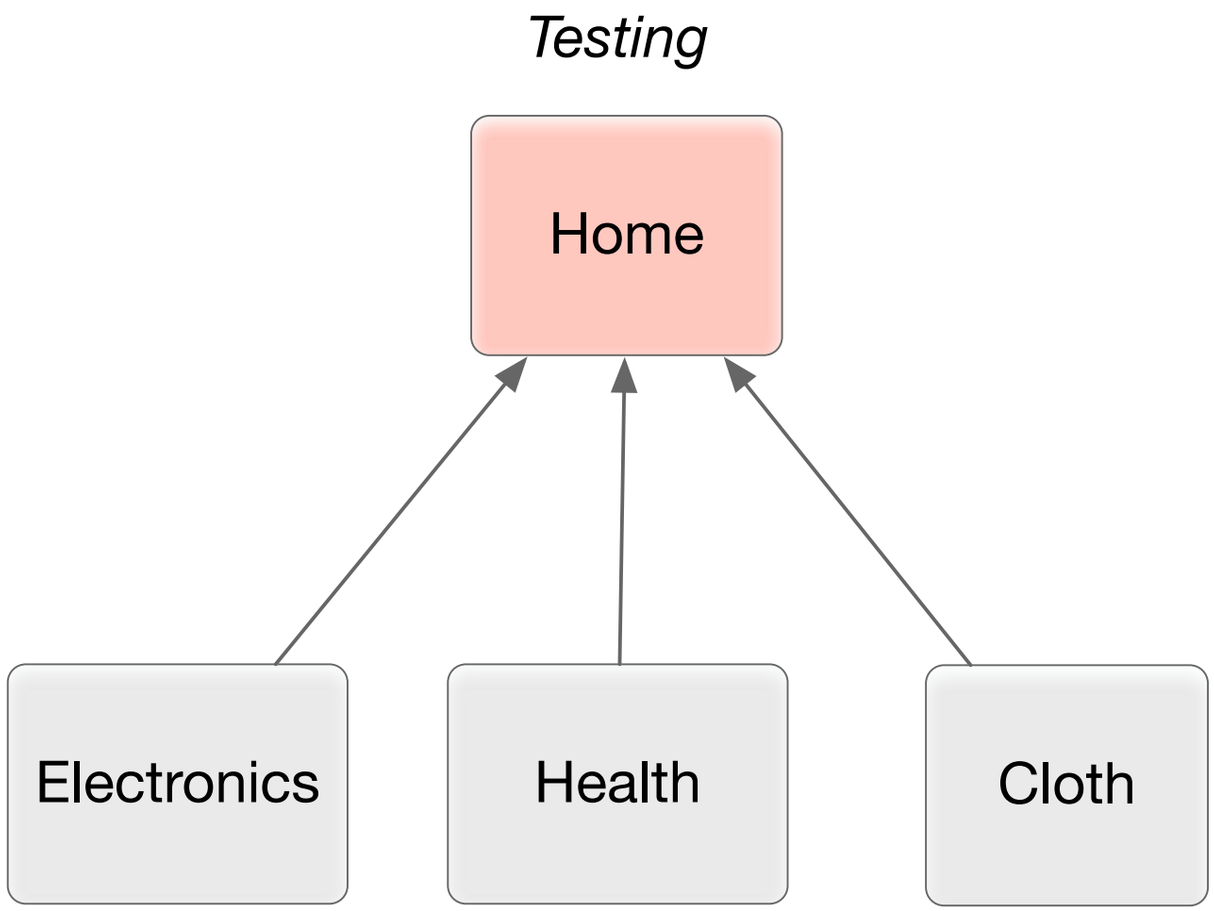


Domains

Cross-domain



Cross-domain



Cross and in-domain

	In-domain	Cross-domain
Avg. ROUGE-L	21.49	21.29

Wrap up

Conclusions

- First application of **few-shot learning** to opinion summarization
- Handful of summaries to switch the **CLM** to a **summarizer**
- Produces **informative** summaries
- Written in the **formal writing style**
- **Outperforms** all other models in **automatic** and **human evaluation**

Code and Data

https://github.com/abrazinskas/FewSum

<END>

Appendix

Content support

- Split summaries by sentences

- Asked AMT workers to assess **support** of their content to the **source reviews**

Content support

	Full (%)	Partial (%)	No (%)
FewSum	43.09	**34.14**	**22.76**
Copycat	**46.15**	27.18	26.67