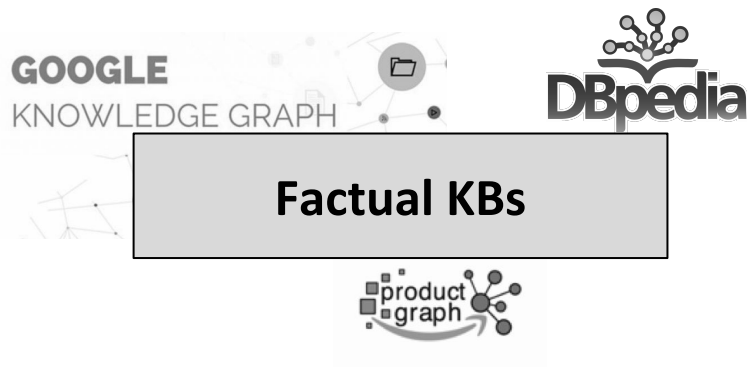# Sampo: Unsupervised Knowledge Base Construction for Opinions and Implications

Nikita Bhutani, Aaron Traylor, Chen Chen, Xiaolan Wang, Behzad Golshan, Wang-Chiew Tan

Megagon Labs

- Motivation

Factual KBs

Reviews

- encyclopedic knowledge
- real-world entities
- relations are explicitly stated

- ~~real-world entities~~ opinions
- ~~relations~~ implications
- relations NOT explicitly stated

- Example opinions and implication

Opinion modifier aspect

Review₁: The movie has everything: fast paced action, romance and complex characters.

fast paced action   complex characters

Review₂: There's some pretty good writing going on.

pretty good writing

complex characters → *implies* → pretty good writing

Use cases: Sentiment Analysis, Question Answering, Retrieval
e.g. "Show me a movie with good writing"
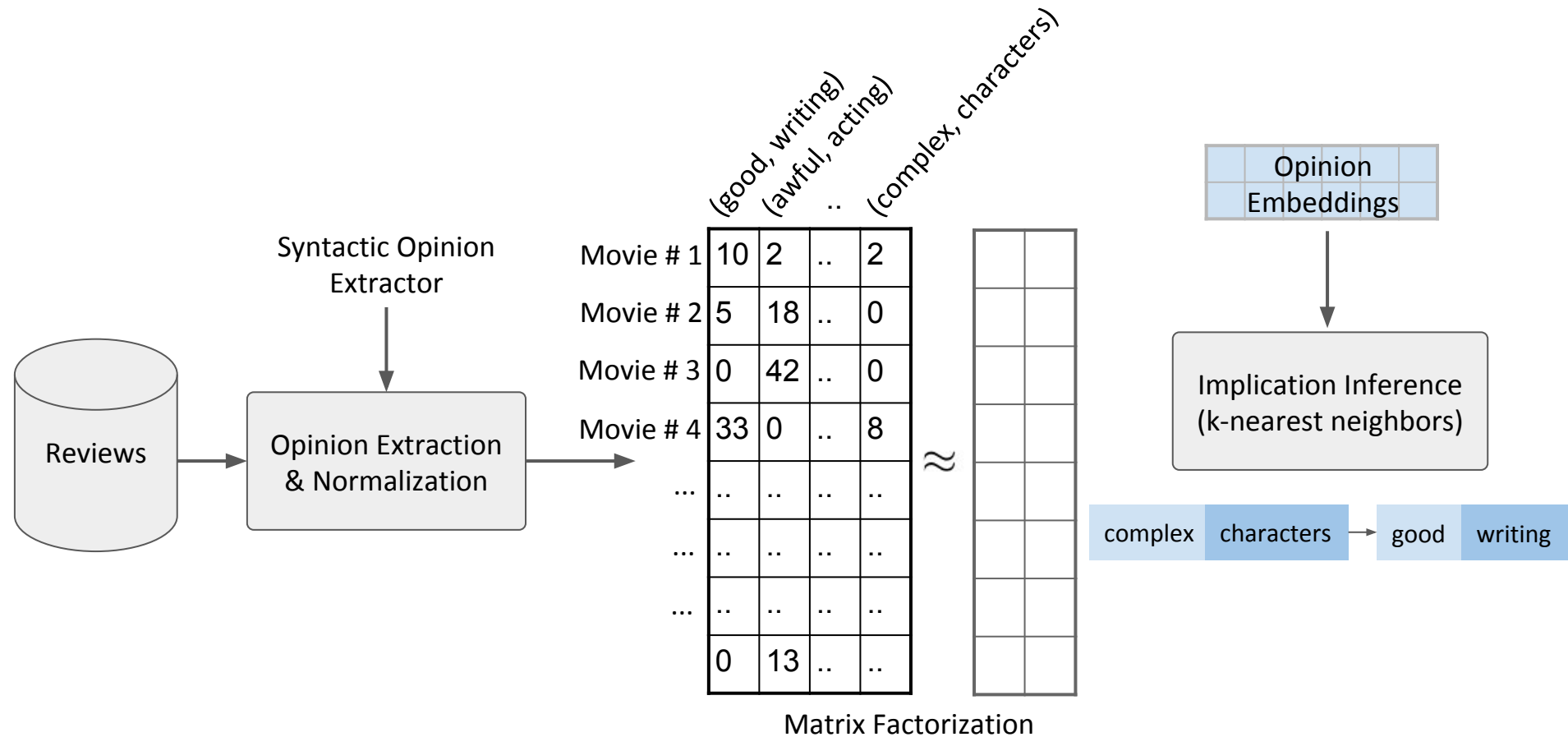
- Existing KBC methods are extractive and target knowledge explicitly stated in text
  - Implication relationships between opinions are implicit
- Supervised or semi-supervised learning models
  - Obtaining annotated data for each review domain is expensive

SAMPO

Domain-independent unsupervised KBC for opinions and implications

- Sampo

Syntactic Opinion Extractor

Reviews → Opinion Extraction & Normalization →

	(good, writing)	(awful, acting)	..	(complex, characters)
Movie # 1	10	2	..	2
Movie # 2	5	18	..	0
Movie # 3	0	42	..	0
Movie # 4	33	0	..	8
...	..	..	..	..
...	..	..	..	..
...	..	..	..	..
	0	13	..	..

≈

Matrix Factorization

Opinion Embeddings

Implication Inference
(k-nearest neighbors)

complex characters → good writing

- Experimental Results

Dataset	Best Dim	Sampo	Sampo$_{Basic}$	PMI	Univ. Schema	BERT
Movies	30	0.4926	**0.5047**	0.2548	0.0983	0.3794
Electronics	30	**0.4477**	0.4243	0.2142	0.0609	0.2796
Books	40	**0.3422**	0.2955	0.1656	0.0919	0.3073
CD Vinyl	50	**0.2692**	0.2575	0.1881	0.1702	0.2506
Hotels	30	0.3611	**0.4256**	0.0993	0.1750	0.3519
Restaurants	30	0.3869	0.4116	0.1975	0.1233	**0.4724**

MAP@5 of different KBC methods
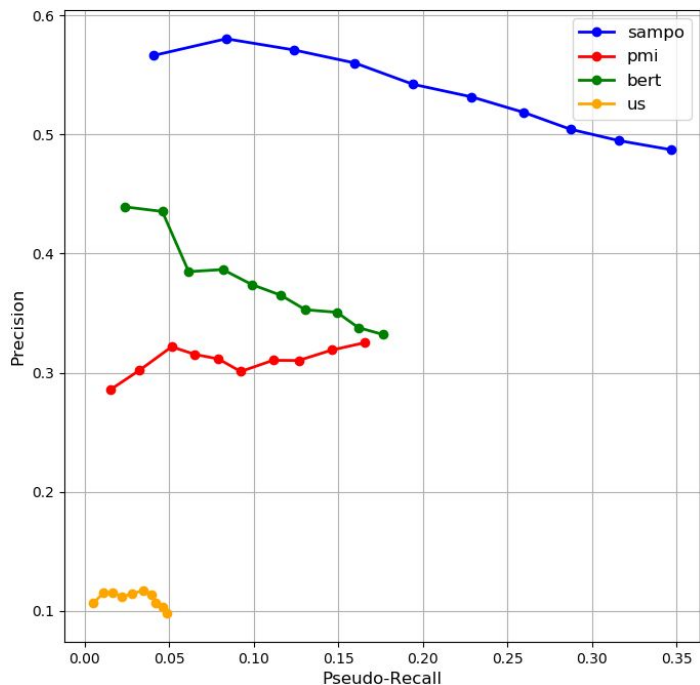
Sampo achieves highest area under the PR curve (Electronics)
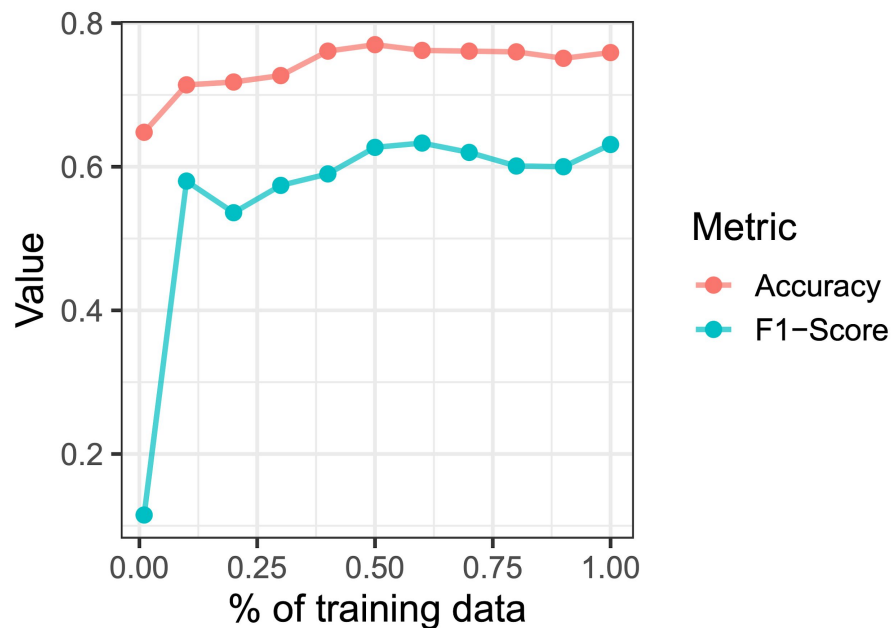
KB built using SAMPO (Electronics) when used as training data improve performance of LM-based KBC methods

- Our tool is publicly available

Sampo: Unsupervised Knowledge-Base Construction from Reviews

This repository implements Sampo, a framework for automatically building Knowledge-bases (KB) from review data. More specifically, Sampo is designed to build KBs that capture how modifier-aspect pairs extracted from reviews relate to each other. For instance in the hotel domain, *"cold coffee"* implies *"bad breakfast"* which further implies *"poor service"*.

1. Installation

You can install Sampo as a python package via the following command:

```
$ pip install -U -e .
```

You also need to download the spacy model for English via the following command:

```
$ python -m spacy download en_core_web_md
```

2. Tools and Scripts

All the scripts are located under the "sampo" folder. What follows is a description of each tool.

2.1 - Micropine

Micropine is a simple rule-based opinion extractor which extracts (modifier, aspect) pairs from reviews. You can find the set of rules used to find the modifiers and aspects here. We recommend using a stronger opinion extractor if such extractor is available for your domain of interest.

- Thank you for listening