CAIR CENTRE FOR ARTIFICIAL INTELLIGENCE RESEARCH

Knowledge Representation in Text using Tsetlin Machine

BIMAL BHATTARAI

ICT

University of Agder

Introduction

- Data is a discrete arrangement of information.
- Data plays a key role in all machine learning problems.
- Machine can extract patterns that represent a particular event, we can say that the machine has learned that information.
- Knowledge representation is used to extract meaningful latent features from data.
- The idea is that a network rids noisy input data of extraneous details as if by squeezing the information through a bottleneck, leaving only the features most relevant to general concepts.

Introduction

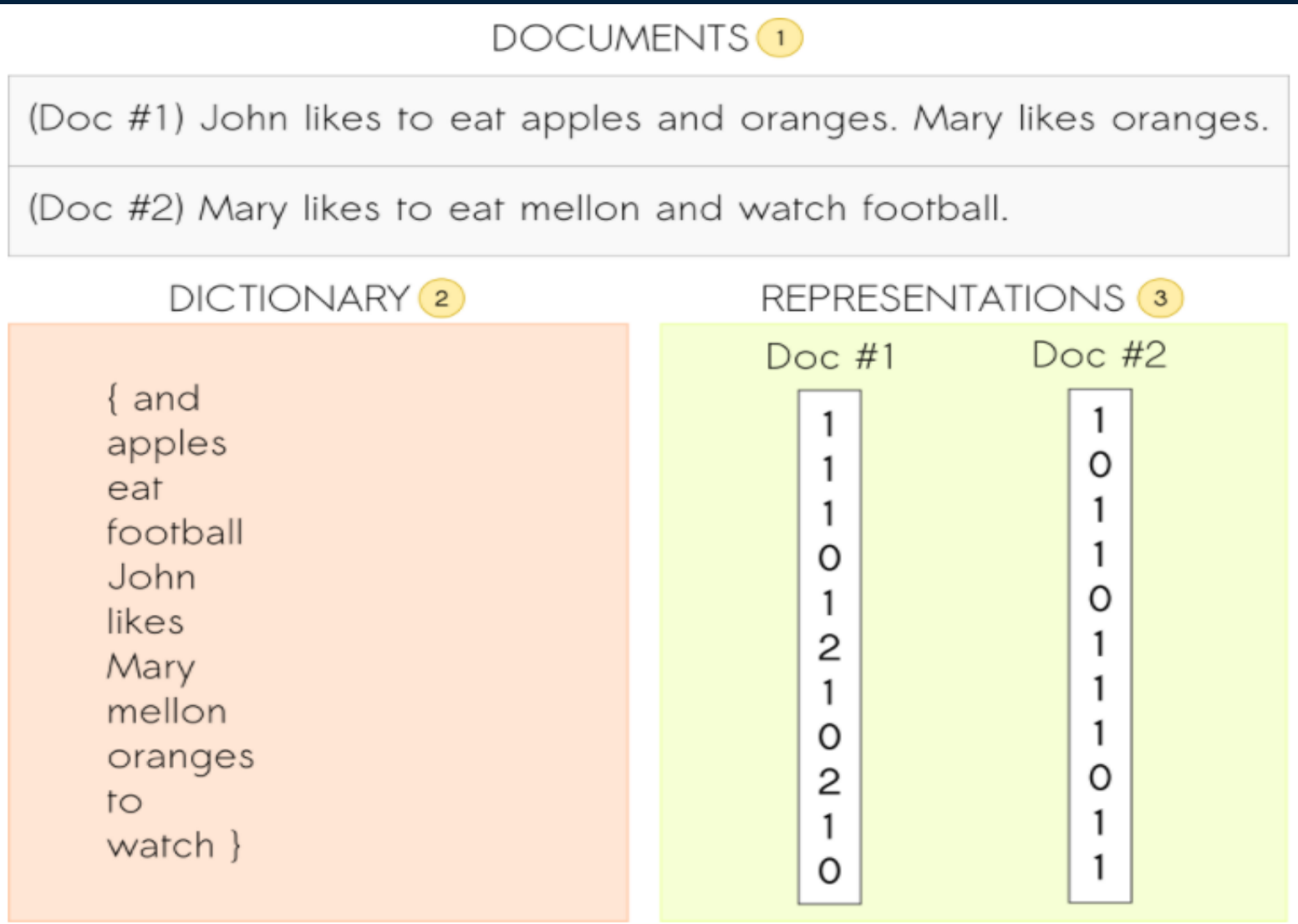


Figure 1: From a set of documents, build a dictionary containing the set of unique words, then each document is represented as a feature vector containing the count (the number of times) of each word in that document.

Introduction

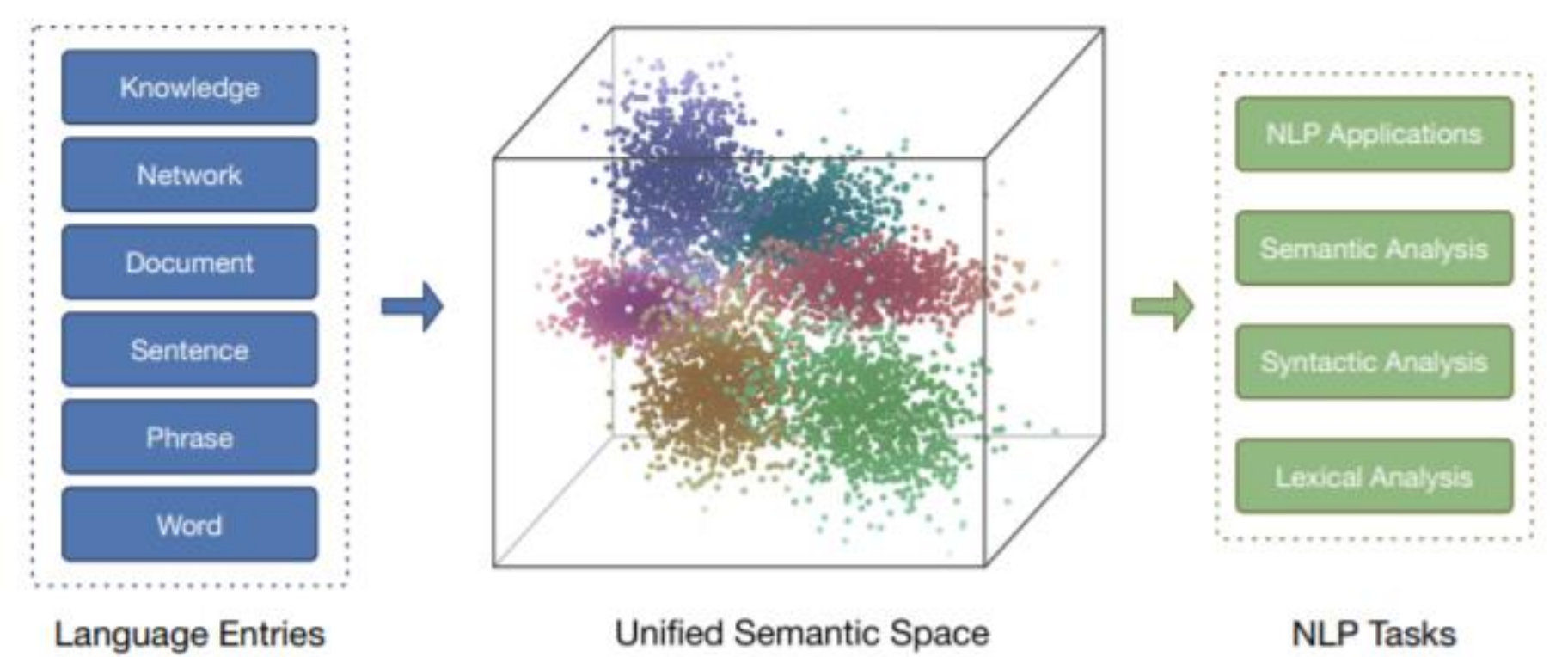


Fig. 1.2 Distributed representation can provide unified semantic space for multi-grained language entries and for multiple NLP tasks

Problem Statement

- Features describes the attributes, and it correlates with different outcomes. In case of noise, or discrepancies, the outcomes can be different.
- Most ML algorithms have superficial understanding of the data.
- For many tasks, it is impossible to know what features should be extracted.

Possible Solution

- Find robust representation of the data
- The problem with representation learning is that it’s very difficult to get representations that can solve a given problem.

Problem Example

- Problem of Polysemy:

  Sentence 1: *He sent me a **present** for my birthday.*

  Sentence 2: *There were 200 people **present** at the meeting.*

- Problem of contextual dilemma:

  Sentence : *I am a _____, and I am in a class.*

  A) Student

  B) Teacher

Tsetlin Machine

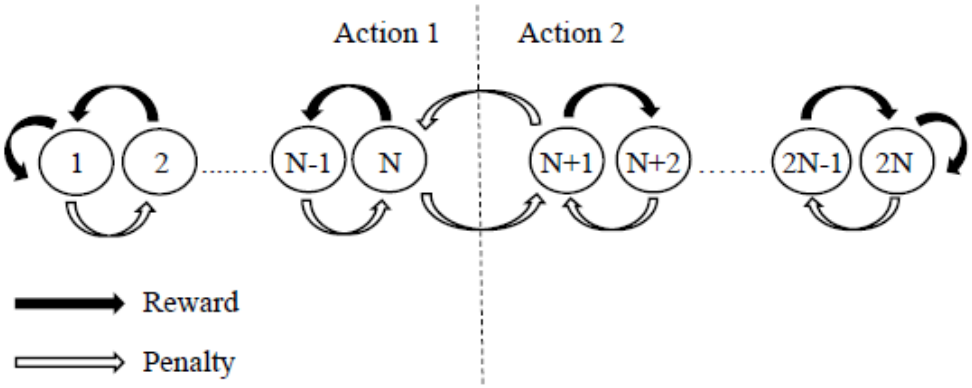
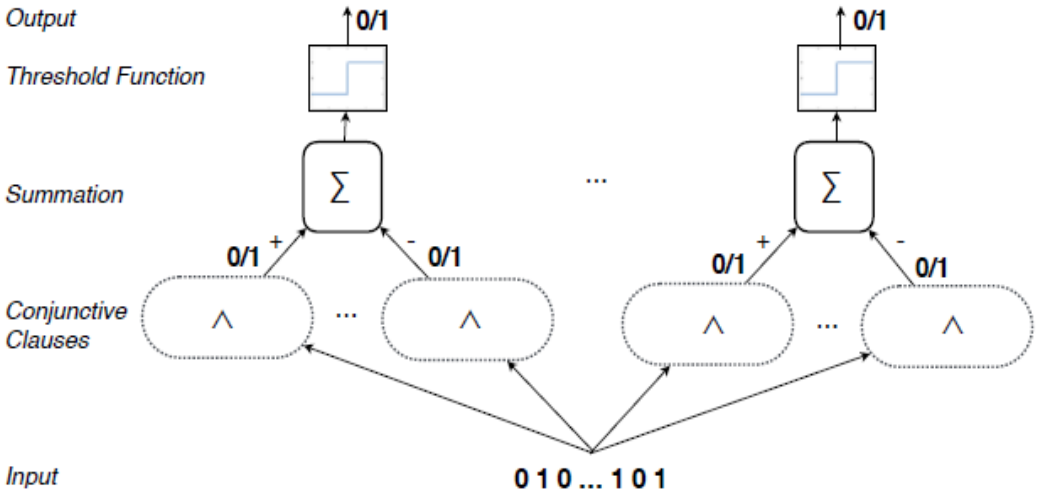
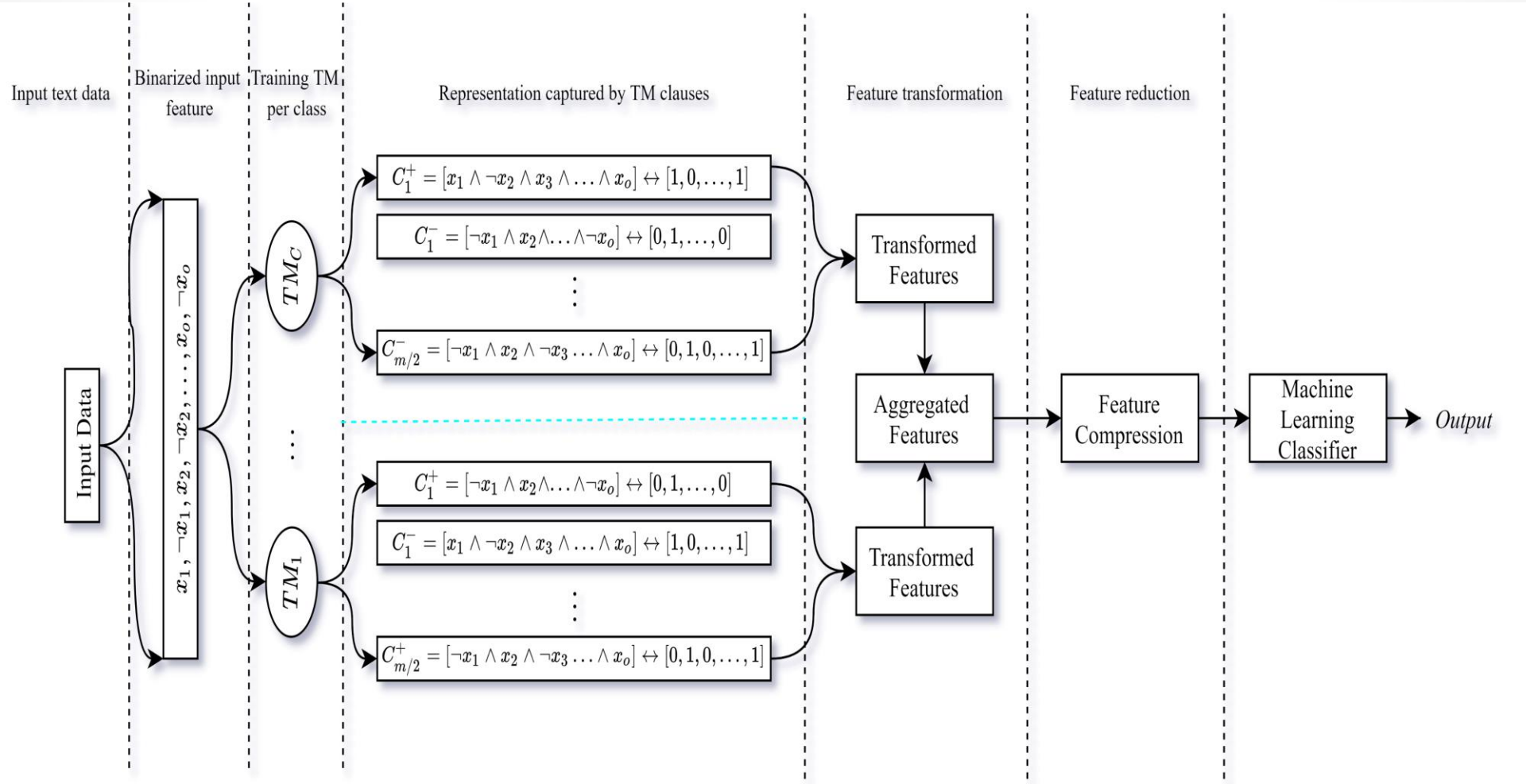


Fig. 1: Transition graph of a two-action Tsetlin Automaton.



Representation Pipeline



Representation Example

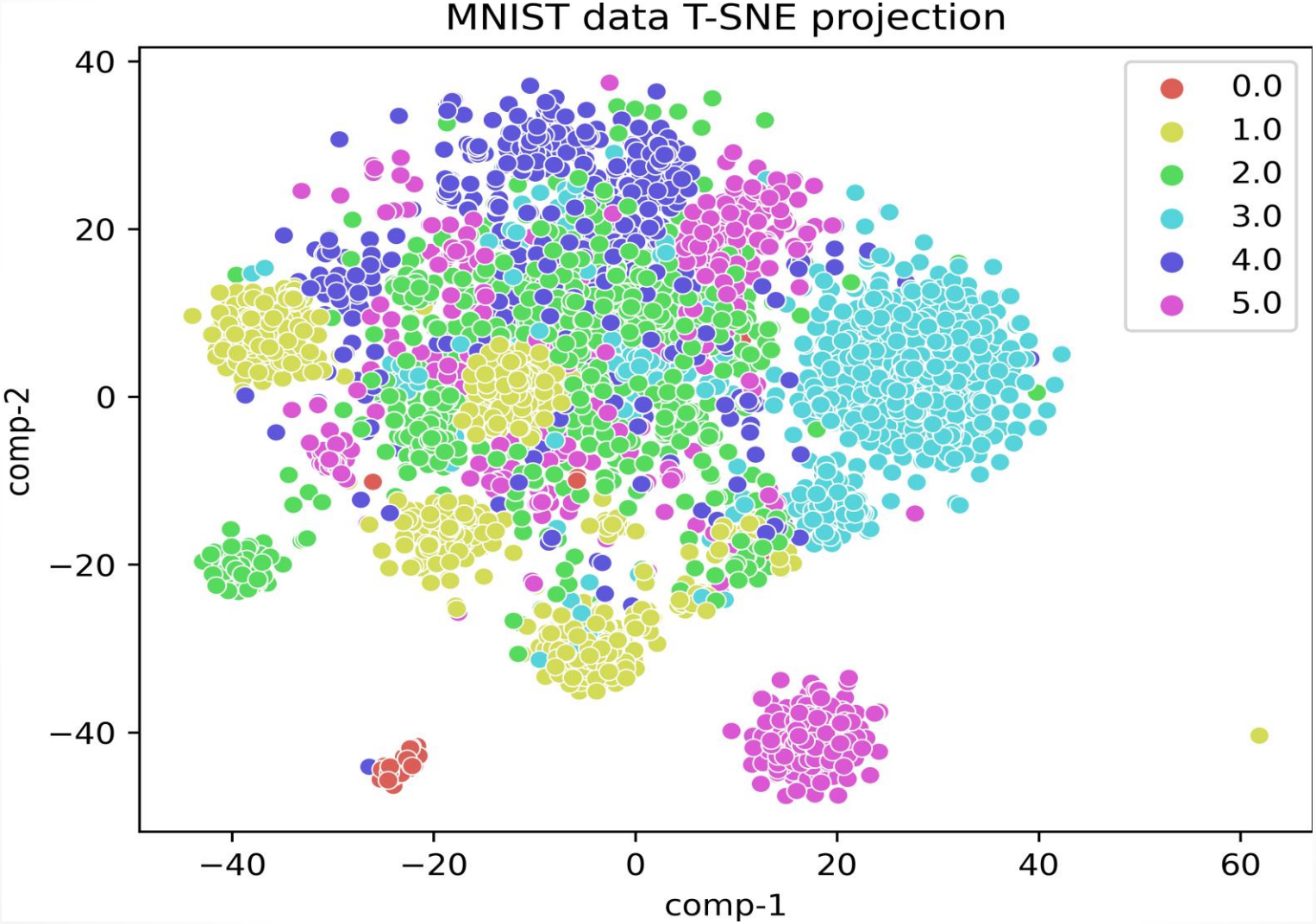
Example:
Feature – dog
Deep learning representation (embedding): *[0.35, 0.86, -0.36, ……., -0.21]*
TM representation: [*canine* ***and*** *loyal* ***and*** *domestic* **…… *and*** *bark***]**
TM Feature space: Binary (i.e., [1, 0, 1, ……, 1])

Practicality:
Class : Cancer
Deep learning representation: Embedding
If Noise is present, Prediction can be False

TM representation: Clauses Captured Features
Noise tolerant, Robust Prediction

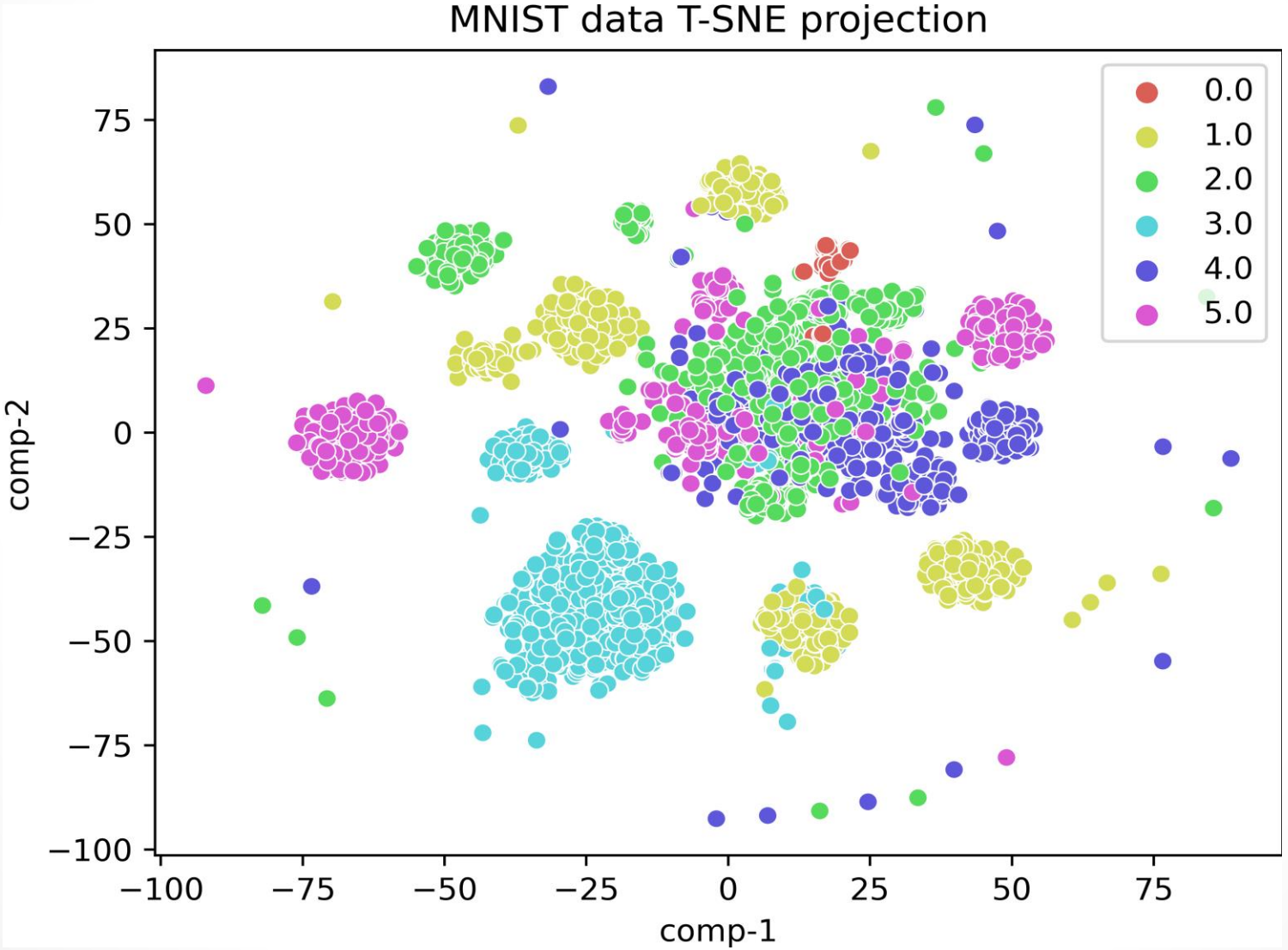
Visualization

Without knowledge representation



Visualization

With knowledge representation



Result

Table 1. Performance comparison of vanilla TM with and without using Knowledge representation.

Datasets	$TM_{vanilla}$	$TM_{representation}$		
		TM_{best}	$TM_{10\ epochs}$	$TM_{3\ epochs}$
TREC	91.6	95	**95.6**	92.2
WebKB	91.69	**93.05**	92.19	92.47
Ohsumed	46.45	**57.77**	56.56	54.58
MPQA	74.55	**87.3**	82.75	81.33
CR	80.55	**83.06**	77.76	81.48
SUBJ	86.8	88.4	89.9	**90.1**
SST-2	74.19	**79.13**	78.41	75.56
R8	95.93	**96.84**	96.71	96.29

CAIR CENTRE FOR ARTIFICIAL INTELLIGENCE RESEARCH

Thank you!