Paradigm Shift in NLP

Tianxiang Sun, Xiangyang Liu, Xipeng Qiu, Xuanjing Huang
Fudan University

txsun19@fudan.edu.cn
11 Oct 2021

https://arxiv.org/abs/2109.12575

Outline

• Introduction
• The Seven Paradigms in NLP
• Paradigm Shift in NLP Tasks
• Potential Unified Paradigms
• Conclusion
Outline

• Introduction

• The Seven Paradigms in NLP

• Paradigm Shift in NLP Tasks

• Potential Unified Paradigms

• Conclusion
What is Paradigm?

• Definition from Wikipedia
  • In science and philosophy, a **paradigm** is a distinct set of **concepts** or **thought patterns**, including theories, research methods, postulates, and standards for what constitutes legitimate contributions to a field.

• Definition in the context of NLP
  • *Paradigm is the general framework to model a class of tasks*
What is Paradigm?

• **Definition from Wikipedia**
  • In science and philosophy, a *paradigm* is a distinct set of *concepts* or *thought patterns*, including theories, research methods, postulates, and standards for what constitutes legitimate contributions to a field.

• **Definition in the context of NLP**
  • *Paradigm is the general framework to model a class of tasks*
What is Paradigm?

• Definition from Wikipedia
  • In science and philosophy, a **paradigm** is a distinct set of **concepts** or **thought patterns**, including theories, research methods, postulates, and standards for what constitutes legitimate contributions to a field.

• Definition in the context of NLP
  • *Paradigm is the general framework to model a class of tasks*

![Diagram of Paradigm and Sequence Labeling Architecture]
Paradigms, Tasks, and Models

• A Rough Illustration
Paradigms, Tasks, and Models

• A Rough Illustration
Outline

• Introduction

• The Seven Paradigms in NLP

• Paradigm Shift in NLP Tasks

• Potential Unified Paradigms

• Conclusion
The Seven Paradigms in NLP

- Seven Paradigms
  - Class
  - Matching
  - SeqLab
  - MRC
  - Seq2Seq
  - Seq2ASeq
  - (M)LM
Classification (Class)

• Paradigm

\[ y = \text{CLS}(\text{ENC}(x)). \]

• Model
  • \( \text{ENC}(\cdot) \): CNN, RNN, Transformers...
  • \( \text{CLS}(\cdot) \): (max/average/attention) pooling + MLP

• Tasks
  • Sentiment Analysis
  • Spam Detection
  • ...

![Diagram of model components with labels: Encoder, Classifier, Text, Label]
Matching

• Paradigm

\[ Y = \text{CLS}(\text{ENC}(X_a, X_b)) \]

• Model
  - \text{ENC}(\cdot) : encode the two texts separately or jointly
  - \text{CLS}(\cdot) : capture the interaction, and then prediction

• Tasks
  - Natural Language Inference
  - Similarity Regression
  - ...
**Sequence Labeling (SeqLab)**

- **Paradigm**
  \[ y_1, \cdots, y_n = \text{DEC}(\text{ENC}(x_1, \cdots, x_n)) \]

- **Model**
  - \text{ENC}(\cdot) : sequence model (RNN, Transformers…)
  - \text{DEC}(\cdot) : conditional random fields (CRF)

- **Tasks**
  - Named Entity Recognition (NER)
  - Part-Of-Speech Tagging
  - …
Machine Reading Comprehension (MRC)

**Paradigm**

\[ y_k \cdots y_{k+l} = \text{DEC} (\text{ENC}(x_p, x_q)) \]

**Model**

- **ENC(·)** : CNN, RNN, Transformers...
- **DEC(·)** : start/end position prediction

**Tasks**

- Machine Reading Comprehension
Sequence-to-Sequence (Seq2Seq)

- **Paradigm**

\[ y_1, \cdots, y_m = \text{DEC}(\text{ENC}(x_1, \cdots, x_n)) \]

- **Model**

  - \( \text{ENC}(\cdot) \) : CNN, RNN, Transformers...
  - \( \text{DEC}(\cdot) \) : CNN, RNN, Transformers...

- **Tasks**

  - Machine Translation
  - End-to-end dialogue system
  - ...
Sequence-to-Action-Sequence (Seq2ASeq)

• Paradigm

\[ A = \text{CLS}(\text{Enc}(\mathcal{X}), \mathcal{C}) \]

• Model
  
  • \( \text{Enc}(\cdot) \) : CNN, RNN, Transformers...
  
  • \( \text{CLS}(\cdot) \) : predict an action conditioned on a configuration and the input text

• Tasks
  
  • Dependency Parsing
  
  • Constituency Parsing
  
  • …
(Masked) Language Model (\textbf{(M)LM})

- **Paradigm**
  - LM: \( x_k = \text{DEC}(x_1, \ldots, x_{k-1}) \)
  - MLM: \( \bar{x} = \text{DEC}(\text{ENC}(\bar{x})) \)

- **Model**
  - \( \text{ENC}(\cdot) \) : CNN, RNN, Transformers...
  - \( \text{DEC}(\cdot) \) : simple classifier, or a auto-regressive decoder

- **Tasks**
  - Language Modeling
  - Masked Language Modeling
  - ...
Compound Paradigm

• Complicated NLP tasks can be solved by combining multiple fundamental paradigms

• An Example
  • HotpotQA

---

**Paragraph A, Return to Olympus:**
[1] Return to Olympus is the only album by the alternative rock band Malfunkshun. [2] It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990. [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

**Paragraph B, Mother Love Bone:**

**Q:** What was the former band of the member of Mother Love Bone who died just before the release of “Apple”?

**A:** Malfunkshun

**Supporting facts:** 1, 2, 4, 6, 7

---

HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering, EMNLP 2018
Compound Paradigm

- Complicated NLP tasks can be solved by combining multiple fundamental paradigms

- An Example
  - HotpotQA = Matching + MRC

---

**Paragraph A, Return to Olympus:**
[1] Return to Olympus is the only album by the alternative rock band Malfunkshun. [2] It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990. [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

**Paragraph B, Mother Love Bone:**

Q: What was the former band of the member of Mother Love Bone who died just before the release of “Apple”? A: Malfunkshun

**Supporting facts:** 1, 2, 4, 6, 7

Outline

• Introduction

• The Seven Paradigms in NLP

• Paradigm Shift in NLP Tasks

• Potential Unified Paradigms

• Conclusion
Paradigm Shift in NLP
Paradigm Shift in NLP

<table>
<thead>
<tr>
<th>Task</th>
<th>Input Output Example</th>
<th>Class</th>
<th>Matching</th>
<th>SeqLab</th>
<th>MRC</th>
<th>Seq2Seq</th>
<th>Seq2ASeq</th>
<th>(M) LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC</td>
<td>$X$, $Y$</td>
<td>Devlin et al. (2019)</td>
<td>$X,\mathcal{L}$</td>
<td>$Y \in {0, 1}$</td>
<td>Chai et al. (2020)</td>
<td>$X$</td>
<td>$y_1, \ldots, y_m$</td>
<td>Yang et al. (2018a)</td>
</tr>
<tr>
<td>NLI</td>
<td>$\mathcal{X}_a \oplus \mathcal{X}_b$, $\mathcal{Y}$</td>
<td>Devlin et al. (2019)</td>
<td>$\mathcal{X}_a, \mathcal{X}_b$, $\mathcal{Y}$</td>
<td>Chen et al. (2017b)</td>
<td>$f_{prompt}(\mathcal{X}_a, \mathcal{X}_b)$</td>
<td>$\mathcal{Y}$</td>
<td>$\mathcal{L}$</td>
<td>McCann et al. (2018)</td>
</tr>
<tr>
<td>NER</td>
<td>$X_{span}$, $\mathcal{Y}$</td>
<td>Fu et al. (2021)</td>
<td>$x_1, \ldots, x_n$, $y_1, \ldots, y_m$</td>
<td>Ma and Hovy (2016)</td>
<td>$X$, $Q_y$</td>
<td>$\mathcal{X}<em>{span}$, $\mathcal{Y}</em>{sent}$</td>
<td>Li et al. (2020)</td>
<td>$X$, $\mathcal{X}<em>{sent}$, $\mathcal{Y}</em>{sent}$</td>
</tr>
<tr>
<td>ABSA</td>
<td>$\mathcal{X}_{as}$, $\mathcal{Y}$</td>
<td>Wang et al. (2016)</td>
<td>$X$, $\mathcal{S}_{aux}$</td>
<td>Sun et al. (2019)</td>
<td>$X$, $\mathcal{Q}<em>{as}$, $Q</em>{opinion}$, $Q_{sent}$</td>
<td>$\mathcal{X}<em>{as}$, $\mathcal{X}</em>{opin}$, $\mathcal{Y}_{sent}$</td>
<td>Mao et al. (2021)</td>
<td>$X$, $\mathcal{X}<em>{as}$, $\mathcal{Y}</em>{sent}$</td>
</tr>
<tr>
<td>RE</td>
<td>$X$, $\mathcal{Y}$</td>
<td>Zeng et al. (2014)</td>
<td>$X$, $\mathcal{Q}_y$</td>
<td>Levy et al. (2017)</td>
<td>$X$, $\mathcal{Q}_{sent}$</td>
<td>$\mathcal{X}_{sent}$</td>
<td>Levy et al. (2017)</td>
<td>$X$, $\mathcal{Q}_{sent}$</td>
</tr>
<tr>
<td>Summ</td>
<td>Input Output Example</td>
<td>$(\mathcal{X}, \mathcal{S}<em>{summ}, \mathcal{Y}</em>{sent})^{m-1}_{i=0}$</td>
<td>$\mathcal{S}_{summ}$</td>
<td>Zhong et al. (2020)</td>
<td>$X_1, \ldots, X_n$</td>
<td>$y_1, \ldots, y_n \in {0, 1}$</td>
<td>Cheng and Lapata (2016)</td>
<td>$X$, $\mathcal{Q}_{summ}$</td>
</tr>
<tr>
<td>Parsing</td>
<td>Input Output Example</td>
<td>$(x_1, \ldots, x_n)$, $\mathcal{Y}$</td>
<td>Strzysz et al. (2019)</td>
<td>$X$, $\mathcal{Q}_{child}$</td>
<td>Gan et al. (2021)</td>
<td>$X$, $\mathcal{Q}_{parent}$</td>
<td>Vinyls et al. (2015)</td>
<td>$X$, $\mathcal{Q}_{child}$</td>
</tr>
</tbody>
</table>

Table 1: Paradigms shift in natural language processing tasks. TC: text classification. NLI: natural language inference. NER: named entity recognition. ABSA: aspect-based sentiment analysis. RE: relation extraction. Summ: text summarization. Parsing: syntactic/semantic parsing. $f$ and $g$ indicate pre-processing and post-processing, respectively. In $(M)$ LM, $f(\cdot)$ is usually implemented as a template and $g(\cdot)$ is a verbalizer. In parsing tasks, $g(\cdot)$ is a function that reconstructs the structured representation (e.g., dependency tree) from the output sequence. $\mathcal{L}$ means label description. $\oplus$ means concatenation. $\mathcal{X}_{as}$, $\mathcal{X}_{opin}$, $\mathcal{Y}_{sent}$ mean aspect, opinion, and sentiment, respectively. $\mathcal{S}_{aux}$ means auxiliary sentence. $\mathcal{X}_{sub}$, $\mathcal{X}_{obj}$ stand for subject entity and object entity, respectively. $\mathcal{S}_{sum}$ means candidate summary. $\mathcal{C}_t$ is configuration $t$ and $A$ is a sequence of actions. More details can be found in Section 3.
Paradigm Shift in Text Classification

• Traditional Paradigm: *Class*
• Shifted to...
  • Seq2Seq
  • Matching
  • (M)LM
Paradigm Shift in Text Classification

• Traditional Paradigm: **Class**
• Shifted to...
  • Seq2Seq
  • Matching
  • (M)LM

Convolutional Neural Networks for Sentence Classification. EMNLP 2014
Paradigm Shift in Text Classification

• Traditional Paradigm: Class
• Shifted to...
  • Seq2Seq
  • Matching
  • (M)LM

SGM: Sequence Generation Model for Multi-label Classification. COLING 2018
Paradigm Shift in Text Classification

- **Traditional Paradigm:** Class
- **Shifted to...**
  - Seq2Seq
  - Matching
  - (M)LM

Paradigm Shift in Text Classification

- **Traditional Paradigm:** Class
- **Shifted to...**
  - Seq2Seq
  - Matching
  - (M)LM

Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference. EACL 2021
Paradigm Shift in Text Classification
Paradigm Shift in NLI

• Traditional Paradigm: **Matching**
• Shifted to...
  • Class
  • Seq2Seq
  • (M)LM
Paradigm Shift in NLI

• **Traditional Paradigm:** Matching
• **Shifted to...**
  - Class
  - Seq2Seq
  - (M)LM

Enhanced LSTM for Natural Language Inference. ACL 2017
Paradigm Shift in NLI

- **Traditional Paradigm:** Matching
- **Shifted to...**
  - Class
  - Seq2Seq
  - (M)LM

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL 2019
Paradigm Shift in NLI

- **Traditional Paradigm:** Matching
- **Shifted to...**
  - Class
  - Seq2Seq
  - (M)LM

---

Paradigm Shift in NLI

- **Traditional Paradigm:** Matching
- **Shifted to...**
  - Class
  - Seq2Seq
  - (M)LM

Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference. EACL 2021
Paradigm Shift in NLI

Diagram:
- SeqLab
- Class
- Matching
- MRC
- (M)LM
- Seq2Seq
- Seq2ASeq

Arrows indicate relationships and dependencies between the concepts.
Paradigm Shift in NER

**Flat NER**
Barack Obama was born in the US

**Nested NER**
The Lincoln Memorial

**Discontinuous NER**
have much muscle pain and fatigue
Paradigm Shift in NER

• **Traditional Paradigm:**
  - SeqLab (Flat NER)
  - Class (Nested NER)
  - Seq2ASeq (Discontinuous NER)

• **Shifted to / Unified in...**
  - Class (Flat&Nested NER)
  - MRC (Flat&Nested NER)
  - Seq2Seq (All)
Paradigm Shift in NER

**Traditional Paradigm:**
- SeqLab (Flat NER)
- Class (Nested NER)
- Seq2ASeq (Discontinuous NER)

**Shifted to / Unified in...**
- Class (Flat&Nested NER)
- MRC (Flat&Nested NER)
- Seq2Seq (All)

End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. ACL 2016
Paradigm Shift in NER

- **Traditional Paradigm:**
  - SeqLab (Flat NER)
  - **Class** (Nested NER)
  - Seq2ASeq (Discontinuous NER)

- **Shifted to / Unified in...**
  - Class (Flat&Nested NER)
  - MRC (Flat&Nested NER)
  - Seq2Seq (All)

Multi-Grained Named Entity Recognition. ACL 2019
Paradigm Shift in NER

• **Traditional Paradigm:**
  - SeqLab (Flat NER)
  - Class (Nested NER)
  - **Seq2ASeq** (Discontinuous NER)

• **Shifted to / Unified in...**
  - Class (Flat&Nested NER)
  - MRC (Flat&Nested NER)
  - **Seq2Seq** (All)

An Effective Transition-based Model for Discontinuous NER. ACL 2020
Paradigm Shift in NER

- **Traditional Paradigm:**
  - SeqLab (Flat NER)
  - Class (Nested NER)
  - Seq2ASeq (Discontinuous NER)

- **Shifted to / Unified in...**
  - Class (Flat&Nested NER)
  - MRC (Flat&Nested NER)
  - Seq2Seq (All)
Paradigm Shift in NER

- **Traditional Paradigm:**
  - SeqLab (Flat NER)
  - Class (Nested NER)
  - Seq2ASeq (Discontinuous NER)

- **Shifted to / Unified in...**
  - Class (Flat&Nested NER)
  - MRC (Flat&Nested NER)
  - Seq2Seq (All)

### Matrix $(l \times l \times c)$ Labeling:

\[
\begin{array}{ccc}
\text{The} & 0 & 0 \\
\text{Lincoln} & -1 & 1 \\
\text{Memorial} & -1 & -1 \\
\end{array}
\]

\[
h_s(i) = \text{FFNN}_s(x_{si})
\]

\[
h_e(i) = \text{FFNN}_e(x_{ei})
\]

\[
r_m(i) = h_s(i)^T U_m h_e(i) + W_m(h_s(i) \oplus h_e(i)) + b_m
\]

Named Entity Recognition as Dependency Parsing. ACL 2020
Paradigm Shift in NER

• Traditional Paradigm:
  • SeqLab (Flat NER)
  • Class (Nested NER)
  • Seq2ASeq (Discontinuous NER)

• Shifted to / Unified in...
  • Class (Flat&Nested NER)
  • MRC (Flat&Nested NER)
  • Seq2Seq (All)

Barack Obama was born in the US.

<table>
<thead>
<tr>
<th>Entity</th>
<th>Natural Language Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Find locations in the text, including non-geographical locations, mountain ranges and bodies of water.</td>
</tr>
<tr>
<td>Facility</td>
<td>Find facilities in the text, including buildings, airports, highways and bridges.</td>
</tr>
<tr>
<td>Organization</td>
<td>Find organizations in the text, including companies, agencies and institutions.</td>
</tr>
</tbody>
</table>
Paradigm Shift in NER

- **Traditional Paradigm:**
  - SeqLab (Flat NER)
  - Class (Nested NER)
  - Seq2ASeq (Discontinuous NER)

- **Shifted to / Unified in...**
  - Class (Flat&Nested NER)
  - MRC (Flat&Nested NER)
  - Seq2Seq (All)

A Unified Generative Framework for Various NER Subtasks. ACL 2021
Paradigm Shift in NER

- Traditional Paradigm:
  - SeqLab (Flat NER)
  - Class (Nested NER)
  - Seq2ASeq (Discontinuous NER)

- Shifted to / Unified in:
  - Class (Flat & Nested NER)
  - MRC (Flat & Nested NER)
  - Seq2Seq (All)

A Unified Generative Framework for Various NER Subtasks. ACL 2021
Paradigm Shift in NER
Paradigm Shift in ABSA

A Unified Generative Framework for Aspect-Based Sentiment Analysis. ACL 2021
Paradigm Shift in ABSA

- **Traditional Paradigm:**
  - SeqLab (AE, OE, AOE, ...)
  - Class (ALSC...)

- **Shifted to / Unified in...**
  - Matching (ALSC)
  - MRC (All)
  - Seq2Seq (All)
  - (M)LM (All)
Paradigm Shift in ABSA

• Traditional Paradigm:
  • SeqLab (AE, OE, AOE, ...)
  • Class (ALSC...)

• Shifted to / Unified in...
  • Matching (ALSC)
  • MRC (All)
  • Seq2Seq (All)
  • (M)LM (All)

Attention-based LSTM for Aspect-level Sentiment Classification. EMNLP 2016
Paradigm Shift in ABSA

- **Traditional Paradigm:**
  - SeqLab (AE, OE, AOE, ...)
  - Class (ALSC...)

- **Shifted to / Unified in...**
  - Matching (ALSC)
  - MRC (All)
  - Seq2Seq (All)
  - (M)LM (All)

X: LOC1 is often considered the coolest area of London.
Aspect: Safety

- QA-M What do you think of the safety of LOC1? [X]
- NLI-M LOC1 - safety. [X]
- QA-B The polarity of the aspect safety of LOC1 is positive. [X]
- NLI-B LOC1 - safety - positive. [X]
Paradigm Shift in ABSA

• Traditional Paradigm:
  • SeqLab (AE, OE, AOE, ...)
  • Class (ALSC...)

• Shifted to / Unified in...
  • Matching (ALSC)
  • MRC (All)
  • Seq2Seq (All)
  • (M)LM (All)

Original training example:
  • input text: The ambience was nice, but service was not so great.
  • annotations: (ambience, nice, positive), (service, no so great, negative)

Converted training example 1:
  • query-1: Find the aspect terms in the text.
  • answer-1: ambience, service
  • query-2: Find the sentiment polarity and opinion terms for ambience in the text.
  • answer-2: (nice, positive)

Converted training example 2:
  • query-1: Find the aspect terms in the text.
  • answer-1: ambience, service
  • query-2: Find the sentiment polarity and opinion terms for service in the text.
  • answer-2: (not so great, negative)
Paradigm Shift in ABSA

- **Traditional Paradigm:**

A Joint Training Dual-MRC Framework for Aspect Based Sentiment Analysis. AAAI 2021
Paradigm Shift in ABSA

- **Traditional Paradigm:**
  - SeqLab (AE, OE, AOE, ...)
  - Class (ALSC...)
- **Shifted to / Unified in...**
  - Matching (ALSC)
  - MRC (All)
  - Seq2Seq (All)
  - (M)LM (All)

A Unified Generative Framework for Aspect-Based Sentiment Analysis. ACL 2021
Paradigm Shift in ABSA

Traditional Paradigm:
- SeqLab (AE, OE, AOE, …)
- Class (ALSC…)

Shifted to / Unified in:
- Matching (ALSC)
- MRC (All)
- Seq2Seq (All)
- (M)LM (All)

A Unified Generative Framework for Aspect-Based Sentiment Analysis. ACL 2021
Paradigm Shift in ABSA

• Traditional Paradigm:
  • SeqLab (AE, OE, AOE, ...)
  • Class (ALSC...)

• Shifted to / Unified in...
  • Matching (ALSC)
  • MRC (All)
  • Seq2Seq (All)
  • (M)LM (All)

Consistency prompt
Polarity prompt

The owners are great fun and the beer selection is worth staying for.

The owners are great fun? [MASK].

This is [MASK].
Paradigm Shift in ABSA

- Traditional Paradigm:
  - SeqLab (AE, OE, AOE, …)
  - Class (ALSC…)
  - Shifted to / Unified in…
  - Matching (ALSC)
  - MRC (All)
  - Seq2Seq (All)
  - (M)LM (All)

Paradigm Shift in ABSA
Paradigm Shift in Relation Extraction

- **Traditional Paradigm:**
  - SeqLab (entity extraction)
  - Class (relation classification)

- **Shifted to / Unified in...**
  - Seq2Seq
  - MRC
  - (M)LM
Paradigm Shift in Relation Extraction

• **Traditional Paradigm:**
  • SeqLab (entity extraction)
  • Class (relation classification)

• **Shifted to / Unified in...**
  • Seq2Seq
  • MRC
  • (M)LM

Relation Classification via Convolutional Deep Neural Network. COLING 2014
Paradigm Shift in Relation Extraction

- **Traditional Paradigm:**
  - SeqLab (entity extraction)
  - Class (relation classification)

- **Shifted to / Unified in...**
  - Seq2Seq
  - MRC
  - (M)LM

---

Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism. ACL 2018
Paradigm Shift in Relation Extraction

- **Traditional Paradigm:**
  - SeqLab (entity extraction)
  - Class (relation classification)

- **Shifted to / Unified in:**
  - Seq2Seq
  - MRC
  - (M)LM

Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism. ACL 2018
Paradigm Shift in Relation Extraction

• Traditional Paradigm:
  • SeqLab (entity extraction)
  • Class (relation classification)

• Shifted to / Unified in...
  • Seq2Seq
  • MRC (entity prediction)
  • (M)LM

<table>
<thead>
<tr>
<th>Relation</th>
<th>Question Template</th>
<th>Sentence &amp; Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>educated_at(x,y)</td>
<td>Where did x graduate from? In which university did x study? What is x’s alma mater?</td>
<td>Albert Einstein was awarded a PhD by the University of Zürich, with his dissertation titled...</td>
</tr>
<tr>
<td>occupation(x,y)</td>
<td>What did x do for a living? What is x’s job? What is the profession of x?</td>
<td>Steve Jobs was an American businessman, inventor, and industrial designer.</td>
</tr>
<tr>
<td>spouse(x,y)</td>
<td>Who is x’s spouse? Who did x marry? Who is x married to?</td>
<td>Angela Merkel’s second and current husband is quantum chemist and professor Joachim Sauer, who has largely...</td>
</tr>
</tbody>
</table>
Paradigm Shift in Relation Extraction

- **Traditional Paradigm:**
  - SeqLab (entity extraction)
  - Class (relation classification)

- **Shifted to / Unified in...**
  - Seq2Seq
  - MRC (triplet extraction)
  - (M)LM

Formulate RESUME dataset as Multi-turn QA:

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 Person: who is mentioned in the text?</td>
<td>A: $e_1$</td>
</tr>
<tr>
<td>Q2 Company: which companies did $e_1$ work for?</td>
<td>A: $e_2$</td>
</tr>
<tr>
<td>Q3 Position: what was $e_1$’s position in $e_2$?</td>
<td>A: $e_3$</td>
</tr>
<tr>
<td>Q4 Time: During which period did $e_1$ work for $e_2$ as $e_3$</td>
<td>A: $e_4$</td>
</tr>
</tbody>
</table>
Mark Twain was the father of Langdon.

[p] the **person** Langdon [p] ‘s parent was
[p] the **person** Mark Twain [p].

**Paradigm Shift in Relation Extraction**

- **Traditional Paradigm:**
  - SeqLab (entity extraction)
  - Class (relation classification)

- **Shifted to / Unified in...**
  - Seq2Seq
  - MRC (triplet extraction)
  - (M)LM

Paradigm Shift in Relation Extraction

Traditional Paradigm:
- SeqLab (entity extraction)
- Class (relation classification)

Shifted to Unified in...
- Seq2Seq
- MRC (triplet extraction)

Paradigm Shift in Relation Extraction
Paradigm Shift in Text Summarization

- **Traditional Paradigm:**
  - SeqLab (extractive)
  - Seq2Seq (abstractive)

- **Shifted to / Unified in...**
  - Matching (extractive)
  - (M)LM (abstractive)
Paradigm Shift in Text Summarization

- **Traditional Paradigm:**
  - SeqLab (extractive)
  - Seq2Seq (abstractive)

- **Shifted to / Unified in...**
  - Matching (extractive)
  - (M)LM (abstractive)

SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents. AAAI 2017
Paradigm Shift in Text Summarization

- **Traditional Paradigm:**
  - SeqLab (extractive)
  - Seq2Seq (abstractive)

- **Shifted to / Unified in...**
  - Matching (extractive)
  - (M)LM (abstractive)

Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond. CoNLL 2016
Paradigm Shift in Text Summarization

- **Traditional Paradigm:**
  - SeqLab (extractive)
  - Seq2Seq (abstractive)

- **Shifted to / Unified in...**
  - Matching (extractive)
  - (M)LM (abstractive)

Extractive Summarization as Text Matching. ACL 2020
Paradigm Shift in Text Summarization

- **Traditional Paradigm:**
  - SeqLab (extractive)
  - Seq2Seq (abstractive)

- **Shifted to / Unified in...**
  - Matching (extractive)
  - (M)LM (abstractive)

Paradigm Shift in Parsing

Dependency Parsing

I prefer the morning flight through Denver

Semantic Parsing

which country had the highest carbon emissions last year

SELECT country.name
FROM country, co2_emissions
WHERE country.id = co2_emissions.country_id
AND co2_emissions.year = 2014
ORDER BY co2_emissions.volume DESC
LIMIT 1;

Constituency Parsing

S
  NP
    Pronoun
      l
    Det
      shot
    Nominal
      Noun
        in my pajamas
  VP
    Pronoun
      l
    Verb
      shot
    Det
      Nominal
        Noun
          elephant
    PP
        an
          Nominal
            in my pajamas
        PP
          Noun
            elephant
Paradigm Shift in Parsing

• Traditional Paradigm:
  • Class (graph-based)
  • Seq2ASeq (transition-based)

• Shifted to / Unified in...
  • SeqLab
  • Seq2Seq
  • (M)LM
  • MRC
Paradigm Shift in Parsing

• Traditional Paradigm:
  • Class (graph-based)
  • Seq2ASeq (transition-based)

• Shifted to / Unified in…
  • SeqLab
  • Seq2Seq
  • (M)LM
  • MRC

Paradigm Shift in Parsing

- **Traditional Paradigm:**
  - Class (graph-based)
  - Seq2ASeq (transition-based)
- **Shifted to / Unified in...**
  - SeqLab
  - Seq2Seq
  - (M)LM
  - MRC

Paradigm Shift in Parsing

• Traditional Paradigm:
  • Class (graph-based)
  • Seq2ASeq (transition-based)

• Shifted to / Unified in…
  • SeqLab
  • Seq2Seq
  • (M)LM
  • MRC

Linearize a parsing tree:

```
John has a dog.  \rightarrow  NP  
                  |   VP
                  NNP  VBZ

John has a dog.  \rightarrow  (S (NP NNP)_{NP} (VP VBZ (NP DT NN)_{NP} )_{VP} . )_{S}
```
Paradigm Shift in Parsing
Trends of Paradigm Shift

Trends of Paradigm Shift

• More General and Flexible Paradigms are Dominating
  • Traditional: Class, SeqLab, Seq2ASeq
  • General: Matching, MRC, Seq2Seq, (M)LM

• The Impact of Pre-trained LMs
  • Formulate a NLP task as one that PLMs are good at!
Outline

• Introduction

• The Seven Paradigms in NLP

• Paradigm Shift in NLP Tasks

• Potential Unified Paradigms

• Conclusion
Why Unified Paradigm?

- **Data Efficiency**
  - Task-specific models usually required large-scale annotated data, while unified models can achieve considerable performance with much less data.

- **Generalization**
  - Unified models can easily generalize to unseen tasks.

- **Convenience**
  - Unified models are easier and cheaper to deploy and serve. They are born to be commercial black-box APIs.
Potential Unified Paradigms

- (M)LM
- Matching
- MRC
- Seq2Seq
Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference. EACL 2021
Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference. EACL 2021
• **Prompt**
  - Manually designed
  - Mined from corpora
  - Generated by paraphrasing
  - Generated by another PLM
  - Learned by gradient search/descent

• **Verbalizer**
  - Manually designed
  - Automatically searched
  - Constructed and refined with KB
• **Parameter-Efficient Tuning**
  - Only tuning prompts can match the performance of fine-tuning
  - Mixed-task inference

Matching

Matching

Entailment head
(Feed Forward Layer)

Predict

Entail
- Not entail

Science:

- Entail
- Not entail

Business:

- Entail
- Not entail

Sports:

- Entail
- Not entail

[CLS] The IAU downgrade Pluto as a dwarf planet

Textual Entailment

This is science news [EOS]

This is business news [EOS]

This is sports news [EOS]

Label descriptions

Label Description

Matching

- **Label Description**
  - Manually designed (can be the same as prompt)
  - Generated by reinforcement learning ([Chai et al.](https://arxiv.org/abs/2104.14690))

Matching

- **Label Description**
  - Manually designed (can be the same as prompt)
  - Generated by reinforcement learning ([Chai et al.](https://arxiv.org/abs/2104.14690))

- **The Entailment Model**
  - Fine-tuning a PLM on MNLI
(M)LM or Matching?

- **(M)LM**
  - [MASK] -> MLM head, instead of randomly initialized classifier
  - Require modifications of input (prompt) and output (verbalizer)
  - Pre-trained LMs can be directly used (even zero-shot)
  - Compatible with generation tasks

- **Matching**
  - [CLS] -> MNLI/NSP head, instead of randomly initialized classifier
  - Only label descriptions are required (less engineering!)
  - Contrastive learning can be applied
  - Suffer from domain adaption (due to the requirement of supervised data)
  - Only support NLU tasks
• **A Highly General Paradigm**

  A task can be solved as a MRC one as long as its input can be formulated as `[context, question, answer]`.

**Examples**

<table>
<thead>
<tr>
<th>Question</th>
<th>Context</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is a major importance of Southern California in relation to California and the US?</td>
<td>...Southern California is a major economic center for the state of California and the US...</td>
<td>major economic center</td>
</tr>
<tr>
<td>What is the translation from English to German?</td>
<td>Most of the planet is ocean water.</td>
<td>Der Großteil der Erde ist Meerwasser</td>
</tr>
<tr>
<td>What is the summary?</td>
<td>Harry Potter star Daniel Radcliffe gains access to a reported £320 million fortune...</td>
<td>Harry Potter star Daniel Radcliffe gets £320M fortune...</td>
</tr>
<tr>
<td>Hypothesis: Product and geography are what make cream skim milk work. <strong>Entailment</strong>, neutral, or contradiction?</td>
<td>Conceptually cream skimming has two basic dimensions – product and geography.</td>
<td>Entailment</td>
</tr>
<tr>
<td>Is this sentence positive or negative?</td>
<td>A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film.</td>
<td>positive</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question</th>
<th>Context</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What has something experienced?</td>
<td>Areas of the Baltic that have experienced eutrophication.</td>
<td>eutrophication</td>
</tr>
<tr>
<td>Who is the illustrator of Cycle of the Werewolf?</td>
<td>Cycle of the Werewolf is a short novel by Stephen King, featuring illustrations by comic book artist Bernie Wrightson.</td>
<td>Bernie Wrightson</td>
</tr>
<tr>
<td>What is the change in dialogue state?</td>
<td>Are there any Eritrean restaurants in town?</td>
<td>Bernie Wrightson, food: Eritrean</td>
</tr>
<tr>
<td>What is the translation from English to SQL?</td>
<td>The table has column names... Tell me what the notes are for South Australia</td>
<td>SELECT notes from table WHERE 'Current Slogan' = 'South Australia'</td>
</tr>
<tr>
<td>Who had given help?</td>
<td>Joan made sure to thank Susan for all the help she had given.</td>
<td>Susan</td>
</tr>
</tbody>
</table>

MRC

• A Highly General Paradigm
  • A task can be solved as a MRC one as long as its input can be formulated as [context, question, answer].

• MRC has been applied to many tasks...
  • entity-relation extraction, coreference resolution, entity linking, dependency parsing, dialog state tracking, event extraction, aspect-based sentiment analysis...

• How to Utilize the Power of Pre-Training?
  • All NLP tasks as open-domain QA?
  • Dense Passage Retriever (DPR) may help (REALM, RAG, ...)
Seq2Seq

- A Highly General and Flexible Paradigm
  - Suitable for complicated tasks (e.g. structured prediction, discontinuous NER, triplet extraction, etc.)

Structured prediction as translation between augmented natural languages. ICLR 2021
Seq2Seq

- A Highly General and Flexible Paradigm
  - Suitable for complicated tasks (e.g. structured prediction, discontinuous NER, triplet extraction, etc.)

- Powered by Pre-training
  - MASS, BART, T5...

- Compatible with (M)LM and MRC

- However...
  - High Latency at Inference Time (Non-autoregressive? Early exiting?)
Outline

• Introduction
• The Seven Paradigms in NLP
• Paradigm Shift in NLP Tasks
• Potential Unified Paradigms
• Conclusion
Conclusion

- (M)LM, aka prompt-based tuning, is exploding in popularity...
  - Does the power come from the pre-trained MLM head?
  - What if the classification head can be replaced with the NSP head, entailment head, or other classification/generation heads?
  - What if pre-training can also boost other paradigms?

- More attention is needed on other promising paradigms
  - **Matching**: less engineering, benefit from supervised data and contrastive learning
  - **MRC**: general, interpretable
  - **Seq2Seq**: compatibility, flexible to handle very complicated tasks
Thank You!

Any question or suggestion is welcome!

txsun19@fudan.edu.cn

https://arxiv.org/abs/2109.12575