

CSCI E-89B Introduction to Natural Language Processing

Harvard Extension School

Dmitry Kurochkin

Fall 2025
Lecture 13

Contents

1 Introduction to Machine Translation (MT)

- Overview
- Challenges in MT
- Contemporary MT Models
- BLEU Scores

2 Sequence-to-Sequence Models (Seq2Seq)

- Seq2Seq Models
- Advancements and Applications

3 Transformers for MT

- Transformer Architecture
- Applications and Impact of Transformers

4 Hybrid Models

Contents

1 Introduction to Machine Translation (MT)

- Overview
- Challenges in MT
- Contemporary MT Models
- BLEU Scores

2 Sequence-to-Sequence Models (Seq2Seq)

- Seq2Seq Models
- Advancements and Applications

3 Transformers for MT

- Transformer Architecture
- Applications and Impact of Transformers

4 Hybrid Models

Overview of Machine Translation

- **Definition:**

- ▶ Machine Translation (MT) refers to the use of software to automatically translate text or speech from one language to another with the aim of replicating human translation quality.

- **Goal:**

- ▶ To achieve translations that are not only grammatically and syntactically correct but also preserve the original text's meaning and cultural nuance.
- ▶ Foster cross-lingual communication and understanding in diverse global contexts:
 - ★ Enhance intercultural dialogue by accurately interpreting idiomatic expressions (e.g., "hit the nail on the head") and colloquial expressions (e.g., "shoot the breeze").
 - ★ Support international business and diplomacy with precise document translation.
 - ★ Facilitate access to educational resources and research across language barriers.
 - ★ Enable real-time communication in multilingual settings, such as conferences and meetings.

Historical Approaches to Machine Translation

- **Rule-Based Machine Translation (RBMT) (1950s–1980s):**

- ▶ Relies on manually crafted linguistic rules and comprehensive dictionaries to perform translation tasks.
- ▶ Effective in specific language domains (e.g., weather reports, technical manuals, legal contracts), capable of precise translations when rules are comprehensive, though challenging to scale or adapt quickly.
- ▶ Notable use during early phases, including the Georgetown-IBM experiment in 1954 (using 250 vocabulary items and 6 grammar rules).

- **Statistical Machine Translation (SMT) (1990s–2010s):**

- ▶ Employs statistical methods derived from bilingual corpora, focusing on probability-based word and phrase alignment.
- ▶ Revolutionized translation with IBM Alignment Models and Phrase-Based approaches, leveraging parallel texts like those from the European Union.
- ▶ Allows for more flexible, contextually adaptive translations than rule-based systems.

Historical Approaches to Machine Translation (Continued)

- **Neural Machine Translation (NMT) (2010s–Present):**

- ▶ Replaces SMT's multi-component pipeline with a unified model that maps whole sentences to whole sentences.
- ▶ Utilizes deep learning architectures, such as Sequence-to-Sequence (Seq2Seq) models and Transformers, to capture intricate patterns within vast datasets.
- ▶ Features end-to-end training, significantly enhancing fluency and coherence by considering complete sentence contexts.
- ▶ Continuously learns and improves with new data, enhancing translation accuracy and contextual sensitivity.
- ▶ Dominant approach due to its ability to handle complex linguistic subtleties and deliver high-quality translations in real-time applications.

Contents

1 Introduction to Machine Translation (MT)

- Overview
- Challenges in MT
- Contemporary MT Models
- BLEU Scores

2 Sequence-to-Sequence Models (Seq2Seq)

- Seq2Seq Models
- Advancements and Applications

3 Transformers for MT

- Transformer Architecture
- Applications and Impact of Transformers

4 Hybrid Models

Challenges in Machine Translation

- **Language Ambiguity:**

- ▶ Words or phrases can have multiple meanings, requiring context-specific interpretation for accurate translation.

- **Structural Differences:**

- ▶ Diverse grammar and syntax across languages make direct translation difficult, necessitating advanced linguistic processing.

- **Idiomatic Expressions:**

- ▶ Idioms often don't translate directly, requiring an understanding of their underlying meanings to convey accurately.

- **Context Preservation:**

- ▶ Maintaining meaning and nuance over long passages is complex, especially in ensuring coherence and consistency.

Contents

1 Introduction to Machine Translation (MT)

- Overview
- Challenges in MT
- Contemporary MT Models
- BLEU Scores

2 Sequence-to-Sequence Models (Seq2Seq)

- Seq2Seq Models
- Advancements and Applications

3 Transformers for MT

- Transformer Architecture
- Applications and Impact of Transformers

4 Hybrid Models

Contemporary Models of Machine Translation

- **Neural Machine Translation (NMT):**

- ▶ An overarching framework involving deep learning for translating languages with continuous improvement and adaptation.
- ▶ Includes various architectures such as Seq2Seq and Transformers.

- **Sequence-to-Sequence Models (Seq2Seq):**

- ▶ The first successful application of neural networks for MT using RNNs and attention mechanisms.
- ▶ Led to innovations like the Transformer model.

- **Transformers:**

- ▶ Advances beyond Seq2Seq with self-attention and parallel processing for superior translation quality.
- ▶ Forms the foundation for modern models like Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT).

- **Hybrid Models:**

- ▶ Enhance performance by integrating elements from neural and statistical approaches.

Contents

1 Introduction to Machine Translation (MT)

- Overview
- Challenges in MT
- Contemporary MT Models
- BLEU Scores

2 Sequence-to-Sequence Models (Seq2Seq)

- Seq2Seq Models
- Advancements and Applications

3 Transformers for MT

- Transformer Architecture
- Applications and Impact of Transformers

4 Hybrid Models

Understanding BLEU Scores in Machine Translation

• Definition:

- ▶ BLEU (Bilingual Evaluation Understudy) measures the quality of machine-translated text by comparing it to one or more human reference translations.

• Core Components:

▶ N-gram Precision:

- ★ Measures how many n-grams in the candidate translation match with the reference translations.
- ★ $\text{Precision}_n = \frac{\sum \text{Matches}(n\text{-gram})}{\sum \text{Total}(n\text{-gram})}$

▶ Brevity Penalty (BP):

- ★ Ensures score is not biased towards shorter translations.
- ★ Calculated as:

$$\text{BP} = \begin{cases} 1, & \text{if } \text{Length}_{\text{candidate}} \geq \text{Length}_{\text{reference}} \\ e^{1 - \frac{\text{Length}_{\text{reference}}}{\text{Length}_{\text{candidate}}}}, & \text{otherwise} \end{cases}$$

• Score Calculation:

- ▶ BLEU score = $\text{BP} \times \exp \left(\sum_{n=1}^N w_n \ln \text{Precision}_n \right)$
- ▶ Typically averages over n-grams up to $N = 4$ (unigram to four-gram).
- ▶ w_n are weights, often set equally (e.g., $w_n = \frac{1}{4}$).

Understanding BLEU Scores in Machine Translation

- **Advantages:**

- ▶ Standardized, automatic, reproducible comparison metric for translation systems.

- **Limitations:**

- ▶ Does not capture semantic meaning — reliant on exact matches.
- ▶ May not grab appropriate synonyms or varied but valid expressions.
- ▶ Struggles with highly inflected languages or flexible word order, where exact matches are less likely even for good translations.

- **Applications:**

- ▶ Benchmarking performance of machine translation models in academia and industry.

Contents

1 Introduction to Machine Translation (MT)

- Overview
- Challenges in MT
- Contemporary MT Models
- BLEU Scores

2 Sequence-to-Sequence Models (Seq2Seq)

- Seq2Seq Models
- Advancements and Applications

3 Transformers for MT

- Transformer Architecture
- Applications and Impact of Transformers

4 Hybrid Models

Contents

1 Introduction to Machine Translation (MT)

- Overview
- Challenges in MT
- Contemporary MT Models
- BLEU Scores

2 Sequence-to-Sequence Models (Seq2Seq)

- Seq2Seq Models
- Advancements and Applications

3 Transformers for MT

- Transformer Architecture
- Applications and Impact of Transformers

4 Hybrid Models

Sequence-to-Sequence Models (Seq2Seq)

- **Overview:**

- ▶ Transform input sequences into output sequences; foundational in Neural Machine Translation (NMT).
- ▶ Developed to surpass limitations of statistical and rule-based methods.

- **Architecture:**

- ▶ **Encoder:** Converts input into a fixed-length context vector, initially facing the "bottleneck" issue.
- ▶ **Decoder:** Generates output sequences; uses context vector autoregressively.
- ▶ **Attention Mechanism:** Dynamically attends to input, addressing bottleneck limitations by focusing on relevant parts.

Contents

1 Introduction to Machine Translation (MT)

- Overview
- Challenges in MT
- Contemporary MT Models
- BLEU Scores

2 Sequence-to-Sequence Models (Seq2Seq)

- Seq2Seq Models
- Advancements and Applications

3 Transformers for MT

- Transformer Architecture
- Applications and Impact of Transformers

4 Hybrid Models

Seq2Seq: Advancements and Applications

- **Key Advantages:**

- ▶ **Contextual Understanding:** Enhanced handling of long sequences with attention.
- ▶ **Flexibility:** Suitable for diverse NLP tasks beyond translation, like summarization and chatbots.

- **History and Evolution:**

- ▶ Rooted in information theory, Seq2Seq became foundational in machine translation and NLP.
- ▶ Developed in the early 2010s. Key papers in 2014 by Sutskever et al. and others pioneered its use with LSTMs.
- ▶ Enabled advancements like Transformers, enhancing efficiency and parallelism.

- **Broader Applications:**

- ▶ Widely applicable for complex sequence mapping tasks, including advanced symbolic computation, such as symbolic integration, and sophisticated AI-driven conversational models.

Contents

1 Introduction to Machine Translation (MT)

- Overview
- Challenges in MT
- Contemporary MT Models
- BLEU Scores

2 Sequence-to-Sequence Models (Seq2Seq)

- Seq2Seq Models
- Advancements and Applications

3 Transformers for MT

- Transformer Architecture
- Applications and Impact of Transformers

4 Hybrid Models

Contents

1 Introduction to Machine Translation (MT)

- Overview
- Challenges in MT
- Contemporary MT Models
- BLEU Scores

2 Sequence-to-Sequence Models (Seq2Seq)

- Seq2Seq Models
- Advancements and Applications

3 Transformers for MT

- **Transformer Architecture**
- Applications and Impact of Transformers

4 Hybrid Models

Transformer Architecture

- **Introduction:**

- ▶ Introduced in the seminal paper "Attention is All You Need" (2017) by Vaswani et al.
- ▶ Revolutionary move from sequential processing (RNNs/LSTMs) to fully-attentional architecture.

- **Core Components:**

- ▶ **Self-Attention Mechanism:** Allows the model to weigh the significance of different words, capturing long-range dependencies with parallel processing capabilities.
- ▶ **Encoder-Decoder Structure:** Both encoder and decoder employ attention layers, enhancing contextual understanding.
- ▶ **Positional Encoding:** Introduces information about token sequence order within the input.

- **Key Advantages:**

- ▶ **Scalability:** Highly parallelizable, leading to efficient training and inference workflows.
- ▶ **Flexibility:** Effective for a broad range of NLP tasks due to its modular architecture.

Contents

1 Introduction to Machine Translation (MT)

- Overview
- Challenges in MT
- Contemporary MT Models
- BLEU Scores

2 Sequence-to-Sequence Models (Seq2Seq)

- Seq2Seq Models
- Advancements and Applications

3 Transformers for MT

- Transformer Architecture
- Applications and Impact of Transformers

4 Hybrid Models

Applications and Impact of Transformers

- **Transformational Impact:**

- ▶ Set new benchmarks for NLP performance, facilitating advancements in machine translation, summarization, and sentiment analysis.
- ▶ Replaced RNNs/LSTMs as the dominant architecture in many NLP applications.

- **Broad Applications:**

- ▶ Forms the backbone of state-of-the-art models like BERT, GPT, T5, and more, impacting tasks from language modeling to general AI.
- ▶ Utilized in domains beyond NLP, including image processing with Vision Transformers (ViTs), and protein folding prediction.

- **Future Directions:**

- ▶ Ongoing research focused on improving efficiency, reducing resource usage, and expanding Transformers' adaptability to multimodal tasks.
- ▶ Transformers inspire new architectures emphasizing efficiency and scalability in various computational fields.

Contents

1 Introduction to Machine Translation (MT)

- Overview
- Challenges in MT
- Contemporary MT Models
- BLEU Scores

2 Sequence-to-Sequence Models (Seq2Seq)

- Seq2Seq Models
- Advancements and Applications

3 Transformers for MT

- Transformer Architecture
- Applications and Impact of Transformers

4 Hybrid Models

Hybrid Models in Machine Translation

- **Definition and Purpose:**

- ▶ Hybrid models integrate multiple MT approaches, leveraging the strengths of neural, statistical, and sometimes rule-based systems.
- ▶ Aimed at enhancing translation accuracy, fluency, and domain adaptability.

- **Key Components:**

- ▶ **Neural Networks:** Provide contextual understanding and adaptability, handling fluent and natural translations.
- ▶ **Statistical Methods:** Offer robust alignment and probability-based translations, especially useful for phrase alignment.
- ▶ **Rule-Based Elements:** Apply for specific language rules and syntax corrections to improve grammatical precision.

Hybrid Models in Machine Translation

- **Advantages:**

- ▶ **Improved Accuracy:** Combines the beneficial aspects of each approach to minimize weaknesses.
- ▶ **Domain Flexibility:** Adapts easily to specialized vocabulary and structure in varied applications.

- **Challenges:**

- ▶ **Complexity in Integration:** Blending different systems requires careful design to ensure compatibility and optimal performance.
- ▶ **Resource Intensive:** May demand more computational resources and expertise in multiple techniques.

- **Applications:**

- ▶ Commonly used in enterprise solutions requiring high accuracy across diverse languages and domains, such as legal and medical translations.

Famous Hybrid Models in Machine Translation

- **SYSTRAN:**

- ▶ **Background:** Founded in 1968, initially based on rule-based translation. Evolved by incorporating statistical methods in 2010 and neural methods subsequently.
- ▶ **Key Feature:** Combines linguistic rules with neural networks to enhance translation fluency and accuracy, especially in domain-specific contexts.

- **Omniscien (Asia Online):**

- ▶ **Approach:** Originally founded as Asia Online, rebranded to Omniscien Technologies in 2016. Integrates statistical and neural approaches for robust phrase alignment and contextual translation.
- ▶ **Application:** Widely used for large-scale enterprise solutions demanding high-quality multilingual outputs.

Famous Hybrid Models in Machine Translation (Continued)

- **Microsoft Translator:**

- ▶ **Integration Strategy:** Introduced in 2007, combines neural networks with statistical components to improve translation precision and context relevance.
- ▶ **Innovation:** Leverages AI advancements for real-time translation across numerous languages and platforms, with significant updates in 2018 to include neural machine translation as the default.

- **Custom Enterprise Solutions:**

- ▶ **Example:** IBM's Watson Language Translator, which emerged with hybrid capabilities in the mid-2010s, integrating proprietary rules with neural translation models for specific industry applications.
- ▶ **Specific Use Cases:** Designed to cater to specialized needs in sectors like healthcare and legal, where terminology consistency is crucial.

Advantages of Transformers

- **Scalability and Performance:**
 - ▶ Parallelization allows for efficient training and fast inference.
 - ▶ Superior speed due to handling full sequences concurrently.
- **State-of-the-Art Accuracy and Fluency:**
 - ▶ Achieves high benchmarks in NLP tasks with robust contextual understanding.
 - ▶ Consistently outperforms older models in translation accuracy.
- **Flexibility Across Tasks:**
 - ▶ Unified architecture suited for various NLP tasks like summarization and text generation.
- **Reduced Feature Engineering:**
 - ▶ Learns directly from large datasets, minimizing the need for handcrafted features.
- **Continuous Improvement:**
 - ▶ Models like BERT, GPT, and T5 showcase ongoing advancements.

Remaining Significance of Hybrid Models

- **Domain-Specific Applications:**

- ▶ Effective where adherence to domain-specific grammar and terminology is critical.

- **Resource-Constrained Environments:**

- ▶ Utilizes existing linguistic and statistical resources when robust training data for Transformers is sparse.

- **Integration Flexibility:**

- ▶ Hybrid models can tailor translations to specific industry needs by integrating pre-defined rules with neural approaches.

- **Customization Potential:**

- ▶ Allows for specialized customizations and optimizations based on specific application requirements.