



Online Ethics Center
FOR ENGINEERING AND SCIENCE

Machine Translation: Analyzing Gender

Description

This study published by Gendered Innovations discusses gender stereotypes that are perpetuated by machine translation programs that cannot use the context of a sentence or phrase to determine if the actor is male or female.

Abstract

Body

The Challenge

Machine translation (MT) becomes increasingly important in a global world. Although error rates are still high, MT system accuracies are improving incrementally. Some errors in current systems, however, are based on fundamental technological challenges that require non-incremental solutions. One such problem is related to gender: State-of-the-art translation systems like Google Translate or Systran massively overuse masculine pronouns (he, him) even where the text specifically refers to a woman (Minkov et al., 2007). The result is an unacceptable infidelity of the resulting translations and perpetuation of gender bias.

Method: Analyzing Gender

The reliance on a "masculine default" in modern machine translation systems results from current systems that do not determine the gender of each person mentioned in a text. Instead, the translation is produced by finding all the possible matches for a given phrase in large collections of bilingual texts, and then choosing a match based on factors such as its frequency in large text "corpora" (or bodies of text). Masculine pronouns are over-represented in the large text corpora that modern systems are trained on, resulting in over-use in translations. In July 2012 the Gendered Innovations project convened a workshop to discuss potential solutions. Improving feminine-masculine pronoun balance in these corpora, for example, would still not fix the problem, since it will simply cause both women and men to be randomly referred to with the wrong gender. Instead, it is crucial to develop algorithms that explicitly determine the gender of each person mentioned in text and use this computed gender to inform the translation. Such algorithms could avoid the masculine default and also increase the quality of translation overall

Gendered Innovations:

1. **Studying the Male Default in Machine Translation**
2. **Detecting the Gender of Entities to Improve Translation Algorithms (research in progress)**
3. **Integrating Gender Analysis into the Engineering Curriculum**

ExternalURL

<http://genderedinnovations.stanford.edu/case-studies/nlp.html#tabs-2>

Rights

Use of Materials on the OEC

Resource Type

Instructor Materials

Topics

Artificial Intelligence and Robotics

Discipline(s)

Computer, Math, and Physical Sciences

Computer Sciences