

Conference on ENTERprise Information Systems / International Conference on Project  
MANagement / Conference on Health and Social Care Information Systems and Technologies,  
CENTERIS / ProjMAN / HCist 2015 October 7-9, 2015

## Neural-based machine translation for medical text domain. Based on European Medicines Agency leaflet texts

Krzysztof Wołk\*, Krzysztof Marasek

*Department of Multimedia, Polish Japanese Academy of Information Technology, Koszykowa 86, 02-008 Warsaw, Poland*

---

### Abstract

The quality of machine translation is rapidly evolving. Today one can find several machine translation systems on the web that provide reasonable translations, although the systems are not perfect. In some specific domains, the quality may decrease. A recently proposed approach to this domain is neural machine translation. It aims at building a jointly-tuned single neural network that maximizes translation performance, a very different approach from traditional statistical machine translation. Recently proposed neural machine translation models often belong to the encoder-decoder family in which a source sentence is encoded into a fixed length vector that is, in turn, decoded to generate a translation. The present research examines the effects of different training methods on a Polish-English Machine Translation system used for medical data. The European Medicines Agency parallel text corpus was used as the basis for training of neural and statistical network-based translation systems. The main machine translation evaluation metrics have also been used in analysis of the systems. A comparison and implementation of a real-time medical translator is the main focus of our experiments.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of SciKA - Association for Promotion and Dissemination of Scientific Knowledge

**Keywords:** machine translation; statistical machine translation; nlp; text processing.

---

### 1. Introduction

Machine Translation (MT) is a computer's translation of text with no human assistance. MT systems have no knowledge of language rules. Instead, they translate by analyzing large amounts of text in language pairs. They can be trained for specific domains or applications using additional data germane to a selected domain. MT systems

---

\* Corresponding author. Tel.: +48 606 918 623.

E-mail address: [kwolk@pja.edu.pl](mailto:kwolk@pja.edu.pl)

typically deliver translations that sound fluent, although they tend to be less consistent than human translations. Statistical machine translation (SMT) has rapidly evolved in recent years. However, existing SMT systems are far from perfect, and their quality decreases significantly in specific domains. The scientific community has been engaged in SMT research. Among the greatest advantages of statistical machine translation is that perfect translations are not required for most applications [1]. Users are generally interested in obtaining a rough idea of a text's topic or what it means. However, some applications require much more than this. For example, the beauty and correctness of writing may not be important in the medical field, but the adequacy and precision of the translated message is very important. A communication or translation error between a patient and a physician in regard to a diagnosis may have serious consequences on the patient's health. Progress in SMT research has recently slowed down. As a result, new translation methods are needed. Neural networks provide a promising approach for translation [2].

Machine translation has been applied to the medical domain due to the recent growth in interest in and success of language technologies. As an example, a study was done on local and national public health websites in the USA with an analysis of the feasibility of edited machine translations for health promotional documents [3]. It was previously assumed that machine translation was not able to deliver high quality documents that can be used for official purposes. However, language technologies have been steadily advancing in quality. In the not-too-distant future, we expect that machine translation will be capable of translating any text in any domain at the required quality.

In our opinion, the medical data field is a bit narrow, but very relevant and a promising research area for language technologies. Medical records can be translated by use of machine translation systems. Access to translations of a foreign patient's medical data might even save their life. Direct speech-to-speech translation systems are also possible. An automated speech recognition (ASR) system can be used to recognize a foreign patient's speech. After it is recognized, the speech could be translated into another language with synthesis in real time. As an example, the EU-BRIDGE project intends to develop automatic transcription and translation technology. The project desires innovative multimedia translation services for audiovisual materials between European and non-European languages [<http://www.eu-bridge.eu>].

Making medical information understandable is relevant to both physicians and patients [4]. As an example, Healthcare Technologies for the World Traveler emphasizes that a foreign patient may need a description and explanation of their diagnosis, along with a related and comprehensive set of information. In most countries, residents and immigrants communicate in languages other than the official one [5].

Karliner et al.[6] talks of the necessity of human translators obtaining access to healthcare information and, in turn, improving its quality. However, telemedicine information translators are not often available for either professionals or patients. Machine translation should be further developed to reduce the cost associated with medical translation [7]. In addition, it is important to increase its availability and overall quality.

Patients, medical professionals, and researchers need adequate access to the telemedicine information that is abundant on the web [8]. This information has the potential to improve our health and well-being. Medical research could also improve the sharing of medical information. English is the most used language in medical science, though not the only one.

Due to its complexity, Polish is considered to be one of the most challenging West-Slavic languages. This makes Polish [9] translation very difficult for an SMT system. Polish grammar includes the complications of language elements and rules, along with an immense vocabulary. Word order in sentences is also a problem. In addition, the language has seven cases and fifteen gender forms of nouns and adjectives.

The goal of this paper is to present our experiments on neural based machine translation in comparison to statistical machine translation. The adaptation of translation techniques as well as proper data preparation for the need of PL-EN translation was also necessary. Such systems could possibly be used in real time speech to speech translation systems and aid foreign travelers that would require medical assistance. Combining such system with OCR and augmented reality tool could bring to mobile devices a real time translator as well. Human interpreters with a proper medical training are extremely rare and costly. Machine translation could also assist in the evaluation of medical history, diagnoses, proper medical treatments, health related information, and the findings of medical researchers from entire word. Mobile devices, the Internet and web applications can be used to boost delivery of machine translation services for medical purposes even as real time speech-to-speech services.

Download English Version:

<https://daneshyari.com/en/article/489425>

Download Persian Version:

<https://daneshyari.com/article/489425>

[Daneshyari.com](https://daneshyari.com)