Machine Learning

2a. Provide information on your computing innovation and computational artifact.
Name the computing innovation that is represented by your computational artifact.
Describe the computing innovation intended purpose and function.
Describe how your computational artifact illustrates, represents or explains the computing innovation intended purpose, its function or its effect.
(Approximately 100 words)

My computing innovation is machine translation using IBM models. This innovation uses an advanced set of algorithms to translate from one language to another by various methods of optimization (Jurafsky). Its purpose is to provide reliable translations using computers. My artifact illustrates that my innovation takes words or phrases from one language, processes it through a computer algorithm, and makes it understandable for someone who speaks another language.

2b. Describe your development process, explicitly identifying the computing tools and techniques you used to create your artifact. Your description must be detailed enough so that a person unfamiliar with those tools and techniques will understand your process.
(Approximately 100 words)

I used photoshop to create my artifact. I took clipart from google images, got the IBM logo from their website, drew simple shape vectors, and wrote out text using the text tool. The images I used were the person listening and the person talking, and the vectors included make up the half-circle and box with the arrow inside of it. I separated all of these elements from the background and applied a gradient overlay where a faded red flowed into a light blue. I then applied a separate gradient to the IBM logo and background using the same colors.

2c. Explain at least one beneficial effect and at least one harmful effect the computing innovation has had, or has the potential to have, on society, economy, or culture.
(Approximately 250 words)

Beneficial effects of machine translation include that it saves time, costs less than a human translator, and gives access to information for people from all over the world (Gubler). The money and time it would take to hire a translator and have a piece of writing translated are cut down extremely by computers. Computers have the ability to make decisions about translations in a split-second and software is basically free at this point for anyone to use. China has almost one-fifth of the world's population but only three percent of the information on the internet is in languages from China (Jessie). Machine translation can instantly translate an English webpage into something readable for this significant portion of the world. While the positive aspects of this innovation are very good, the threat of a mistranslation is dangerous. The nuances of different languages pose as barriers to computer logic and make the threat of miscommunication a lot

more plausible (Huddleston). The harm a miscommunication could cause is immeasurable. For example, take an English sentence using hyperbole to demonstrate an ironic stance towards a foreign politician. If a computer translated this sentence for someone in this foreign country without correctly identifying the figurative language, the outcome would have the opposite effect than what was intended.

2d . Using specific details, describe: The data your innovation uses. How the innovation consumes (as input), produces (as output), and/or transforms data. At least one data storage concern, data privacy concern, or data security concern directly related to the computing innovation.
(Approximately 250 words)
The IBM methods use millions and millions of sentences from sets of parallel corpus, or text correctly translated into two languages (Collins). An example of this would be the proceedings of the European Parliament where the same piece of text (roughly one million sentence pairs) is translated into 21 different languages (Loewen). When translating a sentence (ex: from French to English) the computer will use what is called a Noisy Channel Model where the computer will, instead of trying to find the best translation of the sentence, try to find the English sentence that has generated the observed French sentence (Collins). This is done by using an informed search (A*) to find the best compromise between faithfulness (how well the translation represents the original text) and fluency (how correct syntactically the sentence is) (Loewen). Because there is a set algorithm for this innovation and all the recourses are open-source there are little to none privacy, security, or storage concerns. This is unlike Google Translate whos translation algorithms are based on neuro networking and are essentially black boxes. Nothing is left in the dark with IBM models so there are no direct threats that could stem from this innovation.

2e.
For each online source, include the permanent URL. Identify the author, title, source, the date you retrieved the source, and, if possible, the date the reference was written or posted.
For each print source, include the author, title of excerpt/article and magazine or book, page number(s), publisher, and date of publication

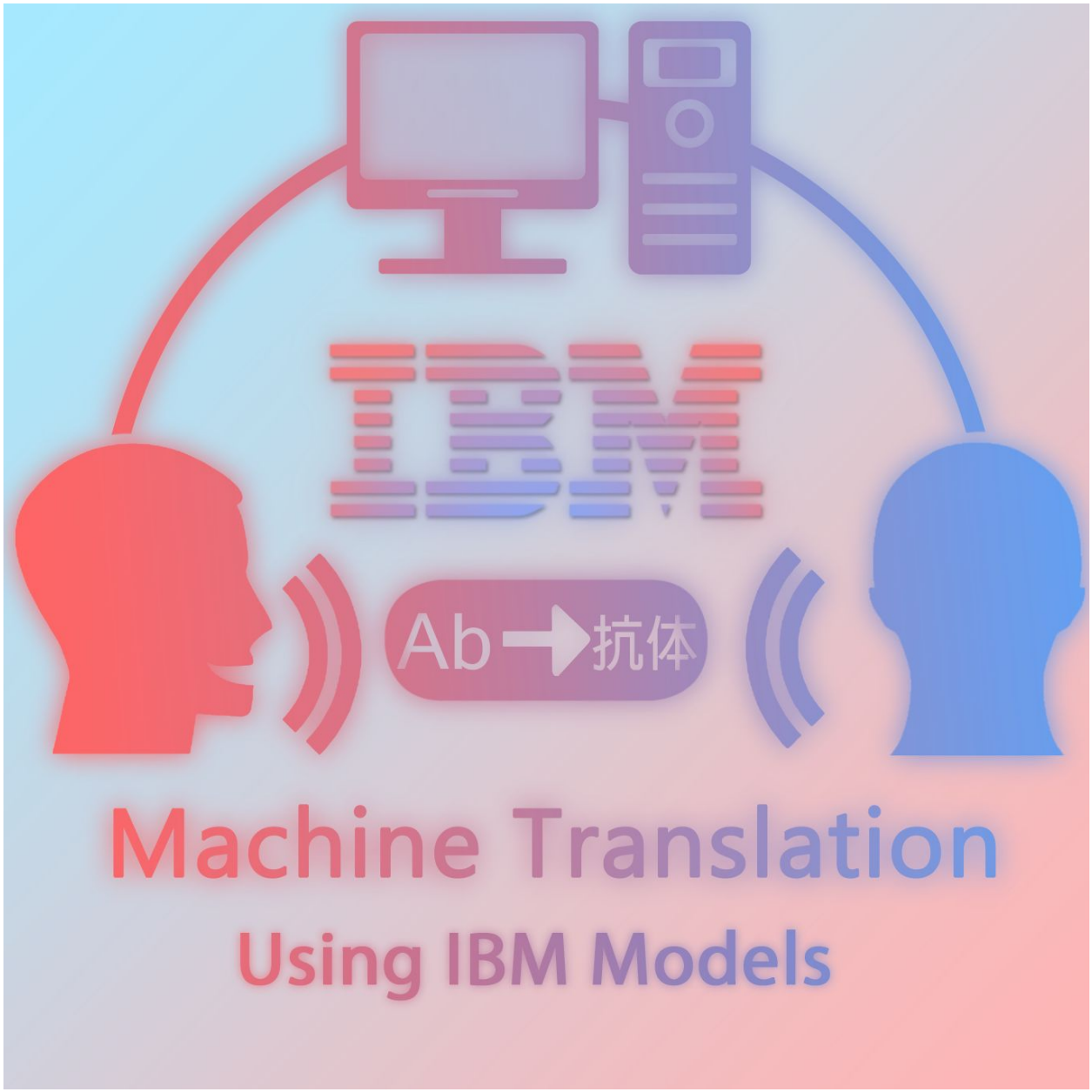
[1] Collins, Michael. “Statistical Machine Translation: IBM Models 1 and 2.” 2004, pp. 1–22., http://www.cs.columbia.edu/~mcollins/ibm12.pdf.

[2] Gubler, Kennedi. “The Benefits of Machine Translation.” InWhatLanguage, Publisher Name InWhatLanguage, 2 Apr. 2016, www.inwhatlanguage.com/blog/the-benefits-of-machine-translation/.

[3] Huddleston, Gunilla. “The Dangers of Using Free Machine Translation.” Language Connect, 20 Apr. 2017, www.languageconnect.net/blog/retail-and-ecommerce/the-dangers-of-using-free-machine-translation

[4] Loewen, Benjamin. Personal Interview. 12 December 2018.

[5] Jessie. “Top Languages of the Internet, Today and Tomorrow.” Unbabel, Building Universal Understanding, 11 Jan. 2018, unbabel.com/blog/top-languages-of-the-internet/.


IBM
Ab
抗体
Machine Translation
Using IBM Models