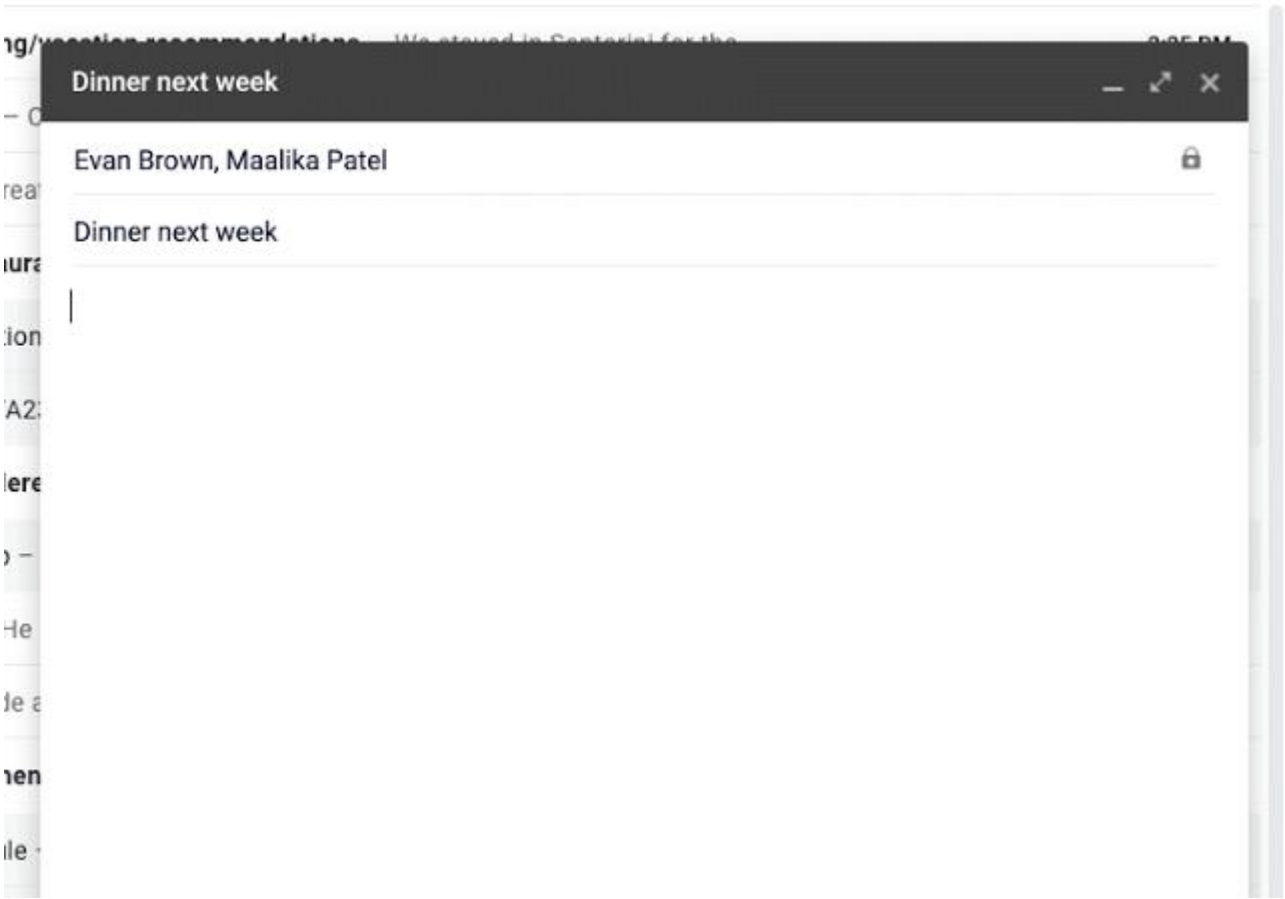
Google 智能编撰：使用神经网络协助编写电子邮件

原创：Google谷歌开发者8 月 30 日
文 / 吴勇辉，谷歌大脑团队首席工程师

在 Google I / O 上，我们推出了 Smart Compose，这是 Gmail 中的一项新功能，它利用机器学习，通过交互方式为正在编写邮件的用户提供补全句子的预测建议，从而让用户更快地撰写邮件。 Smart Compose 基于智能回复技术，提供了一种全新的方式来帮助您撰写邮件 - 无论您是在回复邮件亦或是从头开始起草一封新邮件。



在开发 Smart Compose 过程中，遇到了一些关键性的挑战，其中包括：

- 延迟：由于 Smart Compose 需基于用户的每一次按键输入来作出预测，如若想让用户察觉不到任何延迟，必须在 100 毫秒内作出理想的预测。这时候，平衡模型复杂性和推理速度就成了重中之重。

- 规模：Gmail 拥有超过 14 亿的用户。为了面向所有 Gmail 用户提供自动组句预测功能，该模型必须具有足够强大的建模能力，以便能够在细微差异的文本环境中为用户提出量身定制的建议。

- 公平与隐私：在 Smart Compose 开发过程中，我们需要在训练过程中处理潜在偏倚的来源，并且遵循与 Smart Reply 同样严格的用户隐私标准，以确保我们的模型不会暴露用户的隐私信息。 此外，研究人员也不具备访问和查看用户电子邮件的权限，这就意味着他们不得不在一个自己都无法查看的数据集上开发和训练一个机器学习系统。

寻找合适的模型

典型的语言生成模型，例如 ngram，神经词袋（BoW）和 RNN 语言（RNN-LM）模型，是在以前缀词序列为条件的基础上学习预测下一个单词。 然而，在电子邮件中，用户在当前电子邮件中键入的单词成为模型可用于预测下一单词的“信号”，模型将利用该信号来预测下一个单词。 为了结合更多有关用户想要表达的内容，我们的模型还会参考电子邮件的主题和先前的电子邮件正文内容（假设用户正在回复一封刚刚收到的电子邮件）。

注：ngram 链接
https://en.wikipedia.org/wiki/Language_model
神经词袋（BoW）链接
http://www.jmlr.org/papers/volume3/bengio03a/bengio03a.pdf
RNN 语言 链接
https://www.isca-speech.org/archive/interspeech_2010/i10_1045.html

包含和利用附加语境的一个方法是将问题转换成 sequence-to-sequence（seq2seq）机器翻译任务，其中源序列是邮件主题和先前电子邮件正文（如有）的串联，而用户正在撰写的邮件作为目标序列。尽管这种方法在预测质量方面表现良好，但它未能满足我们严苛的延迟标准。

为了改善这一点，我们将 BoW 模型与 RNN-LM 结合起来，结合后的模型比 seq2seq 模型更快，而对模型的预测质量影响极小。在这种混合法中，我们通过把单词嵌套平均分配在每个区域内，对邮件主题和之前的电子邮件内容进行编码。然后我们将这些平均分配后的嵌套连接在一起，并在每次执行解码步骤时将它们馈送到目标序列 RNN-LM，过程如下面的模型图所示。

Smart Compose RNN-LM 模型架构。通过对每个字段中的单词嵌套平均分配到每个区域内，将邮件主题和先前的电子邮件信息进行编码。随后，平均后的嵌套会在每次执行解码步骤时提供给目标序列 RNN-LM。

加速模式培训与服务

当然，一旦我们决定采用这种建模方法，我们就必须调整各种模型超参数，并使用数十亿个示例对模型进行训练，所有这些操作都相当费时。 为了加快速度，我们使用了一个完整的 TPUv2 Pod 来进行实验。如此，我们能够在一天之内将一个模型训练至收敛状态。

即便训练出了速度更快的混合模型，初始版本的 Smart Compose 在标准 CPU 上运行时，依旧存在着几百毫秒的平均服务延迟，这似乎与 Smart Compose 努力帮助用户节省时间的初衷依旧相去甚远。幸运的是，在推断期间可以使用 TPU 来大大加快用户体验。 通过将大部分计算分流到 TPU 上，我们将平均延迟时间缩短至几十毫秒，与此同时还大幅增加了单台计算机可处理的服务请求数量。

公平与隐私

由于语言理解模型会反映人类的认知偏差，导致得到多余的单词关联和组句建议，因此在机器学习中实现公平性至关重要。 正如 Caliskan 等人在他们近期的论文 “Semantics derived automatically from language corpora contain human-like biases 从语料库中自动导出的语义包含类似人类的偏见”中指出，这些关联深深隐藏在自然语言数据中，这对于构建任一语言模型来说都是相当大的挑战。 我们正在积极研究如何继续减少训练程序中的潜在偏见问题。此外，由于 Smart Compose 是基于数十亿的短语和句子进行训练，类似垃圾邮件机器

学习模型的训练方式，我们已经进行了广泛的测试，确保模型只记忆各类用户使用的常用短语。

未来研究方向

我们一直致力于通过遵循最先进的架构（例如，Transformer，RNMT+等），并尝试最新和最先进的训练技术，不断提高语言生成模型的预测质量。 一旦模型的实验结果能够满足严苛的延迟约束条件，我们就会将这些更先进的模型部署到产品中。此外，我们还在努力整合个人语言模型，旨在使它能够在系统中更加准确地模拟不同用户的个性化写作风格。

鸣谢

Smart Compose 语言生成模型由 *Benjamin Lee，Mia Chen，Gagan Bansal，Justin Lu，Jackie Tsay，Kaushik Roy，Tobias Bosch，Yinan Wang，Matthew Dierker，Katherine Evans，Thomas Jablin，Dehao Chen，Vinu Rajashekhar，Akshay Agrawal， Yuan Cao，Shuyuan Zhang，Xiaobing Liu，Noam Shazeer，Andrew Dai，Zhifeng Chen，Rami Al-Rfou，DK Choe，Yunhsuan Sung，Brian Strope，Timothy Sohn，Yonghui Wu* 等开发。