



Santas Digitales Dilemma: Die Bewältigung des Ansturms von Wünschen mit der Magie von Sprachmodellen

Autor: Christoph Graczyk (ZIB)

Projekt: IOL

Aufgabe

Inmitten des endlosen Schnees und der funkelnden Sterne des Nordpols nahm der Weihnachtsmann eine moderne Wendung. Um weltweit besser erreichbar zu sein, richtete er einen dedizierten E-Mail-Posteingang für die herzlichen Weihnachtswünsche der Kinder ein. Das Verfassen personalisierter Antworten auf Millionen von E-Mails scheint jedoch selbst für den Weihnachtsmann zu weit zu gehen.

Die Elfen, immer bereit, den Weihnachtsmann zu unterstützen, beriefen einen Notfallrat ein, um das Problem zu lösen. Die jüngeren Elfen, wie Elvin, befürworteten den Einsatz moderner Technikwunder wie Large Language Models. „Stellt euch vor“, rief er aus, „automatisches Kartenschreiben mit dem neuen Computer, den wir für E-Mails verwenden! Über der Weihnachtsmann zögerte, besorgt darüber die persönliche Note zu verlieren, die seine Karten ausmachte.

Hier schlug Elara, die älteste der Elfen, einen Mittelweg vor. „Warum verwenden wir nicht etwas Ähnliches wie ein Bigram Model? Es ist die Grundlage von sprachgenerierenden Werkzeugen. Es kann helfen, die einführenden und abschließenden Sätze der Karten auszuwählen, basierend auf unserer reichen Geschichte personalisierter Karten. Auf diese Weise könnte der Weihnachtsmann immer noch die Hauptbotschaft verfassen. Das Modell würde nur Santas eigene Phrasen als Vokabular verwenden, um die Karten zu generieren, anstatt nur die einzelnen Buchstaben des Alphabets, die in diesen Modellen sonst verwendet werden.“

Elaras Idee war einfach. Durch die Analyse vergangener Karten könnten sie vorhersagen, wie wahrscheinlich es ist, dass eine Phrase erscheint, basierend auf den vorherigen Phrasen. Dadurch würde die Essenz von Santas Botschaften bewahrt und der Prozess beschleunigt werden. Um dies zu demonstrieren, tauchte Elara in die Archive ein und zog eine zufällige Stichprobe von 10.000 Karten heraus.

Jede Karte, die der Weihnachtsmann schrieb, hatte typischerweise die gleiche Struktur: einen **Eröffnungssatz** gefolgt von seiner persönlichen Botschaft, dann einem **Abschlusswunsch** und einem **Abschiedssatz** vor Santas Unterschrift. Elara analysierte die Proben aus den Archiven akribisch. Nachfolgend sind einige exemplarische Statistiken für die

Häufigkeit der Verwendung bestimmter Phrasen aufgeführt.

Eröffnungssätze:

- „Wenn der Winterwind flüstert,“ - 220 Vorkommen.
- „Unter dem schimmernden Nordlicht“ - 180 Vorkommen.

Abschlusswünsche: Die Auswahl des Abschlusswunsches hängt nur von den **gegebenen** Eröffnungssätzen ab:

- Für Karten, die mit „Wenn der Winterwind flüstert“beginnen:
 - Anfänge:
 - * „Eine Schneeflocke halt durch die frostige Luft und “- 70 Vorkommen.
 - * „In der einzigartigen Reise jeder Schneeflocke“- 150 Vorkommen.
 - Enden:
 - * „spielt eine Melodie aus Freude und Hoffnung.“- 100 Vorkommen.
 - * „ist die Geschichte von tausend Sternen.“- 120 Vorkommen.
- Für Karten, die mit „Unter dem schimmernden Nordlicht“beginnen:
 - Anfänge:
 - * „Wenn der Kamin leise knistert,“- 80 Vorkommen.
 - * „Mit der Ruhe einer winterlichen Nacht,“- 100 Vorkommen.
 - Enden:
 - * „möge dein Herz fröhlich und leicht sein.“- 90 Vorkommen.
 - * „lass, Wärme und Geborgenheit in deinem Herzen wohnen.“- 90 Vorkommen.

Abschiedssatz:

- „Aus dem Winterwunderland“- 245 Vorkommen.
- „Mit festlicher Fröhlichkeit“- 155 Vorkommen.

Jedoch übersah Elvin in seiner Aufregung einen entscheidenden Aspekt des Modells. Im Gegensatz zum tatsächlichen Bigram-Modell, das die Wahrscheinlichkeit der Phrasen in Abhängigkeit von der vorherigen Phrase lernt, war Elvins Modell viel einfacher. Sein Modell wählt Phrasen auf Grundlage ihrer relativen Häufigkeit aus, mit dem zusätzlichen Kriterium, dass der Beginn der Abschlusswünsche vom Eröffnungssatz abhängt. Aber anders als Elara erklärt hatte, behandelt sein Modell die gegebenen Enden der Abschlusswünsche unabhängig von den gegebenen Anfängen der Abschlusswünsche. Dieser Fehler könnte zur Erstellung ungewöhnlicher oder sogar sinnloser Kombinationen führen. Für den Weihnachtsmann waren die schlimmsten dieser Kombinationen die, die ein fehlendes Subjekt im Abschlusswunsch haben.

Frage:

Mit dem auf der Stichprobe basierenden Modell und unter Berücksichtigung von Elvins Versäumnis, wie hoch ist die Wahrscheinlichkeit, dass eine Karte, die der Weihnachtsmann in diesem Jahr erstellt hat,

- mit einem der gegebenen Eröffnungssätzen beginnt,
- einen grammatikalisch inkorrekten Abschlusswunsch enthält, bei dem das Subjekt fehlt,
- und mit einem der gegebenen Abschiedssätze endet?

Hinweis: Bei dem Imperativsatz „.....lass Wärme und Geborgenheit in deinem Herzen wohnen.“ darf angenommen werden, dass ein Subjekt (implizit) gegeben ist.

Bei den zwei falschen Sätzen handelt es sich um:

- ”Wenn der Kamin leise knistert, ist die Geschichte von tausend Sternen”(”Wer oder was ist die Geschichte von tausend Sternen?”)
- ”Mit der Ruhe einer winterlichen Nacht, ist die Geschichte von tausend Sternen”(”Wer oder was ist die Geschichte von tausend Sternen?”)

Antwortmöglichkeiten:

1. 6.075%
2. 2.73375%
3. 0.0216%
4. 0.00972%
5. 0.243%
6. 0.436%
7. 0.78%
8. 10.45%
9. 3.34%
10. 0.89%

Projekt Referenz:

Das IOL-Forschungslabor widmet sich der Erkundung der Schnittstelle von mathematischer Optimierung und maschinellem Lernen mit dem Schwerpunkt auf der Entwicklung innovativer Techniken für das Lernen und die Optimierung. Durch die Integration dieser beiden Bereiche strebt die Gruppe an, neue Ansätze zur Lösung komplexer Probleme zu schaffen, die die Stärken sowohl der Optimierung als auch des maschinellen Lernens nutzen.