



Facebook liest und lernt

Vor kurzem stellte Facebook DeepText vor. Das neue Tool erkennt nach eigenen Angaben die Inhalte mehrerer Tausend Facebook-Seiten in nur wenigen Sekunden.

"Es regnet! Ich brauche ein Taxi!" Sätze wie dieser, via Facebooks Messenger geteilt, könnten in nicht allzu ferner Zukunft tatsächlich zum gewünschten Taxi verhelfen. Denn Facebook liest mit und versteht das Gelesene immer besser. In einem möglichen nächsten Schritt könnte der Messenger via Uber oder MyTaxi einfach eines ordern.

Tatsächlich kann Facebook das bereits. Das neueste System, genannt DeepText, wurde vor kurzem vorgestellt und kann laut Unternehmensangaben Texte mit nahezu menschlicher Genauigkeit in mehr als 20 Sprachen erfassen.

Sascha Pfeiffer, Sicherheitsberater der Firma Sophos, erläutert: „Verbraucher sollten sich immer darüber im Klaren sein, dass sie von großen Internetunternehmen als Kunde gesehen werden, nicht als Nutzer. Das gilt auch, wenn der eigentliche Dienst, wie etwa bei Facebook, kostenfrei ist. Von einem Kunden wird Umsatz in Form von Käufen erwartet, ein Nutzer zahlt mit seinen Daten. Hier ist grundsätzlich Vorsicht geboten, denn in Situationen, wie diesen, kann keine Firewall oder Antiviren-Software das Abfließen der persönlichen Informationen verhindern.“

Deep Learning – der Kontext macht den Unterschied

Zum Verstehen der Texte wird sogenanntes Deep Learning verwendet. Damit eine Maschine Text verstehen kann, muss immens viel erklärt werden. Es gibt die allgemeine Klassifikation, wie etwa die Tatsache, dass es sich um einen Post zum Thema Basketball handelt. Zusätzlich muss das Verständnis für sinnvolle Kontextinformationen, wie Spielernamen oder Statistiken eines Spiels, vorhanden sein.

Um überhaupt in die Nähe eines menschlichen Textverständnisses zu gelangen, müssen auch Slang und Begrifflichkeiten verstanden werden. So kann es sich bei Apple oder Blackberry um Obst oder technische Geräte handeln – der Kontext macht den Unterschied. Zudem muss die Technik skalierbar sein: Wie [TechCrunch](#) kürzlich berichtet, werden auf Facebook pro Minute rund 400.000 neue Stories und 125.000 Kommentare zu öffentlichen Beiträgen veröffentlicht. Das ist bereits jetzt immens viel und wird in absehbarer Zeit kaum weniger werden.

DeepText, das Tool, das all diese kleinen Textteilchen zu etwas Verständlichem zusammenfügen soll, basiert auf Deep Learning. Dabei handelt es sich um einen Zweig des maschinellen Lernens basierend auf einer Reihe von Algorithmen, die durch die Verwendung mehrerer Verarbeitungsschichten Abstraktionen in Daten modellieren, entweder mit komplexen Strukturen oder zusammengesetzt aus nichtlinearen Transformationen.

Ein Beispiel für die Verwendung von Deep Learning: In der traditionellen Verarbeitung natürlicher Sprache werden Wörter in ein Format umgewandelt, das einen Computer-Algorithmus erlernen kann. Das Wort "Bruder" könnte also einer ID wie 4598 zugewiesen werden, während das Wort "Bro" die 986665 erhält. Aber das funktioniert nur, wenn jedes Wort genau richtig geschrieben ist.

Deep Learning nutzt stattdessen "Wort Einbettungen", ein mathematisches Konzept, das die semantische Beziehung zwischen Wörtern bewahrt. Das ermöglicht einem Computer zu erkennen, dass "Bruder" und "Bro" eigentlich nahe beieinander sind, und es ermöglicht die Erfassung der tieferen semantischen Bedeutung.

Word-Einbettungen können Sprachen verständlich machen. Zum Beispiel, "alles Gute zum Geburtstag" und "feliz cumpleaños" sollte in dem gemeinsamen Einbettungsraum sehr nahe beieinander sein, erklärt Facebook. DeepText kann daher Wörter und Phrasen, die sich in einem gemeinsamen Einbettungsraum befinden, abbilden, um so Modelle zu bauen, die sprachunabhängig sind.

Aus Facebooks Ankündigung:

Die Facebook Gemeinde ist wirklich global, daher ist es ist enorm wichtig für DeepText, so viele Sprachen wie möglich zu verstehen. Traditionelle NLP-Techniken erfordern eine umfangreiche Vorverarbeitungslogik, die auf komplizierte Technik und Sprachkenntnisse baut. Es gibt zudem viele Variationen innerhalb jeder Sprache, da die Menschen umgangssprachliche Begriffe und verschiedene Schreibweisen verwenden, um die gleiche Idee zu kommunizieren.

Mit Deep Learning können wir die Abhängigkeit von sprachbezogenem Wissen reduzieren, da das System den Text mit wenig oder keiner Vorbereitung lernen kann. Dies hilft uns, mit wenig technischem Aufwand schnell mehrere Sprachen zu verarbeiten. Für mehr Informationen zum Thema lesen Sie eines der Dokumente der Facebook Artificial Intelligence Researcher Ronan Collobert und Yann LeCun.

Man will uns nur helfen

Facebook verwendet bereits jetzt DeepText Modelle, um Menschen Tools anzubieten, die sie vielleicht brauchen könnten. Möchte jemand beispielsweise ein Fahrrad für 200 Euro verkaufen erkennt DeepText, zuerst einmal, dass etwas verkauft werden soll. Es erkennt außerdem das angebotene Produkt, die Art des Verkaufs und den Preis. Zu diesem Zweck könnte es dann besondere Tools anbieten, die solche Transaktionen erleichtern.

Neben dem besseren Verständnis der Posts können Absichten und Stimmungen besser entschlüsselt werden. Das gilt auch für gemischte Inhalte, die aus Bild und Text bestehen. Auf diese Weise könnte qualitativ hochwertiger Inhalt gefördert und unerwünschter Inhalt, wie etwa Hass-Kommentare, gelöscht werden.

Für den Fall, dass uns diese allzu menschliche Fähigkeit von Facebook ein wenig gruselig vorkommen, empfiehlt Sascha Pfeiffer, weiterhin beim Teilen vorsichtig zu sein: „Auch wenn Sie denken, dass Sie nur mit Freunden reden: Facebook hört mit und versteht von Tag zu Tag besser, was wir sagen.“

Pressekontakt:

Sophos

Jörg Schindler, PR Manager CEEMEA

joerg.schindler@sophos.com, +49-721-25516-263

TC Communications

Arno Lucht, +49-8081-954619

Thilo Christ, +49-8081-954617

Christiane Capps, +49-174-3335550

Ulrike Masztalerz, +49-30-55248198

sophos@tc-communications.de