

Entwicklung neuer Algorithmen für FoodCASE¹

Bestimmung der Inhaltstoffzusammensetzung von Nahrungsmitteln

- FoodCASE berechnet die Zusammensetzung von Nahrungsmitteln auf Mikronährstoffebene.
- Neuentwicklung eines Algorithmus zur Berechnung der prozentualen Zusammensetzung eines Nahrungsmittels.
- Nutzung von LLMs und Vektordatenbanken in der FoodCASE Software für automatisiertes Food Matching.
- Die getesteten Ansätze sind vielversprechend, bedürfen allerdings umfangreichere Testung.

Hintergrund und Fragestellung

Die Erfassung von Nährstoffdaten dient der Berechnung von verpflichtenden Nährwertkennzeichnungen, sowie als Datengrundlage für ernährungs- und gesundheitspolitische Maßnahmen. Dazu hat das **Max-Rubner-Institut** den **Bundeslebensmittelschlüssel (BLS)** erarbeitet, welcher mit Hilfe von Laboranalysen und Literaturrecherchen erstellt wird. Die **FoodCASE Software** ermöglicht verschiedene Funktionalitäten aufbauend auf dem BLS. Durch neue Algorithmen und Maschinellem Lernen soll FoodCASE verbessert werden. Zum einen soll die **Ermittlung von Prozentanteilen einzelner Nährstoffe** optimiert werden. Andererseits soll die **Auffindung von Lebensmitteln** in der Datenbank mithilfe von Large Language Models (LLMs) verbessert werden.

Vorgehensweise

Nahrungsmittelprodukte listen ihre **Inhaltsstoffe** in der Reihenfolge ihres Gewichtsanteils auf, jedoch **ohne genaue Prozentangaben**. Zusätzlich enthalten sie eine **Nährwerttabelle** mit Angaben zu den Big 7 (wie z.B. Eiweiß). Mithilfe des BLS und einem Optimierungsverfahren lassen sich auch die Mikronährstoffe (wie z.B. Vitamine) berechnen. Im Rahmen der Optimierungsuntersuchungen wurde zunächst eine **Optimierungsfunktion formuliert**, die relevanten Einschränkungen definiert und anschließend **der COBYLA-Algorithmus** aus der Scipy-Bibliothek angewandt. Erste Ergebnisse deuteten darauf hin, dass dieser Ansatz eine bessere Leistung zeigt als das auf Linear Programming basierende Verfahren das bereits in FoodCase implementiert ist. Aufgrund des Fehlens eines Validierungsdatensatzes konnte jedoch kein endgültiger Vergleich zwischen den Verfahren vorgenommen werden. Ein weiteres Problem stellte sich im Umgang mit Produkten, deren Inhaltsstoffe einen

hohen Wasseranteil aufweisen, der beispielsweise während der Zubereitung verdampft. Diese Dynamik konnte in unserem bestehenden Lösungsansatz nicht berücksichtigt werden. Abgesehen von diesen Einschränkungen fielen die Ergebnisse des COBYLA-Algorithmus insgesamt jedoch positiv aus.

Für das automatisierte Food-Matching wurden **LLMs in Verbindung mit einer Vektordatenbank** (ChromaDB) getestet. Dabei wurden die BLS-Zutatenverzeichnisse auf Englisch und Deutsch verwendet: Die Zutaten wurden mithilfe von **Embeddingmodellen** in Vektoren umgewandelt und in der Vektordatenbank gespeichert. Die Abfragen (z.B. „Zucker“) an die Datenbank wurden ebenfalls als Vektoren umgewandelt und mit den Vektoren der Datenbank verglichen, um die Ähnlichkeit zu berechnen. Die Ergebnisse der Suche sind die Zutaten, die der Abfrage am ähnlichsten sind. Damit kann man Synonyme von Zutaten suchen, was mit einer herkömmlichen Suche nicht immer möglich ist.

Ergebnisse und Schlussfolgerungen

Der **COBYLA-Algorithmus zeigt vielversprechende Ergebnisse**, jedoch könnten **auch alternative Verfahren** in Betracht gezogen werden. Solche Methoden könnten, unter der Voraussetzung einer geeigneten Datenbasis und eines klar definierten Optimierungskriteriums, ebenfalls getestet werden, um die Effektivität des Systems weiter zu steigern.

Die ersten Tests mit dem BLS-Zutatenverzeichnis zeigten **positive Ergebnisse**. Allerdings gab es bei einigen Zutaten **Schwierigkeiten**, die ähnlichsten Einträge aus der Datenbank zu finden. Bei manchen Abfragen fehlte noch eine Entsprechung in BLS. In solchen Fällen müssen die Modelle dennoch die ähnlichsten Lebensmittel finden. Auch weitere Embeddingmodelle könnten in Betracht gezogen werden. Der entwickelte Code ist öffentlich verfügbar².

Kontakt	Informationen
KI-Beratung (KIDA): kida@bmleh.bund.de	KIDA-Bearbeitende: Steffen Albrecht (KIDA, FLI), Cristina Ortiz Cruz (KIDA, MRI), Tim Wessels (KIDA, BVL) Team-Leitung: Micha Schneider (KIDA KI-Beratung, Thünen) Software/Code: ¹ https://foodcase.org/ ² https://github.com/kida4bmel/foodcase