

テキストマイニングに基づいたカーボンフットプリントの計算方法の提案

その他（別言語等）のタイトル	A method based on text mining for calculating carbon footprint
著者	趙 培江, 永野 宏治
雑誌名	第11回日本LCA学会研究発表会講演要旨集
巻	11
ページ	292-293
発行年	2016-03-02
URL	http://hdl.handle.net/10258/00009201

テキストマイニングに基づいたカーボンフットプリントの計算方法の提案

A method based on text mining for calculating carbon footprint

○趙 培江*1)、永野 宏治 1)

Zhao Peijiang and Koji Nagano

1) 室蘭工業大学

* 14096014@mmm.muroran-it.ac.jp

1. はじめに

カーボンフットプリント(CFP)とは、製品やサービスの原料・製造・輸送・消費・廃棄の全過程で発生する温室効果ガスの量を二酸化炭素(CO₂)の量に換算して、ラベル付けする社会的な仕組みである¹⁾。現在、日本で展開されている CFP の取組みは、カーボンフットプリントコミュニケーションプログラム(CFPプログラム)、コープさっぽろの食品 CFP の取組み等がある。現在の CFP の取組には以下の問題がある。

- (1) CFP が公開されている商品の数が少ない。2015 年 12 月の時点で、CFP プログラムでは 1153 品の商品を公開し、127 品を CFP 対象製品としている。
- (2) CFP の認知度が低い。筆者らが 2014 年 9 月と 2015 年 8 月に札幌においてアンケート調査した結果 CFP を知っている一般消費者は 14%であった²⁾。
- (3) CFP の大きさを把握しにくい。CFP の大きさの基準がない。また、CFP 計算されている商品が少ないため、CFP の大小を相対的に評価しにくい。

これらの問題は相互に関連している。これらの問題の中で、CFP が計算された商品数が増えるならば、CFP の認知度の向上、CFP の大小の商品間比較が可能になると考えられる。

一方、現代は商品に関する情報が多量にウェブに公開されている。このようなウェブに公開されている情報から CFP を推定できるようになると、メーカーに直接頼らずに、商品の CFP を求められるようになる可能性がある。本研究では、テキストマイニングを用いて、ウェブに公開されているデータを使って、商品の CFP の範囲を計算する方法を提案する。

2. 推定方法の概要

図 1 にテキストマイニングを用いた CFP の推定方法の概略を示す。本手法はウェブからデータを収集する部分と CFP を計算する部分からなる。ウェブから収集するデータは商品の原材料データ、原材料産地データ、生産データである。本手法でウェブから収集されるデータには、オープンデータベースとして統一した表現形式に整理されているデータと、一般のウェブサイトにおいて公開されているテキストデータがある。オープンデータベースから取得したデータは整理されたデータのため、直接 CFP の計算に容易に使用できる。一方、一般のウェブサイトにおいて公開されているテキスト

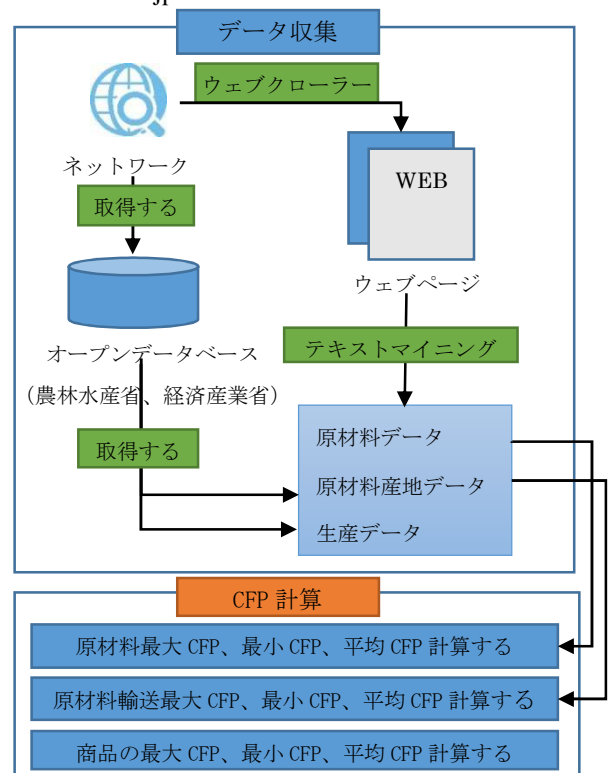


図 1 テキストマイニングを用いた CFP の推定方法

データは、表現がウェブサイト毎に異なるため、直接使用できない。そこで、著者らは、一般のウェブサイトから CFP に関するデータを自動的に抜き出すテキストマイニングの方法を開発している。CFP の計算では、収集した商品データを用いて、該当商品種の CFP の最大値、最小値、平均値を計算する。収集した商品データから商品種の CFP を計算する時、その商品種を構成する個々商品の原材料にばらつきがあるため、商品種の CFP 計算結果はばらついた推測値になる。このような本手法の特徴は、ウェブに公開されているデータを使って CFP の分布範囲を推定し、商品の CFP の基準を示す点にある。

3. データの収集

本手法ではウェブからデータを収集する時、一般のウェブサイトから商品の原材料のデータを収集し、オープンデータベースから原材料産地データと生産データを収集する。

原材料データを収集するため、本研究では、原材料データが記録されるウェブサイトを検索するウェブクローラーと商品に関するテキストデータを抽出するデ

キストマイニング用の再帰自己符号化器を開発している。このウェブクローラーでは、計算する商品の商品名を含むウェブサイトを自動的に検索し、商品の原材料に関するテキストを収集する。本研究の再帰自己符号化器は、テキストデータから単語を抽出し、その抽出した単語が商品に関するデータである確率を計算する。本手法で、収集する原材料データは、原材料名、原材料の出現頻度比、原材料と商品の重量比の最大値、最小値、平均値である。原材料の出現頻度比とは、その商品が現れるウェブページの数と、そのウェブページの中で、その原材料が現れるウェブページの数の比である。

本手法でオープンデータベースから収集される原材料産地データは、その原材料の日本国内で産地名、産量と海外の輸入先、輸入量である。収集する先は農林水産省と経済産業省が公開しているオープンデータベース³⁾である。本研究では 6962 品の原材料の産地データを収集した。

本手法で収集された生産データは、各生産メーカーが公開している工場の CO₂ 排出量と工場所在地である。

4. CFP の計算

本手法でウェブから収集したデータには、以下の性質がある。

(1) 商品の原材料データにばらつきがある。本手法では、原材料データを複数のウェブサイトから収集するため、同じ商品種でも個々の商品の重量、原材料の種類、その原材料が商品に占める重量比が異なる。

(2) 原材料の真の産地情報は不明である。この二つの性質のため、本手法では次に述べるように CFP を推定する。

原材料データにばらつきがあるため、本手法では、各原材料の出現頻度比を考え、原材料 CFP の最大値、最小値、平均値を計算する。また、個々の商品によって重量が変わるため、本手法では、商品の重量 m を 100g に固定する。

原材料 A の原材料 CFP の最大値 CFP_{A-max} 、最小値 CFP_{A-min} 、平均値 \overline{CFP}_A はそれぞれ(1),(2),(3)式で計算式する。 R_{A-max} , R_{A-min} , \overline{R}_A は原材料と商品の重量比の最大値、最小値、平均値である。 P_A は原材料の出現頻度比である。原材料 CFP の最小値を計算する(2)式では原材料の出現頻度が 50%以上の原材料を使う。CO₂ 原単位 U_A は味の素グループ版「食品関連材料 CO₂ 排出係数データベース」に基づいた数値である。

$$CFP_{A-max} = R_{A-max} * P_A * U_A * m \quad (1)$$

$$CFP_{A-min} = R_{A-min} * P_A * U_A * m \quad (2)$$

$$\overline{CFP}_A = \overline{R}_A * P_A * U_A * m \quad (3)$$

表 1 カップヌードル CFP 計算結果

CFP 種類	最大 CFP (g-CO ₂ / 100g)	最小 CFP (g-CO ₂ / 100g)	平均 CFP (g-CO ₂ / 100g)
原材料 CFP	172.4	58.2	108.8
輸送 CFP	17.4	5.4	14.2
生産 CFP	32.2	32.2	32.2
総 CFP	222.0	95.8	155.2

原材料 A の輸送 CFP の最大値 CFP_{At-max} 、最小値 CFP_{At-min} 、平均値 \overline{CFP}_{At} はそれぞれ(4),(5),(6)式で計算する。本手法では、原材料 A を各産地からメーカーの工場までの最長輸送距離 L_{A-max} 、最短距離 L_{A-min} を用いて、輸送 CFP の最大値、最小値を計算する。(6)式の平均輸送距離 \overline{L}_A は、各産地 k の原材料 A の産量 M_k と日本国内総消費量 M_{all} の比で重み付けした輸送距離 L_{Ak} の重み付け平均値である。なお、日本国内総消費量は国内産量と海外輸入量の和である。輸送原単位 U_t では、国内輸送はトラックの輸送原単位 0.272 kg-CO₂ / (ton km) を使用し、海外輸送は船の輸送原単位 0.0089 kg-CO₂ / (ton km) を使用している。

$$CFP_{At-max} = L_{A-max} * R_{A-max} * P_A * U_t * m \quad (4)$$

$$CFP_{At-min} = L_{A-min} * R_{A-max} * P_A * U_t * m \quad (5)$$

$$\begin{cases} \overline{CFP}_{At} = \overline{L}_A * \overline{R}_A * P_A * U_t * m \\ \overline{L}_A = \sum_k \frac{M_k}{M_{all}} * L_{Ak} \end{cases} \quad (6)$$

5. 計算結果と考察

本研究で提案した方法を用いて、21 個のウェブサイトから日清食品のカップヌードルのデータを収集し、CFP を計算した。現在、日清食品のカップヌードルには 34 種が販売されている。また、本研究のテキストマイニングの開発を十分に完成されていないため、本報告では、商品のデータの一部は著者らが手動で収集した。計算結果を表 1 に示す。

本報告では、テキストマイニングにより、大量な商品データを収集し、商品の CFP の範囲を計算できる可能性を示した。しかし、本手法を使った計算する過程には、再帰自己符号化器の認識率が足りない、収集した CFP の計算に関するデータの量が少ない等の課題がまだあるため、改良を重ねていく。

引用文献

- 1) 永野宏治, 吉田洋一, 小松均, 村上伸吾(2013): 日本 LCA 学会誌, 9(4), 299-305
- 2) 趙培江, 永野宏治, 鈴木昭徳, 佐藤和彦(2016): 日本 LCA 学会誌, 12(1)
- 3) 農林水産省: “農林水産省/統計情報”, <<http://www.maff.go.jp/j/tokei/index.html>>, (参照 2015-11-25)