



# RT法による著作者の識別

—次元（単位・性格）が異なるデータへの規準化技術の適用—

*Authorship Discrimination using the RT Method*

—Application of Normalization for Different Dimension Item Data—

長谷川 良子\*

*Yoshiko Hasegawa*

Authorship discrimination was studied by applying the Recognition Taguchi (RT) method to the works of four prominent Japanese writers. The RT method was proposed on the assumption that it would be applied to data of identical dimensions, such as pixel data or time-series data, so other types of data must be normalized. Two new normalization methods, one using the median and range and the other using the median and standard deviation, were proposed and compared with normalization by the standard deviation alone, which had been proposed previously. With any of the three normalization methods RT analysis successfully identified the four writers, but normalization by median and range was shown to be the best of the three methods. The conclusions reached were that the RT method can be used for authorship discrimination, and that when the RT method is applied to data of different dimensions (different units or characteristics), its discriminative capabilities can be improved by applying one of the newly proposed normalization methods.

**Key words** : literary style, stylometry, authorship discrimination, Taguchi methods, quality engineering, Mahalanobis-Taguchi (MT) system, pattern recognition, RT method (T Method-3), S/N ratio, normalization, dimension

## 1. はじめに

近代の計量文献学は、1851年にド・モルガン (Augustus de Morgan : イギリスの数学者, 1806 ~ 71) が、それぞれの作者の文体は「単語の長さ」(word-length) の平均値で説明できる可能性を示

唆したことに始まるといわれている。計量文献学とは文献の特徴を数値化し、統計学的手法を用いて文献の分析や比較を行う学問である。これまで主に、作者不詳の著者を調べる著作者問題や年代不詳の年代を明らかにする問題に応用されてきた。

ユール (Yule, G.U.) はセンテンス長 (sentence-length; センテンスに含まれる単語の数) を利用して『キリストにならいて』 (*De Imitatione Christi*)

\* 長谷川技術士事務所, 正会員