統計学One Point【10】巻　テキストアナリティクス
金 明哲著
http://www.kyoritsu-pub.co.jp/bookdetail/9784320112612

まえがき

統計的データ処理を行う環境が急速に変化している．従来は，実験・観測・調査で得られたデータを手作業で表型に整理し統計的に分析を行うのがほとんどであった．

昨今，情報機器やインターネットの普及に伴いデータの収集方法が大きく変化した．インターネット上には大量のデータが蓄積されている．その多くはテキスト型データである．たとえば，電子新聞，ブログ，Twitter，メール，電子掲示板の情報，ネット小説，文学作品，コーパス等枚挙に暇がない．このような電子化されたテキスト型データの分析は歴史が浅い．

90 年代から，データマイニングのブームに乗りテキストマイニングという研究応用分野が形成され始め，大きな進展があった．この分野では，構造化されていないテキスト型データを自然言語処理の技術で構造化し，汎用的なデータ分析方法や機械学習方法を用いるのが一般的であったが，テキスト型データの分析を前提とした分析方法の開発も進んでいる．近年，「テキストマイニング」からより汎用的な用語「テキストアナリティクス」に移行する傾向がみられるようになった．本書では，計量的にテキストを分析する主な手法とその行為をテキストアナリティクスという書名でまとめた．

第 1 章ではテキストアナリティクスの基本的な考え方，第 2 章ではテキストの電子化などの前処理，第 3 章ではテキストデータの視覚化，第 4 章ではテキストにおける法則と指標，第 5 章ではテキストの特徴分析，第 6 章ではテキストのクラスター分析，第 7 章ではテキストの分類分析，第 8 章ではテキスト関連の予測や要因分析の主な方法について，例を用いて平易な説明に注意を払った．

テキストアナリティクスには，幅広い統計的データ処理や機械学習の方

統計学One Point【10】巻　テキストアナリティクス
金 明哲著
http://www.kyoritsu-pub.co.jp/bookdetail/9784320112612

法が用いられている．そのすべてについて詳細に説明する紙面がないため，主な方法について説明を行った．深く理解するためには，関連の文献を参考にすることが必要である．本書の内容が日本語のテキストアナリティクスの発展の一助になれば幸いである．

本書の執筆の機会を与えていただいた本シリーズの鎌倉稔成編集長，宿久洋編集委員を含む編集委員の皆様，テキスト計量分析に導いていただいた恩師 現勉誠出版文化情報学研究所村上征勝所長に深く感謝の意を表する．またご丁寧に査読し，有益なコメントをいただいた先生方，本書の校正原稿について，有益なコメントをいただいた名古屋大学人文学研究科中村靖子教授，同志社大学文化情報学部孫昊助手に感謝する．なお，丁寧な編集をしてくださった共立出版編集部に感謝の意を表す．

2018 年夏吉日

金　明哲