3S-07

プレイリスト時系列分析における不規則間隔データ

向山　輝†　　沼尾　幹房†　岸上　順一†

室蘭工業大学情報電子工学系学科†

１．　はじめに

近年，医学や経済学などの分野において時系列データを用いた傾向分析，将来予測を行うことが増加してきている．分析対象となる代表的な時系列データとしては，血液検査データ，為替データや株価データがある．このような時系列データは定間隔でデータサンプリングが行われる場合と，不規則間隔でサンプリングが行われるデータが存在する．時系列分析では，データを定式化するために様々なモデルが提案されており，作成されるモデルを分析し将来予測が行われる．しかしながら，モデル作成するにはデータが等間隔である必要があり，不規則間隔でサンプリングが行われるデータからモデル作成する際には何らかの処理が必要となる．データの前処理を行う際には，処理前と処理後のデータ間の傾向の類似性が重要となってくる．処理後のデータが持つ傾向と元データの持つ傾向と類似するほど，時系列分析において精度の高い分析結果を得ることができる．

本研究では，プレイリスト時系列分析における不規則間隔データの前処理の手法について検討を行う．また，手法の検討には 24 時間放送されているインターネットラジオ局のプレイリストの不規則間隔な楽曲データを用いる．提案する手法でプレイリストデータがもつ傾向を維持することを目標とする．

２．　研究動向

先行研究として，不規則間隔な時系列データに関する研究ではデータ属性値の持続時間の推定から分析をする手法[1]やフラクタル解析を利用したものがある[2]．また，データの補完による分析についての研究も行われている[3]．これらの手法は存在しないデータの補完や推定に注目をしている．

Playlist analysis using non-equal distance series data
Hikaru Mukaiyama, Numao Mikifusa, Jay Kishigami
Department of Information and Electronic Engineering
Muroran Institute pf Technology
E-mail: 11024162@mmm.muroran-it.ac.jp

３．　提案手法

本研究では，プレイリストデータの時系列分析を行うための前処理としてデータ抽出手法を提案する．処理の手順は次の２つの Step に分かれており，元データの傾向を保持したデータの抽出処理を行う．

Step1: 一日分の放送曲数の統一
Step2: 曲データの等間隔化

Step2 の処理では曲ごとに再生時間が異なるため図１に示した例のように曲間隔を等間隔化する．

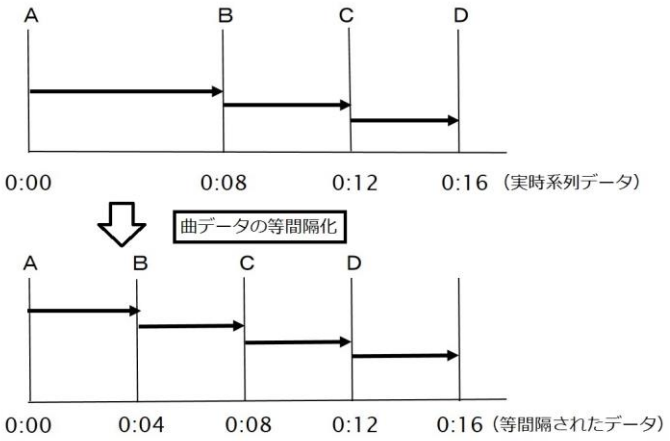

図１　曲データ等間隔化の例

３．１．　Step1: 一日分の放送曲数の統一

扱うデータは 2012 年 4 月の 3564 曲分のデータであり，メタデータには「放送時間」，「編曲者名」，「タイトル名」，「編曲者出生年」がある．また，3564 曲中で「編曲者出生年」が不明なデータが 116 曲存在するので実際に扱うデータは 3448 曲分のデータとなる．プレイリストの時系列分析を行うにはサンプリング数(一日分の放送曲数)の統一が必要となる．本研究では，サンプリング数の統一の方法として月の中で最も放送された楽曲数が少ない日の時の曲数を統一のサンプリング数とする方法を用いた．2012 年 4 月の最小曲数は 73 曲であり(表１)，一日毎にランダムで 73 曲の楽曲の抽出を行った．

表１　日別放送曲数

日付	曲数	日付	曲数	日付	曲数
4/1	141	4/11	120	4/21	73
4/2	112	4/12	121	4/22	119
4/3	116	4/13	122	4/23	124
4/4	124	4/14	101	4/24	121
4/5	111	4/15	123	4/25	108
4/6	120	4/16	117	4/26	110
4/7	86	4/17	110	4/27	120
4/8	135	4/18	121	4/28	91
4/9	118	4/19	115	4/29	122
4/10	111	4/20	125	4/30	111

３．２．　Step2：曲データの等間隔化

Step1 の処理によってサンプリング数が統一されたが，サンプリング間隔にはまだばらつきがある．曲間隔を 1/73 とすることで曲データの等間隔化を行い，時系列分析を行える形にデータを整形した．なお，本研究では 0 時～24 時間を 0～１に置き換えて一日の時間を表すこととする．

４．　評価

Step1，Step2 の処理によって「放送時間」と「編曲者出生年」の傾向の変化見られたかの評価を行った．まず，Step1 の処理前の 3448 曲のデータと処理後 2190 曲(73 曲×30 日)のデータをグラフにプロットした．縦軸に「編曲者出生年」をとり、横軸に「放送時間」とり，図２が処理前で図３が処理後のグラフとなっている．２つのグラフを比べると丸で囲った部分に共通の傾向が見られているのがわかる．この結果より，ランダムを用いた抽出方法で傾向の類似性が見られると考えられる．

Step2 に関しても Step1 と同様にグラフを作成したがあまり類似性を見ることができなかった．これは Step2 で曲データの等間隔化を行う際に図１にあるように実際の放送時間とのずれが生じ，その結果傾向にもずれが生じてしまったと考えられる．

５．おわりに

本研究では不等間隔時系列データにおいて時系列分析を行うための処理として２段階の処理手順を提案した．サンプリング数統一のための Step1 の処理では類似性が見られ，データの等間隔化の処理としての Step2 の処理後では傾向の類似性が見られなくなってしまった．一日分のプレイリストを分析するにあたり重要となってくるのは曲の放送された時間帯であり，今回の

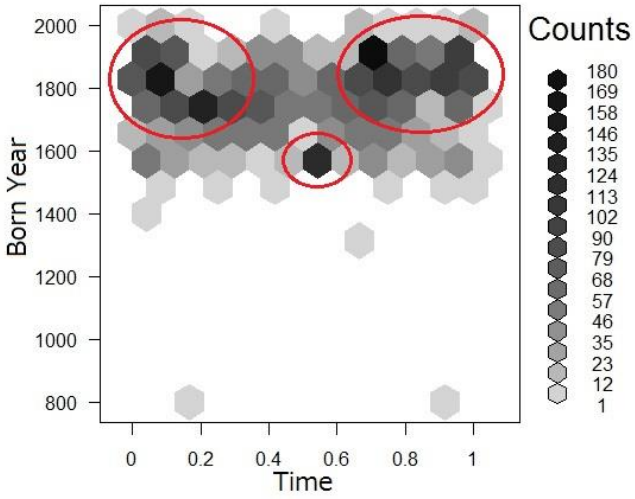

図２　処理前の 3448 曲分の曲データのグラフ

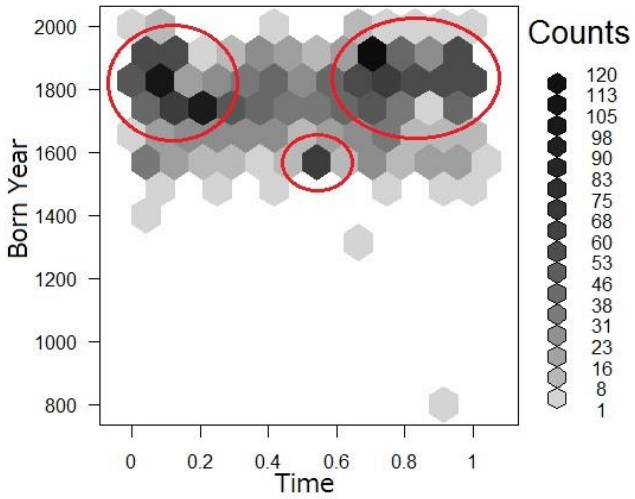

図３　Step1 後の 2190 曲分の曲データのグラフ

手法では曲データの等間隔化を行う際に楽曲が持つ「放送時間」のメタデータの情報が失われ傾向の変化が生じてしまった．

今後は「放送時間」を考慮したデータの抽出方法を考え，より元データの傾向を維持できるような処理方法を考えプレイリストの時系列解析を行えるようにしていきたい．

６．参考文献

[1] 本山真也・市瀬龍太郎・沼尾正行(2005)「間隔不定な時系列データからの知識発見」『知識ベースシステム研究会』69，27-32

[2] 熊谷善彰(2001)「不等間隔時系列のフラクタル解析」『日本応用数理学会論文誌』11(4)，179-186

[3] 辰巳憲一・松葉育雄(2008)「時系列データにおける補完方法の分析と考察」『学習院大学経済経営研究所年報 』22，35-43