

# テレビ番組推薦の精度を向上させる 番組メタデータ補完技術

Metadata Complementation Technology to Improve Accuracy of TV Program Recommendations

小川 修太

西本 喜則

大盛 善啓

■ OGAWA Shuta

■ NISHIMOTO Yoshinori

■ OHMORI Yoshihiro

近年、デジタル放送への移行に伴ってチャンネル数が増え、膨大な数のテレビ（TV）番組の中からユーザーが視聴したい番組を見つけにくくなってきている。それを解決する手段の一つとして番組推薦技術が求められている。番組推薦には番組メタデータが必要であるが、常に十分な番組メタデータが得られるとは限らず、推薦精度が低くなってしまいう課題があった。

東芝は、この課題を解決するため、番組メタデータの補完技術を開発した。この技術では、番組タイトルを照合して、例えば毎週放送される番組（シリーズ番組）を検出し、不足しているメタデータを補完できる。シリーズ番組の検出精度のF値<sup>(注1)</sup>が0.20から実用レベルの0.95に向上し、メタデータを補完できることを確認した。これにより、推薦精度が向上した。この技術を、2014年春に商品化された欧州市場向けTVの自動録画機能<sup>(注2)</sup>に適用した。

With the increase in the number of TV channels accompanying the shift to digital broadcasting, it has become difficult for viewers to find TV programs that they particularly wish to see from among the large number of programs available. Demand has therefore been growing in recent years for the development of TV program recommendation technologies. However, the shortage of program metadata necessary to find programs according to the preferences of each viewer has lowered the accuracy of TV program recommendation.

As a solution to this issue, Toshiba has developed a metadata complementation technology for TV program recommendation. This technology makes it possible to complement missing metadata by comparing program titles to determine, for example, whether a program is part of a series that is broadcast weekly. We have conducted evaluation experiments and confirmed that the metadata complementation technology can complement missing metadata due to improvement of the F-measure of detection accuracy of series type programs from the conventional level of 0.20 to a practicable level of 0.95, thereby improving recommendation accuracy. We have applied this technology to the automatic recording function of TV products for the European market since the spring of 2014.

## 1 まえがき

近年、デジタル放送への移行に伴って衛星放送の多チャンネル化が進み、視聴可能なTV番組の数が膨大になっている。このため、ユーザーが視聴したい番組を簡単に見つけることが従来よりも難しくなってきた。その解決手段として、ユーザーの指定する条件に合致した番組を検索する手法や、ユーザーの嗜好（しこう）に合った番組を推薦する手法が求められている。

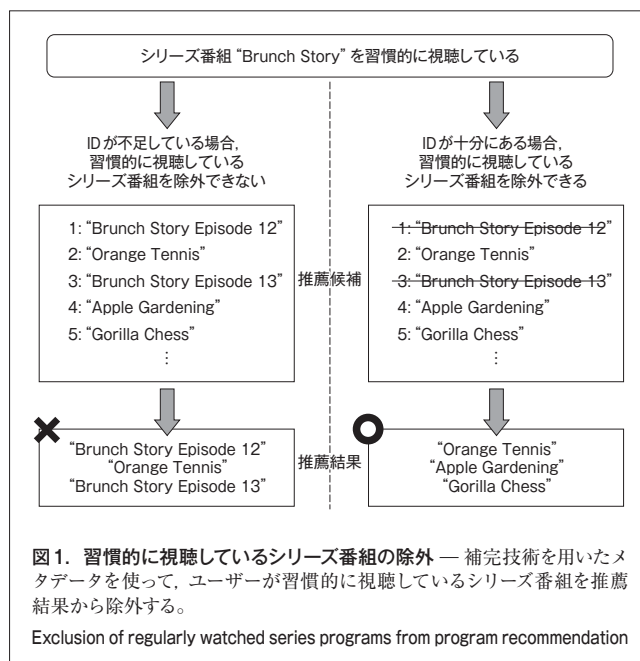
このうち、番組推薦では、ユーザーの嗜好に合った番組が必ずしも常に推薦されるべき番組であるとは限らない。例えば、ユーザーがよく視聴している番組は、どれほどユーザーの嗜好に合っているとしても、推薦する価値は低い。これまでに、ユーザーにとって意外性や新規性のある番組を推薦することが満足度の向上につながる事が指摘されている<sup>(1)</sup>。また、この考えに基づいて、ユーザーの行動履歴から嗜好と習

(注1) 検索性能の評価指標の一つで、検出誤りの少なさを表す適合率（P）と、検出漏れの少なさを表す再現率（R）の調和平均で算出される。

$$F\text{値} = \frac{2 \cdot P \cdot R}{P + R}$$

F値が高いほど、検索性能が良いことを示す。

(注2) 自動録画機能の使用には、ユーザーによる初期設定が必要。



慣のモデルを作り、それらを用いて意外性の高い推薦結果を得る手法も提案された<sup>(2)</sup>。

そこで番組推薦エンジンとして、推薦結果の意外性を高めるための処理を検討した。具体的には、ユーザーが習慣的に視聴しているシリーズ番組を推薦結果から除外する処理である。この除外処理には、同じシリーズに属する番組メタデータのID（識別情報）を用いる。しかし、IDが不足している場合には除外処理が機能せず、ユーザーがよく視聴しているシリーズ番組が推薦結果の中に入り、実用的な推薦精度が得られないという課題があった（図1）。

東芝は、この課題を解決するために、TV番組推薦のためのメタデータ補完技術を開発した。ここでは、開発したメタデータ補完技術の概要について述べる。

## 2 番組推薦におけるIDの必要性

IDの必要性について説明するため、初めに番組推薦エンジンの基本的な処理について述べる。

番組推薦エンジンの構成を図2に示す。TV番組に関するユーザーの嗜好を学習するユーザー嗜好学習処理（以下、学習処理と呼ぶ）と、学習した嗜好に基づいて推薦対象の番組群からユーザーが好む番組を予測する推薦番組予測処理（以下、予測処理と呼ぶ）を順に実行する。

まず学習処理では、過去にユーザーが視聴した番組のメタデータのうち、タイトルや内容説明文から抽出したキーワードをカウントし、ユーザーが習慣的に視聴している番組に多く含まれるキーワードをユーザーの嗜好として学習する。

次に予測処理では、推薦対象である番組について、それぞれのメタデータから抽出したキーワードを調べ、学習処理で学習したキーワードが含まれる個数をカウントして推薦スコアを算出する。推薦対象の番組それぞれについて推薦スコアをひとつと算出した後、推薦スコアの大きい順に並べ替えて上位から一定数の番組を推薦結果とする。

更に、このような基本的な処理のほかに、番組推薦エンジン

は、ユーザーが習慣的に視聴しているシリーズ番組を推薦結果から除外する処理を行う。シリーズ番組は、例えば毎週など一定の時間間隔をおいて放送される番組群で、同じシリーズの番組は出演者が共通していたり、連続ドラマのように内容につながりがあったりする。

ここで、学習処理では、ユーザーが習慣的に視聴しているシリーズ番組のIDを習慣性データベースに記録する。そして予測処理では、推薦結果を選ぶ際に習慣性データベースに記録されたIDを参照して、同じIDを持つ番組を推薦結果から除外する。

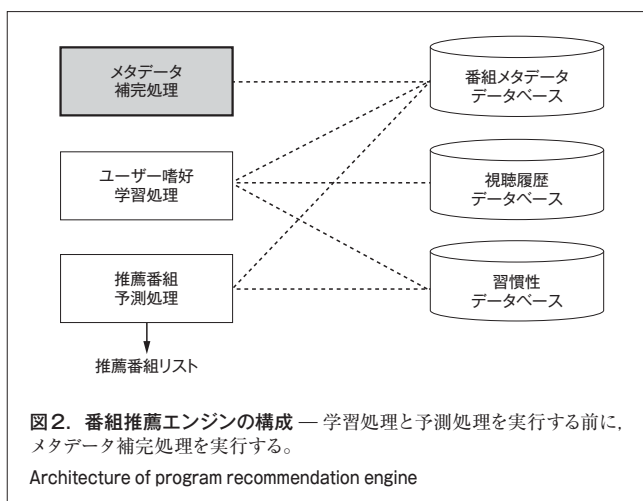
しかし、番組メタデータにIDが付いていないと、習慣的に視聴しているシリーズ番組を推薦結果から除外できず、実用的な推薦精度が得られない。

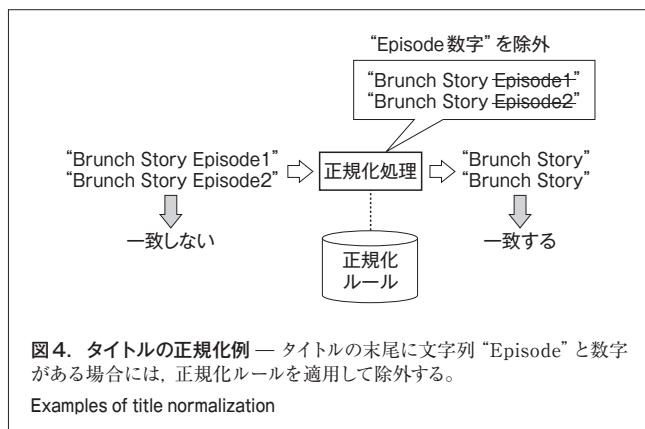
## 3 メタデータ補完技術

不足しているIDを補完するため、学習処理を実行する前にメタデータ補完処理を実行する。メタデータ補完処理では、各番組のタイトルを用いて同じシリーズに属する番組を検出して、IDを補完する。タイトルが前方一致、又は後方一致する番組どうしを同じシリーズに属する番組として検出する。

検出の例を図3に示す。一致を検証する文字数は比較する二つのタイトルで短い方の文字列長とする。この比較により、例えば“Breakfast News”と“Breakfast News Ski Special”を、同じシリーズに属する番組として検出する。

しかし、シリーズ番組のタイトルには様々なものがあり、上記の方法だけではうまく検出できないケースもある。例えば、同じシリーズに属する番組として“Brunch Story Episode 1”と“Brunch Story Episode 2”を比較する場合、これらは文字列長が同じで、末尾の数字だけが異なるため、一致しない。そのため、別のシリーズに属する番組であると認識されてしまい、同じシリーズ番組として検出できない。





このような誤認を避けるため、図4に示すように、一致を検証する前にタイトルの正規化処理を行う。正規化処理では、例えば“Episode”という文字列の後に数字が続く場合など、シリーズ番組のタイトルとして頻出する部分文字列を除外する。

あらかじめ、シリーズ番組のタイトルで頻出する部分文字列について、番組メタデータを対象として実態調査を行う。この調査から得られた部分文字列のパターンに基づいて、正規化処理のためのルールを作成する。

## 4 実験

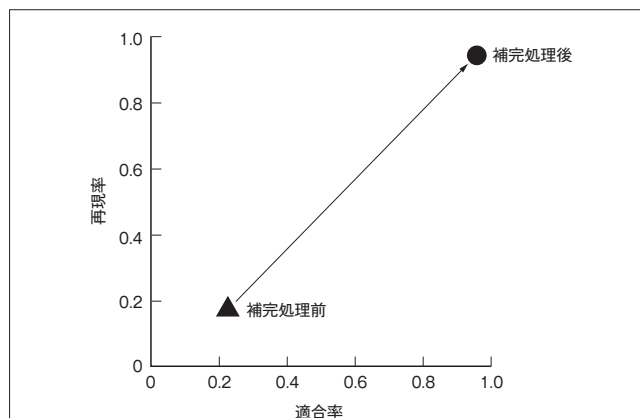
### 4.1 実験方法

実験対象は、欧州の4か国、1か月分のTV番組メタデータである。まず実験データを調査して正解データを作成する。国ごとに、一般的なシリーズ番組タイトルを約30タイトル挙げて、番組メタデータのタイトル、内容説明、クレジット（出演者やスタッフなどのリスト）の目視確認、及び番組ホームページの調査によって、各シリーズに属する番組を集め、正解データとする。

メタデータ補完処理後と比較するため、補完処理前の番組メタデータが持つIDを使ってシリーズ番組を検出し、検出精度を示す適合率と再現率を算出する。IDが付いていない番組メタデータについては、シリーズ番組ではない単独の番組として扱う。

次に、メタデータ補完処理を実行して各番組メタデータに新たなIDを付加する。このIDを使ってシリーズ番組を検出し、適合率と再現率を算出する。

ここで適合率は、検出した番組のうち、正解の番組の割合、すなわち検出誤りの少なさを表す。適合率が低い場合、本来は推薦すべき番組を、習慣的に視聴しているシリーズ番組であると誤認して推薦結果から除外するケースが増える。再現率は、正解の番組のうち検出した番組の割合、すなわち検出漏れの少なさを表す。再現率が低い場合、習慣的に視聴しているシリーズ番組であるとして本来は推薦結果から除くべき番組



を、別のシリーズ番組であると誤認して推薦結果から除外できないケースが増える。

補完処理前の状態では、あるシリーズに属する番組メタデータのいずれにもIDが付いていないケースが頻出した。このケースの適合率と再現率はともにゼロとした。

### 4.2 実験結果

メタデータ補完処理前後の適合率と再現率について、4か国の平均値を図5に示す。補完処理前はF値が0.20で、シリーズ番組の大部分を正しく検出できず、目視でも実用的な推薦精度ではなかった。一方、補完処理後はF値が実用レベルの0.95で、シリーズ番組を正しく検出でき、目視でも実用的な推薦精度が得られた。

## 5 TVへの適用

開発したメタデータ補完技術を用いた番組推薦エンジンを、2014年春に欧州市場向けに商品化したTVの自動録画機能に適用した。この機能では、番組推薦エンジンが推薦した番組を自動録画できる。録画された番組は、録画番組の統合表示アプリケーションのMediaGuide Replayで表示される。

MediaGuide Replayの画面例を図6に示す。各番組のサムネイルを格子状に並べ、1行目に録画予約番組、2行目に今週の録画番組、3行目に先週の録画番組というように、放送日時順に表示する（図6-①）。また、ユーザーが選択した番組（図6-②）の概要を画面上部に表示する（図6-③）。更に、自動録画した番組には“AUTO”のアイコンを表示する（図6-④）。

メタデータ補完技術によって、ユーザーが習慣的に視聴しているシリーズ番組は自動録画対象から除外できる。これによって、ユーザーにとって未知で、嗜好に合った番組が自動録画され、MediaGuide Replayに表示される。



図6. MediaGuide Replayのデモ画面例 — メタデータ補完技術により、よく知っているシリーズ番組は除かれ、未知の番組が自動録画されて画面上に並ぶ。  
Example of MediaGuide Replay display using metadata complementation technology

## 6 あとがき

当社は、番組推薦エンジン向けのメタデータ補完技術を開発し、検証実験によってその有効性を確認した。この技術を適用した番組推薦エンジンを、2014年春に欧州市場向けに商品化したTVの自動録画機能に組み込んだ。

今後は推薦エンジンの更なる精度向上に向けて、推薦結果に対するユーザーからのフィードバックを活用する方法について検討していく。

## 文 献

- (1) Swearingen, K. ; Sinha, R. Beyond Algorithms: An HCI Perspective on Recommender Systems. ACM SIGIR Workshop on Recommender Systems. 13, 5-6, 2001, p.393 - 408.
- (2) 村上知子 他. 推薦の意外性向上のための手法とその評価. 人工知能学会論文誌, 24, 5, 2009, p.428 - 436.



小川 修太 OGAWA Shuta

パーソナル&クライアントソリューション社 ライフスタイルソリューション開発センター クラウド技術開発部主務。推薦エンジンの開発に従事。

Lifestyle Solutions Development Center



西本 喜則 NISHIMOTO Yoshinori

パーソナル&クライアントソリューション社 ライフスタイルソリューション開発センター クラウド技術開発部。クラウドアプリケーションの開発に従事。

Lifestyle Solutions Development Center



大盛 善啓 OHMORI Yoshihiro

パーソナル&クライアントソリューション社 ライフスタイルソリューション開発センター クラウド技術開発部グループ長。スマートTVの開発に従事。電子情報通信学会, 情報処理学会会員。

Lifestyle Solutions Development Center