D-013

Webページへのタグ付与による類似ユーザ群に共通した意味情報の抽出

Extracting An Ontology Common to Similar Users with Tagging to Web Pages

伊藤 真也† 小河 真之‡ 原田 史子† 島川 博光†
Masaya Ito Masayuki Ogawa Fumiko Harada Hiromitsu Shimakawa

1. はじめに

嗜好や環境が同じ人同士であるほど，言葉が頻繁にやり取りされる過程で使用されている言葉が変化したり暗黙的な意味が付与される．そのため，同一の言葉であってもその人が所属するグループによって，言葉の意味は異なる．例えば，“情報推薦”という言葉を考える．甲が所属するグループにおいて“情報推薦”という言葉は，検索エンジンなどの Pull 型のシステムとポップアップ広告などの Push 型のものの双方を含む意味で使用しているとする．一方，乙が所属するグループでは，“情報推薦”という言葉を前者のみを指すものとして使用しているとする．このとき，甲と乙の間で“情報推薦”という言葉を使用すると甲と乙の間で食い違いが発生する．このように，嗜好や環境が異なるグループ間で情報をやり取りする場合，言葉の意味のずれの問題は無視できない．

本論文では，言葉の意味のずれを解決する変換機構を実現するために，ある類似ユーザ群内で用いられる言葉の階層関係および類似関係で定義される意味情報を抽出する手法を提案する．本手法では，類似ユーザ群という比較的小規模なグループから意味情報を抽出することによって，言葉の意味のずれや辞書では定義されていない暗黙的な意味を抽出できる．

2. 言葉の変換機構と意味情報

嗜好や環境が似通ったグループ内では，言葉の意味はほとんど同一である．本論文では，言葉の意味のずれがない嗜好や環境が似通ったグループを類似ユーザ群と呼ぶ．言葉の変換機構を用いることで，異なる類似ユーザ群間で，使用されている言葉のどの言葉とどの言葉が同じ意味であるかを判別できる．1. で示した例において，甲が所属するグループでは Pull 型での情報の推薦を“Pull 推薦”と呼んでいた場合，乙が使用した“情報推薦”という言葉は，変換機構を用いて，甲に“Pull 推薦”と伝えられる．言葉の意味は，そのグループ内で用いられる他の言葉との階層関係や類似関係を用いて表せる．文献 [1] では，知識共有を目的としてオントロジーを自動構築している．ここで分析されている言葉の狭義語，広義語および同義語の関係は，本手法における言葉の階層関係および類似関係に相当し，これらを言葉の意味情報と定義する．変換機構を実現するためには，各グループにおける言葉の意味を把握し，言葉の意味情報を抽出する必要がある．

3. 類似ユーザ群内における意味情報の抽出

3.1 提案モデル

本論文では，言葉の意味のずれを解決する変換機構を実現するために，ある類似ユーザ群内で用いられる言葉の意味情報を抽出する手法を提案する．本手法では，あるWebページを再度閲覧するために，普段からユーザがWebページにタグを付与し，整理している環境を想定する．意味情報は，以下 (1)-(4) にしたがって抽出する．

(1) 各ユーザによる Web ページへのタグ付け
(2) 各ユーザのタグによる Web ページ分類の抽出
(3) 類似ユーザ群内における Web ページの分類統合
(4) 分類された Web ページのタグから言葉の意味情報の抽出

3.2 タグ付けによる Web ページの整理

類似ユーザ群内の言葉の意味を抽出するためには，まず各ユーザが使用する言葉を知る必要がある．本手法では，ユーザがブックマークした Web ページに付与されたタグを用いて，個人が使用する言葉を収集する．このタグは，ソーシャルブックマークサービスのように複数のユーザが各 Web ページに対して協調的にタグを付与するものとは異なり，ユーザ個人が各 Web ページに対して，独立にタグを付与する．Web ページにタグ付けしたイメージを図 1 に示す．ここであるひとりのユーザがタグを付与した Web ページの集合をそのユーザのブックマーク集合と呼ぶ．図 1 は，ユーザ A とユーザ B のブックマーク集合が等しい場合を示している．

ユーザがタグ付けによって Web ページを整理することで，タグ名とそのタグによって整理された Web ページの集合が得られる．個人が整理するために付与しているタグであるので，同句異義語のタグ名は発生しないと考えられる．類似ユーザ群は，嗜好や環境が似通っているので，類似ユーザ群に属する類似ユーザ群には，同一の Web ページが含まれると考えられる．

3.3 Web ページに付与されたタグ情報の整理

個人が Web ページに対して付与したタグを基にタグ情報を整理する．本手法では，ブックマーク集合内の任意のふたつの Web ページからなる組合せに対し，付されたタグ群の一致性を計算する．

Web ページが n あるとき，Web ページの組合せは ${}_nC_2$ できる．各組合せに対して 2 ページ内で一致するタグの個数を一致度と定義する．例えば，図 2 のユーザ A では，Web ページ①と②のタグ“A1”が一致している．よって，一致したタグの数は 1 になる．図 2 に示した，ページの組合せと一致度を整理した表を分類情報と呼ぶ．

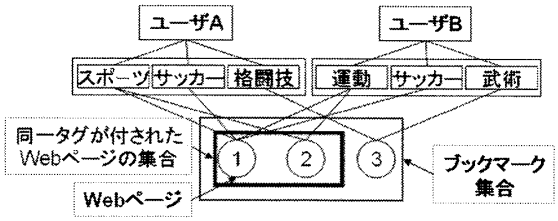

図 1: タグ付けによる Web ページの整理

†立命館大学情報理工学部
‡立命館大学大学院理工学研究科

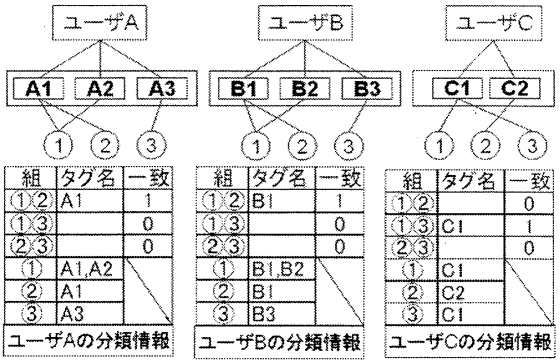

図 2: タグ付けによる整理から得られる個人の分類情報

3.4 個人の分類情報の統合

ユーザの分類情報を統合し，類似ユーザ群の分類情報を生成する．ある類似ユーザ群の任意のふたつの Web ページに対して，一致度の大きいユーザの割合が大きいほど，これらの Web ページの関連性は強いといえる．本論文では，類似ユーザ群内のユーザ同士の分類情報を統合するさいの指標となる値を結合度と呼ぶ．結合度は，ある Web ページの組でのタグの一致度が 1 以上のユーザの割合を示す．ここで，ある Web ページの組の結合度を C，一致度が 1 以上のユーザの数を W，全ユーザ数を U とする．結合度 C は $C = \frac{W}{U}$ と表せる．この結合度を基に個人の分類情報を統合する．そして，統合した分類情報から類似ユーザ群の意味情報を抽出する．分類情報の統合の手順は以下のとおりである．

I. 結合度の低い Web ページの組の発見

結合度が閾値 α 以下のものは結合度が低いと判断する．例えば，図 3 では閾値を 1/3 とすると①と③，②と③の組が閾値以下のため，結合度が低いと判断できる．

II. 結合度の低い組に関連するタグの削除

I において，結合度が低いと判断された組のタグを削除する．例えば，図 3 では，結合度が低いと判断された①と③の組に付与されているタグ “C1”を削除する．

III. 共通タグの発見

結合度が閾値 α より高く，結合度が高いと判断された組に共通するタグ名を見つける．例えば，図 3 の場合，①と②の結合度が閾値 1/3 より高く，共通したタグ “A1”，“B1”を発見できる．

IV. 共通タグの取り出し

III の手順の中で発見された，共通タグを取り出す．

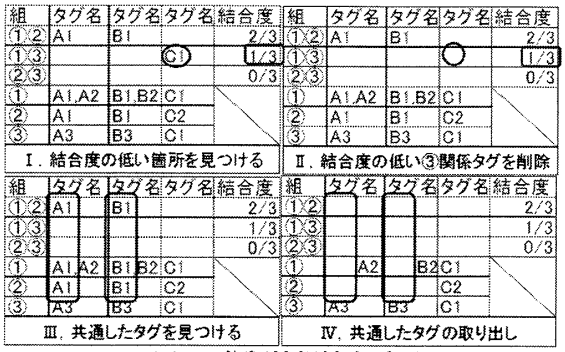

図 3: 分類情報統合手順

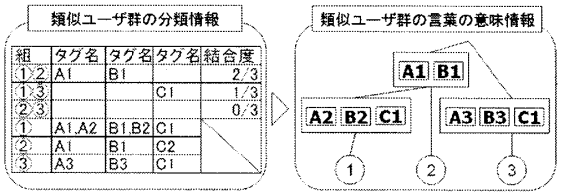

図 4: 分類情報から言葉の意味情報への変換

ここで取り出したタグ名は，類義語として扱われる．例えば，図 3 の場合，“A1”，“B1”が取り出されているので “A1”，“B1”は類義語となる．そして，この処理によって {“A1”,“B1”} がひとつのノードとして意味情報に追加される．このノードによって，①と②がひとつの集合として分割される．

V. 分割した集合に I–IV の処理をする

分割した集合がさらに分類できる場合，I –IV の処理をする．図 3 の場合は分割されたページ群 {①，②} に I –IV の処理をする．このとき，{①，②} で取り出された共通するタグを {“A1”,“B1”} を各ページから取り除き，残ったタグを用いて一致度を予め再計算する．{①，②} への I –IV の処理によってできたノードは，ノード {“A1”,“B1”} の子として意味情報へ追加される．集合がそれ以上分割できなくなったとき，分類情報の統合は完了となる．処理が完了した結果を図 4 に示す．

以上により，類似ユーザ群に共通した意味情報が抽出できる．本手法で抽出した意味情報から言葉の意味のずれを解決できる変換機構を作成できる．

4. 関連研究

Wikipedia と Folksonomy タグによって，オントロジーの構築支援を試みている手法 [2] が提案されている．この手法では，ある分野についての言葉の意味を定義する領域オントロジーを構築できる．しかし，比較的小規模なグループにおける言葉の意味のずれは考慮されていない．本手法では，類似ユーザ群という言葉の意味のずれのないグループから意味情報として言葉の使い方を抽出することによって，比較的小規模なグループにおける言葉の意味のずれに対応できる．

5. おわりに

本論文では，異なる嗜好や環境のグループ間における言葉の意味のずれを解消するために，類似ユーザ群に共通した意味情報を抽出する手法を提案した．今後は提案手法を実現するシステムを実装し，閾値の設定，有用性を検証する予定である．

参考文献

[1] 内田英里，石野武志：オントロジーの自動構築に関する基礎的研究: 人工知能学会，セマンティック Web とオントロジー研究会,pp.05-1–05-10,2003.

[2] 手島拓也ほか:日本語 Wikipedia マイニングと Folksonomy タグに基づく領域オントロジー構築支援, JSAI2007,1D2-5, 2007