

三つ組を構成するデータの類似度の分類と計算

山河絵利奈[†] 田島 敬史[†]

[†] 京都大学大学院情報学研究科 〒 606-8501 京都府京都市左京区吉田本町

E-mail: †yamakawa@dl.soc.i.kyoto-u.ac.jp, ††tajima@i.kyoto-u.ac.jp

あらまし データの類似度は様々なところに活用されている。データにはさまざまな構造があるが、そのすべてに確立された類似度の計算方法が存在するわけではない。本論文では、ユーザ・コンテンツ・評価、あるいは、ワーカ・タスク・回答などの三種類の値が組になる場合の三つ組データを想定し、そのようなデータに対し類似度を求め利用することを考える。三つ組データに対し望ましい類似度は、そのデータを持つ特徴によってさまざまな種類がある。そこで、三つ組データに対して定義する類似度の性質の分類を行い、それぞれに対応する類似度の定義を提案する。さらに、分類の一部に対し、提案する定義に基づく類似度が有用なものであること、また提案する複数の類似度の定義のうち分類上該当する類似度の定義が最も適切であることを示す。

キーワード 類似度、クラウドソーシング、協調フィルタリング、アノテーション

1 はじめに

近年、様々なデバイスの発達やインターネットの普及と共に、数多くのデータの電子化と蓄積が進んでいる。それに伴い、得られるデータの構造も多様化してきた。

様々な構造のデータのうち、二つのオブジェクトの間の関係を表すデータは通常のグラフで表現でき、そのようなデータについては広く研究されているが、二つではなく三つのオブジェクトの間の関係を表すデータも様々な応用において用いられている。本研究では、そのような三つのオブジェクトの間の関係を表すデータを三つ組データと呼び、その利用に注目した。

データから算出できる値の一つとして類似度が挙げられる。類似度の計算方法として、各オブジェクトの特徴を活かして類似度を計算することが挙げられる。また、オブジェクトの種類に関わらずグラフ構造から類似度を計算する方法として、SimRankがある。SimRankは、類似するオブジェクトを近傍に持つオブジェクト同士はやはり類似しているという考え方をもとに類似度を計算する。これにより、直接隣接するオブジェクトに共通部分がないオブジェクト同士でも、他のオブジェクトを経由する共通のオブジェクトへのパスが存在すれば類似度を計算できるという特徴がある。SimRankは、通常のグラフあるいは二部グラフに対して利用することができる。しかし、前述のような三つ組データ構造に対し、構造を活かした類似度の計算方法は確立されていない。本研究では、SimRankの計算方法を拡張し、三つ組データの集合における類似度の定義を提案する。

また、三つ組データは様々なあり、データの収集方法や特徴によって類似度に反映させたい考え方も異なる。本研究では、複数の類似度の計算方法を提案した上で、三つ組データの種類に応じてどの計算方法が適切かを検討する。

さらに、三つ組データの種類により、提案手法による類似度の利用方法も多数考えられる。本研究では、三つ組データの一

例として、映画のレビューを取り上げる。映画のレビューデータは、(レビュー、映画、レビュー値)の三つのオブジェクトで構成されるデータの集合であると考えられる。レビューデータの利用方法の一つとして、レビューへの映画の推薦が挙げられる。推薦には、過去のレビューデータを用いてレビュー値の予測が行われる。推薦手法の一つに協調フィルタリングが挙げられる。協調フィルタリングでは、レビュー同士の相関係数を定義して計算し、相関係数の高いレビューのレビューを重視するようにレビュー値の予測を行う。この相関係数は同じ映画に同じようにレビュー値を与えるレビュー間で高くなるように計算されているため、提案手法で計算するレビューの類似度を利用することでも同様に計算できると考えられる。提案手法では、直接同じ映画をレビューしていなくても類似度を計算できるため、より精度が高く予測できるのではないかと考えられる。

他にも、三つのオブジェクトで構成されるデータとして、クラウドソーシングのデータが挙げられる。例えば、機械学習の教師データにも用いられる画像などのアノテーションタスクであれば、(ワーカ、画像、選択肢)の三つのオブジェクトで構成されるデータの集合であると考えられる。

クラウドソーシングでは、時間的・金銭的成本を鑑み、専門家ではない不特定多数の一般人(ワーカ)に作業が依頼される[1]。不特定多数の人にタスクを依頼するという特性上、回答の精度が保証されず、得られるデータの品質にばらつきが生じてしまう。そのため、単一のタスクを多数のワーカに依頼し、その回答を統合することでデータの品質を保証することがある。このような方法をとる際には、一定の人数ごとにワーカのグループを編成し、そのグループ内で回答の統合、すなわち多数決による最終的な回答の選定を行うことが多い。この場合、グループ編成次第で多数決の結果得られるデータの精度が変わるため、どのようにグループ編成を行うかが重要となる。今回は、あらかじめ全ワーカが正解が既知である複数のタスクに取り組んでいると仮定する。その際、正答率を高くするための最も正確なグループ編成方法は、全ての編成パターンに対してあ

らかじめ取り組んでもらったタスクの回答をもとに正答率の期待値を計算し、その値が最も高かった編成を採用する方法である。しかしこれは、計算量が膨大になってしまい、現実的には使用できない方法である。

そこで、正確に正答率の期待値を計算するのではなく、正答率が高くなるように予測してグループ編成を行うことが求められる。その方法の一つに、ワーカ間の距離を利用して、距離が遠いワーカ、すなわち答え方が似ておらず同じ間違えをする確率が低いワーカを同じグループに入れるようにすることで、正答率の高くなるような編成をするという考え方がある [2]。本研究では、距離として SimRank を拡張した類似度を用いてグループ編成を行うことを提案する。

本論文は以下の構成を取る。まず、第 2 章で関連研究について述べる。続いて第 3 章では三つ組データの類似度を計算する方法を提案し、第 4 章では具体的に三つ組データを想定した上で、提案した複数の手法のうちいずれが適切かを述べる。第 5 章では映画レビューデータにおいて、提案手法を用いてレビュー値の予測を行う実験と実験結果について説明し、第 6 章ではその結果を踏まえた考察を述べる。第 7 章では提案手法を利用し、クラウドソーシングデータを用いてグループ編成を行う実験と実験結果について説明し、第 8 章ではその結果を踏まえた考察を述べる。最後に第 9 章で、結論と今後の展望を述べる。

2 関連研究

2.1 SimRank

本研究では、ワーカ間の類似度を計算する際に、Jeh らによって提案された SimRank [3] の計算方法を拡張して用いる。SimRank の計算方法は、第 3 章にて説明する。

SimRank は近年でも様々な場面で拡張・利用されている。Symeonidis ら [4] は、ニュース推薦システムに対して SimRank を用いている。ニュース推薦システムでは、時間の情報を用いた推薦が重要となる。そこで、セッションと呼ばれるユーザー操作のグループを考慮した、時間の経過とともに変化する異種グラフ (Heterogeneous Graph) に対して使用できるように SimRank を拡張している。

Dey ら [5] は、スケールフリー性に着目した SimRank の拡張を提案している。実際の Web グラフは次数分布がべき乗則に従い、そのようなグラフ上ではスケールフリー性が重要なものの、従来の研究では近接する頂点が重視されておりスケールフリー性については考慮されていなかった。Dey ら [5] は P-SimRank として、そのようなグラフに対して SimRank を適用できるような拡張を提案している。

2.2 類似度に基づく評価値の予測手法

第 5 章で行う、映画レビューデータを用いたレビューの予測実験に関連する研究として、Resnick ら [6] による協調フィルタリングがある。この研究では、ニュースサイトにおいて読者の好むニュースを予測するために使用されている。協調フィルタ

リングは、あるユーザのあるアイテムへの評価値を、他のユーザのアイテムへの評価値をもとに予測する方法である。あるユーザと同じ傾向で評価するユーザが高く評価していること、また反対の傾向で評価するユーザが低く評価していることを反映させて予測を行う。協調フィルタリングにおける予測の方法を説明する。まず、ユーザ間の傾向として、すべてのユーザ間で相関係数を求める。ユーザ A, B 間の相関係数 r_{AB} は以下の通り計算する。

$$\begin{aligned} r_{AB} &= \frac{Cov(A, B)}{\sigma_A \sigma_B} \\ &= \frac{\sum_{i \in I_{AB}} (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_{i \in I_{AB}} (A_i - \bar{A})^2} \sqrt{\sum_{i \in I_{AB}} (B_i - \bar{B})^2}} \end{aligned} \quad (1)$$

ここで、 \bar{A} はユーザ A の評価の平均点、 I_{AB} は A, B 共に評価したアイテム、 U_i はアイテム i に評価をつけているユーザーの集合、 A_i はユーザ A のアイテム i の評価、 B_i はユーザ B のアイテム i の評価、 \bar{B} はユーザ B の評価の平均値である。そして、これらをもとに、以下のようにユーザ A のアイテム i の評価の予測 \hat{A}_i を計算する。

$$\hat{A}_i = \bar{A} + \frac{\sum_{B \in U_i} (B_i - \bar{B}) r_{AB}}{\sum_{B \in U_i} |r_{AB}|} \quad (2)$$

さらに、Amazon.com における推薦では、アイテムを基準にした協調フィルタリングが使用されている [7]。Amazon.com は当時 1000 万の顧客と商品を擁しており、この手法は巨大なデータの上で高い精度で推薦を行うために考えられた [8]。また、Resnick らによる協調フィルタリング [6] などのそれまでの推薦手法と異なり、そのアイテムがユーザのどの行動による推薦なのか、すなわちどのアイテムを購入したことにより推薦されているのかがわかる [8]。アイテム間協調フィルタリングではまず、アイテムが同じ人々に購入される傾向をもとに、アイテム間の類似度を計算する。アイテム X を購入したユーザの集合を U_X として、アイテム X, Y 間の類似度 r_{XY} の計算式は次の通りである。

$$r_{XY} = \frac{|U_X \cap U_Y|}{\sqrt{|U_X| |U_Y|}} \quad (3)$$

次に、アイテムを推薦したいユーザが興味を持っているとわかっているアイテムそれぞれに対し、類似するアイテム、すなわち同じ人に購入されやすいアイテムを選出する。そして、選ばれたアイテムを組み合わせて、ユーザへ推薦を行う。

2.3 クラウドソーシングの品質向上

第 7 章では、クラウドソーシングで収集したデータを使用した実験を行う。クラウドソーシングでは不特定多数のワーカがタスクに取り組む関係上、品質を向上させるための工夫が必要不可欠であり、そのため品質向上についての様々な研究がなされている。Dawid ら [9] は複数の医師による診断の結果から正しい診断結果を推定する方法を提案している。推定には EM ア

ルゴリズム [10] を利用した手法が提案されており、この研究はクラウドソーシングの品質管理の研究の基礎ともなっている。

多数決と回答状況に注目した研究もなされている。Joglekar ら [11] は、多数決において、回答の一致状況をもとに、ワーカの誤り率とその信頼区間を推定する方法を提案し、精度の悪いワーカを除くことにより品質の向上に寄与した。松田ら [2] は、同じく多数決を行うケースで、回答の傾向が異なるワーカ同士をグループにすることで精度を上げる方法を提案している。回答傾向の近い人が同じクラスタになるようにクラスタリングした上で、異なるクラスタのワーカ同士とグループとすることで精度の高い多数決を行う。クラスタリングの手法としては、ワーカの回答データをベクトル化しその距離を基にしたクラスタリングと、ワーカ間の親和度という概念を定義し利用したクラスタリングの 2 種類が提案されている。

Welinder ら [12] は、画像アノテーションタスクのデータをもとに、画像のグルーピングを行う方法を提案している。提案手法は、画像の形成プロセスのモデルと画像アノテーションのプロセスに基づいている。Welinder らは、質的に異なる画像のセットを発見するほか、異なるスキルや知識を持つワーカのグルーピングも行っている。本研究では、SimRank の計算方法を拡張することにより画像のラベリングタスクには特化しない類似度の推定方法を提案する。

3 提案手法

本章では、SimRank の計算方法を拡張して三つ組データの類似度を求める方法を提案する。

3.1 想定するデータ構造

まず、本研究で想定するデータ構造を定める。本研究では、データが三つのオブジェクトが組となり構成されていると仮定する。さらに、区別のため、その三つのオブジェクトを (ワーカ, タスク, 回答) であると仮定する。例えば、既定のアノテーションプリセットから選択する形式の画像アノテーションタスクは、(ワーカ, 画像, 選んだアノテーション) という形で当てはまり、映画の 5 段階レビュータスクは、(レビュー, 映画, レビュー値) という形で当てはまる。

3.2 SimRank

SimRank [3] は、Jeh らによって提案された有向グラフ上でノード間の類似度を計算する方法である。まず有向グラフ $G = (V, E)$ を考える。 $v \in V$ に対し、 $I(v)$ を v への入ノード集合とし、個々の v への入ノードを $I_i(v), 1 \leq i \leq |I(v)|$ とする。 2 頂点 $x_1, x_2 \in V$ の類似度を $s(x_1, x_2)$ と表すと、 $s(x_1, x_2)$ は以下のように定義されている。

$$s(x_1, x_2) = \begin{cases} 0 & (|I(x_1)| \neq 0 \text{ or } |I(x_2)| \neq 0) \\ 1 & (x_1 = x_2) \\ C \times \frac{\sum_{i=1}^{|I(x_1)|} \sum_{j=1}^{|I(x_2)|} s(I_i(x_1), I_j(x_2))}{|I(x_1)||I(x_2)|} & (x_1 \neq x_2) \end{cases} \quad (4)$$

ここで、 C は $C \in [0, 1]$ の定数である。 SimRank は、収束するまで反復して計算することにより値を求める。 C は計算を収束させるための収束定数である。反復計算に関しては第 3.5 節で述べる。

SimRank は、三つの条件式により計算式が分かれている。まず、一つ目の条件は $(|I(x_1)| \neq 0 \text{ or } |I(x_2)| \neq 0)$ である。これは、頂点 x_1 もしくは x_2 がグラフ上で孤立点であることを表す。この時、 $s(x_1, x_2) = 0$ となり、全く類似しないという定義となっている。

次に、二つ目の条件である $x_1 = x_2$ は、頂点自身との類似度の定義を表しており、 $s(x_1, x_1) = 1$ 、すなわち自分自身との類似度は 1 となる。

最後に、三つ目の条件式である。これは、上記二つの条件に当てはまらない場合に適用される。

3.3 構造をもとにした拡張

SimRank は通常のグラフに対する類似度計算方法であるが、異なる構造のグラフに対する拡張の一つとして、二部グラフに対する計算方法が定義されている。

二部グラフは、グラフ全体を $G = (V, E)$ として $V_1 \cup V_2 = V, V_1 \cap V_2 = \phi$ である。グラフは、 $x_1 \in V_1, y_1 \in V_2$ であるような (x_1, y_1) の形のデータの集合であると捉えられる。 $I(x)$ を x との組 (x, y) がデータとして存在するような y の集合、 $O(y)$ を y との組 (x, y) がデータとして存在するような x の集合とする。また、 $I(x)$ の個々のデータ y を $I_i(x), 1 \leq i \leq |I(x)|$ 、 $O(y)$ の個々のデータ x を $O_i(y), 1 \leq i \leq |O(y)|$ とする。この時の類似度は次のように定義されている。

$$s_1(x_1, x_2) = \begin{cases} 0 & (|O(x_1)| \neq 0 \text{ or } |O(x_2)| \neq 0) \\ 1 & (x_1 = x_2) \\ C_1 \times \frac{\sum_{i=1}^{|O(x_1)|} \sum_{j=1}^{|O(x_2)|} s_2(O_i(x_1), O_j(x_2))}{|O(x_1)||O(x_2)|} & (x_1 \neq x_2) \end{cases} \quad (5)$$

$$s_2(y_1, y_2) = \begin{cases} 0 & (|I(y_1)| \neq 0 \text{ or } |I(y_2)| \neq 0) \\ 1 & (y_1 = y_2) \\ C_2 \times \frac{\sum_{i=1}^{|I(y_1)|} \sum_{j=1}^{|I(y_2)|} s_1(I_i(y_1), I_j(y_2))}{|I(y_1)||I(y_2)|} & (y_1 \neq y_2) \end{cases} \quad (6)$$

C_1, C_2 はそれぞれ $C_1, C_2 \in [0, 1]$ の定数であり、収束のための定数である。

式 (5) 及び式 (6) を見ると、二部グラフにおける類似度は、 V_1 に属する頂点同士、及び、 V_2 に属する頂点同士に対して計算するように定義されていることがわかる。また、この時、 V_1 に属する頂点同士の類似度をそれらの頂点が接続する V_2 に属する頂点の類似度で、 V_2 に属する頂点同士の類似度をそれらの頂点に接続する V_1 に属する頂点の類似度で計算するような構造となっていることがわかる。

本研究で類似度を計算したいデータは、(ワーカ、タスク、回答) の三つ組で構成されていると考える。この時、(ワーカ、タスク、回答) を (ワーカ、(タスク、回答)) という形で捉え、(タスク、回答) という組を 1 つのデータと見ることで、三つ組を二部グラフの構造に落とし込むことができる。これにより、二部グラフの SimRank 計算の式 (5)(6) を利用した類似度計算が可能となる。

しかしこの方法では、データ数が少ない時に疎なグラフ構造となり、精密に類似度を計算できない。また、ワーカの類似度の計算に、タスクの類似度、回答の類似度の値を用いないこととなり、計算に使用する情報量も少なくなってしまう。よって、本研究では異なる拡張を提案する。

二部グラフにおける SimRank 計算式 (5)、及び式 (6) に戻って考える。二部グラフの場合前述の通り、 V_1 に属する頂点同士の類似度をそれらの頂点が接続する V_2 に属する頂点の類似度で、 V_2 に属する頂点同士の類似度をそれらの頂点に接続する V_1 に属する頂点の類似度で計算する。同様に、三つ組データ (ワーカ、タスク、回答) における類似度の計算でも、(ワーカ、タスク、回答) を (x_1, y_1, z_1) として、ワーカの類似度 $s_1(x_1, x_2)$ 、タスクの類似度 $s_2(y_1, y_2)$ 、回答の類似度 $s_3(z_1, z_2)$ をそれぞれ定義し、各類似度の計算の中で残り 2 種類の類似度を使用することを考える。これにより、疎なグラフ構造を作る必要もなく、計算にもより多くの情報量を含めることができると考えられる。

さらに具体的な計算式を考える。式 (5)、及び式 (6) はどちらも三つの条件式に分かれて定義されている。まず一つ目は、どちらかの頂点が孤立点の場合の計算方法である。この条件は、提案手法でもそのまま踏襲することとする。すなわち、ワーカの類似度計算の場合、全く回答データのないワーカは、その他全てのワーカとの類似度は 0 となる。次に、二つ目は自分自身との類似度を定義している。この条件も、そのまま踏襲する。すなわち、ワーカの類似度計算の場合、ワーカのそのワーカ自身との類似度は 1 となる。

拡張が必要なのは、それ以外の場合に当てはまる三つ目の条件式である。まず、 $I(x)$ に当てはまるものが明確ではない。次に、新たな $I(x)$ の定義に合わせ $s((y_1, z_1), (y_2, z_2))$ についても検討する必要がある。

まず、 $I(x), I(y), I(z)$ を考える。二部グラフの場合、グラフ全体を $G = (V, E)$ とすると、頂点集合は $V_1 \cup V_2 = V, V_1 \cap V_2 = \phi$ となるように分けられる。 $x \in V_1, y \in V_2$ とすると、グラフは (x, y) の形のデータの集合であると捉えられる。この時、 $I(x)$ は (x, y) というデータが存在するような y の集合である。同様に、

$O(y)$ も (x, y) というデータが存在するような x の集合を表す。これを三つ組データに拡張することを考える。三つ組データの場合、頂点集合が $V_1 \cup V_2 \cup V_3 = V, V_1 \cap V_2 = \phi, V_2 \cap V_3 = \phi, V_1 \cap V_3 = \phi$ となるように分けられ、エッジ E が V_1, V_2, V_3 から 1 つずつノードを選んで繋ぐようなグラフを作るデータであると考えることができる。 $x \in V_1, y \in V_2, z \in V_3$ として、(ワーカ、タスク、回答) を (x, y, z) と表す。 $I(x)$ は (x, y, z) というデータが存在するような (y, z) の集合、 $I(y)$ は (x, y, z) というデータが存在するような (x, z) の集合、 $I(z)$ は (x, y, z) というデータが存在するような (x, y) の集合と定義する。

ここまでの検討をもとに、ワーカ同士の類似度を $s_w(x_1, x_2)$ 、タスク同士の類似度を $s_t(y_1, y_2)$ 、回答同士の類似度を $s_a(z_1, z_2)$ 、(タスク、回答) のような組同士の類似度を計算する関数を $s'((y_1, z_1), (y_2, z_2))$ で表すと、以下ようになる。

$$s_w(x_1, x_2) = \frac{C_w \times \sum_{(y_1, z_1) \in I(x_1)} \sum_{(y_2, z_2) \in I(x_2)} s'((y_1, z_1), (y_2, z_2))}{|I(x_1)| |I(x_2)|} \quad (7)$$

$$s_t(y_1, y_2) = \frac{C_t \times \sum_{(x_1, z_1) \in I(y_1)} \sum_{(x_2, z_2) \in I(y_2)} s'((x_1, z_1), (x_2, z_2))}{|I(y_1)| |I(y_2)|} \quad (8)$$

$$s_a(z_1, z_2) = \frac{C_a \times \sum_{(x_1, y_1) \in I(z_1)} \sum_{(x_2, y_2) \in I(z_2)} s'((x_1, y_1), (x_2, y_2))}{|I(z_1)| |I(z_2)|} \quad (9)$$

なお、 C_w, C_t, C_a はそれぞれ $C_w, C_t, C_a \in [0, 1]$ の定数であり、収束のための定数である。

3.4 データの特徴に対応した拡張

前節では、通常のグラフにおける SimRank の計算式と二部グラフにおける SimRank の計算式の構造と概念をもとに、三つ組データに対する類似度計算のベースとなる式を提案した。本節では、さらに、データの特徴を考慮して $s'((y_1, z_1), (y_2, z_2))$ の計算方法を複数提案する。なお本節では、ワーカの類似度を計算する際に使用する式 (7) をもとに拡張を考えるが、タスクの類似度を計算する際の式 (8) と回答の類似度を計算する際の式 (9) にも同様の拡張を考えられる。

$s_w(x_1, x_2)$ でワーカの類似度を計算する際、計算中に現れる関数 s' は (タスク、回答) の組み合わせ同士の類似度を計算する関数となる。(タスク、回答) の組み合わせが類似しているということは、タスクが類似しており、かつ回答も類似していることと考えることができる。タスクの類似度と回答の類似度はそれぞれ計算することができるため、これを利用してこの条件が成り立つように定めれば良い。これは提案するすべての手法で成り立つべき条件とする。

タスクのみが類似している場合、回答のみが類似している場

合、あるいはどちらも類似していない場合に関しては、さまざまなケースを想定することができる。これらのケースに合わせて、複数の計算方法を提案する。

なお、前述の通り、この $s'((y_1, z_1), (y_2, z_2))$ の計算では、タスク間の類似度とワーカ間の類似度を使用したい。そこで、

$$s'((y_1, z_1), (y_2, z_2)) = s''(s_t(y_1, y_2), s_a(z_1, z_2)) \quad (10)$$

として、(タスクの類似度, 回答の類似度) を引数に取るような関数 s'' について考える。

3.4.1 提案手法 1

最も単純な方法として、(タスク, 回答) の組み合わせを、タスクの類似度と回答の類似度の積で定める。これは、次のように数式化できる。

$$s''_1(y, z) = y * z \quad (11)$$

3.4.2 提案手法 2

式 (11) は、タスクの類似度 $s_t(y_1, y_2)$ と回答の類似度 $s_a(z_1, z_2)$ の両方が低い時にはワーカの類似度が低くなるように作用する。また、 $s_t(y_1, y_2)$ と $s_a(z_1, z_2)$ のどちらか一方のみが低い場合にも、積を取る関係上ワーカの類似度が低くなるように作用する。しかしこれは、必ずしも適切であるとはいえない。

例えば、ワーカが取り組むタスクを選べないような状況の場合、取り組んだタスクが似ていないことでワーカの類似度が低いとすることは不適切だと考えられる。よって、そのようなケースに対応するよう、タスクの類似度が低いときにはワーカの類似度への影響を減らすような式を考える。なお、タスクの類似度が高いときには、これまで通りラベルの類似度によってワーカの類似度にどのように作用させるかを定める。

類似度への影響を減らすとは、その類似度の平均値をとるということが考えられる。ここまですを踏まえると、以下の4つの条件が得られる。なお、ワーカの類似度の平均を $\overline{s_w}$ 、タスクの類似度の平均を $\overline{s_t}$ 、回答の類似度の平均を $\overline{s_a}$ とする。

- (1) $s''_2(1, 1) = 1$
- (2) $s''_2(1, 0) = 0$
- (3) $s''_2(0, 1) = \overline{s_w}$
- (4) $s''_2(0, 0) = \overline{s_w}$

ここで、シグモイド関数 $S(x)$ を使用することを考える。

$$S(x) = \frac{1}{1 + e^{-ax+b}} \quad (12)$$

a, b はパラメータである。標準シグモイド関数は $a = 1, b = 0$ であり、定義域としては全ての実数を取り値域が $0 < S(x) < 1$ となる、座標点 $(0, 0.5)$ を変曲点とした点対象な関数である。

s_w の平均は、孤立点との類似度が 0、自分自身との類似度が 1 という定義から、 s_w の平均が 0 や 1 になることは実際のデータではほぼ起こり得ず、 $0 < \overline{s_w} < 1$ であると考えられる。よって、 s'_2 の値域は $0 \leq s'_2 \leq 1$ であると考えられる。 x の値が何の時に $S(x) = \overline{s_w}$ となるようにするかはパラメータによって決められる。よってひとまず $s''_2(y, z) = S(f(y, z))$ という形にすることとし、新たに置いた関数 $f(y, z)$ の挙動を考える。

シグモイド関数を使用することにより、関数 f の取るべき条

件は s''_2 の条件よりも弱まって、次のようになる。

- (1) $f(0, 1) = f(0, 0)$
- (2) $f(1, 1) > f(0, 0)$
- (3) $f(0, 0) < f(1, 0)$

これは、シグモイド関数が $a > 0$ であれば単調増加すること、 $s''_2(1, 0) < s''_2(0, 1) = s''_2(0, 0) < s''_2(1, 1)$ であることから得られる。

まず、関数 f の一つ目の条件を満たすために、 $f(y, z) = y \times g(z)$ とする。これにより、 $f(0, z) = 0$ となるため、関数 f の一つ目の条件は満たされる。

$f(1, z) = g(z)$ より、関数 f の二つ目、三つ目の条件はそれぞれ、

2. $g(1) > 0$
3. $g(0) < 0$

となる。簡単のために

- 2'. $g(1) = 0.5$
- 3'. $g(0) = -0.5$

とすると、値域の幅が 1 となり、シグモイド関数を使用できる。シグモイド関数を使用する場合、パラメータの調整により $g(z) = 0$ となるような z の値を指定することができる。今回は、 $g(\overline{s_a}) = 0$ という形にすることとした。

ここまです十分に条件を定められたため、使用する 2 つのシグモイド関数のパラメータを計算する。まず、 $g(z)$ について考える。 $g(z)$ にはシグモイド関数を使用するので、

$$g(z) = \frac{1}{1 + e^{-a_g z + b_g}} \quad (13)$$

とおき、条件からパラメータを決定する。 $g(z)$ の取る条件は、 $g(1) = 0.5$ 、 $g(0) = -0.5$ 、 $g(\overline{s_a}) = 0$ であるが、この内 $g(1) = 0.5$ 、 $g(0) = -0.5$ はシグモイド関数の値域の関係上同時に満たすことができない。そこで、整数 c を用いて、 $g(1) > 0.5 - 10^{-c}$ 、 $g(0) > -0.5 + 10^{-c}$ という条件とすることで、同時に満たせるようにする。整数 c の取るべき値については後述する。シグモイド関数の値域が $0 < S(x) < 1$ で、 $g(z)$ の値域が $-0.5 < S(x) < 0.5$ であることから、 $g(z) = \frac{1}{1 + e^{-a_g z + b_g}} - 0.5$ とする。これにより、次の 3 式が得られる。

$$g(1) = \frac{1}{1 + e^{-a_g + b_g}} > 1 - 10^{-c} \quad (14)$$

$$g(0) = \frac{1}{1 + e^{b_g}} > 10^{-c} \quad (15)$$

$$g(\overline{s_a}) = \frac{1}{1 + e^{-a_g \times \overline{s_a} + b_g}} = 0.5 \quad (16)$$

まず、式 (16) より、

$$b_g = \overline{s_a} \times a_g \quad (17)$$

が得られる。これらを代入した上で、式 (14) より

$$a_g > \frac{1}{1 - \overline{s_a}} \log_e(10^c - 1) \quad (18)$$

が、式 (15) からは

$$a_g > \frac{1}{\overline{s_a}} \log_e(10^c - 1) \quad (19)$$

が得られる．式 (18) と式 (19) はいずれも a_g に関する条件式であるが， s_a の平均が 0.5 の時

$$\frac{1}{1-s_a} \log_e(10^c - 1) = \frac{1}{s_a} \log_e(10^c - 1) \quad (20)$$

であるので， s_a の平均が 0.5 より大きい時は式 (18) が，それ以外は式 (19) が採用される．

次に， $s_2''(y, z)$ を考える． $s_2''(y, z)$ は，

$$s_2''(y, z) = S(f(y, z)) \quad (21)$$

$$S(w) = \frac{1}{1 + e^{-a_s \times w + b_s}} \quad (22)$$

であり， $f(x)$ の値域が $-0.5 < f(x) < 0.5$ であることから，条件を式に代入すると次のとおりとなる．

$$(1) \quad S(0.5) = 1$$

$$(2) \quad S(-0.5) = 0$$

$$3.4. \quad S(0) = \overline{s_w}$$

これにより次の 3 式が得られる．

$$S(0.5) = \frac{1}{1 + e^{-a_s \times 0.5 + b_s}} > 1 - 10^{-c} \quad (23)$$

$$S(-0.5) = \frac{1}{1 + e^{a_s \times 0.5 + b_s}} > 10^{-c} \quad (24)$$

$$S(0) = \frac{1}{1 + e^{b_g}} = \overline{s_w} \quad (25)$$

まず，式 (25) より，

$$b_s = \log \frac{1 - \overline{s_w}}{\overline{s_w}} \quad (26)$$

が得られる．これらを代入した上で，式 (23) より

$$a_s > -2 \log \frac{\overline{s_w}}{(1 - \overline{s_w})(10^c - 1)} \quad (27)$$

が，式 (24) からは

$$a_s > 2 \log \frac{\overline{s_w} \times (10^c - 1)}{1 - \overline{s_w}} \quad (28)$$

が得られる．式 (27) と式 (28) はいずれも a_s に関する条件式であるが， s_w の平均が 0.5 の時 2 つは同じ式となる． s_w の平均が 0.5 より大きい時は式 (27) が，それ以外は式 (28) が採用される．

ここまでの議論により，次のような式を提案する．

$$s_2''(y, z) = \frac{1}{1 + e^{-a_s \times f(y, z) + b_s}} \quad (29)$$

$$f(y, z) = y \times \left(\frac{1}{1 + e^{-a_g \times (z - \overline{s_a})}} - 0.5 \right) \quad (30)$$

$$b_s = \log \frac{1 - \overline{s_w}}{\overline{s_w}} \quad (31)$$

$$a_s = \max \left(-2 \log \frac{\overline{s_w}}{(1 - \overline{s_w})(10^c - 1)}, 2 \log \frac{\overline{s_w} \times (10^c - 1)}{1 - \overline{s_w}} \right) \quad (32)$$

$$a_g = \max \left(\frac{1}{1 - \overline{s_a}} \log_e(10^c - 1), \frac{1}{\overline{s_a}} \log_e(10^c - 1) \right) \quad (33)$$

なお，整数 c が大きくなればなるほど， $s_2''(1, 1)$ は 1 に近づく．一方で，シグモイド関数の変曲点付近で急激に傾きが変化するので，変曲点周辺を除く多くの範囲で両端で同じような値を取るようになってしまう．本研究ではこのラバンスを鑑み， $c = 3$ とした．

3.4.3 提案手法 3

第 3.4.2 小節では，タスクの類似度が低いときにはワーカの類似度への影響を減らし，タスクの類似度が高いときには，ラベルの類似度によってワーカの類似度にどのように作用させるかを定めるような方法を提案した．これは逆に，ラベルの類似度が低いときにはワーカの類似度への影響を減らし，ラベルの類似度が高いときには，タスクの類似度によってワーカの類似度にどのように作用させるかを定めるという方法も考えることができる．この場合数式としては手法 2 の計算上でのタスクとラベルを入れ替える形で実現することができる．

$$s_3''(y, z) = \frac{1}{1 + e^{-a_s \times f(y, z) + b_s}} \quad (34)$$

$$f(y, z) = z \times \left(\frac{1}{1 + e^{-a_g \times (y - \overline{s_t})}} - 0.5 \right) \quad (35)$$

$$b_s = \log \frac{1 - \overline{s_w}}{\overline{s_w}} \quad (36)$$

$$a_s = \max \left(-2 \log \frac{\overline{s_w}}{(1 - \overline{s_w})(10^c - 1)}, 2 \log \frac{\overline{s_w} \times (10^c - 1)}{1 - \overline{s_w}} \right) \quad (37)$$

$$a_g = \max \left(\frac{1}{1 - \overline{s_t}} \log_e(10^c - 1), \frac{1}{\overline{s_t}} \log_e(10^c - 1) \right) \quad (38)$$

3.4.4 提案手法 4

第 3.4.2 小節及び第 3.4.3 小節では，タスクの類似度とラベルの類似度の一方をシグモイド関数にかけ，もう一方との積を取った．これらを組み合わせることで，双方ともにシグモイド関数にかけた上で積を取るという計算手法も考えることができる．まず提案手法となる式を構成した上で，その式を持つ意味を検討する．

まず， $f(y, z)$ を

$$f(y, z) = \left(\frac{1}{1 + e^{-a_1 \times (y - \overline{s_t})}} - 0.5 \right) \times \left(\frac{1}{1 + e^{-a_2 \times (z - \overline{s_a})}} - 0.5 \right) \quad (39)$$

$$a_1 = \max \left(\frac{1}{1 - \overline{s_t}} \log_e(10^c - 1), \frac{1}{\overline{s_t}} \log_e(10^c - 1) \right) \quad (40)$$

$$a_2 = \max \left(\frac{1}{1 - \overline{s_a}} \log_e(10^c - 1), \frac{1}{\overline{s_a}} \log_e(10^c - 1) \right) \quad (41)$$

とする．

この時， $0 \leq \frac{1}{1 + e^{-a_1 \times (a - \overline{s_t})}} \leq 1$ ， $0 \leq \frac{1}{1 + e^{-a_2 \times (b - \overline{s_a})}} \leq 1$ であり， $-0.25 \leq f(y, z) \leq 0.25$ となる．これを， $-0.5 \leq f(y, z) \leq 0.5$ とすれば， s_2'' や s_3'' の計算式に現れる $f(y, z)$ にそのまま当てはめて利用できるため， $f(y, z)$ を 2 倍して使用することとする．

これらをまとめると，次のような式が得られる．

$$s_4''(y, z) = \frac{1}{1 + e^{-a_s \times f(y, z) + b_s}} \quad (42)$$

$$f(y, z) = 2 \times \left(\frac{1}{1 + e^{-a_1 \times (y - \bar{s}_t)}} - 0.5 \right) \times \left(\frac{1}{1 + e^{-a_2 \times (z - \bar{s}_a)}} - 0.5 \right) \quad (43)$$

$$a_s = \max \left(-2 \log \frac{\bar{s}_w}{(1 - \bar{s}_w)(10^c - 1)}, 2 \log \frac{\bar{s}_w \times (10^c - 1)}{1 - \bar{s}_w} \right) \quad (44)$$

$$b_s = \log \frac{1 - \bar{s}_w}{\bar{s}_w} \quad (45)$$

$$a_1 = \max \left(\frac{1}{1 - \bar{s}_t} \log_e(10^c - 1), \frac{1}{\bar{s}_t} \log_e(10^c - 1) \right) \quad (46)$$

$$a_2 = \max \left(\frac{1}{1 - \bar{s}_a} \log_e(10^c - 1), \frac{1}{\bar{s}_a} \log_e(10^c - 1) \right) \quad (47)$$

これは、タスクの類似度とラベルの類似度の両者が低いときにワーカの類似度が高くなるように作用する式である。三者を入れ替えると、ワーカの類似度とタスクの類似度の両者が低いときにラベルの類似度が高くなる・ワーカの類似度とラベルの類似度の両者が低いときにタスクの類似度が高くなる、という式となる。

3.4.5 提案手法5

また、他の例として、タスクの類似度 $s_t(y_1, y_2)$ と回答の類似度 $s_a(z_1, z_2)$ の両方が低い時にはワーカの類似度が低くなるように、タスクの類似度 $s_t(y_1, y_2)$ と回答の類似度 $s_a(z_1, z_2)$ のどちらか一方のみが低く他方が高い場合は一定の値をとるようにしたいということが考えられる。このような場合に対し、 $s_t(y_1, y_2)$ と $s_a(z_1, z_2)$ すわなち $s''(y, z)$ の引数 y, z の和をとる形の式を提案する。

$$s_5''(y, z) = p \times y + (1 - p) \times z \quad (48)$$

なお、この式にはパラメータ $0 < p < 1$ を含む。この値を調節することで、タスクの類似度とラベルの類似度のどちらかの影響を大きくする、あるいは両者同じ影響とすることができる。

3.5 実際の計算について

SimRank は、反復によって計算される。 k 回目の試行により計算される SimRank を $R_k(*, *)$ と表すこととする。この時、初期値 $R_0(*, *)$ を

$$R_0(a, b) = \begin{cases} 0 & (a \neq b) \\ 1 & (a = b) \end{cases} \quad (49)$$

として、 $R_{k+1}(*, *)$ を $R_k(*, *)$ から以下のように計算できる。

$$R_{k+1}(a, b) = \frac{C}{|I(a)||I(b)|} \sum_{i=1}^{|I(a)|} \sum_{j=1}^{|I(b)|} R_k(I_i(a), I_j(b)) \quad (50)$$

二部グラフにおける SimRank に関しても同様である。本研究で提案した手法の計算も、SimRank と同様に反復によって行う。

4 類似度の特徴と分類

第3章で、三つ組データの SimRank の計算方法として複数の種類を提案した。これらの式はそれぞれ結果として算出される類似度の持つ意味が異なると考えられる。同様に、三つ組データの持つ特性によって、求めたい類似度は異なる。本章では、複数の三つ組データを対象に、それらに対しどのような類似度をもとめることが適切かを示した上で、対応する類似度の計算式を示す。

4.1 類似度の場合分け

データの特性や類似度の使用方法をもとにどの手法を使用するかを決める前に、類似度の求め方は二通り考えることができる。

一つは、「似ているタスクに似ている回答をつけるワーカは似ている」「似ているワーカが似ているタスクにつける回答は似ている」「似ているワーカが似ている回答をつけるタスクは似ている」という概念を定義であると考えられる方法である。つまりこれを命題であると捉え、論理的に考えて適用する手法を選ぶ。

もう一つは、実際の「類似する」を表せるように数式を使う方法である。例えば映画のレビューデータであれば、映画に対しては人間の想定する「類似する」という概念が存在する。同じようなジャンルであり、同じようなストーリー展開が行われる映画は似ていると考えられるだろう。このような一般に想定される「類似する」を推定できるように、どの提案手法を使うかを決定する。

4.2 類似を定義する方法

まず、類似するを定義する方法で考える。

すべての手法で共通して採用した考え方として、

- 類似するタスクに類似する回答をつけるワーカは類似
 - 類似するワーカが類似するタスクにつける回答は類似
 - 類似するワーカが類似する回答をつけるタスクは類似
- というものがある。これを定義であると考え、論理的に導けるその他の条件を検討する。

まずこれらの定義を、次のように書き換える。

- タスクが類似 \wedge 回答が類似 \Rightarrow ワーカは類似
- ワーカが類似 \wedge タスクが類似 \Rightarrow 回答は類似
- ワーカが類似 \wedge 回答が類似 \Rightarrow タスクは類似

これらを命題とし、それぞれ対偶をとると、

- ワーカは非類似 \Rightarrow タスクが非類似 \vee 回答が非類似
- 回答は非類似 \Rightarrow ワーカが非類似 \vee タスクが非類似
- タスクは非類似 \Rightarrow ワーカが非類似 \vee 回答が非類似

という命題が得られる。この時、タスク x_1, x_2 に対して「似ていないワーカが似ている回答をつけた」というデータがあった場合を考える。「ワーカは似ていない \Rightarrow タスクが似ていない \vee 回答が似ていない」に着目すると、ワーカが似ていないため、タスクか回答のどちらかは似ていないことになる。回答は似ているというデータなので、結果としてタスク a, b が似ていない方に作用するように計算すべきである。同様に、回答 z_1, z_2 に

対して「似ていないワーカが似ているタスクにつけていた」という場合、「ワーカは似ていない \Rightarrow タスクが似ていない \vee 回答が似ていない」に着目して、回答 z_1, z_2 が似ていない方に作用するように計算すべきである。残りの二つの命題にも同じことが考えられ、これより次の6つの命題を得ることができる。

- ワーカが非類似 \wedge 回答が類似 \Rightarrow タスクは非類似
 - ワーカが非類似 \wedge タスクが類似 \Rightarrow 回答は非類似
 - タスクが非類似 \wedge 回答が類似 \Rightarrow ワーカは非類似
 - タスクが非類似 \wedge ワーカが類似 \Rightarrow 回答は非類似 \dots (*)
 - 回答が非類似 \wedge ワーカが類似 \Rightarrow タスクは非類似 \dots (*)
 - 回答が非類似 \wedge タスクが類似 \Rightarrow ワーカは非類似
- ここまでの議論により、ワーカの類似に関する定義は、

- タスクも回答も似ている時
- タスクは似ているが回答は似ていない時
- 回答は似ているがタスクは似ていない時

の3つが得られており、パターンとして不足している定義は「タスクも回答も似ていない時」である。ここで、

- タスクが非類似 \wedge 回答が非類似 \Rightarrow ワーカが非類似
- とすると、その対偶は

- ワーカが類似 \Rightarrow タスクが類似 \vee 回答が類似

となるが、これは(*)のついた条件と矛盾する。

- タスクが非類似 \wedge 回答が非類似 \Rightarrow ワーカが類似

とすればこのような矛盾は生じないため、これを採用する。タスク、回答にも同じ議論ができる。

これらの定義をまとめると、ワーカの類似度、タスクの類似度、回答の類似度の計算式はすべて、式(10)が同じ以下の条件を満たす式であることが求められる。

- $s'(1, 1) = 1$
- $s'(1, 0) = 0$
- $s'(0, 1) = 0$
- $s'(0, 0) = 1$

これを満たす式は提案手法4のみである。よって、この考えから得られる最終的な類似度の計算式には、ワーカの類似度、タスクの類似度、回答の類似度の計算全てに提案手法4を用いる。なお、挙げられている式はワーカの類似度計算中に用いるタスクの類似度と回答の類似度からの計算式であるが、式中の各種平均値を対応するものに変更することで、タスクの類似度計算や回答の類似度計算に用いるものが得られる。

また、この考え方は、議論の最中に三つ組データの中身が実際何であるかは考慮していない。すなわち実際に三つ組がどのようなデータであるかは問わない手法であり、どのようなデータに対しても使用できると考えられる。

4.3 想定する類似を表す方法

次に、実際に想定される「類似する」を推定する方法を考える。このためには、計算に使用するデータと、類似度をどのように使用するかを明らかにした上で考える必要がある。

本研究では、第5章と第7章で使用することを前提として、そのような状況で使用できるようにワーカ、タスク、回答それぞれに対し用いる提案手法を検討する。

4.3.1 映画レビューデータと、それを用いたレビュー値予測
第5章で用いることを前提に、映画のレビューデータに対して類似度を計算することを考える。レビューデータは、レビューが映画に対し5段階評価を行うというものを想定する。実験では、協調フィルタリングを応用した手法で、レビューが未視聴の映画に対して評価値の予測を行う。実験の詳細は第5章で説明する。なお、類似度の計算の際、レビューをワーカ、映画をタスク、レビュー値を回答と考えて計算する。

実験において、最終的な評価値予測にはレビューの類似度のみが使用され、映画や回答の類似度はレビューの類似度を求めるために間接的にのみ使用される。そこでまず、レビューの類似度を考える。今回は映画のレビュー値の予測を行う。協調フィルタリングでは、共通して評価している映画に同じような評価をつけている人を探し、それらの人と同じように評価するだろうということ、逆に正反対の評価をつけている人とは正反対の評価をするだろうという考え方で予測を行なっている。今回の実験では協調フィルタリングで使用する評価の一致度合いを類似度で代用する。そのため、レビューの類似度としては、映画の嗜好が似ている人を高い類似度であるとするのが好ましいと考えられる。

嗜好という観点からレビューの行動を考える。レビューは自分の好みの映画を選んで視聴する。そのため、見ている映画が類似していないレビューは似ていないと考えられる。また、似ている映画を見ている場合でも、評価が似ていない場合はそのレビューは似ていないと考えられる。この考えに対応する手法は提案手法1である。

次に、映画の類似度について考える。似ていないレビューが似ている評価をつけた場合を考える。評価が似ているということは、その映画に対し好きであるか否かが似ていると考えられる。レビューの類似は好みの類似であるとしたので、似ていないレビューが同じような好みであることを示している映画は似ていないと考えられる。似ているレビューが似ていない評価をつけた場合を考える。この時、上記で定めたレビューの類似度の検討から考えると、似ているレビューが行うのは似ている映画に対し似ている評価をつけることだけである。すなわち、似ているレビューが似ていない評価をつけることは起こり得ないはずである。よって、似ているレビューが似ていない評価をつけたというデータが存在した場合は一種のイレギュラーであると考え、映画の類似度に影響させないことが望ましい。似ていないレビューが似ていない評価をつけたことについて考える。似ていないレビューの行動も上述のレビューの類似度の検討から読み取れるが、似ていないレビューには「似ていない映画を見ているので(つけた評価に関わらず)似ていない」場合と、「似ている映画に似ていない評価をつけたので似ていない」場合がある。ここから考えると、似ていないレビューは似ている映画にも似ていない映画にも似ていない評価をつけ得るので、似ていないレビューが似ていない評価をつけたことも映画の類似度に影響させないことが望ましい。この考えに対応する手法は提案手法3である。なお、第3章で示している数式はワーカの類似度に対するものである。式中の s_w の平均値は s_t の平均値に、

s_t の平均値は s_w の平均値に変更して計算する。

最後に、評価の類似度について考える。似ていないレビューは嗜好が違うので、似ている映画には似ていない評価をつけることが期待される。一方、似ているレビューが似ていない映画に対してつける評価は、同じような嗜好をしていても映画が違うのであれば評価は似ていることも似ていないことも想定され得るので、評価の類似度に影響させないことが望ましい。最後に、似ていないレビューが似ていない映画につけた場合であるが、これも同様に評価値が似ているケースも似ていないケースも考えられるため、類似度にてできるだけ影響させないことが望ましい。この考えに対応する計算式は提案手法3である。なお、第3章で示している数式はワーカの類似度に対するものである。式中の s_w の平均値は s_a の平均値に、 s_t の平均値は s_w の平均値に変更して計算する。

4.3.2 犬科の画像ラベリングデータと、それをを用いた多数決の精度向上

第7章で用いることを前提に、犬科の画像ラベリングデータに対して類似度を計算することを考える。タスクは、ワーカが、提示される複数のイヌ科の画像それぞれに対し、示されたいくつかのイヌ科の名称(ラベル)から一つを選択することを想定する。このとき与えられるラベルは、Alaskan Malamute, Siberian Husky, Samoyed, German Shepherd, Gray Wolf, Coyote, Dhole の7種類である。実験では、ワーカを複数のグループに分けてタスクに取り組み、グループ内で多数決を取ることで一つの選択肢を選出することを想定し、より正答率の向上するようなグルーピングを計算する。実験の詳細は第7章で説明する。なお、類似度の計算の際、画像をタスク、選ばれたラベルを回答と考えて計算する。

実験において、最終的な評価値予測にはワーカの類似度のみが使用され、映画や回答の類似度はワーカの類似度を求めるために間接的にのみ使用される。そこでまず、ワーカの類似度を考える。今回はワーカの解答をもとにグループの編成を行う。実験は、松田ら[13]らの研究による考え方を元に行う。この研究では、グループを編成する際にワーカが同じ間違えの選択肢に偏らないように編成することで、精度が向上することを示している。よって、今回の実験で使用したいワーカの類似度は、ワーカの回答傾向の類似性であり、これを求められるようにどの提案手法を用いるか検討する。

ワーカはどの画像にラベルをつけるかを選ぶことはできない。そのため、画像が似ているか否かのみでワーカの類似度は判断できない。また、ワーカはどのようなラベルもつける可能性があるため、ラベルの類似度では判断できない。そのため、組み合わせの判断が必要となる。まず、似ている画像に似ているラベルをつけるワーカは似ていると考えられ、逆に似ている画像に似ていないラベルをつけるワーカは似ていないと考えられる。似ていない画像に対しては、似ているラベルがつけられることも似ていないラベルがつけられることも考えられるため、類似度にてできるだけ影響させないことが望ましい。この考えに対応する計算式は提案手法2である。

次に、画像の類似度について考える。似ていないワーカが似

ているラベルをつけている場合を考える。ワーカが似ていないということは回答傾向が似ていないということであるが、もしこの画像が似ているならこのワーカからは似ている画像に似ているラベルをつけたということになり、回答傾向が似ていないことに矛盾する。よって、似ていないワーカが似ているラベルをつけている場合その画像は似ていない。似ているワーカが似ていないラベルをつけることを考える。似ていないラベルは似ている画像にも似ていない画像にもつけることが想定されるので、画像の類似度に影響しないようにする。最後に、似ていないワーカが似ていないラベルをつけた場合を考えるが、同様に、似ていないワーカが似ていないラベルをつけたことから情報は得られないので、画像の類似度に影響しないようにする。この考えに対応する計算式は提案手法3である。なお、第3章で示している数式はワーカの類似度に対するものである。式中の s_w の平均値は s_t の平均値に、 s_t の平均値は s_w の平均値に変更して計算する。

最後に、ラベルの類似度について考える。似ていないワーカは回答傾向が似ていないので、似ている画像には似ていないラベルをつけるはずである。似ていない画像に対しては、似ているラベルをつける可能性も似ていないラベルをつける可能性も考えられる。そのため、レビューの類似度を問わず、似ていない画像に対してラベルをつけたデータは、ラベルの類似度に影響させないこととする。この考えに対応する計算式は提案手法3である。なお、第3章で示している数式はワーカの類似度に対するものである。式中の s_w の平均値は s_a の平均値に、 s_t の平均値は s_w の平均値に変更して計算する。

5 映画レビューデータを用いたレビュー値予測実験

本節では、第4章で述べた提案手法を用いて行なった、映画レビューデータを用いたレビュー値予測実験の内容とその結果について述べる。本実験の目的は、提案手法の有効性を検証することである。

5.1 実験の概要

この実験は、映画レビューデータをもとに、レビュー値の予測を行う。予測は、5.3.1節で説明する提案手法と5.3.2節で説明する比較手法の2種類を行う。評価は平均絶対誤差(MAE)を利用する。平均絶対誤差は次の式(51)で計算される値である。ここで、 n はテストデータ数、 r_i は i 番目のテストデータの実際のレビュー値、 \hat{r}_i は i 番目のテストデータの予測値である。

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{r}_i - r_i| \quad (51)$$

5.2 使用したデータ

実験は、以下のデータに対して行った。

5.2.1 全データ

利用したデータは、Yahoo!映画[14]より収集した、映画に対するレビューデータである。このデータは、249,550人が33,795本の映画に対して行った1,288,705件のレビューが含まれる。

表 1 データセット A, B, D, D, E, F の説明

	対象映画	対象人	映画 (本)	人 (人)	レビュー (件)	内テスト (件)
A	～100 位	～100 位	109	100	5393	1000
B	～20 位	～100 位	24	91	1365	270
C	～100 位	～20 位	148	21	2218	450
D	80 位～100 位	～100 位	35	99	216	40
E	～100 位	80 位～100 位	166	20	880	170
F	～50 位	～50 位	52	52	1448	289

今回の実験では、Yahoo!映画より収集できるレビュー情報のうち、映画の ID、プロフィール URI、レビュー値を使用し、映画の ID を映画の識別に、プロフィール URI をワーカの識別に使用した。

5.2.2 データの選択

本データは非常に大きく、全データを使つての実験は困難であるため、各映画、各レビューのレビュー件数を求め、レビュー数の順位が一定の順位以上の映画とレビューを選ぶことで 6 個のデータセット A, ..., F を作成した。各データセットの作成に用いた映画とレビューの順位の範囲を表 1 に示す。例えば、データセット B は、レビュー数が 20 位以上 (20 位タイを含む) の映画とレビュー数が 100 位以上 (100 位タイを含む) のレビューからなるデータセットである。映画・レビューはそれぞれレビュー件数を計算し、表 1 に書かれている順位のデータを抽出した。映画の抽出は対象映画のカラムの順位を、レビューの抽出は対象人のカラムの順位を使用している。

5.3 予測手法

5.3.1 提案手法

第 3 章で述べた手法による類似度計算方法を使用して、レビュー値の予測を行う。

まず、テストデータを除いたデータを使用し、類似度の計算を行う。類似度は、レビューをワーカ、映画をタスク、レビュー値を回答と考えて計算する。収束定数 C_1, C_2, C_3 は 0.8 とする。ワーカ、タスク、レビューの計算に第 3 章で述べた提案手法のいずれを使用するかは、第 5.4 小節にて行う実験ごとに述べる。

類似度の計算後、その値をもとにテストデータに対しレビュー値の予測を行う。予測は、協調フィルタリングによる予測式に類似度を当てはめた次の式 (52) により行う。

$$\hat{A}_m = \bar{A} + \frac{\sum_{B \in U_m} (B_m - \bar{B}) \times sim(A, B)}{\sum_{B \in U_m} sim(A, B)} \quad (52)$$

ここで、 U を全レビューとして $A, B \in U$ であり、 U_i はアイテム i に評価をつけているユーザーの集合、 A_m はレビュー A が映画 m に対して行ったレビュー、 \bar{A} は A のレビューの平均値、 B_m はレビュー B が映画 m に対して行ったレビュー、 \bar{B} は B のレビューの平均値である。 $sim(A, B)$ は式 (7) で得られる類似度である。

5.3.2 比較手法

提案手法の有効性を示すため、協調フィルタリングによる推

定も行う。協調フィルタリングによる予測の方法は、第 2 章で説明している。

5.4 実験の詳細

5.4.1 実験 1: 提案手法間の比較

第 3 章で、計 5 通りの類似度計算手法を提案した。また、この 5 通りの特徴を踏まえ、第 4 章においてワーカ、タスク、回答にどの手法を適用するのが相応しいかを検討した。

本実験では、ワーカ、タスク、回答に対してそれぞれ計 5 通りの類似度計算手法を適用した、全 125 の計算方法の組合せの精度を比較する。これにより、第 4 章で検討した組合せがふさわしいことを示す。

実験には、計算速度を考慮しデータセット F を使用した。また、計 5 回実験を行った上で結果を平均して比較することとした。各回では同じテストデータを使用して 125 通りの組合せに対し実験を行い、回が変わるごとにテストデータをランダムに選択し直した。

5.4.2 実験 2: 提案手法と既存手法の比較

本実験は、データセット A, ..., E に対し、F を用いた実験 1 で上位となった 5 通りの組合せと比較手法を使用して行う。これにより、提案手法が既存の手法と比較して有効であることを示す。

実験は、データセットごとにそれぞれ計 5 回実験を行った上で結果を平均して比較することとした。各回では同じテストデータを使用して 5 通りの組合せ及び比較手法に対し実験を行い、回が変わるごとにテストデータをランダムに選択し直した。

6 映画レビューデータを用いたレビュー値予測実験に対する考察

本研究で提案した手法を用いた第 5 章の実験について考察する。

実験の結果は、表 2、表 3 の通りである。表中の数字は平均絶対誤差であり、太字は各データセットで最も誤差が小さかった項目を表している。なお、表 2 は上位の組合せ 5 通りと下位の組合せ 5 通り、及び 4.2 で示した組合せであるワーカ、タスク、回答全てに提案手法 4 を用いた組合せのみを記載している。

6.1 提案手法間の比較

実験 1 では、第 3 章で提案した計 5 通りの類似度計算手法をワーカ、タスク、回答に対してそれぞれ適用し、出来た 125 の組合せの精度を測定した。

最も性能が良かったのは、ワーカ、回答に提案手法 1 を、タスクに提案手法 3 を適用した手法であった。一方、第 4 章において提案した、映画レビューデータを用いて映画レビューの予測を行うというケースに合わせた組合せは、第 2 位であった。最も良い成果が得られたとは言えないが、全部で 125 通りあることを考えると、一定の精度が得られたと考えられる。

また、上位データと下位データを比較すると、上位に入って

表 2 実験 1 の結果

順位	ワーカ	タスク	回答	平均絶対誤差
1	提案手法 1	提案手法 3	提案手法 1	0.7048
2	提案手法 1	提案手法 3	提案手法 3	0.7051
3	提案手法 1	提案手法 4	提案手法 3	0.7052
4	提案手法 1	提案手法 1	提案手法 3	0.7052
5	提案手法 1	提案手法 3	提案手法 2	0.7052
106	提案手法 4	提案手法 4	提案手法 4	0.7110

121	提案手法 4	提案手法 1	提案手法 2	0.7120
122	提案手法 4	提案手法 2	提案手法 3	0.7122
123	提案手法 4	提案手法 2	提案手法 5	0.7123
124	提案手法 4	提案手法 2	提案手法 1	0.7123
125	提案手法 4	提案手法 2	提案手法 2	0.7123

表 3 実験 2 の結果

手法	提案手法の組合せ順位					比較手法
	1 位	2 位	3 位	4 位	5 位	
データセット A	0.715	0.715	0.715	0.715	0.715	0.725
データセット B	0.728	0.729	0.730	0.731	0.730	0.748
データセット C	0.624	0.623	0.622	0.621	0.620	0.638
データセット D	0.792	0.789	0.783	0.784	0.784	0.820
データセット E	0.858	0.856	0.852	0.851	0.850	0.901

いる手法に共通する点として、予測に使用するワーカの類似度が第 4 章において提案した提案手法 1 であるということがあり、これは下位には見られない傾向であった。さらに、第 4 章において提案した、ワーカ、タスク、回答全てに提案手法 4 を使う手法は 106 位であった。これらを踏まえると、予測に直接は使用されないタスクや回答の類似度よりも、実際に予測に使うワーカの類似度を適切に計算することが重要であると考えられる。

さらに、実験 2 では、上位 5 つの組合せのみではあるが、複数のデータセットで比較した場合の順位の変化も確認できる。表 3 を見ると、データセットによって誤差の順位は入れ替わることがわかった。しかし、その差は比較手法との差と比べると非常に小さい。

これらは、平均誤差の数値の差が小さいことが原因であると考えられる。平均誤差の数値自体を比較すると、手法間で大きな差が得られたとは言いがたかった。これは、今回提案した手法を構成する際に着目したポイントが類似度が 0 や 1 という最大最小の値を取るケースであったことが挙げられる。例えばワーカの類似度であれば、タスクの類似度および回答の類似度が 0 もしくは 1 である時の計 4 通りを中心に考えており、この時にある程度の差が生まれることで異なる手法であるとしていた。しかし、データが取りがちな中央の範囲では大きく変化が見られないことで、全体を通した性能の差分は少なくなってしまうのではないかと考えられる。データの取り得る値の端点だけでなく、タスクの類似度および回答の類似度が 0.3~0.7 のあたりでも変化が見られるように式を構成することで、より特徴のある手法が得られ、データの特徴としてふさわしい手法を用いれば

表 4 データセット F で実験 1 で 1 位となった組合せによる類似度計算を行った際レビュー値の類似度

		レビュー値				
		1	2	3	4	5
レビュー値	1	1	-	-	-	-
	2	0.0025249	1	-	-	-
	3	0.0015875	0.0017589	1	-	-
	4	0.0014451	0.0014967	0.0017073	1	-
	5	0.0014811	0.0013089	0.0013109	0.0015671	1

さらに良い精度で予測できるのではないかと考えられる。

6.2 提案手法の精度について

実験 2 では、様々なサイズのデータセットに対して提案手法と比較手法による予測を行った。結果として、全てのデータセットで提案手法による予測の方が比較手法による予測よりも少ない誤差で予測できた。

比較手法である協調フィルタリングは、予測したいレビューとの相関係数が計算できないレビューは予測の算出に使用することができない。一方提案手法の類似度計算の場合、データ全体が完全に分割されていない限り全てのレビュー間で類似度が計算できる。つまり、提案手法は比較手法多くのデータを使用することができるため、より少ない誤差での予測が可能になったと考えられる。

一方、データセット A~C と、データセット D 及びデータセット E では、提案手法と比較手法の誤差の差が異なり、データセット A~C では比較的誤差の差が小さかった。データセット A C は、データセット D 及びデータセット E に比べて多くのレビューを含む密なデータセットである。データが十分にある場合、協調フィルタリングで使用するレビュー間の相関係数を十分な数計算できるため、前述の点での優位性が少なくなる。そのため、結果にあまり差がつかなかったと考えられる。

6.3 レビュー値の類似度について

今回の実験では、(ワーカ・タスク・回答) の解答に該当するデータとして、レビュー値を利用した。レビュー値は数値データであり得られた類似度そのものに対する考察が可能であることから、レビュー値の類似度について考察する。

次の表 4 は、データセット F で、実験 1 で 1 位となった組合せによる類似度計算を行った際の、レビュー値の類似度である。この表を見ると、レビューの値が 1 だけ違うものは確かに高い類似度となっており、レビュー値の差が広まるにつれて類似度の差も大きくなる傾向がある。

しかし、レビューの値としては同じだけの差であったとしても、類似度の差も同程度であるとは限らない。例えば、レビュー値 1 とレビュー値 2 は、他の差が 1 のレビュー値間の類似度に比べて非常に高い。これは、1 も 2 も気に入らない映画につけられる指標であり、使われる場面が非常に限られていることが原因ではないかと考えられる。一方、レビュー値 3~5 はレビューでもよく見られる値であり一般的にレビューの基準となる点数だが、レビューによりレビューの基準値に差があり、そ

の映画を特に気に入った時につける値なのか、普通である時につける値なのか異なる。これにより、同じレビュー値の差が1でも類似度は異なると考えられる。

このように、レビュー値に対して類似度を計算することで、単純に差を類似度の代用とするよりも多くの情報を保つことができると考えられる。

7 クラウドソーシングデータを用いた多数決精度実験

本節では、第3章で述べた提案手法を用いて行なったクラウドソーシングデータを用いたグループ編成による多数決精度の比較実験の内容とその結果について述べる。本実験の目的は、提案手法の有効性を検証することである。

7.1 実験の概要

この実験は、クラウドソーシングデータをもとにした、多数決の精度を測る実験である。データをもとにワーカのグループ編成を行い、さらにグループ内でラベルの多数決をおこなってその正解率を比較する。グループ編成は、第3章で述べた提案手法による類似度を用いて行うものと、松田ら[13]による手法を使用する。さらに比較手法として、無作為な選出によるグループ編成も行う。詳細は7.3節で述べる。

7.2 提案手法

本節では、ワーカ間の類似度を利用してクラスタリングしグループ編成を行う方法を説明する。

7.2.1 グループ編成手法

クラウドソーシングデータを利用したグループ編成手法として、松田ら[2][13]による研究で提案されている考え方によるものを利用する。

多数決を行う場合、不正解の選択肢の一つにワーカの回答が偏ってしまうと、多数決が失敗する可能性がある。そこで、各グループのワーカが同じ選択肢に偏らないように、回答傾向の異なるものを同じグループに選出することで、多数決の精度を上げることができる。松田らの研究では、回答をベクトル化し、回答傾向が似ているか否かを数値化している。本研究で算出されるワーカ間の類似度も、ワーカの回答の傾向が近いものが高類似度になると考えられる。そのため、提案手法による類似度を使用することでも高い正答率を得られると考えられる。

松田らの提案するワーカの回答の類似を利用した手法は、距離としてワーカの回答をベクトル化し、そのユークリッド距離を用いている。しかし我々は類似度として値のみを直接求めているため、K-means法は使用できない。そのため、クラスタリングは、ワーカの類似度を距離関数として、K-medoids法[15]をもとに行う。K-medoids法は、クラスタの代表にメドイドを用いるクラスタリング手法である。まず、クラスタ数分のランダムな代表値、メドイドを選ぶ。次に、それ以外の点を、選ばれた最も近いメドイドのクラスタに割り当てる。そして、各クラスタごとに、メドイドの更新を行う。クラスタCのメドイドの計算方法は以下の通りである。

$$\text{medoid}(C) = \arg \min_{x \in C} \sum_{y \in (C - \{x\})} d(x, y) \quad (53)$$

ここで、 d は距離関数である。本研究では、ワーカの類似度によってクラスタ形成を行うこととした。距離関数 d の類似度を用いた定義は複数考えられ、それらを別の手法として比較する。メドイドの更新とクラスタへの分割を繰り返し、メドイドが変化しなくなればそこでクラスタリング終了となる。

なお、K-medoids法では、クラスタのサイズは均等とは限らない。しかし、本研究ではグループ編成を行うために各クラスタの人数が等しくなる必要がある。よって、最適化によるBalanced K-medoids法[16]によってクラスタリングを行う。

7.2.2 提案手法1

$d(x, y)$ を式(7)で定義した s_w を用いて

$$d(x, y) := s_w(x, y)$$

とする。この場合、類似度は距離が近いすなわち似ている人ほど大きな値となるので、クラスタリングを行った時点で類似する人は別のクラスタに属すると考えられる。そのためクラスタからワーカを一人ずつ選出しグループとするステップは行わず、直接生成したクラスタをグループとして扱った。

7.2.3 提案手法2

式(7)で定義した s_w を用いることは7.2.2と同様である。しかし前述の通り、 s_w は似ている人ほど大きな値となり、そのまま使用すると距離関数の通常の大小の持つ意味とは逆の関係となる。

そこで、 $d(x, y)$ を式(7)で定義した s_w を用いて

$$d(x, y) := 1 - s_w(x, y)$$

とする。これにより、距離が小さいほどワーカが類似していることになり、クラスタリングの結果として類似しているワーカを同じクラスタに組み込むことができる。これ以外の部分は、松田らの論文[13]で行われている手法をそのまま利用する。

7.2.4 使用した類似度計算手法

第3章にて、類似度の計算方法を5種類提案している。これらをワーカの類似度・タスクの類似度・ラベルの類似度にそれぞれ当てはめると、全部で125通りの計算方法がある。本実験では、以下の5つの組合せに対して、実験を行う。収束定数 C_1, C_2, C_3 は0.8とする。

組合せ1 第4章で提案した、全ての類似度に提案手法4を用いる方法

組合せ2 第4章で提案した、ワーカに提案手法3、画像に提案手法3、ラベルに提案手法1を用いる方法

組合せ3~組合せ5 第5章で行った実験1で上位となった手法3つ

7.3 比較手法

提案手法の有効性を示すため、以下の二つの比較手法に対しても実験を行う。

表5 実験結果

手法		正解率
提案手法 1	組合せ 1	0.8056
	組合せ 2	0.7804
	組合せ 3	0.7950
	組合せ 4	0.7950
	組合せ 5	0.7918
提案手法 2	組合せ 1	0.8882
	組合せ 2	0.8958
	組合せ 3	0.8912
	組合せ 4	0.8944
	組合せ 5	0.8940
比較手法 1		0.8928
比較手法 2		0.8783

7.3.1 親和度を用いた手法 (比較手法 1)

比較手法の一つは、松田ら [13] による方法である。松田らは、ワーカの回答を one-hot ベクトル化し、ベクトルのユークリッド距離が遠いもの同士がグループとなるような編成方法を提案している。これを、比較手法 1 と表記する。

7.3.2 無作為な選出 (比較手法 2)

二つ目の比較手法は、ワーカを無作為に選出する方法である。全ワーカの中から、 v 人無作為に選出しグループを繰り返す。これにより、 n/v グループを作成する。これを、比較手法 2 と表記する。

7.4 使用したデータ

実験は、京塚ら [17] が行った犬種類の画像分類タスクの回答データに対しておこなった。これは、Amazon Mechanical Turk [18] でワーカを募集し、画像アノテーションタスクを行ってもらったものである。画像は、Alaskan Malamute, Siberian Husky, Samoyed, German Shepherd, Gray Wolf, Coyote, Dhole の 7 種類のうちいずれかである。ワーカは画像に対し、この 10 種類の中から適切と思われるものを選ぶ。実験は 56 人のワーカに全 800 枚の画像に対して回答してもらった。

実験では、800 種類のタスクから 400 タスクを無作為に選択してグループ編成を行うための学習データとして用い、残り 400 タスクを作成したグループでの多数決の正答率を測定するために使用した。

8 クラウドソーシングデータを用いた多数決精度実験に対する考察

本研究で提案した手法を用いた第 7 章の実験について考察する。実験結果は、表 5 の通りである。正答率は、5 回実験をおこなった平均値を小数点第 4 位で四捨五入した値である。

8.1 提案手法 1 の精度について

表 5 より、提案手法 1 は提案手法 2 及び比較手法に比べ低い正答率となった。

理由として考えられたのはまず、類似度の精度の低さによるということや、類似度が回答傾向を適切に表していないことが

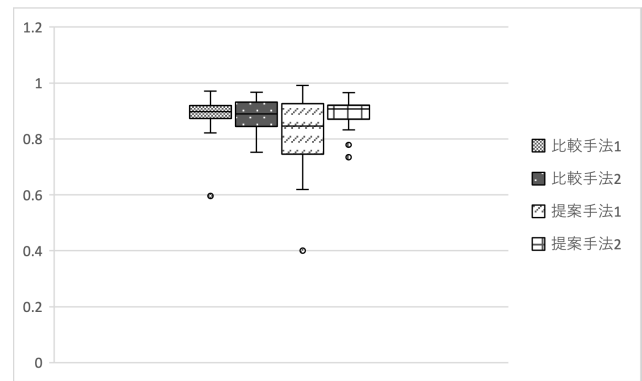


図 1 各手法におけるグループの正答率の分布

ある。しかし、同じ類似度を使った提案手法 2 では比較手法よりも精度が得られたため、これが主たる理由であるとは考えにくい。

それを踏まえ考えられたのは、回答傾向が異なるワーカを直接組み合わせることにより意図せず正答率の低いグループを編成してしまったということである。本提案手法では、同じ誤答に偏ることのないように類似度の低いワーカを組み合わせるグループ編成を行った。しかし、同じ誤答に偏りにくいということは、同じ問題に対し同じ選択肢を選びにくいということであり、それは両者ともに正解の選択肢を選ぶ可能性が低いということにつながる。つまり、類似度が低いワーカの編成は、意図せず正答率も低い編成となってしまう可能性がある。図 1 は、比較手法と、提案手法 1, 2 をワーカ・タスク・回答の類似度全てに 3.4.1 で説明した式を利用する組合せで計算したものである。編成されたグループのそれぞれの正答率である。提案手法 1 によるグループ編成は、正答率の著しく低いグループを含み、それ以外にも正答率が比較的低いグループがあることがわかり、これが正答率の低下を招いたと考えられる。

8.2 提案手法 2 の精度について

表 5 より、提案手法 2 は、提案手法 1 や比較手法に比べ高い正答率となった。同じ類似度を使用しても提案手法 1 に比べ正答率が高かったのは、松田ら [13] の手法に乗り取り、一度類似度が高いワーカをクラスタリングした上で再度グループを編成したことによるものであると考えられる。

また、第 4 章で画像のラベリングタスクに適切な方法として提案した組合せ 2 と、組合せ 4 及び組合せ 5 は、いずれの比較手法よりも平均正答率が高くなるようにグループの編成を行うことができた。特に比較手法 1 は既存の研究に基づく手法であり、これに対し正答率が高かったことから、提案した類似度計算手法に基づくグループ編成は有効であると考えられる。さらに、最も正解率が高かったのは組合せ 2 であり、組合せが適切であることも確認できた。

提案手法が高い正答率となったことについて考察する。これは、既存の手法では同じ問題に同じ解答をした人に着目していたが、同じではなく類似する問題と回答に着目したことによると思われる。同じ問題に対し似たような解答をする人は、今後その問題と同様の問題が出た際に、今度は似ているだけでな

く同じ選択肢を選ぶ可能性があると考えられる。類似する問題に同じ選択肢・あるいは類似する選択肢をつけた人も同様である。また、今回提案した類似度計算手法では、松田らの研究で重視されていた同じ問題に同じ解答をした人についても類似度が高くなるように設定されている。同じ問題に同じ解答をした人を異なるグループにした上で、さらに同じ解答をし得る人も異なるグループに配置することができ、そのためより同じ選択肢に偏る割合を減らすことができたと考えられる。

9 結 論

本研究でのこれまでの流れを振り返る。まず、三種類の値を一組として構成されているデータに対し、(ワーク、タスク、回答)を例として、SimRank を拡張した計算方法によってワークの類似度、タスクの類似度、回答の類似度を求める方法を5種類提案した。また、具体的な三つ組データを想定した上で、提案した複数の計算式のうちのいずれが該当するのかを示した。さらに、これらの類似度の利用方法として、映画レビューデータを用いたレビュー値の予測と、クラウドソーシングデータを用いたグループ編成を行った。レビュー値の予測を行う実験では、Yahoo!映画より収集したデータから複数のデータセットを生成して手法を適用し、予測と実際のデータの平均絶対誤差の比較を行った。予測は協調フィルタリングによるものと類似度によるものを行った結果、類似度による予測の方がより小さい誤差となるように予測できることがわかった。グループ編成を行う実験では、クラウドソーシングで集めたデータに対し手法を適用し、正答率の比較を行った。結果、提案手法では、より平均正答率が高くなるようなグループ編成が行えることがわかった。

今後の展望としては、更なる精度向上のためにワーク・タスク・回答の類似度を組み合わせて使用すること、およびより高速な類似度計算方法の考案が挙げられる。

謝 辞

本研究は JSPS 科研費 21H03446 の助成を受けたものです。

文 献

- [1] Callison-Burch, C. and Dredze, M.: Creating Speech and Language Data with Amazon's Mechanical Turk, *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, CSLDAMT '10, Association for Computational Linguistics, pp. 1–12 (2010).
- [2] 松田浩幸, 田島敬史: クラウドソーシングにおける多数決の精度向上のためのチーム編成手法, 第 13 回データ工学と情報マネジメントに関するフォーラム (2021).
- [3] Jeh, G. and Widom, J.: SimRank: A Measure of Structural-Context Similarity, *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, Association for Computing Machinery, pp. 538–543 (2002).
- [4] Symeonidis, P., Kirjackaja, L. and Zanker, M.: Session-Based News Recommendations Using SimRank on Multi-Modal Graphs, *Expert Systems with Applications*, Vol. 180, No. C (2021).
- [5] Dey, P., Goel, K. and Agrawal, R.: P-Simrank: Extending Simrank to Scale-Free Bipartite Networks, *Proceedings of The Web Conference 2020*, WWW '20, Association for Computing Machinery, pp. 3084–3090 (2020).
- [6] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews, *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, CSCW '94, New York, NY, USA, Association for Computing Machinery, pp. 175–186 (1994).
- [7] Linden, G. D., Jacobi, J. A. and Benson, E. A.: Collaborative recommendations using item-to-item similarity mappings (2001).
- [8] Linden, G., Smith, B. and York, J.: Amazon.com recommendations: item-to-item collaborative filtering, *IEEE Internet Computing*, Vol. 7, No. 1, pp. 76–80 (2003).
- [9] Dawid, A. P. and Skene, A. M.: Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 28, No. 1, pp. 20–28 (1979).
- [10] Dempster, A. P., Laird, N. M. and Rubin, D. B.: Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 39, No. 1, pp. 1–38 (1977).
- [11] Joglekar, M., Garcia-Molina, H. and Parameswaran, A.: Evaluating the Crowd with Confidence, *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pp. 686–694 (2013).
- [12] Welinder, P., Branson, S., Perona, P. and Belongie, S.: The Multidimensional Wisdom of Crowds, *Proceedings of the 23rd International Conference on Neural Information Processing Systems* (Lafferty, J., Williams, C., Shawe-Taylor, J., Zemel, R. and Culotta, A.(eds.)), NIPS'10, Vol. 2, Curran Associates, Inc., pp. 2424–2432 (2010).
- [13] 松田浩幸, 田島敬史: クラウドソーシングにおける多数決の精度向上のためのワーク組合せの選出手法, 第 12 回データ工学と情報マネジメントに関するフォーラム (2020).
- [14] Yahoo: Yahoo!映画, <https://movies.yahoo.co.jp/>.
- [15] Kaufman, L. and Rousseeuw, P. J.: *Finding Groups in Data: An Introduction To Cluster Analysis*, Vol. 344 (1990).
- [16] Bernábe-Loranca, B., Gonzalez-Velázquez, R., Olivares-Benítez, E., Ruiz-Vanoye, J. and Martínez-Flores, J.: Extensions to K-Medoids with Balance Restrictions over the Cardinality of the Partitions, *Journal of Applied Research and Technology*, Vol. 12, No. 3, pp. 396–408 (2014).
- [17] 京塚萌々, 田島敬史: アイテムへのワーク逐次割当てと各ワークの複数ラベル付与によるマルチクラス分類タスク精度向上手法, 第 12 回データ工学と情報マネジメントに関するフォーラム (2020).
- [18] Amazon: Amazon Mechanical Turk, <https://www.mturk.com/>.