

類似性評価知識を用いた レシピ検索システム mineRecipe の実装について

大平 峰子 大園 忠親 新谷 虎松

名古屋工業大学知能情報システム学科

e-mail: {mine,ozono,tora}@ics.nitech.ac.jp

1 はじめに

本研究は、類似性評価知識を用いたレシピ検索システムの実装研究である。ここでの類似性評価知識とは、料理に関する類似性を評価するための知識である。類似性評価知識を用いることにより、レシピ検索の精度を落とすことなく再現率を向上させることが可能になる。

本システムは、インターネットから収集したレシピのテキストデータを用いた検索システムである。キーワードの全文検索による検索では、ユーザは要求するレシピを検索するため試行錯誤をしながら検索する必要がある。本システムは、与えた検索条件に一致する解がない場合に類似解を提示し、ユーザのレシピ検索に関する希望の再現率を向上させることで効率よいレシピ検索を支援する。

類似性評価知識として材料の分類階層情報を用いることで、同一分類の材料によって検索条件を変形し、類似解を得ることができる。しかし、料理において、分類は異なるが食感や色が似ている材料や、ある条件では代替可能であるが、別の条件下では代替できないという材料が存在するため、成分による分類のみで条件の代替を行うことは検索精度の低下を招く。本研究では、料理に特化した材料間の類似性、代替可能性に関する領域知識を利用することで、検索精度を下げずに再現率を向上させることを目指す。しかし、類似性評価知識をユーザから獲得することは非現実的である。本研究では、知識発見技術を用いて既存のレシピから類似性評価知識を自動的に獲得し [1]、得られた類似性評価知識を用いて類似検索を行う。

知識発見技術の情報検索分野への応用として、文献データにおける関連属性を利用して関連するキーワードを発見し、検索式を改善することで関連情報を取り出す研究 [2] がある。文献データは題名、著者名等データの属性が明確であるのに対して、料理のレシピは、データの構造や属性の定義が不明瞭であるため、[2] におけるアプローチをそのまま適用するのは難しい。本研究では、このような属性の定義が不明瞭なデータに対する自動的な知識の獲得および検索を対象とする。

本論文では、レシピ検索システム mineRecipe における類似性評価知識の利用方法について述べ、実験によりその有効性を示す。

2 レシピ検索システム mineRecipe

本レシピ検索システムでは、概念一致検索、および類似検索を行うことで、ユーザの入力条件に対する再現率の向上を図る。概念一致検索では、入力やレシピ中の表記のゆれ、つまりある概念に対する複数の表記に対応し、ユーザから与えられた語が表す概念を含むレシピや、下位概念を含むレシピを返す。概念一致検索を行うことにより、再現率を向上させることができる。しかし実際には、入力条件とレシピに含まれる概念は不一致であるが、類似しており参考のできるレシピが存在する。このようなレシピまで検索する方法の1つとして、材料の分類階層において同じ上位概念をもつ材料同士を代替させる方法がある。しかし、料理においては同じ上位概念を持つ材料同士が代替可能であるとは限らないので、再現率が向上する代わりに精度が低下する。より高精度な類似検索を行うために、本研究では類似性評価知識を用いる。このような概念一致検索、類似検索をテキストデータに対して実現するために、本研究では入力された検索条件に対して、検索条件文の拡張 [3] を導入する。概念一致検索では概念の情報の付加を、類似検索では類似した材料の組合せによる条件の代替を行う。

本レシピ検索システムでは、まず概念一致検索を行い、解が得られなかった場合、もしくは解が得られたがユーザがさらに類似解を望む場合、類似性評価知識を用いた類似検索を行う。類似性評価知識を用いた類似検索によっても解が得られなかった場合、材料の分類階層を用いた類似検索を行う。このように解の提示を段階にわけることによって、ユーザの希望に対する精度をできるだけ下げないようにしながら再現率を向上させる。

2.1 システムの構成

本研究で作成するレシピ検索システム mineRecipe のシステム構成図を図1に示す。問い合わせ文作成機構では、ユーザから与えられた検索条件文を解析し、レシピデータベースに対する問い合わせ文を作成する。本システムでは、領域知識を利用して検索条件文を拡張することで、ユーザの検索要求の再現率の向上を図る。本システムでは、領域知識として類似性評価知識と食品成分表の2つを主に利用する。利用される類似性評価知識は、類似性評価知識作成機構が食品成分表などの基本的な領域知識を参照することにより、保持しているレシピデータから生成する。レシピデータベースには、インターネット上から得られたレシピデータをそのままテキストデータとして格納する。インターネット上からのレシピの自動獲得機構と組み合わせる

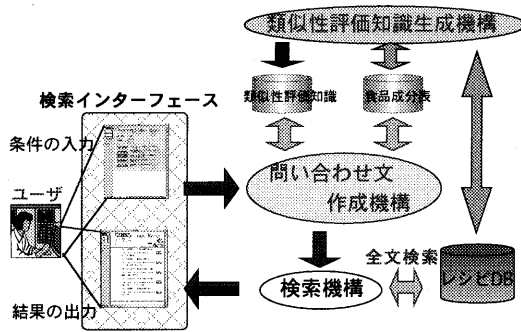


図 1: レシピ検索システム構成図

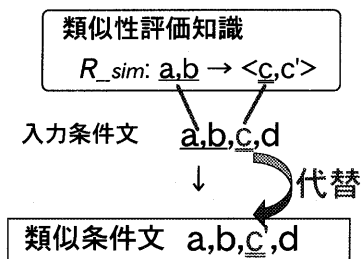


図 2: 類似性評価知識を用いた類似条件文の生成

ことで、より新しいデータを用いた検索システムを実現することができる。

3 類似性評価知識を用いた検索条件文の拡張

ユーザからの検索条件の入力は、ユーザが検索を要求する語の集合と、その語同士を *and* もしくは *or* で結びつける記号を用いて行う。

概念一致検索では、分類階層情報などの背景知識を参照し、対象とする概念の複数の表記を入力条件中に *or* 文として付加することで、入力条件文の拡張を行う。

3.1 類似性評価知識

本研究における類似性評価知識は、 $R_{sim} : S \Rightarrow \langle x, y \rangle$ (x, y はアイテム、 S はアイテムの集合) と表現する。 R_{sim} は、「アイテム集合 S がデータに含まれ、アイテム x もしくは y が含まれるデータは、もう一方のアイテムと入れ替えたデータと類似している」ということを意味する。類似性評価知識は、アприオリアルゴリズムに基づくアルゴリズムで発見できる [1]。

3.2 知識を用いた検索条件文の拡張

類似性評価知識を用いた類似検索条件生成のモデルを図 2 に示した。類似性評価知識 R_{sim} を用いた類似条件文は、類似性評価知識 R_{sim} の条件部 S と x, y のどちらかが入力検索文に含まれているときには、 x, y を他方と交換することで生成することができる。

3.3 評価

類似性評価知識を用いることによるレシピの検索結果の改善について検証するため、簡単な実験を行った。

	実験 1	実験 2	実験 3	実験 4
ヒット数	11	21	31	57
再現率	27.6% (8/29)	62.1% (8/29)	86.2% (25/29)	100% (29/29)
精度	100% (11/11)	100% (21/21)	100% (31/31)	80.7% (46/57)

表 1: 実験結果

類似性評価知識の一例として R_{sim1} : カレー $\Rightarrow$ <牛肉, 鶏肉> を、入力検索条件として q : カレー, 牛肉 を与え、次の 4 つの検索を行い、結果の比較を行った (‘&’ は *and* を、‘|’ は *or* を表現する)。実験 1. 類似性評価知識 R_{sim1} を用いず、検索条件 q に対する一致検索、実験 2. 類似性評価知識 R_{sim1} を用いた類似検索 (q_1 : カレー, (牛肉; 鶏肉))、実験 3. 材料の分類階層 (食品成分表) を用いた類似検索 (q_2 : カレー, (牛肉; 鶏肉; 豚肉))、実験 4. ワイルドカードを用いた検索 (q_3 : カレー)。それぞれの検索結果に対する評価項目は、再現率と精度である。再現率を求めるために、ここではユーザ定義の類似検索条件式 (q_4 : カレー, (牛肉; 鶏肉; 豚肉; 魚介類)) を用いた検索の結果と比較を行った。

2,587 件のテストデータに対する検索結果は表 1 の通りである。実験 4 は、ワイルドカードによって検索した結果であり、最も多い検索結果が得られている。ここでは 1 つの類似性評価知識のみに注目して検証を行っているため、実験 4 の結果の方が、類似性評価知識を用いた実験 2 よりも再現率が高い。しかし、実験 4 では同時に精度が下がっており、参考にできないレシピも含まれてしまうことがわかる。このことから、ワイルドカードによる検索を行うよりも、類似性評価知識を複数組み合わせることで、よりユーザの要求に適した結果を返すことができるといえる。

4 おわりに

本研究では、自動的に獲得した類似性評価知識を利用して、有効な類似解を返すレシピ検索システム mineRecipe の実装を行った。本論文では、テキストデータからの全文検索において、精度を下げずに再現率を向上させるための技術として、類似性評価知識を用いた検索条件文の拡張方法を提案し、実験によりその有効性を示した。

参考文献

- [1] 大平 峰子, 大園 忠親, 新谷 虎松, “料理における類似性評価のための知識の発見手法について”, 第 14 回人工知能学会全国大会論文集, 2000.
- [2] 川原稔, 河野浩之, 長谷川利治, “文献データベース情報検索に対するデータマイニング技術の適用”, 情報処理学会誌 Vol.39, No.4, pp-878-887, 1998.
- [3] Catholijn M. Jonker and Arjen Vollebregt, “ICE-BERG: Exploiting Context in Information Brokering Agents”, Cooperative Information Agents 2000, pp-27-38, 2000.