

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2011-215897

(P2011-215897A)

(43) 公開日 平成23年10月27日(2011.10.27)

(51) Int.Cl.		F 1		テーマコード (参考)
G06N 5/04	(2006.01)	G06N 5/04	550N	5B075
G06F 19/00	(2011.01)	G06F 19/00	130	
G06F 17/30	(2006.01)	G06F 17/30	220Z	

審査請求 未請求 請求項の数 7 O L (全 13 頁)

(21) 出願番号 特願2010-83635 (P2010-83635)
 (22) 出願日 平成22年3月31日 (2010. 3. 31)
 特許法第30条第1項適用申請有り 平成22年3月19日 日本知能情報ファジィ学会発行の「第35回ファジィ・ワークショップ講演論文集」において発表

(71) 出願人 801000027
 学校法人明治大学
 東京都千代田区神田駿河台1-1
 (74) 代理人 100092820
 弁理士 伊丹 勝
 (72) 発明者 高木 友博
 神奈川県川崎市多摩区東三田1-1-1
 明治大学生田校舎内
 Fターム(参考) 5B075 ND03 NR12 QT05

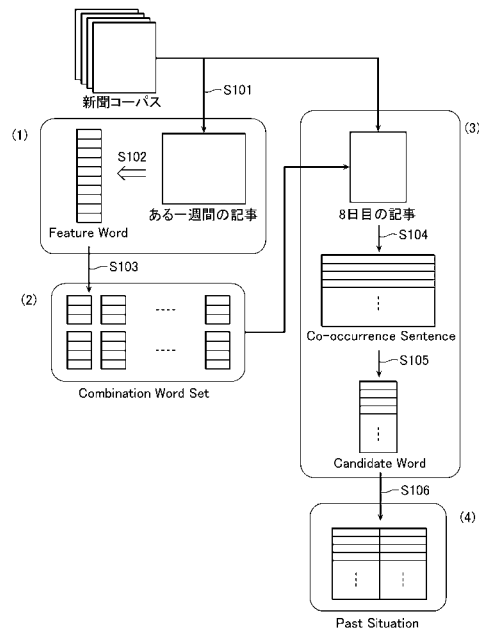
(54) 【発明の名称】 データベース、類推エンジン及び類推システム

(57) 【要約】

【課題】 言語情報に基づく類推システムを提供する。

【解決手段】 類推システムのデータベースは、複数の前提文章に含まれる単語のうち重要度の上位から抽出された複数の単語の集合である特徴語と、結論文章に含まれる単語のうち、前記特徴語に含まれる複数の単語を組み合わせた連結語に共起する単語の集合である候補語との組からなる事象データが複数記憶されていることを特徴とする。

【選択図】 図4



【特許請求の範囲】

【請求項 1】

複数の前提文章に含まれる単語のうち重要度の上位から抽出された複数の単語の集合である特徴語と、

結論文章に含まれる単語のうち、前記特徴語に含まれる複数の単語を組み合わせた連結語に共起する単語の集合である候補語と

の組からなる事象データが複数記憶されている

ことを特徴とするデータベース。

【請求項 2】

請求項 1 記載のデータベースを参照する類推エンジンであって、

複数の入力文章に含まれる単語のうち重要度の上位から抽出された複数の単語の集合である対象ベース語を生成する対象ベース語生成部と、

前記対象ベース語と前記特徴語のマッチングを行い、マッチングスコアの上位の特徴語と事象データを構成する候補語を抽出するマッチング部と、

前記マッチング部で抽出された候補語を統合して前記入力文章を前提とする結論である予測語を生成する予測語生成手段と

を備えることを特徴とする類推エンジン。

10

【請求項 3】

前記マッチング部は、

前記対象ベース語に含まれる複数の単語の組み合わせを生成し、これら複数の単語の重要度を掛け合わせてスコアを算出し、

前記データベースを参照し、前記生成された単語の組み合わせを含む前記特徴語が属する事象データを検索し、この検索した事象データのスコアに前記単語の組み合わせから算出されたスコアを加算し、新たな事象データのスコアを算出し、

前記新たな事象データのスコアの上位から複数の前記候補語を抽出する

ことを特徴とする請求項 2 記載の類推エンジン。

20

【請求項 4】

前記マッチング部は、

前記対象ベース語に含まれる複数の単語の組み合わせを生成し、

前記組み合わせられた単語全てがデータベース上で出現する確率 Q_0 、及び前記組み合わせられた各単語がデータベース上で出現する確率をそれぞれ全て掛け合わせた値 Q_n を求め、前記単語の組み合わせの重さを $\log(Q_0 / Q_n)$ で算出し、

前記データベースを参照し、前記生成された単語の組み合わせを含む前記特徴語が属する事象データを検索し、この検索した事象データのスコアに前記単語の組み合わせから算出された重みを加算し、新たな事象データのスコアを算出し、

前記新たな事象データのスコアの上位から複数の前記候補語を抽出する

ことを特徴とする請求項 2 記載の類推エンジン。

30

【請求項 5】

文章、前記文章のインデックス、及び前記文章に含まれる複数の単語の重要度に基づいて算出されたスコアからなる複数の文書データが記録されたデータベースと、

前記データベースに記録されたスコアに基づいて決定する所定数の単語からなる前提単語群を入力とし、前記前提単語群に基づいて前記データベースに記録された文書データの中から所定数の文書データを選択して出力する類推エンジンと

を備えることを特徴とする類推システム。

40

【請求項 6】

前記類推エンジンは、前記データベースの文書データの文書に含まれる単語からファジィクラスタリングを用いて複数のクラスタを生成し、これら複数のクラスタ及び前記前提単語群との類似度に基づいて前記データベースに記録された文章データの中から所定数の文書データを選択する

ことを特徴とする請求項 5 記載の類推システム。

50

【請求項 7】

前記類推エンジンから出力された所定数の文書データは、ユーザによって採否判断され

、
前記データベースは、過去に前記類推エンジンに入力された前記前提単語群、その前提単語群によって選択され出力された所定数の文書データからなる過去履歴が記録され、

前記類推エンジンは、前記過去履歴を参照し、新たに入力された前記前提単語群と前記過去履歴の前提単語群とを比較して類似度の高い過去履歴を抽出し、抽出された過去履歴の文章データ及び前記類似度に基づいて前記データベースに記録された文書データの中から所定数の文書データを選択して出力する

ことを特徴とする請求項 4 記載の類推システム。

10

【発明の詳細な説明】

【技術分野】

【0001】

この発明は、言語情報に基づくデータベース、類推エンジン及び類推システムに関する

【背景技術】

【0002】

従来ある類推システムの多くは、様々な要因を数値化した数学モデルによって実現している。この様な類推システムの具体例として、特許文献 1 記載の予測システムがある。この予測システムは、過去の損失額、過去の気象、及び経済指標データに基づいて事業リスクを予測している。

20

【0003】

しかし、経済分野に関して言えば、投資家やエコノミストは、何らかの経済的イベントが他に与える影響を感覚的に判断し、将来の経済動向を予測していることが多い。例えば、「現在の経済状態は、 $\times \times$ 年（過去）の状況に似ているため、今後 となる公算が強い」などと予測する。実際、株価は、投資家の市場に対する心理（センチメント）、つまり期待感や不安感の膨らみによって大きく変動する。このような心理変動は言語で表わされていることが多い。

【0004】

経済分野に限らず、一般的に、ある与えられた前提に対してある結果を類推する場合、言語情報に基づいて処理することは有効であるが、このような言語情報は数式に載せることが困難である。したがって、コンピュータシステムへの実装は困難である。

30

【先行技術文献】

【特許文献】

【0005】

【特許文献 1】特開 2004 - 272674 号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

この発明は、言語情報に基づくデータベース、類推エンジン及び類推システムを提供することを目的とする。

40

【課題を解決するための手段】

【0007】

本発明の一態様に係るデータベースは、複数の前提文章に含まれる単語のうち重要度の上位から抽出された複数の単語の集合である特徴語と、結論文章に含まれる単語のうち、前記特徴語に含まれる複数の単語を組み合わせた連結語に共起する単語の集合である候補語との組からなる事象データが複数記憶されていることを特徴とする。

【0008】

本発明の一態様に係る類推エンジンは、前記データベースを参照する類推エンジンであ

50

って、複数の入力文章に含まれる単語のうち重要度の上位から抽出された複数の単語の集合である対象ベース語を生成する対象ベース語生成部と、前記対象ベース語と前記特徴語のマッチングを行い、マッチングスコアの上位の特徴語と事象データを構成する候補語を抽出するマッチング部と、前記マッチング部で抽出された候補語を統合して前記入力文章を前提とする結論である予測語を生成する予測語生成手段とを備えることを特徴とする。

【0009】

本発明の一態様に係る類推システムは、文章、前記文章のインデックス、及び前記文章に含まれる複数の単語の重要度に基づいて算出されたスコアからなる複数の文書データが記録されたデータベースと、前記データベースに記録されたスコアに基づいて決定する所定数の単語からなる前提単語群を入力とし、前記前提単語群に基づいて前記データベースに記録された文書データの中から所定数の文書データを選択して出力する類推エンジンとを備えることを特徴とする。

10

【発明の効果】

【0010】

この発明は、言語情報に基づくデータベース、類推エンジン及び類推システムを提供することができる。

【図面の簡単な説明】

【0011】

【図1】概念ファジィ集合の概念図である。

【図2】本発明の第1の実施形態に係る予測システムの概念図である。

20

【図3】本予測システムの機能ブロック図である。

【図4】本予測システムのデータベースの構築方法を示す図である。

【図5】本予測システムの予測手法を説明する図である。

【図6】本予測システムの予測手法を説明する図である。

【図7】本予測システムに対する評価基準となる株価の推移を示す図である。

【図8】本予測システムに対する実験結果を示す図である。

【図9】本発明の第3の実施形態に係る推薦システムの機能ブロック及び処理のフローを示す図である。

【図10】本推薦システムの過去の操作履歴を示す図である。

【図11】本推薦システムの推薦手法を説明する図である。

30

【図12】本推薦システムの他の推薦手法を説明する図である。

【発明を実施するための形態】

【0012】

以下、添付の図面を参照して本発明の実施形態について説明する。

【0013】

[第1の実施形態]

本発明の第1の実施形態に係る類推システムは、過去の事例に基づいて経済動向を予測するシステムである。

【0014】

[概念ファジィ集合]

40

先ず、具体的な実施形態を説明する前に、本予測システムの基礎となる概念ファジィ集合 (Conceptual Fuzzy Sets、以下「CFS」と呼ぶ) について経済動向予測を念頭に置いて簡単に説明する。

【0015】

人間は、「A (前提) B (結論)」という事例に基づいて、「C X？」という問題の解 (X) を類推によって求めることができる。しかし、この解は、述語論理を用いた通常の人工知能システムでは求めることができない。また、事例ベース類推によって解を得ようとする試みもあるが成功には至っていない。一方、ファジィ論理研究の歴史においては、近似類推というテーマで長く研究されてきた。これは、ファジィ論理を他の知的システムと差別化をする重要な点であった。

50

【0016】

しかし、ファジィ論理における近似類推にも解決されていない大きな問題がある。それは、上記のAやBが、形容詞のように物事の程度を表す場合、例えば、「 x が大きい場合、 x が小さくなる」といった場合を除き、処理できない点である。しかし、実社会では、人間が類推に使用する材料には名詞で表現される事実が多い。例えば、「テロ（前提）経済不安（結論）」といったものが挙げられる。このようなイベント同士の関係は経済に大きな影響を与えるが、ファジィ論理ではその影響を類推することはできない。

【0017】

そこで、文脈に影響を受けさまざまに変化する語の意味を表現する方法として考えられたのがCFSである。

【0018】

図1は、CFSの概念図である。

CFSでは、予め、単語の意味と、文脈を示す直前の単語とのペアを文脈情報としてデータベースに蓄えておく。新たな文脈で単語（図1に示す“Context”）が出現したとき、データベースの文脈情報と適合するかの計算を行う。そして、その結果得られた適合度合いを重みにすることによって、蓄えられた単語の意味情報を統合し、新たな単語の意味（図1に示す“meaning”）を表したCFSを生成する。

【0019】

このCFSの考え方を予測システムに応用すると、図2のように表わすことができる。つまり、過去の経済状況とその後に起こった経済状況を事象データとしてデータベースに蓄えておき、新たな経済状況（図2に示す“condition”）が出現したとき、過去の経済状況と適合するかの計算を行い、新たな経済状況（図2に示す“result”）、つまり、将来の経済状況を予測する。

【0020】

[予測システムの構成]

次に、本発明の第1の実施形態に係る予測システムの機能ブロックについて説明する。

【0021】

本予測システムは、図3に示すように、過去の事例に基づくデータが記憶されるデータベース110と、このデータベース110を参照し、指定された日にちの経済動向を予測する予測エンジン120からなる。

【0022】

[データベースの構築]

次に、データベース110のデータの構築方法を図4を参照しながら説明する。

【0023】

ある日の出来事は、前の出来事に影響を受けて起こる。そこで、本実施形態では、例えば、ある日の新聞の記事（結論文章）から求めた単語と、その日前一週間の新聞の記事（前提文章）から求めた単語とを抽出し、これら単語を組にしてデータベース110に格納しておく。以下の説明において、ある日の新聞の記事とある日前一週間の新聞の記事とを合わせたものを「ウィンドウ」と呼ぶ。

【0024】

データベース110の構築は、ウィンドウを1日ずつずらしてスキャンしながら各ウィンドウにおいて以下の処理を実行する。

【0025】

先ず、図4中(1)に示すように、新聞コーパスからある一週間分の新聞記事を取り出し（図4のステップS101）、全ての記事について形態素解析して単語の重要度としてTF-IDF値（単語の出現頻度・逆出現頻度）を求める。その中からTF-IDF値の上位の単語（例えば、上位10個）を抽出する（図4のステップS102）。この抽出された単語の集合を特徴語（Feature Word）と呼ぶ。

【0026】

続いて、図4中(2)に示すように、例えば、特徴語から数個の単語を選択し、これら

10

20

30

40

50

単語の組み合わせを作成する(図4のステップS103)。この単語の組み合わせを連結語(Combination Word)と呼び、1つの特徴語から生成される連結語の組を連結語セット(Combination Word Set)と呼ぶ。

【0027】

続いて、図4中(3)に示すように、ある日、つまりウィンドウにおける8日目の新聞の記事と連結語に共起する文章(Co-occurrence Sentence)を抽出し(図4のステップS104)、全ての共起する文章に含まれている単語(名詞)を抜き出す(図4のステップS105)。これらの操作を連結語セットの全ての連結語に対して行う。このように抜き出した名詞の集合を候補語(Candidate Word)と呼ぶ。

10

【0028】

最後に、図4中(4)に示すように、候補語毎に次式(1)に示す関係レベル(Relation Level、以下「RL」と呼ぶ。)を付与する(図4のステップS106)。

【0029】

【数1】

$$RL(CA, CO) = \sum_{i=0}^n \left(\frac{\text{COUNT}(CO_i \cup CA)}{\text{COUNT}(CO_i) \times \text{COUNT}(CA)} + 1 \right) \quad (1)$$

20

ここで、「CO」は、連結語を示し、「CA」は、候補語を示す。また、「COUNT」は、全文章(記事)数(AS)に対する、所定の単語(words)が含まれた文章(記事)数(W)の割合を示す尺度であり、次式(2)で求めることができる。

【0030】

【数2】

$$\text{COUNT}(\text{words}) = \frac{W}{AS} \quad (2)$$

この式(1)に示す計算を候補語全てに対して行う。

30

【0031】

以上によって求めた特徴語と候補語を組にしたものが事象データである。この事象データを過去状況(Past Situation)と呼ぶ。この過去状況が1つのウィンドウのデータとなる。例えば、以上の操作を2年分の新聞について実行した場合、データベース110には、730ウィンドウ分程度のデータが記憶される。

【0032】

[予測エンジンと予測手法]

次に、予測エンジン120の機能ブロックについて説明する。

【0033】

予測エンジン120は、図3に示すように、予測する日にちの新聞からTF-IDF値の高い単語を抽出する対象ベース語生成部121、対象ベース語生成部121によって抽出された単語と過去の新聞とのマッチングを行うマッチング部122、このマッチング部122のマッチング結果に基づいて将来の経済動向を示す単語を生成する予測語生成部123からなる。

40

【0034】

対象ベース語生成部121は、予測する日にちの前の一週間分の記事を新聞コーパスから取り出し、図4中(1)に示す処理と同様、一週間分の新聞の記事に含まれる単語のTF-IDF値を求める。そして、全ての単語の中からTF-IDF値の上位の単語(例えば上位10個)を抽出する。この単語の集合は、図5中(1)に示すように対象ベース語(Target Base Word)となる。

50

【 0 0 3 5 】

次に、予測エンジン 1 2 0 の予測手法を図 5、図 6 を参照しながら説明する。

図 6 中 (2 . 1) に示すように、図 3 に示したマッチング部 1 2 2 が、対象ベース語から選択した、例えば、2 個の単語の全ての組み合わせを作成する (図 6 のステップ S 1 2 1)。そして、これら単語の組み合わせ毎に対象ベース語それぞれの単語に付与された T F - I D F 値を掛け合わせてスコア T T を付与する。

【 0 0 3 6 】

続いて、図 6 中 (2 . 2) に示すように、マッチング部 1 2 2 が、データベース 1 1 0 上の各特徴語を参照し (図 5 のステップ S 1 1 1)、ステップ S 1 2 1 で作成された単語の組み合わせが含まれているかを調べる (図 6 のステップ S 1 2 2)。ここで、もし含まれている場合、データベース 1 1 0 が記録する過去状況のスコア P S に対象ベース語のスコア T T を加算し、新たなスコア P S を求める。この処理をデータベース 1 1 0 に含まれている全ての過去状況に対して実行する。

10

【 0 0 3 7 】

続いて、図 6 中 (2 . 3) に示すように、マッチング部 1 2 2 が、過去状況のスコア P S の上位の候補語 (例えば上位 5 個) を抽出する (図 5 のステップ S 1 1 2、図 6 のステップ S 1 2 3)。この抽出された候補語を選択候補語と呼び、これら 5 つの選択候補語を選択候補語セット (Selected Candidate Word Set) と呼ぶ。

【 0 0 3 8 】

最後に、図 5 中 (3) に示すように、図 3 に示した予測語生成部 1 2 3 が、マッチング部 1 2 2 で求めた選択候補語を統合する (図 5 のステップ S 1 1 3)。統合する際に候補語の R L と過去状況のスコア P S を掛け合わせる。この統合したものを予測語 (Predict Word) と呼ぶ。ユーザは、この予測語生成部 1 2 3 から出力された予測語に基づいて、将来の経済動向を予測する。

20

【 0 0 3 9 】

[評価]

次に、本予測システムに対する評価を行った。

【 0 0 4 0 】

始めに、評価方法について説明する。

30

ここでは、1 9 8 9 年、1 9 9 0 年の 2 年分の新聞の記事に基づいて構築されたデータベース 1 1 0 を用いて 1 9 9 4 年の経済状況の動向を予測して評価する。評価は、図 7 に示す 1 9 9 4 年の株価のデータを基準にし、予測語と実際の株価を照らし合わせて行う。例えば、株価が上昇している時に予測語に「上昇」、「急騰」などがあった場合、高い評価とし、逆に、予測語に「下落」、「急落」などがあった場合、低い評価とする。

【 0 0 4 1 】

今回の以下の 4 パターンについて評価した。

パターン 1 : 株価が急激に上がる (図 7 の状況 A)。

パターン 2 : 株価が急激に下がる (図 7 の状況 B)。

パターン 3 : 株価が下がり、更に下がる (図 7 の状況 C)。

パターン 4 : 株価が下がり、一転して上がる (図 7 の状況 D、E)。

40

なお、パターン 3 の状況 C とパターン 4 の状況 D については、数値的な下落傾向が似ている状況を選択している。このように似通った状況を選ぶことで、数値上では下がり方が似ていても予測語で違いがあるかどうか検証するためである。また、パターン 4 については、同じように株価が下落した場合でも、予測語上どのような違いが出るかを検証するために 2 箇所の状況 D、E を選択している。

【 0 0 4 2 】

続いて、上記評価方法に対する本予測システムの結果について説明する。図 8 は、図 7 の状況 A ~ E に対するスコア上位 3 0 個の予測語を示している。

【 0 0 4 3 】

50

パターン 1 の状況 A の場合、「安」、「上昇」といった単語が出現しており、株価上昇が予測できる。

【 0 0 4 4 】

パターン 2 の状況 B の場合、「悪影響」、「低迷」、「下落」、「難」、「不振」といった単語が出現しており、株価下落が予測できる。

【 0 0 4 5 】

パターン 3 の状況 C とパターン 4 の状況 D を比較する。状況 C では「安値」、「下落」といった状況 B に似たような単語が出現している。そのため、株価が悪化することが予測できる。逆に、状況 D では「反騰」、「急騰」というような単語が出現している。このように、状況 C、D は前の一週間の株価動向が似ているにも関わらず予測結果が反対の傾向を示している。この点から本予測システムによれば、数値として株価の推移が似ていても、過去の一週間の新聞記事に出現する単語の違いを反映した好ましい結果を得ることができる。

【 0 0 4 6 】

パターン 3 の状況 D と状況 E を比較する。状況 E では「上昇」という単語が出現している。このように、状況 E の場合、状況 D の場合と表現が異なる単語が出現しているが、結果的に同じような状況を予測することができる。

【 0 0 4 7 】

以上の結果から、本予測システムによれば、過去の状況を表わす言語情報に基づいて、将来の経済動向を表わす言語情報（予測語）を得る事ができる。

【 0 0 4 8 】

[第 2 の実施形態]

本発明の第 2 の実施形態に係る類推システムは、第 1 の実施形態に係る予測システムを改良した予測システムである。

【 0 0 4 9 】

本実施形態に係る予測システムは、第 1 の実施形態に係る予測システムと比べ、マッチング部 1 2 2 による処理が異なる。そこで、以下では、第 1 の実施形態との相違点であるマッチング部 1 2 2 による処理を説明し、その他の説明については割愛する。なお、本予測システムのマッチング部の符号は便宜上 1 2 2 ' と表わす。

【 0 0 5 0 】

本実施形態のマッチング部 1 2 2 ' は、先ず、対象ベース語生成部 1 2 1 が生成した対象ベース語の中から、図 6 中 (2 . 1) と同様、例えば、2 個の単語 A、B の組み合わせを作成する (図 6 のステップ S 1 2 1 相当)。そして、これら単語 A、B の組み合わせ毎に次式 (3) のような重み MI を付与する。

【 0 0 5 1 】

【 数 3 】

$$MI(A:B) = \log \frac{Q(AB)}{Q(A)Q(B)} \quad (3)$$

ここで、 $Q(A B)$ は、単語 A、B が両方出現したデータベース 1 1 0 上の確率を示し、 $Q(A)$ 、 $Q(B)$ は、単語 A 又は単語 B が出現したデータベース上の確率を示す。

【 0 0 5 2 】

続いて、図 6 中 (2 . 2) と同様、マッチング部 1 2 2 ' が、データベース 1 1 0 上の各特徴語を参照し (図 5 のステップ S 1 1 1 に相当)、上記ステップで作成された単語 A、B の組み合わせが含まれているかを調べる (図 6 のステップ S 1 2 2 に相当)。ここで、もし含まれている場合、データベース 1 1 0 が記録する過去状況のスコア PS に対象ベース語の重み MI を加算し、新たなスコア PS を求める。この処理をデータベース 1 1 0 に含まれている全ての過去状況に対して実行する。

【 0 0 5 3 】

10

20

30

40

50

続いて、図 6 中 (2 . 3) と同様、マッチング部 1 2 2 が、過去状況のスコア P S の上位の候補語 (例えば上位 5 個) を抽出する (図 5 のステップ S 1 1 2、図 6 のステップ S 1 2 3 に相当)。この抽出された候補語が選択候補語、これら 5 つの選択候補語が選択候補語セットとなる。

【 0 0 5 4 】

その後の予測語生成部 1 2 3 による処理は、第 1 の実施形態と同様である。

本実施形態に係る予測システムによれば、第 1 の実施形態に係る予測システムに比べ、より予測の精度を高めることができる。

【 0 0 5 5 】

[第 3 の実施形態]

本発明の第 3 の実施形態に係る類推システムは、ユーザに対し、DVD タイトルを推薦する C F S を利用した推薦システムである。

【 0 0 5 6 】

[推薦システムの構成]

先ず、本実施形態に係る推薦システムの構成について説明する。

【 0 0 5 7 】

図 9 は、本実施形態に係る推薦システムの機能ブロックを示す図である。なお、図 9 には、推薦処理のフローも併せて示している。

【 0 0 5 8 】

本推薦システムは、データベース 2 1 0、このデータベースを参照してユーザのニーズにあった DVD タイトルを推薦する推薦エンジン 2 2 0、及び現在の操作履歴を記録する操作履歴記録部 2 3 0 とからなる。ここで、現在の操作履歴には、現在登録されているユーザが気に入った DVD のリスト (以下、「お気に入りリスト」と呼ぶ) と、ユーザが気に入らない DVD のリスト (以下、「不要リスト」と呼ぶ) が含まれている。

【 0 0 5 9 】

データベース 2 1 0 には、DVD データと過去の操作履歴が保存されている。DVD データは、DVD タイトルと、この DVD タイトルの内容を示す単語からなる。この単語は、DVD タイトルを紹介するテキストから抽出された名詞となっており、TF - IDF 値を用いてスコア付けされている。なお、過去の操作履歴については後述する。

【 0 0 6 0 】

[推薦手法]

次に、推薦手法について説明する。

【 0 0 6 1 】

始めに、ユーザによって、画面に表示された推薦 DVD に対して、「お気に入り」「要らない」の選好度合いを示す操作を行う (図 9 のステップ S 2 0 1)。この操作は、データベース 2 1 0 にフィードバックされ過去の操作履歴に記録されると共に (図 9 のステップ S 2 0 2)、DVD 毎に付与された単語のスコア付けの修正に用いられる (図 9 のステップ S 2 0 3)。

【 0 0 6 2 】

続いて、スコア付けされた名詞のうち、スコアの上位の単語を抽出する (図 9 のステップ S 2 0 4)。この抽出された単語は、データベース 2 0 1 にフィードバックされ過去の操作履歴に記録されると共に (図 9 のステップ S 2 0 5)、推薦エンジン 2 2 0 に入力される (図 9 のステップ S 2 0 6)。

【 0 0 6 3 】

続いて、推薦エンジン 2 2 0 は、ステップ S 2 0 6 において入力された入力単語 (前提単語) に基づいて推薦する DVD タイトルをスコア付けし、DVD リストとして出力する (図 9 のステップ S 2 0 7)。

【 0 0 6 4 】

続いて、ステップ S 2 0 7 において出力された DVD リストに対して、現在の操作履歴のお気に入りリスト及び不要リストに含まれている DVD タイトルでフィルタリングし (

10

20

30

40

50

図9のステップS208)、推薦DVDとして画面に表示する(図9のステップS209)。

【0065】

最後に、ステップS209で表示された推薦DVDに対して、ユーザが選好度合いを示す操作を行う(図9のステップS201)。

【0066】

以降、上記ステップS201～S209が繰り返される。

なお、一連の処理によって、図10に示すような過去の操作履歴が構築されることになる。この過去の操作履歴は、推薦エンジン220に対する入力単語と、この入力単語に基づいて推薦エンジン220が出力した推薦DVDに対するユーザの選好度合いを示すデータの組となっている。

10

【0067】

次に、推薦エンジン220の処理例について説明する。

第1のタイプの概念を図11に示す。このタイプは、ファジィクラスタリングを用いてアトミックファジィ集合を生成し、その重ね合わせによってCFSを実現している。

【0068】

始めに、全てのDVDデータの単語に対してファジィクラスタリングを行い、入力単語と各クラスタとの類似度を求める。

【0069】

続いて、類似度の高いクラスタ(図11の斜線を示したクラスタ)を用いて拡張語群を作成する。

20

【0070】

最後に、この拡張語群とDVDデータの類似度を求め、類似度の高いDVDタイトルをDVDリストとして出力する。

【0071】

第1のタイプでは、単語の特徴付けをし、単語の意味を形成することでDVDの内容に基づいた推薦をすることができる。このタイプは、ユーザの操作履歴に依らず、入力単語のみを利用するため、次に説明する第2のタイプの操作履歴集めに用いることができる。

【0072】

第2のタイプの概念を図12に示す。このタイプは、入力単語とDVDの内容紹介テキストにおける単語の共起のみで関係を導きだし、推薦を行っている。

30

【0073】

第2のタイプは、データベースに蓄えてある過去の操作履歴の入力単語と、現在の入力単語とを比較(マッチング)する。その結果、類似度が高い過去の操作履歴について、この過去の操作履歴のDVDリストを足し合わせて、DVDリストとして出力する。

【0074】

この第2のタイプでは、過去の操作履歴を利用しているため、ユーザの趣向に沿ったDVDタイトルを推薦することができる。

【0075】

以上から、本推薦システムによれば、過去の操作履歴とDVDタイトルの言語情報からユーザの趣向に沿ったDVDタイトルを推薦できるばかりでなく、過去の操作履歴に他人の操作履歴も含まれていることから、ユーザにとって意外性のある新規なDVDタイトルの推薦が可能になる。

40

【0076】

[その他]

以上、CFSを利用した類推システムの応用例として、経済動向の予測システムとDVD推薦システムについて説明したが、本発明に係る推薦システムは、以上のようなシステムに限定されるものはない。例えば、音楽を推薦するシステム、ユーザの身体の状態や味の好みに基づいた料理レシピ紹介システム、気候や経済状況に基づいた洋服の流行を予測するシステムなど、様々なシステムに応用することができる。

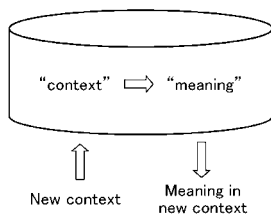
50

【符号の説明】

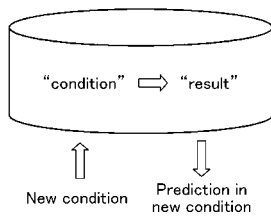
【 0 0 7 7 】

- 1 1 0 データベース
- 1 2 0 予測エンジン
- 1 2 1 対象ベース語生成部
- 1 2 2 マッチング部
- 1 2 3 予測語生成部
- 2 1 0 データベース
- 2 2 0 推薦エンジン
- 2 3 0 操作履歴記録部

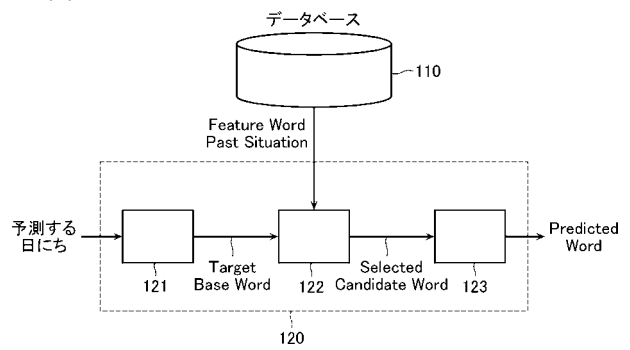
【 図 1 】



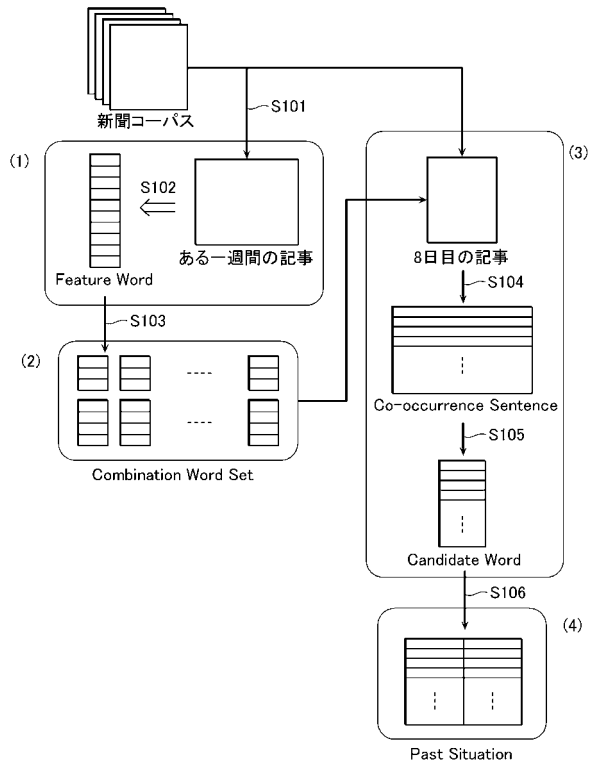
【 図 2 】



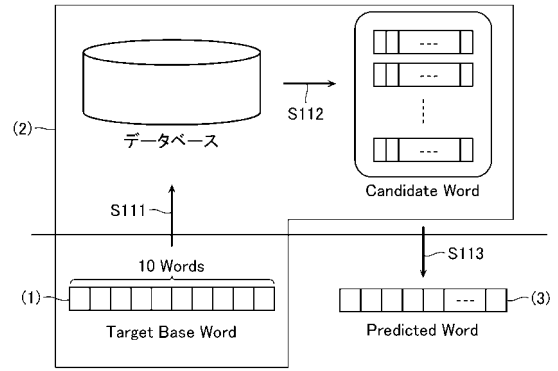
【 図 3 】



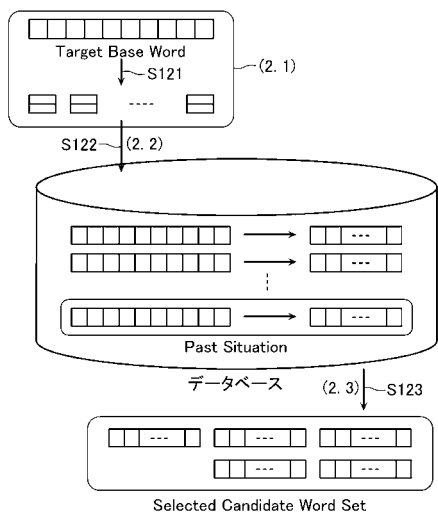
【 図 4 】



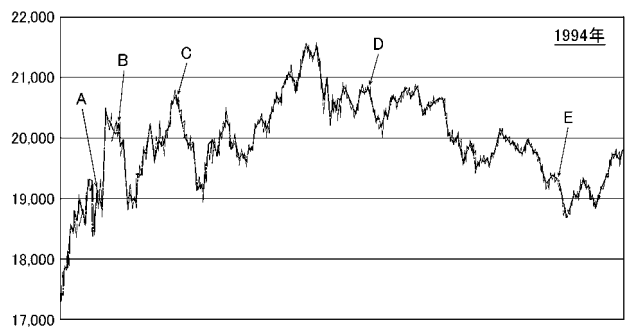
【 図 5 】



【 図 6 】



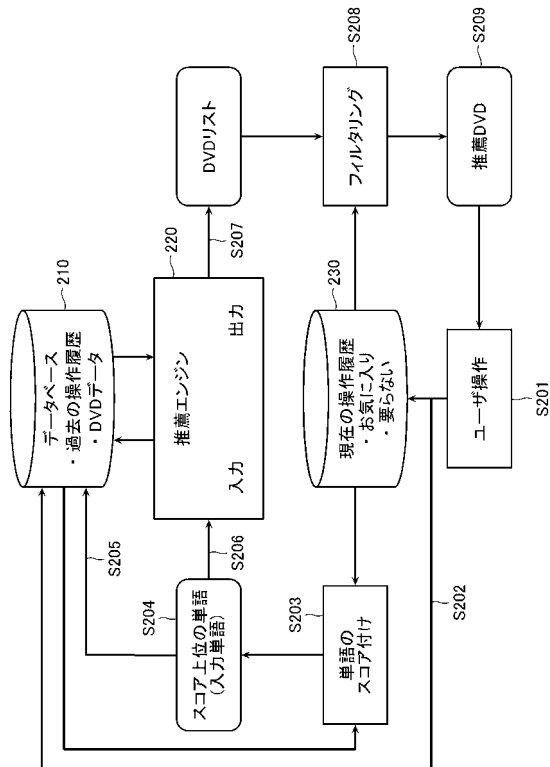
【 図 7 】



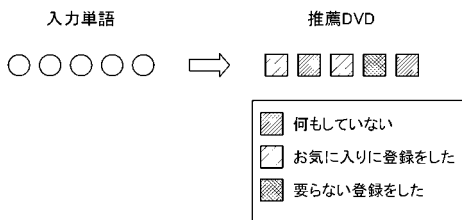
【 図 8 】

順位	A(1)	B(2)	C(3)	D(4)	E(4)
1	相場	株価	安値	株価	金利
2	為替	建	下落	相場	相場
3	金利	投資	小幅	金利	為替
4	材料	悪影響	日数	平均	金融
5	調整	金利	立会	反騰	預金
6	株価	低迷	小口	市場金利	上昇
7	審判	以後	排除	投資	株価
8	圧力	劣便	株式	材料	銀行
9	左右	資金繰り	投資	円	機関
10	物価	景気	開示	銀行	マネーサプライ
11	実	産別	種	台	円安
12	国債	電気労連	余地	仕手株	出来高
13	入札	ドル	決算	利下げ	前日比
14	統計	設備	前日	外国為替市場	安値
15	家	不振	平均	年初渡安値	市場金利
16	総務	好調	意図	急騰	短期プライムレート
17	安	業績	反発	機関投資家	終値
18	理事	前半	狙い	現物	通貨供給量
19	か月	株々	相場	米連邦基準精度理事會	騰落レシオ
20	上昇	株式市場	銀行	裁定取引	景気
21	台	債券	ぶり	金融緩和	調達
22	末	強弱	利益	預金	コスト
23	感	中東	処理	株	大蔵省
24	円	高	海外	下落	定期
25	海外	企業	信託	期待	感
26	中央	先月	円安	連動	資金
27				減益	調達
28				決算	金利目目化
29				減益	
30				決算	

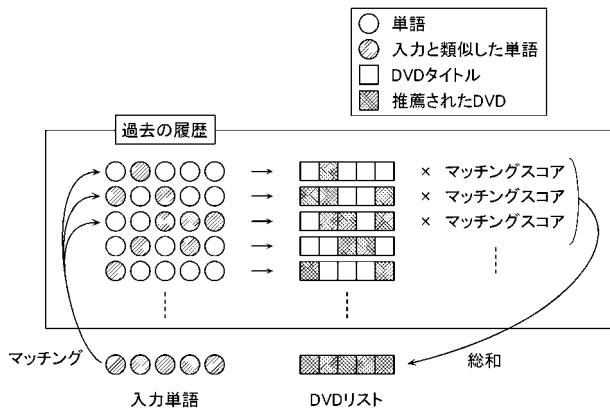
【 図 9 】



【 図 10 】



【 図 12 】



【 図 11 】

