

マンマシンコミュニケーションにおける話題の追跡

3T-8

西村俊和 廣瀬勝一 美濃導彦 池田克夫

京都大学工学部

1 はじめに

人間同士の対話が円滑に行われる要因の一つに、対話者双方が話題の追跡を行っていることが挙げられる。話題という情報の共有を双方で仮定することにより、適度の省略や指示語等の使用が可能であり、簡潔で当を得た対話が可能となる。

マンマシンコミュニケーションにおいても、計算機側で話題を追跡することにより、同様の対話が可能となると期待できる。

本論文ではコマンドに複数のキーワードを対応付け、このキーワードを利用してコマンドの集合であるタスクに名前付けを行い、対話の話題を記述して追跡する手法を提案する。キーワードを適切に選ぶことにより、対話の話題をタスク名と対応させて計算機上に記述することが可能となる。この手法では、あらかじめ用意すべき知識はキーワードのみであり、コマンドの数と同じオーダーである。

2 キーワードによるタスクの名前付け

コマンドの全集合の部分集合をタスクと定義する。タスクを構成するコマンドの数をそのタスクの大きさと呼ぶ。ここでは、大きさが2以上のタスクを考える。ある名前付けの方法に基づいて、タスクにつけた名前をタスク名と呼ぶ。

コマンド数は有限とし、コマンドの全集合を U_C とする。コマンドのベキ集合を 2^{U_C} と表記する。キーワードは有限長の文字列とする。キーワードの全集合 U_K は有限集合とする。キーワードのベキ集合を 2^{U_K} と表記する。キーワードからコマンドのベキ集合への写像を $f: U_K \rightarrow 2^{U_C}$ とし、キーワード $k \in U_K$ について、 $f(k)$ はキーワード k と対応づけられたコマンドの集合を表すものとする。ただしキーワードは必ず複数のコマンドに対応しているものとする。

キーワードのベキ集合からコマンドのベキ集合への写像 $f: 2^{U_K} \rightarrow 2^{U_C}$ は以下の条件を満たすものとする。

1. $f(\phi) = U_C$.

2. $f(\{k\}) = f'(k)$.

3. $K \subseteq U_K, K = K_1 \cup K_2$ のとき
 $f(K) = f(K_1) \cap f(K_2)$.

タスク C について、 n 個のキーワード

$k_i (i = 1, 2, \dots, n)$ が $f'(k_i) \supseteq C$ を満たすとき、 k_i は C の各要素に共通して対応するキーワードであるから、 k_i を C の話題と考えることができる。逆に、 $f(K') = C$ となるキーワードの部分集合 K' が存在した場合、 K' はタスク C を一意に決定する話題であると考えられる。

定義 タスク C について、以下の条件を満たす U_K の部分集合 $t(C)$ を C のタスク名あるいは名前という。

1. $f(t(C)) = C$.

2. 任意の $S \subset t(C)$ について $f(S) \supset C$. \square

タスク名 $t(C)$ はタスク C を一意に決定するための十分かつ極小数の U_K の部分集合である。

3 マニュアルページを利用したキーワードの生成

コマンドに対応するキーワードの生成法として、あらかじめシステム設計者がコマンドとキーワードの対応付けを行うことが考えられる。しかし、この方法は、対象とするシステムが拡張され、新たなコマンドが加わるたびに、人手によりコマンドとキーワードの対応を更新する必要がある。そこで、システムの拡張に際して柔軟に対応するため、コマンドに対応するキーワードを、そのマニュアルページから自動的に生成することを試みる。マニュアルページは、UNIX のコマンドを説明するオンラインドキュメントである。マニュアルページを利用したキーワード生成の手順を以下に示す。

1. マニュアルページからコメントや項目名を削除し、一単語毎に切り出す。
2. 切り出した各単語について、動詞の活用形を原型に戻し、接尾語、接頭語のついた単語を原語に戻す。また、冠詞、助動詞、接続詞等、コマンドに特有でないと思われる単語を削除する。

Tracking the focus of attention in man-machine communication

Toshikazu NISHIMURA, Shoichi HIROSE, Michihiko MINOH, Katsuo IKEDA

Kyoto University

- 各マニュアルページに高々1回しか現れない単語は重要でない単語と見なし、削除する。
- 他のコマンドのマニュアルページに現れていない単語は、キーワードとして不適当なので削除する。残った単語をそのコマンドのキーワードとする。

表1にキーワードの例を挙げる。これらは、後述の実験で用いるコマンドを対象として得られたものである。

表1 キーワードの例

コマンド	キーワード
vi	character classification command contain edit editor environ ex file flag input option set tag test text variable vi view
lpr	character contain copy data example file font job name option page print produce send specify spool standard use

4 キーワードの評価

キーワードによるタスクの名前付けにおいては、コマンドに対応付けるキーワードが非常に重要である。キーワードの対応付けの方法によっては、互いに何ら関連のないコマンドの集合に対してタスク名がつく場合も考えられるが、このようなタスク名は対話の話題を示さない。

人手によりキーワードを評価する場合、定量的な評価は難しい。ここでは統計的手法を用いて、各コマンドに前節の手続きで対応付けたキーワードの評価を行う。

ヒストリは、ユーザが計算機に入力したコマンドの履歴である。ヒストリ中の連続したコマンドからなるタスク(これをヒストリ中のタスクと呼ぶ)に注目し、これがタスク名を持つかを調べることにより、キーワードの妥当性について論ずる。ヒストリ中の連続したコマンド列は、ある話題のもとで入力されたと考えられるので、キーワードとコマンドの対応が妥当であれば、ヒストリ中のタスクは多くの場合タスク名を持つと考えられる。

タスク名とヒストリ中のタスクの関係について検定を行う。以下の試行について、帰無仮説 H_0 をたてる。

試行 すべてのタスク N 個のうち、 T_h 個にタスク名が存在した。ヒストリ中の T_h 個のタスクのうち T_{nh} 個にタスク名が存在した。

帰無仮説 H_0 タスク名の有無とヒストリ中のタスクは無関係、すなわち試行の結果は偶然である。

18人のユーザから得られた約40万行のヒストリ中で二回以上用いられ、マニュアルページを持つコマンドは280個であった。このコマンドと前節の方法で得られたキーワードを用いて、ヒストリ中のタスクおよびタスク名をもつタスクの数を求めた。タスクの大きさ別に表2にまとめる。ただし、タスクの大きさが5以上では $T_{nh} = 0$ であったので、大きさが4以下のタスクを対象とした。

帰無仮説 H_0 における試行では、タスクを T_h 個取り出したときのタスク名の付いているタスクの個数の期待値 m 、標準偏差 σ はそれぞれ

$$m = T_h T_n / N, \sigma = \sqrt{T_h T_n (N - T_n) / N}$$
 とみなせる。

試行結果 T_{nh} 、期待値 m 、標準偏差 σ および確率変数 T' をタスクの大きさ別に表3にまとめる。ただし、 $T' = (T_{nh} - m) / \sigma$ である。

表2 ヒストリ中のタスクのタスク名の有無

タスクの大きさ	タスク数	T_h	T_n	T_{nh}
2	39060	2660	4118	513
3	3619560	5806	19512	98
4	250654530	6712	52811	4
合計	254313150	15178	76441	614

T_h : ヒストリ中のタスクの数

T_n : タスク名の存在するタスクの数

T_{nh} : タスク名を持つヒストリ中のタスクの数

表3 試行の期待値、分散および試行結果の差

タスクの大きさ	期待値 m	試行結果 T_{nh}	試行	
			σ	T'
2	280.4	513	15.87	14.65
3	31.50	98	5.579	11.95
4	1.414	4	1.189	2.174

$$T' = (T_{nh} - m) / \sigma$$

タスクの大きさが2または3の場合、確率変数 T' は、非常に大きな値を持っており、非常に小さな危険率 ($\ll 1\%$) で帰無仮説 H_0 を棄却することができる。すなわち、マニュアルページのキーワードを用いたタスク名と、ヒストリ中のタスクは関係があることが明らかになった。タスクの大きさが4の場合には約3.00%の危険率で、帰無仮説 H_0 を棄却することができる。要素数2, 3の場合よりも危険率が大きい、同様にタスク名とヒストリ中のタスクは関係があることがわかる。

このようなことから、マニュアルページから抽出したキーワードによるタスク名は、ヒストリ中のタスクと関係があるという点で妥当であると結論できる。

5 おわりに

本研究では、コマンドに対応する概念をマニュアルページから得られたキーワードで記述し、話題をタスク名として記述して追跡する手法を提案した。今後は、タスク名を利用することにより、省略、指示語の使用が可能なインタフェースの構築を進める。また、文脈からユーザの意図を理解することにより、柔軟な応答をするヘルプシステム、状況に応じて変化するメニュー等の実現についても研究を行う。