

【公報種別】特許法第17条の2の規定による補正の掲載
 【部門区分】第6部門第3区分
 【発行日】平成19年9月27日(2007.9.27)

【公開番号】特開2006-72787(P2006-72787A)
 【公開日】平成18年3月16日(2006.3.16)
 【年通号数】公開・登録公報2006-011
 【出願番号】特願2004-256505(P2004-256505)
 【国際特許分類】

G 0 6 F 17/30 (2006.01)

【F I】

G 0 6 F 17/30 3 4 0 B

G 0 6 F 17/30 1 7 0 A

G 0 6 F 17/30 3 3 0 C

【手続補正書】
 【提出日】平成19年8月10日(2007.8.10)
 【手続補正1】
 【補正対象書類名】明細書
 【補正対象項目名】0036
 【補正方法】変更
 【補正の内容】
 【0036】
 【数6】

$$p\lambda(y|x) = \frac{1}{Z_{\lambda}(x)} \exp\left(\sum_i \lambda_i f_i(x, y)\right)$$

なお、この確率モデルに相当するモデルは、最大エントロピー法に限らず、HMM（隠れマルコフモデル）、SVM（Support Vector Machine）などを用いて構築することもできる。

【手続補正2】
 【補正対象書類名】明細書
 【補正対象項目名】0050
 【補正方法】変更
 【補正の内容】
 【0050】

作成される質問特徴集合の各特徴は以下の通りである。なお、品詞体系は形態素解析ツールChaSenが出力するIPA（Information-Technology Promotion Agency）の最大4階層の品詞体系を用いている。例えば、「多岐川」の品詞は「名詞－固有名詞－人名－姓」であり、助詞「が」の品詞は「助詞－格助詞－一般」である。以下、最左の品詞から順に、品詞1、品詞2、品詞3、品詞4と呼ぶ。

－（qw）質問中の単語のn-gram（ $1 \leq n \leq N$ 、nは整数）の列挙（例：「首相は誰」に対し、 $N=2$ の場合、「qw：首相、qw：は、qw：誰、qw：首相は、qw：は誰」を特徴とする。）

－（qq）質問中の疑問詞（「誰」「どこ」「何」「いつ」等）

－（qm1）質問中の単語の品詞1の異なりの列挙（例：「首相は誰」に対し、「qm1：名詞、qw1：助詞」を特徴とする）

－（qm2）質問中の単語の品詞2の異なりの列挙

－（qm3）質問中の単語の品詞3の異なりの列挙

－ (q m 4) 質問中の単語の品詞 4 の異なりの列挙

本実施の形態では、q w については、4 - g r a m まで作成している。