

テキスト分類支援ツール FLUTE の開発 (2)

4 J-6

— 障害事例分類への適用 —

森本 由起子 間瀬 久雄 辻 洋 絹川 博之
 (株)日立製作所 システム開発研究所

1 はじめに

多種多様のテキスト情報が容易に作成されてネットワーク上を流れるようになった結果、それらを一つの窓口で受信し、カテゴリ別に自動分類したいという要求が出てきた。その要求に応えるべく、我々は、電子テキストを既存カテゴリに分類するシステムの構築を支援するテキスト分類支援ツール FLUTE を開発している [1][2]。本稿では、FLUTE の有効性を検証するために、ソフトウェアの障害事例分類システムの構築に適用した結果について考察する。

2 障害事例分類問題の記述

2.1 問題の記述と従来のアプローチ

ここでいう障害情報分類システムは、個々の障害の状況や対策について記述したテキスト情報を、障害の内容に応じて既定のカテゴリに自動分類して DB に格納し、必要に応じて検索できるシステムであり、以下のような活用方法がある [3]。

- (1) システム開発におけるノウハウ取得
- (2) 障害発見時の類似事例検索
- (3) 教育支援

上記の活用を踏まえた障害情報分類システムを実現するためには、次の課題をクリアする必要がある。

- (1) ユーザの入力が容易
- (2) 必要な情報を瞬時に取得可能
- (3) 分類知識ベースの作成と保守が容易

エキスパートシステムアプローチによる障害情報分類システム SOCKS [3] では、予め用意したキーワード群の中から関連のあるキーワードをユーザに選択させる。予め用意したキーワード群は、それぞれ関連のある事例と重み付きでリンク付けしており、これらの対応情報を関連度計算ルールとして保持する。関連度計算ルールを用いて、ユーザが選択したキーワードとの関連度を事例毎に計算し、関連度の高い事例を出力する。また、キーワードから連想される

別のキーワードを重み付きで定義したキーワード連想ルールを保持する。

しかし、SOCKS には以下の問題点があった。

- (1) 予め提示された 400 以上のキーワード群の中から関連するキーワードをユーザに選択させるため、キーワードの選択洩れなど必要十分な情報を入力させることが困難である。
- (2) ルールの作成に数か月を要している。また、ルールの重み付け方法にノウハウが必要なために、特定の専門家にしか作成・保守ができない。

2.2 提案するシステムイメージ

FLUTE を用いた分類システムでは、上記の問題点を次のように解決する。

- (1) ユーザがワープロ感覚で自由に入力した任意の文章を入力とすることで、キーワード選択方法よりも入力の範囲を広げる。
- (2) 過去の事例からその事例を特徴付けるキーワードを重み付けで抽出して収集することにより、分類のための知識ベースを全自動で作成する。

図 1 に、本システムの応用イメージを示す。

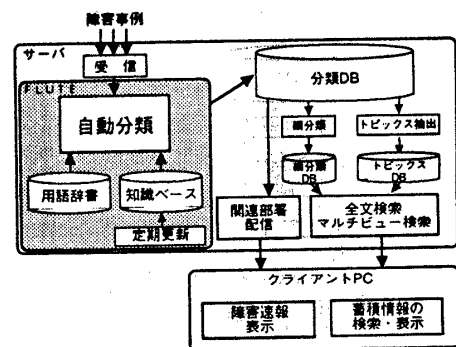


図 1 本システムの応用イメージ

3 シミュレーション実験

2.2 で述べたシステムの実現可能性を検証すべく、FLUTE を用いて実験システムを構築し、評価実験を行なった。実験システムでは、入力された事例を 13 の大分類、161 の小分類にカテゴリ分類する。

*Text Categorization System Building Support Tool.
 †Yukiko MORIMOTO, Hisao MASE, Hiroshi TSUJI, Hiroshi KINUKAWA
 ‡Systems Development Laboratory, Hitachi Ltd.

3.1 ソフトウェア障害事例

知識ベースの作成及び分類精度の評価に用いる障害事例 161 件を用意した。各事例は 161 の小分類カテゴリに 1 対 1 で対応している。各事例は、表 1 に示すように 8 つの項目から構成される。

表 1 事例の構成

1: 事例 ID	事例の識別子	教師 データ
2: 題名	事例から得た教訓を端的に記述	
3: 教訓, 要点	事例の要点、得られる教訓の記述	
4: 品名	事例の対象となる分野、製品名	
5: 事例	障害の現象面から見たタイトル	評価 データ
6: 現象	現象面から見た記述	
7: 原因	現象の発生原因	
8: 対策	現象の解決方法	

3.2 知識ベースの作成

知識ベース作成用の事例（教師データ）と分類精度測定用の事例（評価用データ）は、全く独立であることが望ましい。今回用意した事例数が少ないため、各事例を表 1 に示すように、題名、教訓、要点を教師用データ、事例、現象、原因、対策を評価用データとした。すなわち、実験システムでは、事例、現象、原因、対策の 4 項目の少なくとも 1 つについて記述した文章を入力すると、そこからキーワードを抽出し、知識ベースを参照して関連の深い教訓を持つ事例に対応づけ（分類）、その事例の事例 ID 及びその事例のもつ教訓・要点を出力する仕様にした。

図 2 に知識ベースのサンプルを示す。重みの高いキーワード、あるいは、他のカテゴリに存在しないキーワードは重要なキーワードと見なされる。161 の事例から 2206 レコードからなる知識ベースを自動作成した。

カテゴリ, キーワード, 重み	カテゴリ, キーワード, 重み
11-10, 10進数, 1	07-15, COBOL, 2
13-05, 10進数, 1	02-05, COMMON, 1
06-12, 16進, 2	09-06, CPU, 2
11-10, 16進, 1	12-13, CPU, 1
03-14, 2重, 1	01-07, DO, 1
12-01, 2重, 1	13-09, ELSE, 1

図 2 知識ベースのサンプル

また、FLUTE の一機能であるキーワードのシソーラス展開を用いることが可能である。今回はシソーラス辞書として、SOCKS で用いられているキーワード連想ルールを流用することにした。キーワード連想ルールには、あるキーワードから連想されるキーワードが定義されている。シソーラスを用いることにより、2809 レコードの知識ベースが自動作成できた。なお、知識ベースの作成するのに要した時間は、約 15 分であった。

3.3 分類実験の結果及び考察

FLUTE により、各事例の分類先を決定する。これに要する時間は 1 事例にあたり約 4 秒であった。

表 2 に本実験の分類結果を示す。ここでは、実験システムが順序付けて出力した上位 n 個のカテゴリの中に、最も関連の深いカテゴリが含まれていれば正解とし、その件数の割合を評価尺度とした。n=1 とした場合、大分類ではシソーラスを利用することにより、若干の精度向上が見られた。

また、実験システムの出力した分類結果について、最も関連のあるカテゴリ（正解カテゴリ）が上位から何番目にランクされたかについての平均値を、表 3 に示す。小分類で 4.4 位前後であり、SOCKS に比べると、やや悪い結果となった。キーワードにはなり得ない不要語を除去したり、知識ベースのチューニングにより、SOCKS 以上の精度を実現できると考える。また、知識ベースの開発工数では、FLUTE を用いる方が少なく済んだ。

表 2 本実験の再現率

	シソーラス利用	n = 1	n = 3
大分類	する	75.2 %	93.2 %
	しない	76.4 %	92.6 %
小分類	する	68.3 %	79.5 %
	しない	68.3 %	78.9 %

表 3 カテゴリのランキング平均値

	大分類		小分類		SOCKS
	なし	あり	なし	あり	あり
シソーラス展開					
平均値	2.08	2.06	4.34	4.50	2.60

4 研究開発のまとめ

テキスト分類支援ツール FLUTE を、ソフトウェアの障害事例分類システムの構築に適用した。その結果、人手で作成したルールを用いたエキスパートシステム型の自動分類システムと比較して、分類のための知識ベースを短期間かつ低コストで初期作成し、その分類精度も短期間で検証することができた。

参考文献

- [1] 辻他 3 名: テキスト自動分類エキスパートシステムの一構成法、情報処理学会第 49 回全国大会講演論文集 (3)3-93、('94.9)
- [2] 間瀬他 3 名: テキスト分類支援ツール FLUTE の開発 (1) - 機能と構成 -, 情報処理学会第 52 回全国大会講演論文集、('96.3)
- [3] 橋本他: ソフトウェア常識集知的検索エキスパートシステム、日立評論 (Vol.70, No.11, '88.11)