トップエスイー修了制作

COPY文に着目したCOBOLソースコード依存関係可視化手法の検討

伊藤　秀朗

開発における問題点

■国内の金融・公共分野の基幹系システムは長年の開発運用によりアプリケーションが大規模化・複雑化している.

■上述のような既存システムの再構築案件では，発生する不具合の7割が上流工程に起因したとの報告がある[1].

解決策

■COPY文に着目したCOBOLソースコードの依存関係可視化手法を提案する. 本手法は，同様のCOPY文を複写するCOBOLプログラムは互いに業務的に関連するとの仮説に基づいている. 実システムへの適用を通じ，提案手法が有効に働くことを確認した

研究課題とアプローチ

目的　大規模化・複雑化した業務システムの理解支援

研究課題

■ 業務アプリケーションを対象とした、業務的役割を考慮したソフトウェア可視化手法は知られていない.
・ソフトウェアの構造：Call Graph/Matrix等
・ソフトウェアの振る舞い：シーケンス図再生

■ COBOL言語のCOPY文の利用の仕方に着目して，COBOLプログラム間の業務的関連度に基づいてCOBOLプログラムを分類・可視化する.

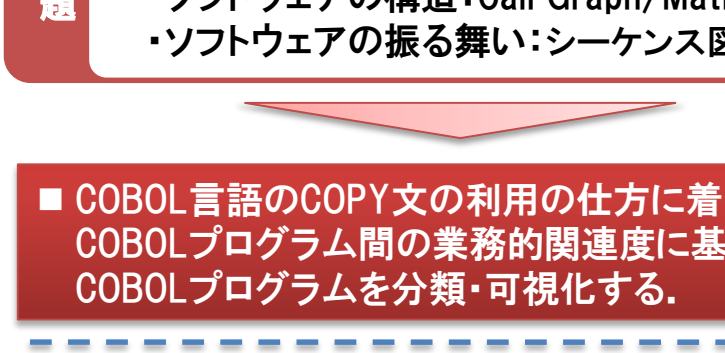
提案手法の概要

■COBOLプログラムP_iの，COPY文C_kの複写有無をd_{ik}として，P_i，P_jの間の業務的関連度s_{ij}を次のように定義する;

$$s_{ij} := \frac{\Sigma_k \, step(d_{ik} \times d_{jk})}{\Sigma_k \, step(d_{ik} + d_{jk})} \Leftrightarrow \frac{Q_i \cap Q_j}{Q_i \cup Q_j} \qquad d_{ik} = \begin{cases} 1 : P_i \text{は} C_k \text{を複写する,} \\ 0 : P_i \text{は} C_k \text{を複写しない.} \end{cases}$$

☆COBOLプログラムP_iが複写するCOPY文集合をQ_iとすると，s_{ij}は集合間類似度の1つであるJaccard係数と等価になる

例）



Q_1=(1,1,1,0,0,0)　Q_2=(1,1,1,1,0,0)　Q_3=(0,0,0,0,1,1)

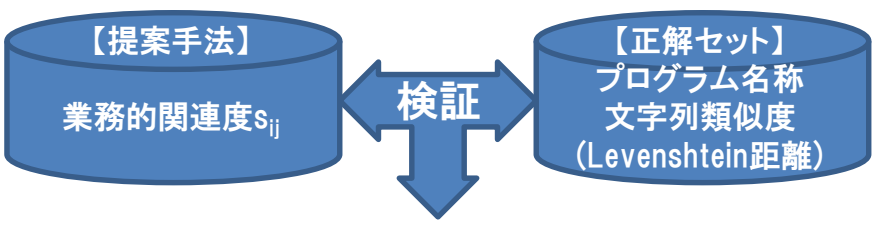
d_{ik}=

	C1	C2	C3	C4	C5	C6
P1	1	1	1	0	0	0
P2	1	1	1	1	0	0
P3	0	0	0	0	1	1

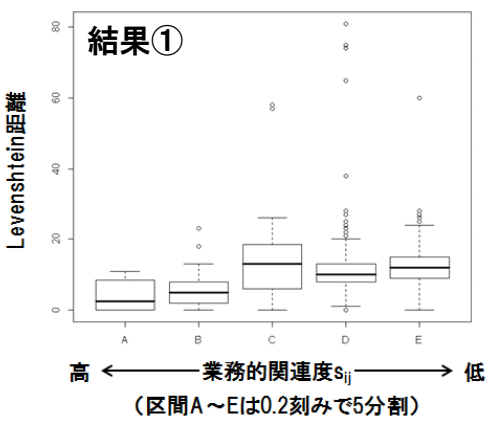
s_{ij}=

	P1	P2	P3
P1	1.00	0.75	0.00
P2	0.75	1.00	0.00
P3	0.00	0.00	1.00

評価実験



結果①



結果②：　s_{ij}=1の場合

	プログラム名称が全一致	部分一致	プログラム名称が異なる	計
正解	14	3	5	22
不正解	0	0	2	2
	14	3	7	24

評価・考察と今後の課題

評価・考察

■業務的関連度s_{ij}によって業務が近いCOBOLプログラムの抽出を確認した（結果①より）
■s_{ij}=1では精度は91.6%(=22/24)となった（結果②）. 特筆すべきはプログラム名称が異っても業務的に関連するプログラムを抽出できた点である.
■COBOLプログラムとCOPY文の依存関係という，ソースコード解析情報のみを入力とする点は提案手法の優位性に繋がると考えられる.

今後の課題

■適切な「正解」を用いた提案方式の再評価実験
■類似度s_{ij}を加味したCOBOLソースコード依存関係グラフを用いた現行システム理解支援手法

[1]NIST Planning report 02-3, *The Economic Impacts of Inadequate Infrastructure for Software Testing*, (2002).

NII 国立情報学研究所 National Institute of Informatics
トップエスイー ～サイエンスによる知的ものづくり教育プログラム～