人工知能学会
第18回知識流通ネットワーク研究会
SIG-KSN-018-01

GSN レビュ実験と評価について

山本　修一郎[1], 森崎　修司[2], 渥美　紀寿[1], 近藤　純平[2], 大林　英晶[2]

[1] 名古屋大学 情報連携統括本部　情報戦略室
[2] 名古屋大学大学院情報科学研究科

An experimental evaluation of GSN review

Shuichiroh Yamamoto[1], Shuji Morisaki[2], Noritoshi Atsumi[1],
Junpei Kondo[2], Hideaki Oobayashi[2]

[1] Strategy Office, Information and Communications Headquarters Nagoya University
[2] Graduate School of Information Science, Nagoya University

概要

保証ケースは，システムの安全性や信頼性を保証するための文書であり，GSN と呼ばれる表記法で記述することができる．しかし，GSN を客観的にレビュする際のレビュ観点は統一されていないため，レビュアーによって異なる結果になる可能性がある．本稿では，2 名のレビュアーによる GSN レビュ実験の結果とその評価について述べるとともに，GSN レビュ指標について考察する．

Abstract

Assurance case is a document to assure safety and reliability of systems. Assurance cases are described in the notation called GSN(Goal Structuring Notation). However, there is a problem that GSN review points are not well defined. Therefore GSN review results tend to vary according to reviewers. In this paper, we describe the review result of an experimental evaluation on different GSN diagrams developed by 14 students for a system. The review is executed by two students. The quantitative GSN review index is also proposed.

1　はじめに

保証ケース（Assurance Case）[1] では，立証したい主張をテスト結果や検証結果などの証拠に基づいてステークホルダ間で議論することにより，主張についての合意を形成することができる．保証ケースの中でも，システムの安全性を確認するものを安全性ケース（Safety Case)，ディペンダビリティを確認するものをディペンダビリティケース（Dependability Case）と呼び，この他にも様々な目的への利用方法が提案されている．保証ケースは，GSN（Goal Structuring Notation）と呼ばれるグラフィカルな表記法によって記述することができる．しかし，GSN を客観的にレビュする際のレビュ観点が統一されていないため，同じ GSN に対してもレビュアーによって異なるレビュ結果となる可能性がある．

そこで本稿では，2 人のレビュアーによる GSN レビュ実験を行うことで，レビュアーが異なっても同じように GSN の問題点を指摘できるかどうかについて調査した結果とその評価について述べるとともに，レビュ観点について考察する．

本稿の構成は次の通りである．2 章でレビュ実験について述べ，3 章で実験結果を示す．4 章では実験結果について考察することで，GSN レビュ指標を明らかにする．最後に 5 章でまとめと今後の課題を明らかにする．

2　レビュ実験

2.1　実験仮説

本実験では,「レビュアーが異なっても，同じように GSN の問題点を指摘できる」という仮説を

表 1: 各 GSN の総ノード数

データ番号	A	B	C	D	E	F	G	H	I	J	K	L	M	N	平均
総ノード数	39	21	53	27	29	33	30	35	21	21	38	26	29	19	30

表 2: 指摘項目種別

種別	レビュ観点	指摘内容例
N	GSN の表記法が誤っている	戦略ノードに証拠ノードが接続されている
P	議論分解パターンが不適切である	戦略 S_x とサブゴールの論点が異なっている
C	必要な前提ノードが不足している	ゴール G_x に前提を示す必要がある
E_g	ゴールノードの記述内容が不適切である	ゴール G_x は特性でなく動作を記述する必要がある
E_c	前提ノードの記述内容が不適切である	接続されている戦略ノードの内容と関係がない
E_e	証拠ノードの記述内容が不適切である	証拠 E_x は，より具体的な証拠を示す必要がある
U	未定義の用語が使用されている	証拠 E_x に未定義の用語が使用されている

立てた．

2.2 実験対象

本実験のレビュ対象は，14 人の作成者が以下に示すような例題に従って記述した GSN である．なお，各 GSN に A〜N のデータ番号を割り振ることでデータの識別を行った．

- 例題：
  電気ポットの加熱制御ソフトウェアに対して，次の加熱安全原則，システム構成，欠陥分析に基づいて，「加熱が安全である」という主張に対する安全性ケースを作成しなさい．
- 加熱安全原則：
  - タンクの水位が不適切（空，満水）のとき，加熱しない
  - 沸騰したら加熱しない
- システム構成：

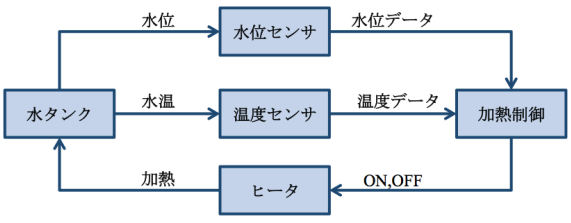

図 1: システム構成図

- 欠陥分析：
  - 水位データ：ない，多い，少ない，その他
  - 温度データ：ない，多い，少ない，その他
  - ON,OFF：ない，その他

レビュ対象の GSN の大きさを示す指標として，各 GSN の総ノード数を表 1 に示す．

2.3 実験手順

本実験は以下の手順に従って行った．

手順 1

2 人のレビュアーがそれぞれ個人でレビュを行う．レビュアーは各自のレビュ観点に従ってレビュを行う．

手順 2

手順 1 のレビュ結果を基にしてレビュアー間でレビュ観点を共有し，2 人のレビュアーが合同でレビュを行う．この手順では，両レビュアーの指摘内容を照らし合わせることにより，手順 1 で得られた指摘項目を両レビュアーが共に指摘している項目とそうでない項目に分類する．

表 2 にレビュ観点を示す．レビュ観点を 7 種類に分類し，それぞれに対して指摘項目種別を表す記号を割り当てた．

3 実験結果

表 2 の指摘項目種別に従って整理したレビュ結果を表 3 に示す．表 3 における「総数」,「共通」,「共通率」の欄は，それぞれ以下のような値を示す．

- 総数：
  実験手順 2 の合同レビュの結果で得られた，

表 3: レビュ実験の結果

	N		P		C		E_g		E_c		E_e		U		計		
GSN 番号	総数	共通	総数	共通	総数	共通	総数	共通	総数	共通	総数	共通	総数	共通	総数	共通	共通率
A	0	0	1	0	0	0	1	0	0	0	1	0	0	0	3	0	0
B	0	0	3	0	0	0	1	0	0	0	0	0	0	0	4	0	0
C	1	0	3	0	0	0	1	0	0	0	1	0	1	0	7	0	0
D	5	4	1	0	1	0	0	0	1	0	1	0	0	0	9	4	0.44
E	1	0	2	0	5	5	3	3	1	0	0	0	5	0	17	8	0.47
F	0	0	4	4	0	0	0	0	1	0	1	0	0	0	6	4	0.67
G	1	0	3	1	0	0	2	0	0	0	1	0	0	0	7	1	0.14
H	1	0	1	0	5	5	0	0	0	0	0	0	0	0	7	5	0.71
I	0	0	4	4	1	0	1	0	1	1	0	0	6	0	13	5	0.38
J	1	0	3	0	3	3	1	0	1	0	0	0	6	0	15	3	0.2
K	0	0	1	0	1	0	0	0	4	4	0	0	0	0	6	4	0.67
L	0	0	0	0	0	0	1	0	1	0	0	0	0	0	2	0	0
M	3	0	0	0	4	4	1	0	1	0	1	0	0	0	10	4	0.4
N	1	1	6	0	1	1	1	0	0	0	0	0	0	0	9	2	0.22
計	14	5	32	9	21	18	13	3	11	5	6	0	18	0	115	40	0.35
共通率	0.36		0.28		0.86		0.23		0.45		0		0		0.35		

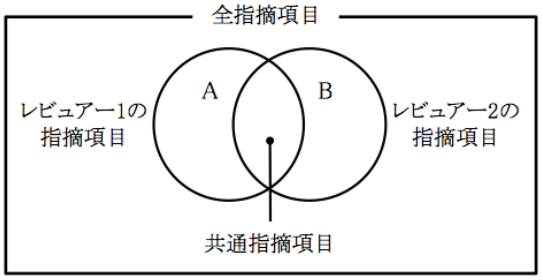

図 2: 指摘項目の分類

指摘項目の総数を示す．なお，両レビュアー間で共通の指摘項目は 1 回のみ数える（図 2 における A ∪ B）．

- 共通：
  実験手順 2 の合同レビュの結果で得られた，共通指摘項目の数を示す（図 2 における A ∩ B）．
- 共通率：
  全指摘項目に含まれる共通指摘項目の割合を示す．なお表 3 では，各指摘項目種別における共通率と各 GSN における共通率を示す．

1 つの GSN に対する指摘項目数の平均は，総数が 8.2 個，共通が 2.9 個となった．

各指摘項目種別の共通率を見ると，もっとも高い共通率となったのは C であり，ほとんどの GSN で全く共通の指摘がなされていた．一方，E_e と U の指摘は全く一致することがなく，共通率は 0 となった．なお，計 6 個の E_e の指摘は全て同じレビュアーによる指摘であり，計 18 個の U の指摘は全てもう一方のレビュアーによる指摘であった．その他の指摘項目種別を見ても，全体的に低い共通率となった．

各 GSN の共通率を見ると，最も高い値でも 0.71 であり，ほとんどの GSN で半数以上の指摘が共通指摘でないという結果となった．また，4 つの GSN で共通率が 0 となったが，いずれも指摘項目総数が比較的少ない GSN であった．

指摘項目全体の共通率は 0.35 であり，指摘の半数以上は共通指摘ではないという結果となった．この結果から，「レビュアーが異なっても，同じように GSN の問題点を指摘できる」という実験仮説は不成立となった．

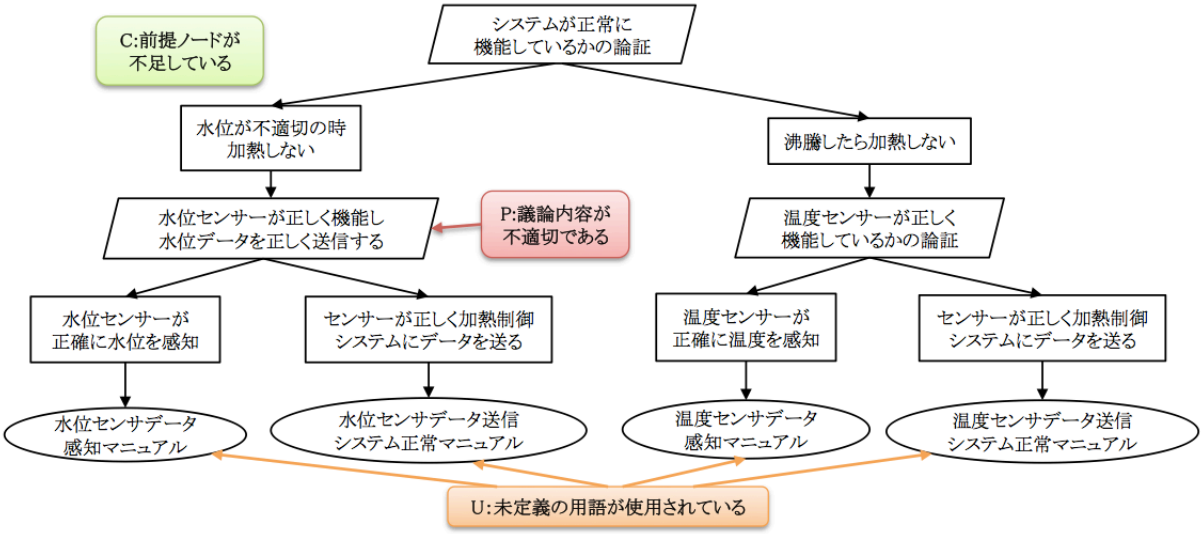

図 3: 問題のある GSN の例 (一部)

4 考察

4.1 指摘項目数に基づく GSN の出来の良さの分類

表 3 から，GSN によって指摘項目数に大きな差が出ることがわかった．そこで，GSN の出来の良さを指摘項目総数に従って表 4 のように 4 段階に分類した（段階 4 の GSN の例を図 3 に示す）．

表 4: 指摘項目総数に従った GSN の分類

段階	指摘項目総数	対応する GSN
1	～4	A，B，L
2	5～8	C，F，G，H，K
3	9～12	D，M，N
4	13～	E，I，J

このうち，出来が良いというレビュ結果となった段階 1 の GSN と，出来が悪いというレビュ結果となった段階 4 の GSN に注目すると，それぞれの段階の GSN 間で以下のような共通の特徴があることがわかった．

- 段階 1 の GSN の特徴：
  C，N の指摘が無く，全ての指摘がノードの記述内容に関する指摘となっている．また，共通率は 0 である．
- 段階 4 の GSN の特徴：
  C，U の指摘が多く，E_e の指摘が無い．また，ほとんど全ての指摘項目種別に対して指摘がある．共通率は 3 個の GSN とも低い値である．

段階 1 の GSN の特徴から，出来の良い GSN は 2 人のレビュアーが共通で指摘してきしやすい誤りがなく，レビュアーの主観に大きく依存するような指摘のみがなされていることがわかった．

段階 4 の GSN は指摘項目数が多いにもかかわらず共通率が低い値であることから，レビュアーの違いによるレビュ観点の差異に依らず指摘項目数が多くなることがわかった．また，U の指摘項目数の合計 18 個のうち 17 個は段階 4 の 3 つの GSN に対する指摘であるため，出来の良くない GSN には未定義の用語が多く使用されていることがわかった．

4.2 指摘項目種別について

実験結果から，各指摘項目種別の特徴と，共通率を向上させるために必要な指摘判断基準について以下のように考察した．

- N：GSN の表記法は明確に定義されているため，表記法が正しいか誤りであるかは個人の主観に依らず一意に判断できる．このことから，共通率が 1 とならなかった原因は一方のレビュアーの見落としであると考えられる．よって，指摘判断基準を統一しなくても，レビュアーの見落としがなければ全ての指摘が共通の指摘となる．
- P：全指摘項目のうち約 3 割が P の指摘であったことから，指摘のしやすいレビュ観点であることがわかったが，共通率は 0.28 と低い値となった．指摘内容は以下の 2 種類に分類で

きる．

1. 戦略ノードを使用する意味を理解していないような記述内容であることに対する指摘（例：記述内容がゴールを分解するような内容となっていない）
2. 戦略ノードの記述内容とサブゴールの記述内容が対応していないことに対する指摘（例：戦略ノードは「リスクごとに議論する」と記述されているが，システムの構成要素ごとにサブゴールが作成されている）

1 の指摘は，C の 1 の指摘と同様の指摘内容である．
2 の指摘は，戦略ノードの記述内容とサブゴールの記述内容を照らし合わせて，対応が取れていないことを指摘判断基準とすることができる．対応が取れていないことを判断するために，保証ケースの議論分解パターン [2],[3] に沿った議論分解が行われていることを確認する．

- C：指摘項目の共通率が高いことから，レビュアーは前提ノードが不足していることに気づきやすいことがわかった．
  このことから，指摘判断基準を統一しなくても複数のレビュアー間で同じような指摘ができると考えられるが，共通率をさらに高めるには前提が不足している基準を明確に定義する必要がある．

- E_g：レビュアーの主観による差異が大きく出る観点であり，多くの指摘内容が一致していなかった．しかし，ほとんどの指摘内容はゴールノードの記述内容の書き方への指摘（例：「〜なので安全である」という書き方にするべきである,「システムは〜しない」ではなく「システムは〜しないので安全である」と書くべきである）であった．
  このことから，ゴールノードの記述内容の書き方が誤っていることを指摘判断基準とすることができるが，共通率を向上させるためには，ゴールノードの記述内容の正しい書き方を明確に定義する必要がある．

- E_c：指摘内容は，以下の 2 種類に分類できる．

  1. 前提ノードを使用する意味を理解していないような記述内容であることに対する指摘（例：「安全原則を定義する」は誤りで,「安全原則」が正しい）
  2. 前提ノードの記述内容が，接続先のゴールノードまたは戦略ノードの記述内容と対応していないことに対する指摘（例：リスク一覧でなく構成要素の一覧を記述する必要がある）

  1 の指摘は GSN のノードの使用方法の誤りに対する指摘であるため，ノードの使用方法を正しく理解していないことを指摘判断基準とすることができる．このような指摘は前提だけでなく他のノードに対しても行われていた（例：戦略ノードがサブゴールのように使用されている）ため，全種類のノード共通の指摘判断基準として統一することができる．
  また，2 の指摘は前提ノードの記述内容と接続先のノードの記述内容を照らし合わせて，対応が取れていないことを指摘判断基準とすることができる．

- E_e：E_e をレビュ観点としていたのは一方のレビュアーのみであり，共通率が 0 となったことから，観点を統一する必要がある．また，指摘は全て「証拠ノードの記述内容を具体的に記述する必要がある」という内容であり，複数のゴールノードに対して同じ内容の証拠ノードが接続されている場合（例：「テスト結果」という内容の証拠ノードが複数存在する）に指摘がされていた．
  このことから，各最下位ゴールに対して固有の証拠ノードが接続されていないことを指摘判断基準とすることができる．

- U：U をレビュ観点としていたのは一方のレビュアーのみであったため共通率が 0 となったことから，観点を統一する必要がある．本実験のように，保証ケースを作成するために使用した文書が明らかであれば，未定義の用語が使用されているかどうかを判断することが可能であるため，指摘判断基準とすることができる．
  また，指摘項目は全て証拠ノードに対する指摘であったため，証拠ノードには未定義の用

表 5: 各指摘項目種別に対する指摘判断基準

種別	指摘判断基準
N	・GSN の図の表記法に誤りがある ・ノードの意味を正しく理解して使用されていない
P	・議論分解が誤っている
C	・必要な前提ノードが不足している
E_g	・ゴールの記述内容の書き方が誤っている
E_c	・前提ノードの記述内容が，接続先のノードの記述内容と対応していない
E_e	・各最下位ゴールに接続されている証拠ノードの記述が固有の内容となっていない
U	・ノードの記述内容に未定義の用語が使用されている

語が使用されやすく，その他のノードには使用されにくいという特徴があると考えられる．

共通率が最も高くなった指摘項目種別は C であり，次いで E_c が高くなったことから，前提ノードに関する指摘は共通して指摘しやすい傾向があることがわかった．

また，E_e，U はそれぞれ別々のレビュアーによる指摘であり，互いに証拠ノードに対する指摘である．E_e の指摘をしたレビュアーは，未定義の用語が使用されていても具体的に内容を示すべきと考え，U の指摘をしたレビュアーは，具体的な内容が示されていても未定義の用語は使用すべきでないと考えていた．このことから，レビュアーによる観点の差が最も大きくなったのは証拠ノードに関する指摘であることがわかった．

4.3 指摘判断基準について

4.2 節から，表 5 のように各指摘項目種別に対する指摘判断基準を整理した．4.2 節で述べた E_c の 1 の指摘と P の 1 の指摘を N の指摘として統一し，全種類のノードに対する指摘とした．

5 おわりに

本稿では，GSN レビュ実験により，レビュアーによってレビュ結果が異なることを明らかにした．また，実験結果から，出来の良い GSN と出来の悪い GSN それぞれの特徴を明らかにするとともに，各指摘項目種別の特徴と指摘判断基準について考察した．

今後の課題としては，4.3 節で述べた指摘判断基準によって実際に共通率が向上することを実験により確認することが挙げられる．

また，GSN をレビュする方法として，システミグラムと呼ばれるモデル図に基づいて GSN をレビュするという手法が提案されており [4]，GSN を直接レビュする方法とシステミグラムに基づいてレビュする方法を比較することも今後の課題として挙げることができる．

謝辞

本研究は，独立行政法人情報処理推進機構技術本部ソフトウェア高信頼化センター（SEC: Software Reliability Enhancement Center）が実施した「2015 年度ソフトウェア工学分野の先導的研究支援事業」の支援を受けたものです．

参考文献

[1] T. Kelly.“ Arguing Safety, a Systematic Approach to Managing Safety Cases ”. PhD Thesis, Department of Computer Science, University of York, 1998

[2] Robin Bloomfield and Peter Bishop, Safety and Assurance Cases: Past, Present and Possible Future - an Adelard Perspective

[3] 松野裕, 山本修一郎, 高井利憲, D-Case 入門, ディペンダビリティ・ケースを書いてみよう! , ダイテックホールディング,ISBN 978-4-86293-079-8, 2012

[4] 山本修一郎, 森崎修司, 渥美紀寿, 構成情報に基づく保証ケースレビュ手法の提案,SIG-KSN-017-03,2015