

ユーザビリティ定量化に向けて

池上輝哉・岡田英彦

要旨

システムの使いやすさ（ユーザビリティ）を定量的に評価する手法として、手順やしきい値を詳細化したチェックリスト評価法を提案、検証実験による精度向上を進めています。実験では、数名で評価を実施し、協議で結果を定めることで高い再現性を得られることを確認しました。本稿では、提案するチェックリスト評価法と実験の詳細、今後の展開について紹介します。

キーワード

●ユーザビリティ ●評価手法 ●定量化 ●チェックリスト ●AHP法

1. はじめに

システムのユーザビリティをアピールする、あるいはさらなる改善につなげるために、様々な評価手法が実践されています。代表的なユーザビリティ評価手法¹⁾の1つであるチェックリスト評価法²⁾には、開発の上流段階での適用に向くといった利点がありますが、一方で、評価結果が評価者のスキルや経験、主観などに依存し、一意で再現性のある評価結果を得ることが難しいという課題があります。

筆者らは、チェックリスト評価法に基づくユーザビリティの定量化に向け、評価者の違いによる結果のぶれをできる限り排除することを目標としたチェックリストの開発を進めています。本稿では、提案手法および検証実験について述べ、最後に今後の展開について記載します。

2. チェックリストの構築

チェックリストによる定量化の具体化にあたり、定量化における課題を整理し方向性を定めた上で開発を進めています。

2.1 定量化における課題

(1) 評価者の違いによる結果のぶれを排除する

チェックリスト評価法では、評価項目に従って対象を確認し、適合度を点数付けすることが一般的です。適合度を5段階評価などに設定した場合、評価者の裁量により結果にぶれが生じることになります。また、評価者によっては項目の意味を理解できない場合や誤った解釈をしてしまう可能性もあります。この課題に対し先行研究³⁾では、評価者がある程度のユーザビリティに関する学習・実践経験を積ん

だ上でチェックリストを使用することで知識や経験の不足を補っています。しかし、すべての評価者が同程度の学習を行うことは困難です。また、たとえばリストボックスやプルダウンメニューなどのUI部品の呼称については熟練者の間でも異なることが多いため、評価者による結果のぶれの最小化はさらなる課題の1つとなっていました。

筆者らは、各項目について評価対象や手順を詳細に記載し、判断基準を規定することで、「可（問題なし）」と「不可（問題あり）」、「該当無し（評価対象が存在しない）」のいずれかで適合度を判定できるようにしました。さらに、評価者によって理解度や解釈にぶれが生じることを抑制するため、チェックリストにおける用語や事例の説明集をあわせて構築しました。

(2) ユーザへの効果を分かりやすく示す

一般的にチェックリストはUI設計の専門家や開発者が使用することを想定しています。このため、評価軸がレイアウトやボタンなど、設計・開発に直結する要素で構成されていることが多く、ユーザにとっての効果が分かり難いという課題があります。また、チェックリストの各項目を満たすことによりユーザに与える効果の内容や程度はそれぞれ異なり、項目への重み付けを妥当性の高い形で行うことが重要となります。

筆者らは、AHP(Analytic Hierarchy Process)法⁴⁾を用いた重み付けを実施、評価結果を「学習しやすさ」、「エラーの少なさ」、「記憶しやすさ」、「効率性」の4つの観点で出力するようにしました。

2.2 チェックリストの構成

筆者らは各種ガイドラインや規格、業務でのノウハウを

ユーザビリティ定量化に向けて

ベースに5セクション、126項目で構成されるチェックリストを構築しました（表1）。

チェックリストの各項目は、評価者の違いによる結果のぶれを排除するために評価手順を詳細化していますが、項目の中には評価対象の業務に関する知識が必要となるものもあります。これに対し、手順通りに評価することで誰でも同程度の結果を出せる「基本項目」と、業務要件を要する「拡張項目」に分離して整備しました。たとえば基本項目の1つである「配色のコントラスト」に関する項目では、テキストと背景色とのコントラストが十分に確保されているかを確認する手順と判定用の数式を記載し、誰でも業務要件に関わらず同じ結果を得ることができます。これに対し、「情報の強調表現」に関する項目では、何が強調すべき重要な情報であるかを判断するために、業務要件が必要となるため拡張項目としています。

2.3 項目の重み付け

チェックリスト各項目のウエイトを決定するとともに、設計・開発の観点により構成されている項目に従って評価した結果を、ユーザ効果の観点で算出するためAHP法を用いました。AHP法では、意思決定を行うにあたり、関連する要素を階層構造によって把握し、階層レベルごとにウエイトを数値化することで全体に対する各要素のウエイトを算出します。本手法の特徴は、ある基準（直接測定することが困難な基準に対しても適用可能）に従って評価対象に対し一対比較を行う点にあり、総合的に各要素のウエイトを決定することに比べ、妥当性の高いウエイトを算出できます。

筆者らは、Nielsenの提唱する5つのユーザビリティ特性¹⁾における、「学習しやすさ」、「エラーの少なさ」、「記憶しやすさ」、「効率性」の4つの観点を基準とし、それぞれにおける全項目のウエイトを決定しました（図）。なお、「主観的満足度」については、前述の4つの他に機能の充実度などの様々な要素を統合して定まるものであり、また、ユーザの嗜好や価値観によって結果が大きく変わるものと考え、本件では除いています。これにより、本チェックリストを用いることで、前述の4つの観点から評価結果を算出することが可能となるとともに、製品のコンセプトに即した観点においてウエイトが高い項目を優先的に改善するといった効率的な改善が可能となります。

表1 チェックリストの構成（基本89項目、拡張37項目）

1. 表示/操作の一貫性	表示や操作方法がシステムを通して一貫しているか。視覚効果、レイアウト・画面遷移、データ出力、操作、応答・通知に関する。
2. 情報の見やすさ・見分けやすさ	情報が見やすいか、異なる情報が見分けやすいか。視覚効果、レイアウト・画面遷移、データ出力、操作に関する。
3. 現在の状態の提示	現在の画面（仕事）や操作の状態に応じた情報を提示しているか。データ出力、操作、応答・通知に関する。
4. ユーザ/環境への適合性	ユーザや環境の特性に柔軟に適應できるか。操作、応答・通知、カスタマイズに関する。
5. 仕事への適合性	ユーザの仕事に即した情報/操作手段を提供しているか。レイアウト・画面遷移、データ出力、操作に関する。

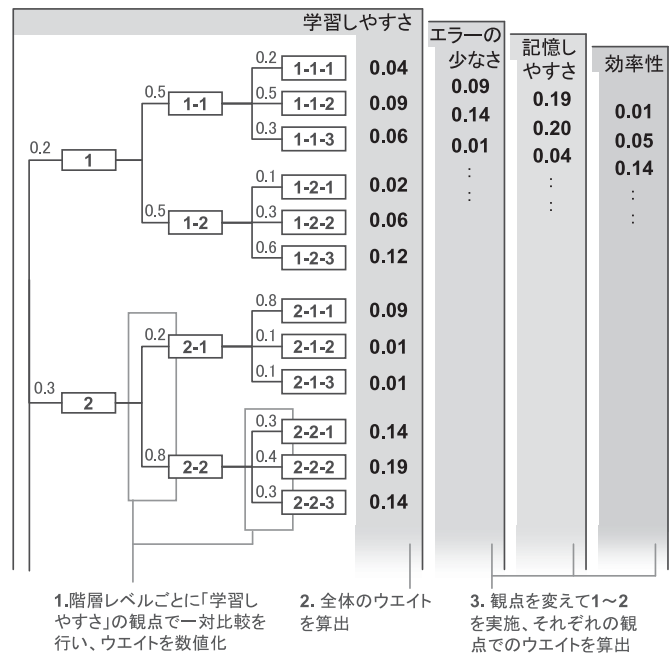


図 チェックリストの重み付け

3. チェックリスト評価実験

開発中のチェックリストを用いた評価を複数の評価者で実施し、結果の一意性の検証および改善のための課題抽出を行いました。

3.1 実験方法

本検討では、2007年9月に第1回の実験を行い、12月までに計5回の実験を行ってきました。1回あたりの実験は以下の方法にて行っています。

(1) 評価するチェックリスト項目

業務要件やスキルを必要としない基本項目を対象とし、4～5項目を実験者があらかじめ選別しました。5回の実験において、全19項目、のべ22項目のチェックリスト項目を評価しました。

(2) 評価者

1回の実験につき、5～6名の評価者の協力を得ました。この評価者のうち3～4名はチェックリスト評価法の経験がこれまでなくユーザビリティの知識も十分でない初心者であり、大学生です。一方、残りの1～2名は当該分野の知識・経験を備えた熟練者であり、NECに所属する研究者です。これらの評価者から報告された結果を比較することにより、初心者であっても熟練者と同様の結果を一様に得ることができるかどうか検証できます。5回の実験において、熟練者は合計3名、初心者は合計10名が参加しました。

(3) 評価対象

ある電子メールソフトのGUIで表示されるウィンドウの中から、4～5つのウィンドウを実験者があらかじめ選別し、これらのウィンドウを評価対象の母集団としました。

(4) 評価方法

個々のチェックリスト項目について、評価対象とするウィンドウ（母集団に含まれるウィンドウすべてもしくはその一部）、評価すべきUI部品の種類（メニュー、ボタン、テキスト入力など）、問題点かどうかの判定方法・基準、チェックリストに適合している/いない例、などを記載した指示書を実験者があらかじめ作成し、評価者に配布しました。評価者はいずれも独立して（他の評価者との協働や情報交換を禁止して）、指定された電子メールソフトを操作しながら、指示書の記載にしたがって評価を行いました。情報の見やすさやUI部品の使い方の妥当性など、ウィンドウ単位で評価可能なチェックリスト項目では、ウィンドウごとに問題の有無を確認し、評価結果を報告させました。また、表示や操作の一貫性など、複数のウィンドウ間の比較に基づいて全体で問題点の有無を判定する必要があるチェックリスト項目では、評価結果はウィンドウごとでな

くウィンドウ全体で1つとしました。

3.2 実験結果

前述の通り、本チェックリストに基づく評価結果は「可」、「不可」、「該当なし」のいずれかとなります。この結果を初心者・熟練者間で比較することにより、初心者であっても熟練者と同じ結果を得ることができたかどうか、また複数の評価者全体で一意的な評価結果を得ることができたかどうかを調べられます。たとえば、表2(a)に示すような結果であった場合には（初1～4は4名の初心者、熟1・2は2名の熟練者を表します）、初心者4名の結果がすべて熟練者と一致しており、このチェックリスト項目が一意的な結果の導出に有効であることを示しています。一方、表2(b)に示すような結果であった場合には、初心者4名のうち1名が熟練者とは異なる結果を報告しており、このチェックリスト項目が一意的な結果の導出において課題を含んでいることを示しています。

ここで、初心者の結果が熟練者の結果とどの程度一致しているかを表す指標として、一致率を以下のように定義します。

たとえば、表2(a)(b)の結果では一致率はそれぞれ100%、75%となります。さらに、一致率の平均を5回の実験別に求めたところ、表3のようになりました。

$$\text{一致率} = 100 * \frac{\text{熟練者と一致した初心者数}}{\text{初心者数}} (\%)$$

表2 評価結果の例

	初心者				熟練者			初心者				熟練者	
	1	2	3	4	1	2		1	2	3	4	1	2
可	●	●	●	●	●	●	可		●				
不可							不可	●		●	●	●	●
該当なし							該当なし						

(a)

(b)

表3 一致率の平均 (実験回数別)

1回目	2回目	3回目	4回目	5回目
62.50%	77.60%	72.00%	61.90%	66.70%

4. 考察

検証実験の結果を、チェックリスト開発にフィードバックすることで、記載内容の精度向上に加え、運用時に考慮すべき事項を抽出し、対策の検討を進めています。

4.1 チェックリスト項目の改善

本チェックリストでは、評価手順をできるだけ詳細化していますが、結果として評価者の負担が大きくなってしまいました。また、評価対象の指定に不備があるなど、評価手順の記述が十分でない項目については、評価者によって解釈が異なり、表2(b)のように、結果にぶれが生じることになりました。

たとえば「配色のコントラスト」に関する項目では、評価対象のすべての色を調べて数式を用いて判定する方法を指示した結果、実験者が想定した時間の数十倍の時間を要した上、一致率が0%となりました。これに対し、評価対象を明確化するとともに、明らかに問題がない箇所については目視での判定を許容し、判定時に使用する図表を提供することとしました。結果、評価に要した時間は大幅に削減され、一致率も83%まで改善されました。

このように、1~3回目の実験において一致率の低かった項目に対しては評価手順や図説の改良を行い、改良後の項目を4~5回目の実験において別の初心者に評価させたところ、一致率の改善が確認できました。このような作業の手間と一致率の両方を考慮した改善は今後も継続して行っていく予定です。

4.2 協議による評価結果の決定

表3を見る限り、一致率は現時点では必ずしも高いとは言えませんが、不一致の原因として、チェック漏れや勘違いなどのヒューマンエラーと考えられたものが多分に含まれていました。これに対し、5回の実験の中からヒューマンエラーによる不一致があったと考えられるチェックリスト項目と評価対象ウィンドウの組合せ12件について、評価に参加した複数の初心者による協議を行わせ、最終的な結果を定める実験を行いました。この結果、12件のうち10件は一致率がすべて100%に向上しました（協議前の段階では一致率0%~70%でした）。このことから、複数評価者が参加して協議により最終的な評価結果を定めることで、ヒューマンエラーによる不一致は解

決できる見込みが高いことが分かりました。なお、一致率が向上した10件のうちの2件について、熟練者の評価結果にも見落としや誤りがあったことが分かりました。評価手順をより詳細化したとしても、評価者が人間である限りはヒューマンエラーが発生することは熟練者であっても避けられません。これに対し、現時点では、複数名での評価を行った上で協議により結果を定めるといった使用形態を推奨していますが、今後は評価者の作業を支援するツールを提供していくことも必要と考えています。

5. まとめ

以上、各項目の評価手順を詳細化し、判断基準を規定することで、評価者の違いによる結果のぶれを排除するチェックリスト構築と検証実験について報告しました。複数の評価者の参加と協議による最終的な評価結果の決定を行うことで一致率が高くなったことから、同様の方法論にて本チェックリストを用いることで、一定の信頼性のある結果を得ることが期待できます。

今後、未検証のチェックリスト項目についても、実験による検証とフィードバックを進めていくとともに、本チェックリストを公開し、誰でも利用できるものとしていきます。また、チェックリスト自体のユーザビリティを高めていくため、より使いやすい評価/配点方法の整備や各種ツールの提供についても取り組んでいきます。

参考文献

- 1) Nielsen, J.; "Usability Engineering", Academic Press 1993.
- 2) Raviden, S., Johnson, G.; "Evaluating Usability of Human-Computer Interfaces: A Practical Method", Prentice Hall 1989.
- 3) 加藤ほか; 「HI設計チェックリストとその有用性の評価」, 情報処理学会論文誌, Vol.36, No.1, pp.61-69, 1995.
- 4) 刀根; 「ゲーム感覚意思決定法」, (株) 日科技連出版社, pp.8-46, 1986.

執筆者プロフィール

池上 輝哉
共通基盤ソフトウェア研究所
ヒューマンインタフェースセンター
主任

岡田 英彦
京都産業大学
准教授