

Effekt unterschiedlicher Visualisierungen wahrscheinlichkeitsbehafteter Daten auf Leistung und mentale Beanspruchung

Sabine THEIS, Rebecca KUMMER, Matthias WILLE, Peter RASCHE,
Christina BRÖHL, Alexander MERTENS, Christopher M. SCHLICK

*Institut für Arbeitswissenschaft, RWTH Aachen
Bergdriesch 27, D-52062 Aachen*

Kurzfassung: Das aktuelle Informationszeitalter geht mit zunehmend wachsenden Datenmengen einher. Informationssysteme werden entwickelt, um relevante Informationen zu finden oder neue Informationen zu extrahieren, für Verbraucher ebenso wie im medizinischen und wissenschaftlichen Kontext. Hier werden wichtige Entscheidungen oft basierend auf den Ergebnissen von Informationssystemen getroffen. Leider sind die verwendeten Daten nicht immer korrekt und Systeme arbeiten nicht immer präzise. Vorliegender Beitrag präsentiert Ergebnisse einer empirischen Studie über den Effekt verschiedener Repräsentationen wahrscheinlichkeitsbehafteter Daten, auf die Effizienz (Zeit, Fehlerrate) und die subjektiv gemessene, kognitive Belastung der Nutzer. Dabei wurden eine Tabelle, eine statische und eine interaktive Visualisierung wahrscheinlichkeitsbehafteter Daten, verglichen. Die Ergebnisse (N = 146) haben gezeigt, dass zwischen den Visualisierungen ein signifikanter Unterschied in der Bearbeitungszeit, nicht aber in den Fehlerraten und der kognitiven Belastung existiert.

Schlüsselwörter: Visualisierung, Leistung, kognitive Beanspruchung, Evaluation.

1. Einleitung und Problem

Daten- und Informationsverarbeitung ist in jedem Anwendungsfeld mit Fehlern behaftet. Formen der Unsicherheit und Fehler entstehen bei der Dateneingabe, Transformation, Weiterverarbeitung und Darstellung. Verfügbar Werte über quantifizierte Unsicherheit so können diese dem Nutzer eines Informationssystems übermittelt werden. Die gleichzeitige Übermittlung eines Wertes zusammen mit dessen Wahrscheinlichkeit oder Unsicherheit, kann das Verständnis des Nutzers und damit die Qualität seiner auf Daten basierenden Entscheidungen verbessern. Unklar ist, wie wahrscheinlichkeitsbehaftete Daten für den Menschen effektiv dargestellt werden können. Nutzerstudien, die Leistung und kognitive Anstrengung einhergehend mit der Visualisierung von wahrscheinlichkeitsbehafteten Daten vergleichen, fehlen. Die meisten Studien konzentrieren sich auf bereichsspezifische Darstellungen von Unsicherheit und evaluieren diese mit Leistungsmessungen, wie Bearbeitungszeit und Fehlerrate (Lundstrom, 2007, Simpson 2006). Ein allgemeiner Ansatz, Unsicherheit darzustellen, indem die Effektivität ein- und zweidimensionaler Datensätze evaluiert wurden, wurde von Sanyal et al. (2009) durchgeführt. Sie generierten systematische und zufällige Unsicherheit in ein- und zweidimensionalen Datensätzen und visualisierten diese mit Symbolen unterschiedlicher Größe und

Farbe, Fehlerbalken und farbigen Oberflächen. Die Nutzer mussten Such- und Zählaufgaben bearbeiten. Dabei wurde ihre Reaktionszeit erfasst. Fehlerbalken führten zu den schlechtesten Leistungen, während skalierte Symbolgrößen und farbige Oberflächen bessere Leistung zur Folge hatten. Eine in der Arbeit vorgeschlagene Anforderung an zukünftige Arbeiten, menschliche Faktoren in die Evaluation von Visualisierungen wahrscheinlichkeitsbehafteter Daten einzubeziehen (Zuk & Carpendale, 2006), wurde bis heute lediglich durch Huang et al. (2009) realisiert, in dem kognitive Beanspruchung erfassten. In vorliegender Arbeit werden diese Ansätze zur Bewertung der Effektivität und des menschlichen Faktors zusammengeführt und auf die Evaluation allgemeiner Darstellungsformen wahrscheinlichkeitsbehafteter Daten angewandt. Ziel der Studie ist es daher, die angemessenste Repräsentation in Bezug auf Bearbeitungszeit, Fehler und kognitiven Aufwand zu finden, indem folgende Hypothesen geprüft werden: (1) H_1 : *Es besteht ein signifikanter Unterschied in der Bearbeitungszeit zwischen Tabelle (RE_t), statisch-grafischer Repräsentation (RE_{sg}) und interaktiv-grafische Repräsentation (RE_{ig}),* (2) H_1 : *Es besteht ein Unterschied in den Fehlerraten zwischen RE_t , RE_{sg} , und RE_{ig} ,* (3) H_1 : *Es besteht ein signifikanter Unterschied in der kognitiven Beanspruchung zwischen RE_t , RE_{sg} , und RE_{ig} .*

2. Methode

Ein Gruppenvergleich untersuchte die Zusammenhänge zwischen der unabhängigen Variable Repräsentationsform mit ihren Ausprägungen Tabelle (RE_{ta}), statisch-grafische Repräsentation (RE_{sg}) sowie interaktiv-grafische Repräsentation (RE_{ig}) und den abhängigen Variablen Aufgabendurchführungszeit, Fehlerrate und subjektiv erfasste kognitive Beanspruchung. Die Experimentalaufgabe war in eine Webseite eingebunden und wurde den VPs über einen Link zugänglich gemacht. Letztere enthielt die Aufgabenbeschreibung, die Versuchsaufgabe mithilfe derer eine Entscheidung getroffen werden mussten, sowie eine strukturierte Abfrage der subjektiven kognitiven Beanspruchung.

2.1 Versuchsaufgabe und Datensatz

Thema der Versuchsaufgabe war eine Wettervorhersage, da dies ein allgemein verständlicher Kontext für die Untersuchung wahrscheinlichkeitsbehafteter Daten ist. Die Eintrittswahrscheinlichkeit einer Wettervorhersage liefert ein Vorhersagemodell. Während der Versuchsaufgabe sollte einer von vierzehn Tagen ausgewählt werden, um segeln zu gehen. Ziel war der Tag an dem die höchste Windkraft mit der größten Wahrscheinlichkeit auftreten würde. Die darzustellenden Daten lieferte ein Vorhersagemodell veröffentlicht von der KNMI (2011). Die Modellvorhersage beinhaltet den geschätzten Wind-Wert pro Tag und sich verändernde Minimum- und Maximum-Werte als Zeitspannen von einem Tag. Je größer das Intervall zwischen geschätztem Minimum und Maximum der Windkraft, desto unsicherer ist die geschätzte Windkraft. Je weiter eine Vorhersage in der Zukunft liegt, desto größer ist der Abstand zwischen Minimum und Maximum. Um zu verhindern, dass dadurch die Entscheidungen zu einfach sein würde, wurden tagesabhängige Schätzwerte des Modells zusammen mit ihren Minimum und Maximum Werten randomisiert angeordnet und auf 14 Tage erweitert. Dabei herrschte an 10 Tagen eine niedrige (2-4 m/s) bis mittlere (4-6 m/s) Windkraft, die als Störgröße dienen sollte. Windkräfte im

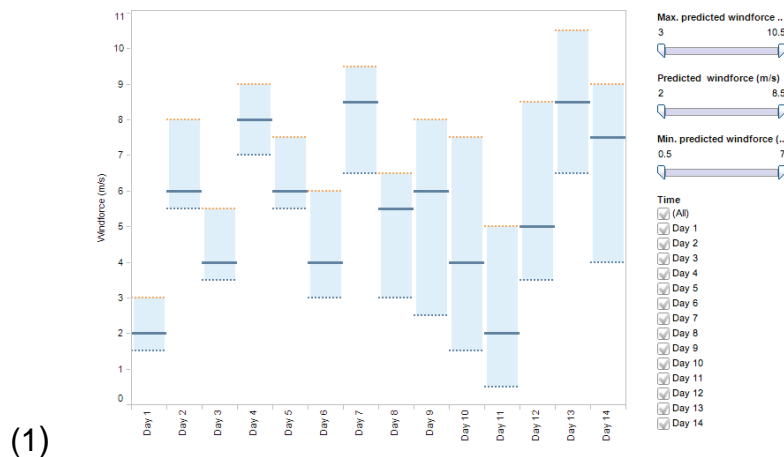
oberen Drittel (6-8,5 m/s, Tag 4,7,13 und 14) sollten die Aufmerksamkeit der Probanden, die sie investieren mussten, um die Aufgabe zu lösen, erhöhen.

Tabelle 1: Antwortmöglichkeiten zur Fehlerratenberechnung, Auswahlmöglichkeit 2 zählt streng genommen nicht als richtige Antwort. Die Fehlerratenberechnung wird unter der Bedingung 2 = richtig und 2 = falsch durchgeführt.

Nr.	Entscheidung	Auswahlmöglichkeit(en)	Bewertung
1	Der Proband wählt den Tag mit der höchsten Windkraft und der kleinsten Unsicherheit.	Tag 4: Wind 8 m/s, Intervall: 2 m/s	Richtig
2	Der Proband wählt einen der Tage mit höherer Windkraft und größerer Unsicherheit.	Tag 13: Wind 8,5 m/s, Intervall: 4 m/s Tag 7: Wind 8,5 m/s, Intervall: 3 m/s	(Richtig)
3	Der Proband entscheidet sich für einem Tag mit hoher Windkraft und kleiner Unsicherheit.	Tag 14: Wind 7,5 m/s, Intervall: 5 m/s	Falsch
4	Der Proband entscheidet sich, für einen Tag mit niedriger oder mittlerer Windkraft und/oder großer Unsicherheit.	Tag 1, 2, 3, 5, 6, 8, 9, 10, 11, 12, untersch. Intervalle	Falsch

2.2 Repräsentationen

Auf der Basis beschriebenen Datensatzes wurden drei Repräsentationsarten ausgewählt die (1) zu den gebräuchlichsten Repräsentationen zählen, und (2) eine denkbare Lösung zur Verwendung in digitalen Informationssystemen darstellen. Im vorliegenden Fall bildet eine Tabelle die Baseline, da ein großer Unterschied zwischen den Ergebnissen der statischen Grafik und denen der Tabelle zu erwarten ist. Für die zweite Repräsentationsform wurden wahrscheinlichkeitsbehaftete Daten mit grafischen Attributen versehen. Die statisch-grafische Repräsentation greift auch hier auf eines der gebräuchlichsten Formate zurück – das Balkendiagramm. Der Wert der geschätzten Windstärke wird durch eine horizontale Linie abgebildet, da sie einen Zeitraum von einem Tag repräsentiert. Die gepunkteten horizontalen Linien stehen für Minimum und Maximum der geschätzten Windkraft.



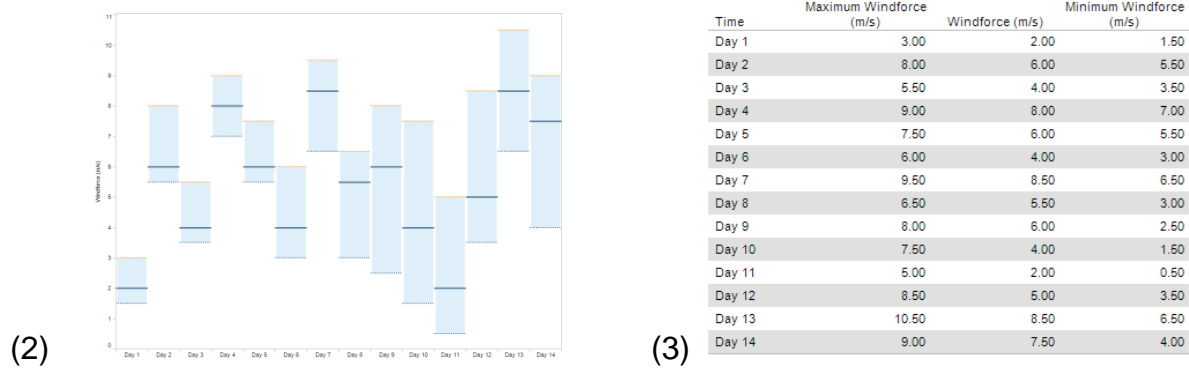


Abbildung 1: Als grundlegendste Form visuell repräsentierter Daten bildet die Tabelle (3) die Baseline des Experiments. Die statische Repräsentation (2) eines Balkendiagramms bildet die zweite Versuchsbedingung. Für eine möglichst hohe Vergleichbarkeit unterscheidet sich die grafisch-interaktive Visualisierung (1) lediglich durch ihre Kontrollelemente von der zweiten Versuchsbedingung.

Wie in Abbildung eins dargestellt, unterscheidet sich die interaktive Visualisierung von der statischen hinsichtlich der Kontrollelemente. Das grafische Erscheinungsbild ist gleich. Den Dimensionen Zeit, geschätzte Windkraft, sowie Maximum und Minimum der geschätzten Windkraft wurde ein Kontrollelement zugeordnet. Jeder geschätzte Wert der Windstärke konnte über Schieber begrenzt werden und bestimmte Tage konnten über ein Kontrollkästchen ausgewählt werden.

2.3 Fragebogen

Nach der Aufgabenbearbeitung beantworteten die VPs einen Fragebogen der die subjektiv wahrgenommene kognitive Beanspruchung entsprechend des SWAT Fragebogens (Subjective Workload Assessment Technique) abfragt (Reid et al., 1988). Die folgenden Fragen bilden die Subskala der kognitiven Beanspruchung evaluieren soll:

Q7: „Die Windvisualisierung hat geringe kognitive Anstrengung verursacht.“

Q8: „Um zu einer Entscheidung zu kommen, war nur wenig Konzentration nötig.“

Q13: „Eine Entscheidung zu treffen war ziemlich automatisch.“

Diese Fragen mittels einer 5-Punkte Likert Skala von „Ich stimme voll zu“ bis „Ich stimme gar nicht zu“ beantwortet werden. Die Reduktion des SWAT auf seine Subskala wurde durch die Erhöhung der Antwortwahrscheinlichkeit motiviert. Zusätzlich wurde die demografische Eigenschaft „Expert“ oder „Novize“ mittels abgefragt.

2.4 Teilnehmer

Die Teilnehmer (N = 148) wurden nach dem Schneeballprinzip im universitären Umfeld akquiriert. 68% der Probanden, die den Fragebogen ausgefüllt haben, waren männlich, 32% waren weiblich, außerdem verfügten die Teilnehmer über einen hohen Bildungsgrad (77% Universitätsabschluss) 56 bezeichneten sich als Experten im Umgang mit Daten und Datenvisualisierungen und 45 als Novizen.

3. Ergebnisse

Die durchschnittliche Aufgabendurchführungszeit unterscheidet sich abhängig von der Repräsentationsart: Probanden mit der interaktiven, grafischen Repräsentation (RE_{ig}) brauchten durchschnittlich am längsten (2:19 Min) um zu einer Entscheidung zu kommen, während es mit RE_{sg} nur 01:18 Minuten und mit der Tabelle VISt RE_t 01:19 Minuten waren. Experten bearbeiteten die Aufgabe im Durchschnitt schneller (01:28 Min.) als Novizen (01:59 Min). Es konnte ein signifikanter Effekt der Repräsentationsart auf die Bearbeitungszeit ($p < .05$) für die drei Konditionen nachgewiesen werden [$F(2, 142) = 5.15, p = .007$]. Damit wird die erste Hypothese bestätigt. Es wurde kein signifikanter Effekt des Grads der Expertise (Experte/Novize) auf die Bearbeitungszeit festgestellt.

3.1 Fehlerrate

Die Fehlerraten (FR) werden sowohl für eine strikte Definition einer richtigen Antwort (nur Tag 4) berechnen, aber auch in der erweiterten Definition mit Tag 13 und 7 als richtigen Antworten. Ein Chi-Quadrat Test zeigte keinen signifikanten Unterschied zwischen Repräsentationsart und Fehlerrate $X^2(2) = 2.19, p = .334$. Ebenso gab es keinen signifikanten Unterschied zwischen Expertenlevel (Experte/Novize) und Fehlerrate $X^2(1) = 2.93, p = .087$. Zählen zusätzlich Tag 13 und 7 als richtige Antwort, konnte ebenfalls keine signifikanten Unterschiede festgestellt werden gefunden. Dementsprechend wird Hypothese zwei verworfen.

3.2 Subjektiv erfasste kognitive Beanspruchung

Die Subskala des SWAT zur Erfassung der subjektiv wahrgenommenen kognitiven Beanspruchung (Q 7, 8, 13) zeigten eine akzeptable Reliabilität, Cronbach's $\alpha = .691$. Die globale Reliabilität des Fragebogens betrug $\alpha = .381$. Streicht man Item Q13, würde das zu einer erhöhten Skalenreliabilität von $\alpha = .691$ führen. Da eine 5-stufige Likertskala verwendet wurde und deren fünf Stufen einzelne Kategorien darstellen, ist die Voraussetzung für eine Varianzanalyse nicht erfüllt. Nach Invertieren der Antworten zu Q7 zeigte ein nicht-parametrischer K-W Test $H(2) = 7.39, p < .05$ einen signifikanten Unterschied in der subjektiven kognitiven Anstrengung (Q7) zwischen den Visualisierungstypen. Der Effekt der Visualisierungsart auf die Konzentration war nicht signifikant $H(2) = 4.81, p > .05$, genauso wie Automatisierung während der Aufgabenbearbeitung (Q 13) $H(2) = 2.04, p > .05$.

Tabelle 2: Mittlere kognitive Beanspruchung abhängig vom Visualisierungstyp

Frage	RE_{ig}	RE_{sg}	RE_t
Q7	2,55 (N = 38)	2,48 (N = 29)	3,62 (N = 31)
Q8	3,00 (N = 40)	2,55 (N = 29)	3,18 (N = 33)
Q13	2,79 (N = 38)	2,95 (N = 29)	3,06 (N = 32)

In Anbetracht dessen, dass alle drei Items zum Faktor der kognitiven Beanspruchung beitragen und zwei von diesen einen nicht-signifikanten Unterschied

zeigen, verwerfen wir Hypothese 3, dass es einen signifikanten Unterschied bezogen auf die kognitive Anstrengung zwischen den verschiedenen Visualisierungstypen gibt.

4. Schlussfolgerung

Vorliegende Studie generiert Wissen zur Bewertung visueller Repräsentationen und unterstützt dadurch die Auswahl von Visualisierungen wahrscheinlichkeitsbehafteter Daten für digitale Informationssysteme auf Basis von Leistungsmaßen und kognitiver Beanspruchung. Die Ergebnisse unterstützen die Hypothese, dass sich die Leistung sich abhängig von der grafischen Repräsentation unterscheidet. Es wurden keine signifikanten Unterschiede in den Fehlerraten oder in der kognitiven Beanspruchung abhängig vom Repräsentationstypen festgestellt. Die interaktive Visualisierung war die zeitintensivste und führte zu einer höheren Fehlerrate als die statische Visualisierung und die Tabelle. Die Teilnehmer haben die kognitive Anstrengung als leicht höher wahrgenommen, wenn sie die interaktive Darstellung oder die Tabelle genutzt haben. Auf der Grundlage unserer Ergebnisse, stellen wir fest, dass die beste Repräsentation in Bezug auf die Bearbeitungszeit, Fehlerrate und kognitive Anstrengung digitale Informationssysteme eine statische grafische Visualisierung wäre. Zusätzliche Dokumentation des subjektiven Feedbacks zu unterschiedlichen Repräsentationsformen sowie detaillierte Diskussion der Ergebnisse finden sich in Theis et al. (2016).

5. Literatur

- Huang, W., P. Eades, and S.-H. Hong, Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, 2009. 8: p. 139-152.
- KNMI, KNMI Verwachtingen, (2011). [Webpage] 2011 [cited 2011 13.07.2011]; Available from: http://www.knmi.nl/waarschuwingen_en_verwachtingen/.
- Lundstrom, C., et al., Uncertainty Visualization in Medical Volume Rendering Using Probabilistic Animation. *IEEE Transactions on Visualization and Computer Graphics*, 2007. 13: p. 1648-1655.
- Reid, G.B.N., Thomas E. and P.A.E.M. Hancock, Najmedin (Ed), *The Subjective Workload Assessment Technique: A scaling procedure for measuring mental workload. Human mental workload. . Human mental workload, Advances in psychology*, 1988. 52: p. 185-218.
- Sanyal, J., et al. (2009), A user study to compare four uncertainty visualization methods for 1D and 2D datasets. *IEEE transactions on visualization and computer graphics*, 2009. 15: p. 1209-18.
- Simpson, A., et al., Using Registration Uncertainty Visualization in a User Study of a Simple Surgical Task, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, R. Larsen, M. Nielsen, and J. Sporring, Editors. 2006, Springer Berlin / Heidelberg. p. 397-404.
- Zuk, T. & S. Carpendale (2006), Theoretical analysis of uncertainty visualizations. *Proceedings of SPIE*, 2006. 6060: p. 606007-606007-14.
- Theis, S., Wille, M., Rasche, P., Bröhl, C., Mertens, A. und Schlick, M, C. (2016) Human factors influencing the design and the evaluation of uncertain data visualizations. Manuscript in preparation.

Danksagung: Diese Veröffentlichung entstand im Rahmen des Forschungsprojekts "TECH4AGE", welches durch das Bundesministerium für Bildung und Forschung gefördert (BMBF, Grant No. 16SV7111) und durch den VDI/VDE Innovation + Technik GmbH beaufsichtigt wird.