# WHERE DO MACHINE LEARNING AND HUMAN-COMPUTER INTERACTION MEET?

VASSILIS S. MOUSTAKIS
Department of Production and Management Engineering, Technical University of Crete, Chania; and Institute of Computer Science, Foundation for Research and Technology–Hellas, Heraklion, Greece

JUERGEN HERRMANN
Universitét Dortmund, Informatik I, Dortmund, Germany

*Implementation of machine learning (ML) in human-computer interaction (HCI) work is not trivial. This article reports on a survey of 112 professionals and academicians specializing in HCI, who were asked to state level of ML use in HCI work. Responses were captured via a structured questionnaire. Analysis showed that about one-third of those who participated in the survey had used ML in conjunction with a variety of different HCI tasks. However, statistically significant differences could not be identified between those who have and those who have not used ML. Using statistics, contingency analysis, and clustering, we modeled interaction between representative HCI tasks and ML paradigms. We discovered that neural networks, rule induction, and statistical learning emerged as the most popular ML paradigms across HCI workers, although intensive learning, such as inductive logic programming, are gaining popularity among application developers. We also discovered that the leading causes for declining use of ML in HCI work are (1) misperceptions about ML, (2) lack of awareness of ML's potential, and (3) scarcity of concrete case studies demonstrating the application of ML in HCI.*

Machine learning (ML) represents one of the fastest growing technologies today with an abundance of prototype and field industrial applications. Robotics, computer vision, manufacturing, medicine, knowledge acquisition, execution and control, design, planning and scheduling, among others, are areas that have uncovered the potential of the technology. Work using new media and networks has also identified a niche for ML in navigation and retrieval of information.

Human-computer interaction (HCI) represents a wide field of research and applications. HCI addresses a variety of areas, such as user interface design or computer-supported cooperative work as well as tasks related to user modeling,

The authors would like to thank the participants of the survey. Opinions, views, and results reported herein represent the authors' opinions and not those of the participants. Also, we wish to thank the referees for their constructive suggestions that improved the quality of the article.

Address correspondence to Vassilis S. Moustakis, Institute of Computer Science, FORTH, P.O. Box 1385, 71110 Heraklion, Greece. E-mail: moustaki@ics.forth.gr

Applied Artificial Intelligence, 11:595–609, 1997

0883-9514/97 $12.00 + .00

system adaptation, and optimization to fit user needs. HCI researchers and professionals find themselves involved in various domains, including but not limited to, aviation, aerospace, and manufacturing. HCI is concerned with the role of humans in complex systems, the design of equipment and facilities for human use, and the development of environments for comfort and safety.

This article reports specifically on the use of ML in the HCI community. Work presented herein complements earlier research related to the application of ML in a wide variety of tasks by Brodley (1993), Kodratoff et al. (1994), Michie et al. (1994), Moustakis et al. (1996), and Sleeman et al. (1995). Brodley (1993) focuses on algorithm selective superiority and relies on empirical comparisons embedded in a model selection system to demonstrate that an ML algorithm is most appropriate for some class of tasks and not equally appropriate for other classes. Moustakis et al. (1996) analyze results of a survey that captured expert perception regarding appropriateness of generic ML methods to a set of generic intelligent tasks. Using statistical, factor, and correlation analysis, the authors group ML methods into overlapping clusters. Results by Moustakis et al. further confirm results by Brodley (1993) and the ML research taxonomy by Michalski and Kodratoff (1990). In a similar vein, Kodratoff et al. (1994) relate ML algorithm selection with characteristics of application domains; such characteristics include attributes (i.e., problem variables), attribute values, size of data set on which learning is performed, role of initial (i.e., input) knowledge in the learning process and types of postlearning interaction. Both Kodratoff et al. (1994) and Sleeman et al. (1995) deal with research methodology and results drawing from the Consultant-II expert system, which was developed during the Machine Learning Toolbox (MLT) project (Kodratoff et al., 1992). MLT includes twelve ML algorithms, and Consultant-II gives advice to the user as to which algorithm(s) should be used according to problem characteristics or specifics. Finally, Michie et al. (1994) focus on classification; authors report results of the STATLOG project and assess applicability of various symbolic, statistical, and neural ML methods in classification tasks.

This article expands on an earlier version, which we presented at a workshop organized during the 13th International Conference on Machine Learning, Bari, Italy, in July 1996. The objective was, and remains, to associate ML with HCI areas of involvement, tasks, and domains of application. We took the side of HCI, and using electronic mail, we conducted a survey to model status and perception of ML application between HCI researchers. Using World Wide Web (WWW) resources, we selected electronic mail addresses of 185 persons involved in HCI both from industry and academia or research. We received back 112 responses. This represents a percentage of more than 60%, which we considered adequate.

In the sections that follow, we give an overview of the protocol used in this study (to capture ML use, or nonuse, in HCI work), present the method and data analysis procedures, and discuss results of the analysis. We conclude by discussing the implications of research results and identifying potential work for combined ML and HCI.

## METHOD AND DATA ANALYSIS

To elicit information about the use of ML in HCI work, we conducted a survey by means of electronic mail. The survey was based on a structured protocol divided into five parts, that is, parts I to V. We used parts I through III to capture HCI participant demographic specifics in terms of area, task, and domain of HCI application involvement. Parts IV and V were used to record ML use or nonuse in HCI work. Specifically, part V gave responders a chance to say why they have not used ML in HCI work.

An initial version of the protocol was critically examined by a small group of HCI experts. During this process, several points both with respect to ML and HCI, were clarified and further refined. Clearly, neither HCI areas of involvement, tasks, or domains of application nor ML methods included in the protocol are orthogonal to each other. Nevertheless, people working in either ML or HCI differentiate between either ML methods or dimensions of HCI work using similar definitions, and the protocol carries on this reality or bias.

Protocol and raw survey data may be retrieved via ftp from the following address: ftp.ics.forth.gr/pub/machine_learning. File survey96.txt includes the protocol, and file survey96.xls includes the raw survey response data.

### Part I: Area of HCI Involvement

We distinguished among eight different areas of HCI involvement and asked responders to indicate those in which they are most involved. We present an overview of responses in Table 1. Most of the responders selected user interface design and evaluation (including mouse display devices). The same area received most of the unique responses, namely, 17 (Table 1). Given that many of the responders indicated multiple areas of HCI involvement, we derived a contingency matrix reporting $2 \times 2$ interactions between areas. A summary of contingency results is provided in Table 2.

Results reported in Tables 1 and 2 demonstrate the following.

1. About 80% of responses focus in four areas of HCI involvement, namely (in decreasing order of priority), user interface design and evaluation (UI), product or services design and evaluation (PS), computer-supported cooperative work (CSCW), and education, training, and intelligent tutoring systems (ITS).

2. Only 25 responders (about 25% of the sample) indicated only one area of involvement. However, the vast majority of those who selected only one area picked

**Table 1.** Areas of HCI Involvement

| Area of HCI involvement | Abbreviation | Number of responses | Involvement in one area only |
|---|---|---|---|
| User interface design and evaluation (including display mouse devices) | UI | 101 | 17 |
| Product or services design and evaluation | PS | 54 | 2 |
| Concurrent engineering | CE | 7 | — |
| Computer-supported cooperative work | CSCW | 37 | 2 |
| Human-robot interaction | HRI | 5 | — |
| Education, training, and intelligent tutoring systems | ITS | 34 | — |
| Virtual reality | VR | 19 | 1 |
| Other | OTH | 24 | 3 |
| Total | | 281 | 25 |

UI; the ratio is 17 out of 25, or, 68%. UI represents the leading area of HCI involvement even when combined with other areas; see Table 2.

3. We only missed 9% of responses, as the OTH entry of Table 1 shows.

## Part II: HCI Task Involvement

The second part of the questionnaire included indicative HCI tasks. We asked responders to select tasks with which they are most involved in professional or academic work. Table 3 summarizes the distribution of responses. Responders were advised to select as many areas as they considered necessary. Average task selection was 4.90 per responder. To facilitate survey participants, we have purposefully avoided creating an orthogonal list of tasks. People tend to refer to the same thing using different terms. The top running tasks that emerged were adapting, customizing, or optimizing systems according to user need (AD), usability engineering (UE), user modeling (UM), multimedia system design and evaluation (MU), and modeling

**Table 2.** Contingency between areas of HCI involvement

| | PS | CE | CSCW | HRI | ITS | VR | OTH |
|---|---|---|---|---|---|---|---|
| UI | 51 | 6 | 34 | 3 | 34 | 17 | 20 |
| PS | | 5 | 17 | 2 | 20 | 10 | 13 |
| CE | | | 5 | 1 | 4 | — | 2 |
| CSCW | | | | 1 | 15 | 7 | 9 |
| HRI | | | | | 1 | 3 | — |
| ITS | | | | | | 10 | 8 |
| VR | | | | | | | 4 |

Entries indicate simultaneous involvement in two areas. For instance, we received 51 responses indicating involvement in both user interface design (UI) and product design (PS). See Table 1 for explanation of abbreviations.

**Table 3.** Distribution of responses between the different areas of HCI task involvement

| HCI task involvement | Abbreviations | Number of responses |
|---|---|---|
| Adapting, customizing, or optimizing systems according to user need and requirements | AD | 80 |
| Interoperability between systems | IO | 17 |
| User modeling | UM | 61 |
| Usability engineering | UE | 66 |
| Modeling of cognitive behavior | CB | 44 |
| Knowledge elicitation of human expertise | KE | 31 |
| Knowledge acquisition | KA | 14 |
| Network access and network services | NW | 23 |
| Work methods and organizational design | WO | 28 |
| Multimedia system design and evaluation | MU | 57 |
| Planning and scheduling | PL | 16 |
| Natural langauge understanding | NL | 9 |
| Classification and prediction | CL | 14 |
| Conflict resolution | CR | 5 |
| Information retrieval | IR | 35 |
| Case based reasoning | CA | 10 |
| Execution and control | EC | 13 |
| Critiquing and error engineering | CE | 11 |
| Safety and risk management | SR | 11 |
| Verification and validation | VV | 6 |
| Other | OTH | 9 |

Each responder could select multiple tasks. The average number of tasks used per responder is 4.90. Total sum across all responders and tasks is 560. Task abbreviations are in parentheses.

of cognitive behavior (CB), followed by information retrieval (IR). Collectively they account for more than 60% of all responses.

Contingency analysis across responses revealed few significant interactions between tasks; see Table 4. Adapting and system optimization (AD) achieved maximum contingency with user modeling (UM), usability engineering (UE), and multimedia system design and evaluation (MU). Also high contingency ratings are noticed between UM and UE, UM and modeling of cognitive behavior (CB), UM and MU, and AD and information retrieval (IR). On the other hand, several other tasks achieved relatively low contingency interaction within the global space of tasks. Nevertheless, Tables 3 and 4 compose an interdisciplinary view for HCI work. Supporting such work with tools coming from ML, or any other technology source, cannot be trivial.

## Part III: HCI Application Involvement

The third part of the questionnaire captured application demographics of responders. We summarize responses in Table 5. Participants could select more than one area of HCI application involvement. Education and training, followed by public

**Table 4.** Contingency across HCI tasks

| | IO | UM | UE | CB | KE | KA | NW | WO | MU | PL | NL | CL | CR | IR | CA | EC | CE | SR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AD | 12 | 51 | 51 | 35 | 28 | 13 | 18 | 23 | 50 | 15 | 8 | 11 | 4 | 31 | 10 | 12 | 11 | 6 |
| IO | | 8 | 11 | 7 | 5 | 3 | 7 | 6 | 12 | 4 | — | 5 | 2 | 6 | 3 | 4 | 3 | 1 |
| UM | | | 36 | 34 | 21 | 8 | 11 | 17 | 33 | 13 | 5 | 8 | 4 | 20 | 9 | 9 | 9 | 6 |
| UE | | | | 33 | 19 | 8 | 12 | 19 | 42 | 9 | 6 | 7 | 3 | 26 | 9 | 7 | 9 | 6 |
| CB | | | | | 18 | 4 | 8 | 15 | 24 | 10 | 5 | 9 | 3 | 18 | 9 | 7 | 9 | 6 |
| KE | | | | | | 6 | 6 | 15 | 19 | 7 | 5 | 5 | 1 | 12 | 7 | 5 | 4 | 2 |
| KA | | | | | | | 4 | 6 | 7 | 1 | 1 | — | — | 12 | 1 | 1 | — | — |
| NW | | | | | | | | 7 | 12 | 4 | — | 4 | 3 | 12 | 2 | 5 | 3 | 1 |
| WO | | | | | | | | | 19 | 7 | 3 | 3 | 1 | 14 | 4 | 4 | 4 | 5 |
| MU | | | | | | | | | | 11 | 6 | 7 | 2 | 22 | 7 | 7 | 7 | 5 |

Task abbreviations are identified in Table 3. Rows that did not include a contingency higher than 10 are omitted.

sector, media or entertainment, and aviation/aerospace lead the list of HCI application domain involvement. Average involvement was 1.88 application domains per responder. Maximum contingency was achieved between education/training and media/entertainment (equal to 16), education/training and public sector (equal to 15), and public sector and media/entertainment (equal to 12). No other contingency was higher than 8. However, responses in application involvement agree with those with respect to area of HCI involvement and HCI task. Altogether they draw an integrative picture of HCI work and research.

**Table 5.** Distribution of responses between the different areas of HCI application invovlement

| Application | Number of responses |
|---|---|
| Aviation/aerospace | 21 |
| Chemical/food industry | 6 |
| Electric power industry | 3 |
| Manufacturing | 16 |
| Shipping industry | 6 |
| Medicine | 14 |
| Education and training | 51 |
| Media or entertainment | 25 |
| Public sector | 31 |
| Software engineering/telecommunications | 12 |
| Other | 25 |
| Total | 210 |

Each responder could assign more than one application.

## Part IV: Machine Learning Expertise and Use (or Nonuse) in HCI Work

The fourth part of the questionnaire captured the degree to which responders are using (or not using) ML in HCI work, the level of awareness with respect to alternative ML paradigms, and satisfaction with ML by those who have used it in HCI work. We asked participants to rate themselves across a range of ML paradigms. Paradigms were motivated from earlier work on ML research and application classification (e.g., Kodratoff et al., 1994; Langley & Simon, 1995; Michalski & Kodratoff, 1990; Moustakis et al., 1996). We also asked participants to indicate whether they have used ML in their work regardless of level of ML expertise and to specify specific ML paradigms used.

To capture level of ML expertise, we asked responders to rate themselves using the following discrete seven-point scale (intermediate points were used to map between positions): 1–2, not at all familiar; 3–4, has some knowledge of; 5–6, used at least once or studied it extensively; and 7, very familiar.

We summarize results in Table 6. Out of 112 participants, 41 indicated that they either have used or are using ML in their work. Frequency of use with respect to each ML paradigm, among the 41 ML users, is listed in Table 6. However, ML users do not usually confine themselves to a single paradigm. Frequency of use estimates, combined with an average of 2.3 paradigms per responder who has used ML and $2 \times 2$ contingency analysis results reported in Table 7, reflect this fact.

Contingency analysis across ML paradigms used in HCI work (see Table 7) exposes the prominent role neural networks (NNL), statistical learning methods (SL), rule induction (RI), and case-based learning (CBL) have played in HCI–ML application development. However, the relatively high interaction between RI and

**Table 6.** Knowledge about machine learning and frequency of use of ML paradigms

| Machine learning paradigm | Abbreviation | Average score ($N$= 112) | Frequency of use |
|---|---|---|---|
| Neural networks (or connectionist models) | NNL | $3.6 \pm 1.4$ | 12 |
| Genetic algorithms | GA | $2.4 \pm 1.3$ | 6 |
| Case-based learning (or instance-based learning) | CBL | $2.8 \pm 1.6$ | 9 |
| Rule induction | RI | $3.0 \pm 1.7$ | 19 |
| Statistical learning models | SL | $2.8 \pm 1.7$ | 18 |
| Reinforcement learning | RL | $2.3 \pm 1.6$ | 4 |
| Knowledge discovery in databases (data mining) | KDD | $2.5 \pm 1.5$ | 3 |
| Knowledge refinement systems | KRS | $2.5 \pm 1.5$ | 6 |
| Conceptual clustering | CC | $2.3 \pm 1.7$ | 10 |
| Inductive logic programming | ILP | $2.3 \pm 1.4$ | 7 |

Average scores are listed using a seven-point discrete scale in which 1 indicates that the responder is not familiar at all (with the associated ML paradigm) and 7 indicates that the responder is exceptionally familiar with the paradigm. Standard deviation estimates are given with the average values.

**Table 7.** Contingency between ML paradigms used in HCI work

| | ML paradigm | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | NNL | GA | CBL | RI | SL | RL | KDD | KRS | CC | ILP |
| NNL | | 4 | 5 | 4 | 7 | 2 | 1 | 2 | 5 | 1 |
| GA | | | 1 | 1 | 2 | 2 | 1 | | 3 | |
| CBL | | | | 4 | 6 | 1 | | 1 | 2 | 2 |
| RI | | | | | 7 | | 3 | 2 | 3 | 6 |
| SL | | | | | | 1 | 1 | 1 | 4 | 3 |
| RL | | | | | | | | 2 | 2 | |
| KDD | | | | | | | | 1 | | 1 |
| KRS | | | | | | | | | 2 | 1 |
| CC | | | | | | | | | | 3 |

Average contingency is 2.20. Abbreviations are introduced in Table 6. Zero entries are omitted.

inductive logic programming (ILP) marks a noteworthy evolution based on the fact that ILP systems provide the user with advanced knowledge representation and inference and learning mechanisms and are likely to empower application development in the future.

Average ML knowledge scores were not that high; they ranged between 3.6 maximum for neural networks (NNL) and 2.3 minimum with respect to reinforcement leaning (RL), conceptual clustering (CC), and inductive logic programming (ILP). However, average scoring conceals real ML skills of participants. If we average values using maximum ratings for each participant, then a different picture emerges. At any rate, we should not expect responders to be experts across all ML paradigms. In fact, this is not even true for ML researchers, as results reported by Moustakis et al. (1996) indicate. Thus we took the maximum rating from each participant and averaged values across all ML paradigms and responders, i.e., those who have or have not used ML in HCI work. The process yielded an average value of $4.8 \pm 1.5$, which comes close to the rating "used it at least once or studied it extensively." We summarize results in Table 8. This means that our sample encom-

**Table 8.** Relation between maximum level of ML awareness with ML usage in carrying out HCI work or research

| | ML rank higher than (or equal to) 4 | ML rank lower than (or equal to) 3 |
|---|---|---|
| Has used ML in some HCI task(s) | 40 | 1 |
| Has not used ML in any HCI task(s) | 44 | 27 |

With respect to ML awareness, we are using 4 as a threshold; namely, 4 lies in between the scale ranks of "has some knowledge of" and "used at least once or studied it extensively." Entries correspond to survey responses and sum up to the total number of participants in the survey ($N = 112$).

passed persons with good level of familiarity with at least one ML paradigm. On the other hand, high rating with respect to at least an ML paradigm seemed to have influenced use of ML in HCI work. Using 4 as a threshold (4 lies in between "has some knowledge of" and "used it at least once or studied extensively" levels, we identified that 40 out of 41 ML users exhibited an ML skill that exceeded the threshold value. In addition, average awareness across all paradigms with respect to all ML users exceeded the level of 6, i.e., ranged between "used it at least once or studied extensively" and "very familiar" levels. Finally, a significant number of responders, i.e., 44, indicated ML awareness higher than or equal to 4; however, they have not used ML in HCI work.

The same part of the questionnaire also included a satisfaction score between those who have used ML in HCI work. To capture level of satisfaction, we used a five-point scale, namely, 1, poor; 2, little; 3, good; 4, very good; and 5, excellent.

Average satisfaction across all ML users is 3.3, which exceeds the "good" level.

## Part VI: Why ML Has Not Been Used in HCI Work

Finally, responders were asked to indicate the reason(s) for not having used ML in their work. Responders were asked to select as many reasons as necessary from a list of five alternatives. We summarize responses in Table 9. Results indicate that among those who have not used ML in HCI work the main reasons were

- they perceived ML as not being necessary to what they were doing,
- they were not aware that ML could be useful to their work, or
- they have not seen enough success stories of using ML in work similar to theirs.

Indeed, contingency analysis between nonuse reasons, not reported in detail herein for sake of brevity, identified significant interaction between the points just raised. Having completed descriptive analysis of survey data, we decided to continue guided analysis according to the following research questions:

**Table 9.** Distribution of reasons for not using (or having used) ML in HCI work

| Reasons for not having used ML | Number of responses |
|---|---|
| Was fairly certain that it was not necessary | 30 |
| Was not aware that ML may be used | 37 |
| Tried but did not work | 5 |
| Did not know how to incorporate ML in my work | 7 |
| Have not read enough success stories about using ML in my work | 28 |
| Because of lack of resources | 9 |
| Total | 116 |

Responders could select more than one reason.

Are there significant differences between users of ML and nonusers?
What paradigms do people using ML use the most in HCI work?

## Do Users of ML Differ from Nonusers?

Focusing on HCI task involvement and clustered responses to examine differences between users and nonusers of ML, AutoClass (Cheeseman & Stutz, 1996) generated three classes out of 112 vectors of binary data extending over 21 attributes. Attributes correspond to tasks, and valuation was either *yes* in the case the responder indicated involvement with the task or *no* otherwise. We decided to focus on tasks alone because they are generic and characterize HCI work regardless of either area or application domain (see Tables 3 and 5, respectively). Use or nonuse of ML was not included to avoid biasing the results. The process converged to three classes, namely, classes A, B, and C. We summarize results in Table 10. Case-based reasoning (CA; see Table 3) emerged as the most significant attribute, i.e., task, in class formation, followed

**Table 10.** Term influence values across classes

| HCI task | Abbreviation | Class A (54) | Class B (48) | Class C (10) | Global |
|---|---|---|---|---|---|
| Adapting, customizing, optimizing systems according to user need and requirements | AD | 0.136 | 0.151 | 0.199 | 0.368 |
| Interoperability between systems | IO | | 0.009 | 0.165 | 0.132 |
| User modeling | UM | 0.021 | | 0.400 | 0.320 |
| Usability engineering | UE | 0.131 | 0.094 | 0.313 | 0.407 |
| Modeling of cognitive behavior | CB | 0.076 | 0.013 | 0.729 | 0.619 |
| Knowledge elicitation of human expertise | KE | 0.065 | 0.018 | 0.334 | 0.316 |
| Knowledge acquisition | KA | 0.065 | 0.082 | 0.042 | 0.143 |
| Network access and network services | NW | | | 0.036 | 0.028 |
| Work methods, organizational design | WO | 0.063 | 0.032 | 0.136 | 0.175 |
| Multimedia system design and evaluation | MU | 0.080 | 0.068 | 0.071 | 0.166 |
| Planning and scheduling | PL | 0.023 | 0.004 | 0.207 | 0.176 |
| Natural language understanding | NL | 0.007 | 0.002 | 0.092 | 0.076 |
| Classification and prediction | CL | | 0.097 | 0.736 | 0.631 |
| Conflict resolution | CR | | 0.022 | 0.340 | 0.274 |
| Information retrieval | IR | 0.180 | 0.128 | 0.073 | 0.288 |
| Case-based reasoning | CA | 0.066 | 0.014 | 1.242 | 1.000 |
| Execution and control | EC | | 0.077 | 0.771 | 0.642 |
| Critiquing and error engineering | CE | 0.026 | 0.007 | 0.745 | 0.588 |
| Safety and risk management | SR | | | 0.003 | 0.003 |
| Verification and validation | VV | 0.026 | 0.031 | 0.002 | 0.044 |
| Other | OTH | 0.030 | 0.032 | 0.012 | 0.057 |
| Has used ML in HCI work | | 18 | 21 | 2 | |

Term influence values give a rough heuristic measure of the relative influence of each attribute in differentiating classes from the overall data set. Global values are normalized. Numbers in parentheses denote number of responses belonging to each class. Zero entries are omitted.

by execution and control (EC), classification (CL), modeling of cognitive behavior (CB), and critiquing and error engineering (CE); Table 10.

This result agrees with the scope of the survey, i.e., to link ML use in HCI work, because leading attributes represent intelligent tasks, which are often combined with use of ML [see also the work of Moustakis et al. (1996)] and are not confined solely to HCI work and research. Being "HCI peripheral," these tasks influenced the most class formation. Class-specific influence values are also listed in Table 10. In both classes A and B, no task exerted a single major influence; however, a bundle of tasks achieved high influence values in class C formation. ML use in HCI work is almost equally split between classes A and B; class A encompassed 18 users of ML and class B, 21 users, while 2 users were classed in C.

Class formation alerted us that significant differences could *not* be identified between those who have used ML and those who have not. In fact, repeated two-tailed pairwise t-tests of means, using HCI area of involvement data (see Table 11), HCI tasks (Table 12), and HCI area of application (Table 13) revealed that all effects and interactions were not significant at the $\alpha = 0.05$ level. Neither did modeling of patterns using CN2 (Clark & Niblett, 1989) or C4.5 (Quinlan, 1992) produce meaningful patterns related to use or nonuse of ML despite the rather high statistical accuracy of learned rules; accuracy ranged between 90% and 96%, approximately. Experiments with CN2 and C4.5 were conducted using the same data we used with AutoClass by adding ML skill to the list of attributes and use of ML as a consequence.

## To Which Tasks Do People Apply ML?

Focusing on responders who have used ML in HCI work, we constructed a $2 \times 2$ contingency table to relate ML paradigms with HCI tasks. We summarize results

**Table 11.** Distribution of responses between those who have and those who have not used ML

| | | Number of responses | | | |
|---|---|---|---|---|---|
| HCI area of involvement | Abbreviation | Yes | No | Yes (%) | No (%) |
| User interface design/evaluation (including display and mouse devices) | UI | 35 | 66 | 32 | 39 |
| Product or services design and evaluation | PS | 24 | 30 | 22 | 18 |
| Concurrent engineering | CE | 2 | 5 | 2 | 3 |
| Computer-supported cooperative work | CSCW | 13 | 24 | 12 | 14 |
| Human-robot interaction | HRI | 2 | 3 | 2 | 2 |
| Education, training | ITS | 14 | 20 | 13 | 12 |
| Virtual reality | VR | 10 | 9 | 9 | 5 |
| Other | OTH | 10 | 14 | 9 | 8 |
| Total | | 110 | 171 | 100 | 100 |

Two-tail, pairwise t-test analysis with $\alpha = 0.05$ confirmed the null hypothesis that the difference of means between those who have used and those who have not used ML is zero. Statistical testing was done using percentage estimates, i.e., last two columns of the table. Yes refers to responders who have used ML and No to those who have not.

**Table 12.** Distrubtion of responses with respect to HCI task invovlement

| HCI task | Abbreviation | Yes | No | Yes (%) | No (%) |
|---|---|---|---|---|---|
| Adapting, customizing, etc. | AD | 32 | 48 | 15 | 14 |
| Interoperability between systems | IO | 4 | 13 | 2 | 4 |
| User modeling | UM | 23 | 38 | 11 | 11 |
| Usability engineering | UE | 20 | 46 | 10 | 13 |
| Modeling of cognitive behavior | CB | 21 | 23 | 10 | 7 |
| Knowledge elicitation | KE | 13 | 18 | 6 | 5 |
| Knowledge acquisition | KA | 4 | 10 | 2 | 3 |
| Network access and network services | NW | 6 | 17 | 3 | 5 |
| Work methods, organizational design | WO | 7 | 21 | 3 | 6 |
| Multimedia system design, etc. | MU | 21 | 36 | 10 | 10 |
| Planning and scheduling | PL | 5 | 11 | 2 | 3 |
| Natural language understanding | NL | 6 | 3 | 3 | 1 |
| Classification and prediction | CL | 9 | 5 | 4 | 1 |
| Conflict resolution | CR | 2 | 3 | 1 | 1 |
| Information retrieval | IR | 12 | 23 | 6 | 7 |
| Case-based reasoning | CA | 4 | 6 | 2 | 2 |
| Execution and control | EC | 3 | 10 | 1 | 3 |
| Critiquing and error engineering | CE | 3 | 8 | 1 | 2 |
| Safety and risk management | SR | 4 | 7 | 2 | 2 |
| Verification and validation | VV | 3 | 3 | 1 | 1 |
| Other | OTH | 5 | 4 | 2 | 1 |
| Total | | 207 | 353 | 100 | 100 |

Two-tail, pairwise t-test analysis with $\alpha = 0.05$ confirmed the null hypothesis that the difference of means between those who have and those who have not used ML is zero. Statistical testing was done using percentage estimates, i.e., last two columns of the table. (Yes relates to those who have used ML, and No to those who have not used it.)

**Table 13.** Distribution of responses with respect to HCI application domain invovlement

| Application | Yes | No | Yes (%) | No (%) |
|---|---|---|---|---|
| Aviation/aerospace | 9 | 12 | 12 | 8.9 |
| Chemical/food industry | 0 | 6 | 0 | 4.4 |
| Electric power industry | 0 | 3 | 0 | 2.2 |
| Manufacturing | 3 | 13 | 4 | 9.6 |
| Shipping industry | 2 | 4 | 2.7 | 3 |
| Medicine | 8 | 6 | 11 | 4.4 |
| Education and training | 17 | 34 | 23 | 25 |
| Media or entertainment | 11 | 14 | 15 | 10 |
| Public sector | 13 | 18 | 17 | 13 |
| Software engineering/telecommunications | 2 | 10 | 2.7 | 7.4 |
| Other | 10 | 15 | 13 | 11 |
| Total | 75 | 135 | 100 | 100 |

Two-tail, pairwise t-test analysis with $\alpha = 0.05$ confirmed the null hypothesis that the difference of means between those who have and those who have not used ML is zero. Statistical testing was done using percentage estimates, i.e., last two columns of the table. (Yes relates to those who have used ML, and No to those who have not used it.)

in Table 14. Obviously, differences both across tasks and ML paradigms are significant. Results reported in Table 14 confirm the following.

- The popularity of rule induction (RI), statistical learning (SL), neural networks (NNL), and case-based learning (CL). However, it is also interesting that people are also relying heavily on knowledge intensive learning methods, i.e., inductive logic programming (ILP). Given that RI, SL, NNL, and quite often, CBL are good only with simple data (i.e., most of these methods fail to represent effectively relational expressions), use of ILP techniques empowers the application of ML to HCI tasks.
- The belief of ML researchers concerning the applicability of ML in real-world tasks. For instance, taxonomies reported by Kodratoff et al. (1994), survey results published by Moustakis et al. (1996), as well as the work of Brodley (1993) on ML algorithm superiority do give similar results; however, we should note that our survey encompassed more tasks than in previous works.
- Synthetic nature of HCI work, which requires concurrent or sequential application of more than one paradigm given a task. (Of course, we should note that results

**Table 14.** Contingency analysis between HCI task involvement and ML paradigm used in HCI work

| | Machine learning paradigm | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Task | NNL | GA | CBL | RI | SL | RL | KDD | KRS | CC | ILP |
| AD | 11 | 5 | 9 | 14 | 14 | 3 | 3 | 5 | 5 | 6 |
| IO | 2 | | 1 | 2 | 3 | | 1 | | | 1 |
| UM | 6 | 2 | 6 | 10 | 8 | 3 | 2 | 5 | 5 | 3 |
| UE | 4 | 4 | 5 | 6 | 9 | 1 | | 2 | 2 | 6 |
| CB | 7 | 3 | 4 | 8 | 9 | 2 | 1 | 2 | 2 | 3 |
| KE | 6 | 3 | 2 | 5 | 4 | 2 | 3 | 5 | 5 | 2 |
| KA | 2 | 1 | 1 | 3 | 1 | | 3 | 1 | 1 | 1 |
| NW | 2 | 1 | 2 | 4 | 3 | 1 | 1 | | | 4 |
| WO | 3 | 2 | 2 | 3 | 2 | | 1 | 2 | 2 | 1 |
| MU | 7 | 3 | 5 | 8 | 9 | 1 | 1 | 3 | 3 | 4 |
| PL | 1 | | 1 | 1 | 4 | | | | | |
| NL | 5 | 3 | 3 | 1 | 4 | 1 | | 1 | 1 | 1 |
| CL | 4 | | 2 | 6 | 7 | | | | | 2 |
| CR | | | | 2 | | | | | | 2 |
| IR | 3 | 3 | 3 | 6 | 3 | 1 | 3 | 2 | 2 | 4 |
| CA | 1 | | 1 | 3 | | | | | | 3 |
| EC | 1 | | 1 | 3 | 1 | | | | | 2 |
| CE | | | | 3 | | | | | | 3 |
| SR | 2 | | 1 | 4 | 1 | | | 1 | 1 | 1 |
| VV | | | | 3 | 3 | | | | | |
| OTH | 1 | 1 | 1 | 3 | 1 | | | 1 | 1 | |
| Sum | 68 | 31 | 50 | 98 | 86 | 15 | 19 | 30 | 30 | 49 |

Abbreviations are explained in Tables 3 and 6. Last line includes the sum of paradigm use across all tasks. Zero entries are omitted.

reported in Table 14 assume that those who have used ML have done so with respect to all HCI tasks with which they work).

## CONCLUDING REMARKS

This article was based on a structured survey aimed at identifying and modeling use of machine learning (ML) in HCI work. Analysis of 112 responses from leading researchers in the HCI field lead us to formulate some critical initial conclusions, namely, the following.

- A significant number of HCI workers (both from industry and academia) are well aware of and knowledgeable about ML.
- More than a third of HCI workers has actually reported that they have, or are, using ML in HCI work, and in doing so, they are satisfied with the result. Certainly, average satisfaction may be improved, and this represents a challenge for the future.
- Implementation of ML in HCI tasks is not trivial. HCI work is synthetic and so has to involve application of ML. However, this poses a great challenge to ML, which has to mature enough before it is able to "play" effectively in synthetic "games." Moreover, it will require effective process models, leading to an agenda for action appropriate for the situation at hand.
- A high percentage, e.g., slightly less than two-thirds, of those who participated in the survey do not use ML in HCI work. A sample of reasons for this is reported in the text; see Table 9. Leading reasons include lack of awareness about the potential, perception that ML is not necessary, and lack of concrete case studies. In fact, these reasons mix with each other as the contingency matrix in Table 15 suggests. Leading contingencies are between lack of concrete case studies and misperception, and between lack of awareness and misperception. To overcome these barriers, people should try to promote their work, and researchers in both fields should cooperate to advance pedagogy about ML. Several years ago,

**Table 15.** Contingency analysis between reasons for not having used ML in HCI work

| | Reason | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| (1) Thought ML was not necessary | | 10 | 1 | | 13 | 2 |
| (2) Was not aware ML is useful | | | 2 | 3 | 14 | |
| (3) Tried ML but did not work | | | | | 3 | |
| (4) Did not know how to use ML | | | | | 4 | 2 |
| (5) Have not seen enough ML success stories | | | | | | 2 |
| (6) Lack of resources to use ML | | | | | | |

Zero entries are omitted. Refer to Table 9 for an extensive description of reasons used in the survey.

> Churchman and Shainblatt (1965) pondered that effective implementation of management science in the workplace would require a cooperative dialectic between the researcher and the manager. Likewise ML and HCI colleagues should join forces and try to understand each other's needs, ideas, biases, and perceptions. Both ML and HCI are cognitive sciences. Both focus on humans and draw from human-centered needs. Thus joining forces is the only way for achieving maximum benefit and synergy, which may also unleash potential in each science alone.

This article explored the potential, yet much ground remains hidden. We hope that future studies will focus on specifics and produce concrete case results, much needed by colleagues in both fields.

## REFERENCES

Brodley, C. 1993. Addressing the selective superiority problem: Automatic algorithm/model class selection. In *Proceedings of the 10th International Conference on Machine Learning,* pp. 17–24. San Mateo, Calif.: Morgan Kaufmann.

Cheeseman, P., and J. Stutz. 1996. Bayesian classification (AutoClass): Theory and results. In *Advances in knowledge discovery and data mining,* eds. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, pp. 153–180. Menlo Park, Calif.: AAAI Press/MIT Press.

Churchman, C. W., and A. H. Schainblatt. 1965. The researcher and the manager: A dialectic of implementation. *Management Science* 11(4):B69–87.

Clark, P., and T. Niblett. 1989. The CN2 induction algorithm. *Machine Learning* 3:261–283.

Kodratoff, Y., V. S. Moustakis, and N. Graner. 1994. Can machine learning solve my problem? *Applied Artificial Intelligence: An International Journal* 8(1):1–36.

Kodratoff, Y., D. Sleeman, M. Uszynski, K. Gausse, and S. Craw. 1992. Building a machine learning toolbox. In *Enhancing the knowledge engineering process—Contributions from ESPRIT,* eds. B. Le Pape and L. Steels. Amsterdam: Elsevier.

Langley, P., and H. Simon. 1995. Applications of machine learning and rule induction. *Communications of the ACM* 38(11): 54–64.

Michalski, R. S., and Y. Kodratoff. 1990. Research in machine learning: Recent progress, classification of methods, and future directions. In *Machine learning: An artificial intelligence approach,* vol. 3, eds. Y. Kodratoff and R. S. Michalski, pp. 3–30. San Mateo, Calif.: Morgan Kaufmann.

Michie, D. D., J. Spiegelhalter, and C. C. Taylor. 1994. *Machine learning: Neural and statistical classification.* Hemel Hempstead, UK: Ellis Norwood.

Moustakis, V. S., M. Lehto, and G. Salvendy. 1996. Survey of expert opinion: Which machine learning method may be used for which task? *International Journal of Human-Computer Interaction* 8(3):221–236.

Quinlan, J. R. 1992. *C4.5: Programs in machine learning.* San Mateo, Calif.: Morgan Kaufmann.

Sleeman, D., M. Rissakis, S. Craw, N. Graner, and S. Sharma. 1995. Consultant-2: Pre- and post-processing of machine learning applications. *International Journal of Human-Computer Studies* 43:43–63.