Revista Brasileira de Computação Aplicada, Novembro, 2024

DOI: 10.5335/rbca.v16i3.15683
Vol. 16, Nº 3, pp. 89–99
Homepage: seer.upf.br/index.php/rbca/index

ARTIGO ORIGINAL

Desvendando a experiência e o humor do usuário: uma avaliação comparativa e uma abordagem de modelagem preditiva

Unraveling user experience and mood: a comparative evaluation and predictive modeling approach

Gabriela Jie Han[1], Erico de Souza Veriscimo[1], João Luiz Bernardes Júnior[1], Luciano Antonio Digiampietri[1]

[1]Escola de Artes, Ciências e Humanidades - Universidade de São Paulo

*gabi.jiehan@usp.br; ericoveriscimo@usp.br; jlbernardes@usp.br; digiampietri@usp.br

Recebido: 27/03/2024. **Revisado:** 11/11/2024. **Aceito:** 30/11/2024.

Resumo

Background A avaliação da experiência do usuário (UX) é crucial para os sistemas de informação, uma vez que está intimamente relacionada com a sua aceitação e desempenho pós-adoção. Implica frequentemente a utilização de um ou mais questionários normalizados. Esta avaliação pode ser dispendiosa, desde a seleção do questionário adequado até à extração de dados de um número significativo de usuários, análise dos dados e comparação de resultados, especialmente com avaliações que utilizam questionários diferentes. Assim como em diferentes áreas de conhecimento, no projeto de interfaces e avaliação da UX, a Inteligência Artificial tem sido cada vez mais adotada. Este trabalho explora vários algoritmos de *Machine Learning* (ML) para correlacionar e prever as respostas de diferentes questionários de UX e de humor. **Resultados** A abordagem desenvolvida previu com sucesso as respostas ao questionário com elevada precisão (erro inferior a 1 numa Escala Diferencial Semântica de 7 pontos), particularmente para o questionário mais longo. Além disso, revelou que o humor contribui significativamente para esta precisão. **Conclusões** As principais contribuições deste trabalho incluem a demonstração de como o ML pode ser utilizado para reduzir os custos de avaliação da experiência do usuário e a exploração do impacto do humor do usuário nesta avaliação.

Palavras-Chave: Avaliação da experiência do usuário; Interação 3D; Experiência do Usuário

Abstract

Background User Experience (UX) evaluation is crucial for Information Systems, as it is closely related to their acceptance and post-adoption performance. It often involves the use of one or more standardized questionnaires. This evaluation can be costly, from selecting the appropriate questionnaire to extracting data from a significant number of users, analyzing the data and comparing results, especially with evaluations that use different questionnaires. As in different areas of knowledge, Artificial Intelligence has been increasingly adopted in interface design and UX evaluation. In this work, we explore several machine learning (ML) algorithms to correlate and predict the answers of different UX and mood questionnaires. **Results** The developed approach successfully predicted questionnaire responses with high accuracy (error less than 1 on a 7-point Semantic Scale), particularly for the longer questionnaire. Additionally, it revealed that mood significantly contributes to this accuracy. **Conclusions** Our main contributions include demonstrating how ML can be utilized to reduce UX evaluation costs and exploring the impact of user mood on this evaluation.

Keywords: 3D interaction, User Experience, UX Evaluation

1 Introdução

A avaliação da Experiência de Usuário (UX), definida como as percepções e respostas de uma pessoa que resultam do uso ou da antecipação do uso de um sistema, produto ou serviço (Law et al., 2009; Bevan, 2009), é crucial para sistemas de informação, uma vez que está intimamente ligada à aceitação dos sistemas e ao desempenho em seu uso após sua adoção. O Modelo de Aceitação de Tecnologia (Szajna, 1996) postula que para que sistemas de informação sejam adotados e de fato aproveitados, a utilidade percebida pelo usuário em sua experiência com o mesmo deve ser maior que a dificuldade percebida. As tarefas e características do sistema do modelo Task-Technology Fit (TTF) (Goodhue e Thompson, 1995) exploram a utilização de sistemas de informação após sua adoção e estão fortemente relacionadas com UX, a tal ponto que o TTF tem sido usado desde sua criação para obter informações sobre esta experiência e para aprimorá-la. De acordo com este modelo, uma UX inadequada devido a uma discordância entre tarefas e características do sistema prejudica, por exemplo, o processo de tomada de decisão (Vessey e Galletta, 1991).

No atual cenário de "*Big Data*", a UX tem se tornado cada vez mais importante. Graças ao crescimento do poder computacional e de coleta e armazenamento de dados, sistemas de informação são capazes de utilizar e produzir conjuntos de dados caracterizados por grandes volumes, disponibilidade e variedade, incluindo não só dados numéricos e textuais estruturados como também dados não-estruturados (Lyytinen e Grover, 2017). Dada a complexidade das tarefas de explorar este grande volume de dados, extrair dele informações relevantes e utilizá-las no apoio a decisão, torna-se necessária a adoção e eficiente utilização de sistemas de informação inovadores que possam auxiliar nestas tarefas, tanto que pesquisa e educação em ciência de dados são parte de um dos grandes desafios de pesquisa em SI (Boscarioli et al., 2017).

Duas das principais estratégias para explorar estes dados são a Visualização de Informação, intrinsecamente multidimensional (LaViola Jr et al., 2017; Schroeder et al., 2006; Araújo et al., 2015; da Costa et al., 2014; Choma et al., 2014; Ferreira et al., 2019), e o uso de Inteligência Artificial (IA). As pesquisas sobre avaliação de UX em sistemas de IA têm crescido recentemente devido à popularização do uso de *chatbots*, mas a análise de UX em sistemas com interação 3D para visualização de informação ainda tem sido relativamente pouco investigada (Veriscimo et al., 2020).

A avaliação da UX frequentemente envolve o uso de instrumentos como questionários padronizados e validados e, atualmente, há vários destes questionários que podem ser utilizados (Veriscimo et al., 2020). No entanto, além de possuírem perguntas muito diferentes (em quantidade, aspectos a serem avaliados e objetivos), esses questionários não procuram analisar a influência do humor do participante durante o processo de avaliação. E esta avaliação pode incorrer em custos consideráveis, desde a seleção dos instrumentos mais apropriados à extração dos dados de um número significativo de usuários, à análise desses dados e comparação dos resultados, especialmente quando se procura comparar avaliações que utilizam instrumentos diferentes. Por isso, a IA tem sido cada vez mais usada para auxiliar nestas tarefas.

Neste contexto, o objetivo deste trabalho é explorar diferentes algoritmos de aprendizado de máquina para correlacionar e prever respostas de diferentes questionários de avaliação de UX e de humor. A principal pergunta de pesquisa é: "É possível inferir, com boa precisão, as respostas de cada pergunta de um questionário de UX com base nas respostas de outro(s)?".

Para alcançar este objetivo, foram realizados 35 testes com usuários em um experimento aprovado pelo Comitê de Ética da instituição, no qual foram aplicados dois questionários padronizados: o User Experience Questionnaire (UEQ-S) (Laugwitz et al., 2008) e o PLEX Framework (Lucero et al., 2013; Arrasvuori et al., 2011), para avaliar a experiência do indivíduo após o mesmo interagir em um jogo 3D, além de um questionário de avaliação do humor antes e depois do experimento, o POMS (Viana et al., 2001).

A Seção 2 segue esta introdução discutindo trabalhos correlatos. Já a Seção 3 descreve em mais detalhes os materiais e métodos utilizados neste estudo. A Seção 4 apresenta e discute os resultados obtidos. Por fim, a Seção 5 apresenta as principais conclusões do trabalho.

2 Trabalhos Correlatos

Em uma revisão sistemática (RS) sobre avaliação de experiência do usuário (Veriscimo et al., 2020) foi identificado que o questionário é o método de avaliação mais utilizado em avaliação de UX e, em relação aos critérios de avaliação, foi possível perceber que os critérios pragmáticos foram privilegiados em relação aos hedônicos. Visto que os critérios pragmáticos são os relacionados ao desempenho do usuário na realização de tarefas no sistema e os critérios hedônicos são aqueles mais relacionados às emoções e prazer dos usuários, é perceptível que a preocupação maior não está voltada ao humor, sentimentos e prazer do usuário. Contudo, nesta RS não é mencionado nenhum uso de questionários e/ou avaliações do humor do usuário no auxílio da avaliação de UX.

O questionário padronizado mais utilizado em avaliação de UX discutido por Veriscimo et al. (2020), é o User Experience Questionnaire (UEQ) que tem como objetivo avaliar UX com 26 itens dispostos em seis categorias: atratividade, perspicuidade, eficiência, confiabilidade, estímulo e novidade (Laugwitz et al., 2008), existe também uma versão curta do mesmo questionário denominada User Experience Questionnaire Short (UEQ-S) com apenas oito perguntas, tanto a versão curta como a tradicional possuem os seus itens em estrutura de Escala Diferencial Semântica de sete pontos entre antônimos, como difícil e fácil, e nenhuma questão do questionário aborda algo relacionado ao humor do usuário.

Uma alternativa para avaliação de UX voltada para a parte lúdica da experiência é o PLEX Framework (Lucero et al., 2013; Arrasvuori et al., 2011) organizado em 22 itens customizados conforme o contexto da aplicabilidade. Ainda que a ludicidade pode manifestar-se de muitas maneiras diferentes, uma vez que os humanos são inerentemente brincalhões por natureza, e a ludicidade está intimamente relacionada ao humor, o PLEX Framework não possui nenhum item relacionado ao humor do participante.

Laugwitz et al. (2008) propõem um método para a construção de um questionário de UX e concluem que o processo deve garantir que o maior número possível de características relevantes do produto sejam levadas em consideração. Porém, deste modo o alvo principal se torna o produto, deixando o usuário em segundo plano. Além disso, o humor não é levado em consideração para a avaliação.

Vários trabalhos fazem a avaliação da UX em diferentes áreas de aplicação, como Quek e See (2015), que propõem e analisam a experiência em um jogo simples com Realidade Aumentada (RA), Dupont et al. (2018), que propõem uma estrutura para sistematizar a avaliação de desempenho imersivo e colaborativo e Yangguang et al. (2014), que projetaram e avaliaram o Sistema de Treinamento Colaborativo Multijogador (MCTS), porém nenhum deles considera ou utiliza o humor em suas avaliações.

Bertão e Joo (2021) analisam o uso de Inteligência Artificial (IA) em diferentes práticas relacionadas ao desenvolvimento de interfaces com o usuário e UX no Brasil, destacando as perspectivas dos profissionais, taxas de adoção e expectativas futuras. O estudo revelou uma baixa adoção de ferramentas de IA no país: apenas 18% utilizam ferramentas de design com IA e 5% usam ferramentas integradas. Esses números contrastam com taxas mais altas de adoção em países como EUA, Reino Unido e Escandinávia. O principal desafio identificado foi a falta de oportunidades para trabalhar com IA nos ambientes de trabalho brasileiros. Esse estudo também mostrou que os designers brasileiros veem a IA principalmente como uma ferramenta funcional para otimizar processos como planejamento, pesquisa e testes.

Souza et al. (2019) desenvolveram o *Artificial Intelligence and Mouse Tracking-based User eXperience Tool (AIMT-UXT)*, uma plataforma gratuita e de código aberto para monitorar e analisar interações de usuários em sistemas web com o objetivo de avaliar a UX. O AIMT-UXT integra métodos de inteligência artificial, como lógica fuzzy e algoritmos de agrupamento, para obter *insights* quantitativos a partir de dados de desempenho dos usuários. A ferramenta foi validada em um estudo de caso no site da Receita Federal do Brasileira. Foram monitoradas interações dos usuários, incluindo movimentos do mouse e parâmetros de navegação, e os dados foram analisados com técnicas de IA. Os resultados foram comparados com métodos tradicionais de avaliação de UX, como o System Usability Scale (SUS). As pontuações de UX do AIMT-UXT se alinharam fortemente com as do SUS, indicando que técnicas de inteligência artificial podem ser utilizadas para mapear diferentes ações dos usuários nas respectivas respostas a questionários de UX ou, diretamente, para avaliar a experiência do usuário.

Não foi encontrado nenhum trabalho na literatura que utilize ou relacione o humor ou avaliação de humor na avaliação de UX.

3 Materiais e Métodos

Os principais objetos de estudo utilizados neste trabalho consistem em 146 respostas a questionários realizadas por 35 pessoas que interagiram com um jogo 3D desenvolvido por um dos autores do artigo.

Foram feitos convites para alunos de curso técnico e superior em computação, destes, 35 aceitaram participar do experimento. Os participantes na sua grande maioria possuem idades entre 16 e 18 anos devido ao número de alunos do curso técnico terem a maior participação no teste, a Fig. 1 apresenta a idade dos participantes separadas em quatro grupos: de 16 a 18 anos; de 19 a 21 anos; de 22 a 24 anos; 25 anos ou mais. A proporção entre homens e mulheres é de 66% (23) homens para 34% (12) mulheres. Outra informação importante é que a maioria dos participantes já tiveram algum contato com jogos tridimensionais (83%).

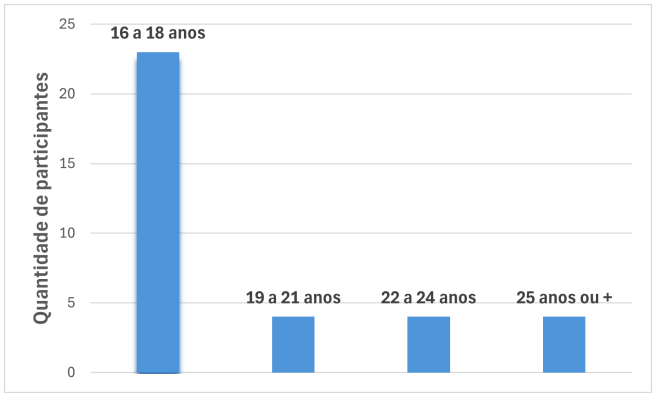

Figura 1: Idade dos participantes

O jogo consiste em o usuário conseguir capturar três pontos que estão espalhados pelo cenário com uma bola que rola conforme a inclinação do terreno, e o usuário controla as inclinações fazendo com que a bola mude de direção, acelere ou desacelere no cenário. Existem três fases com diferentes dificuldades, a fase inicial é a mais fácil, seguida da intermediária e a terceira e última é a fase mais complexa. A Fig. 2 apresenta a imagem inicial da primeira fase. O jogo termina quando o usuário captura os três pontos (vence o jogo) ou quando o usuário sai do cenário/terreno, pois as laterais são abertas e a bola pode cair (perde o jogo).

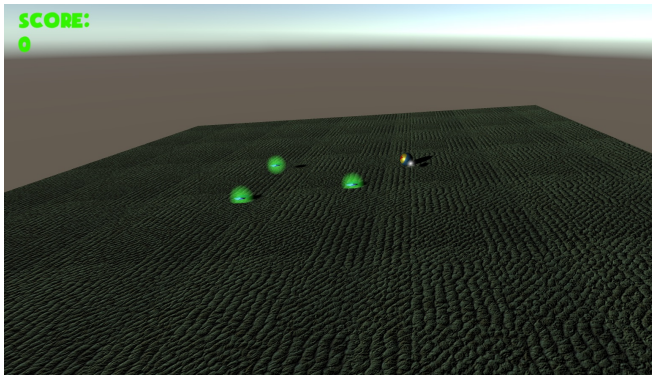

Figura 2: Fase 1 do jogo

Cada pessoa jogou as três fases utilizando o teclado e

de uma a três fases utilizando o *leap motion*, totalizando 146 fases jogadas. Antes de começar a jogar, cada pessoa preencheu o questionário de humor POMS (Viana et al., 2001), apresentado na Tabela 1, que será referenciado neste texto como Humor Antes (HA). Após jogar cada fase, cada um dos participantes preencheu dois questionários de experiência do usuário (UX): o questionário UEQ-S (Laugwitz et al., 2008), ilustrado na Tabela 2 e o questionário PLEX (Lucero et al., 2013; Arrasvuori et al., 2011), ilustrado na Tabela 3. Após terminar de jogar todas as fases, cada pessoa preencheu novamente o questionário de humor (este preenchimento será chamado no presente texto de Humor Depois - HD).

Tabela 1: Perguntas do Questionário de Humor Antes (HA) e Humor Depois (HD)

QUESTIONÁRIO PERFIL DE ESTADOS DE HUMOR				
Humor Antes	**Humor Depois**	Humor		
HA1	HD1	Cansado	o o o o o o o	Descansado
HA2	HD2	Irritado	o o o o o o o	Tranquilo
HA3	HD3	Triste	o o o o o o o	Alegre
HA4	HD4	Mal-humorado	o o o o o o o	Bem-humorado
HA5	HD5	Desanimado	o o o o o o o	Animado
HA6	HD6	Impaciente	o o o o o o o	Paciente
HA7	HD7	Ansioso	o o o o o o o	Calmo

Conforme apresentado na Tabela 1, o questionário de humor é composto por sete perguntas, cada uma considerando uma Escala Diferencial Semântica com sete opções, indicando se seu humor está mais próximo ou mais afastado dos valores opostos de um dado sentimento (por exemplo, cansado ou descansado).

Tabela 2: Perguntas do Questionário UEQ-S

	QUESTIONÁRIO UEQ-S		
UEQS1	Obstrutivo	o o o o o o o	Condutor
UEQS2	Complicado	o o o o o o o	Fácil
UEQS3	Ineficiente	o o o o o o o	Eficiente
UEQS4	Confuso	o o o o o o o	Evidente
UEQS5	Aborrecido	o o o o o o o	Excitante
UEQS6	Desinteressante	o o o o o o o	Interessante
UEQS7	Convencional	o o o o o o o	Original
UEQS8	Comum	o o o o o o o	Vanguardista

O questionário UEQ-S, que corresponde à versão curta do questionário UEQ, é composto por oito questões, conforme apresentado na Tabela 2. Assim como para o questionário de humor, o participante deve assinalar, dentre sete opções, como foi sua experiência em relação a um aspecto (por exemplo, obstrutivo e condutor).

O questionário PLEX utilizado é composto por 14 questões, conforme apresentado na Tabela 3. Diferentemente dos demais questionários apresentados, o usuário deverá

Tabela 3: Perguntas do Questionário Plex

	QUESTIONÁRIO PLEX
Plex1	Cativante
Plex2	Desafiante
Plex3	Competitivo
Plex4	Divertido
Plex5	Frustrante
Plex6	Se sentiu no controle
Plex7	Se sentiu terminando uma tarefa importante
Plex8	Encontrou algo novo ou desconhecido
Plex9	Expressivo
Plex10	Relaxante
Plex11	Simula algo da vida real
Plex12	Adrenalina (derivado do risco e/ou perigo)
Plex13	Faz parte de uma estrutura maior
Plex14	Experiência que precisa da imaginação

dar uma nota, de um a cinco, em relação a sua experiência considerando cada um dos aspectos do questionário. Por exemplo, para o aspecto *cativante*, uma nota baixa indica que a experiência foi pouco cativante e uma nota mais alta indica que a experiência foi mais cativante.

Para padronizar as respostas como valores numéricos inteiros, todas as respostas foram armazenadas como números inteiros, de um a sete para o questionário de humor e UEQ-S e de um a cinco para o questionário PLEX.

A Fig. 3 apresenta o boxplot dos valores das respostas fornecidas pelos usuários para os questionários de experiência de usuário (UEQ-S e PLEX). Conforme esperado, diferentes usuários tiveram, no geral, experiências diferentes no uso de um mesmo sistema. É possível observar pelo boxplot que para algumas perguntas os valores das respostas não variaram muito, por exemplo, para as duas primeiras questões do questionário PLEX, enquanto outras tiveram uma variação bem maior, por exemplo, a primeira pergunta do questionário UEQ-S.

Os métodos do presente trabalho incluem o cálculo e análise das correlações entre os diferentes questionários, de forma a verificar a relação entre diferentes perguntas e entre o humor e a experiência do usuário. A investigação das correlações em si não está diretamente relacionada à questão de pesquisa deste trabalho, porém julgou-se relevante analisar as correlações entre as respostas de forma a evidenciar as potenciais sobreposições entre os questionários.

Adicionalmente, diferentes algoritmos de regressão e classificação foram utilizados para, a partir das respostas de um usuário a um subconjunto dos questionários utilizados, tentar predizer a resposta dos questionários de UX. A solução foi desenvolvida na linguagem de programação Python usando classificadores e regressores da biblioteca *sklearn*. Foram utilizados o classificador *dummy* que escolhe a resposta mais frequente do conjunto de treinamento

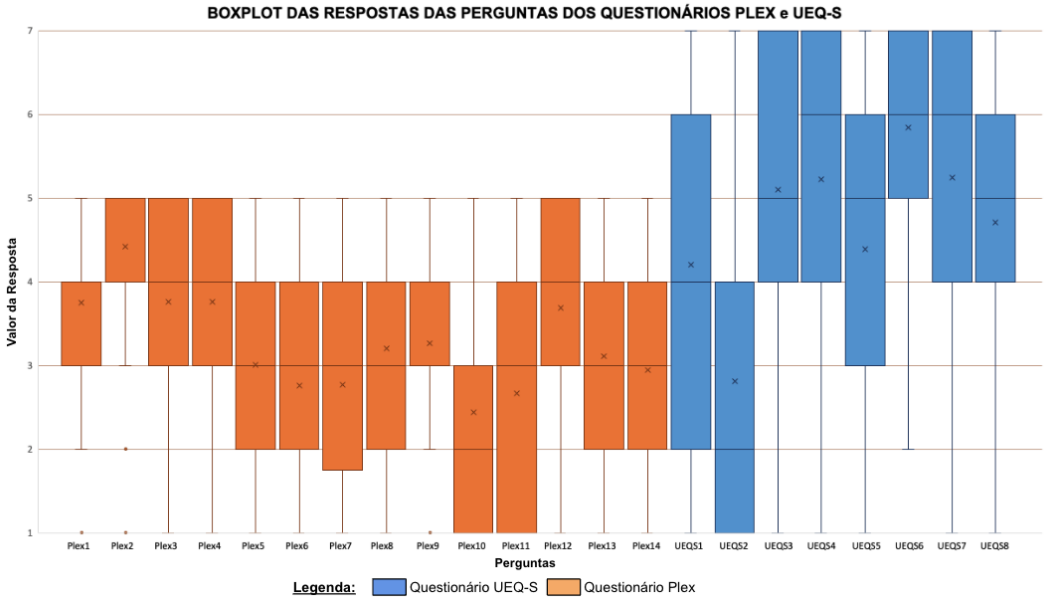

Figura 3: Boxplot das Respostas das Perguntas dos Questionários Plex e UEQ-S

e foi usado como *baseline*; regressão linear; regressão logística; rede neural MLP; Naïve Bayes gaussiano; K-vizinhos mais próximos; árvore de decisão; floresta aleatória e SVM. Todos os classificadores utilizaram valores padrão para seus hiperparâmetros, exceto, o SVC que além de utilizar valores padrão (com *kernel* linear) também foi executado com *kernel* polinomial.

Os resultados foram avaliados considerando-se o erro absoluto médio, calculado por meio da validação cruzada com dez subconjuntos.

4 Resultados

Esta seção apresenta os resultados divididos em duas seções. Na primeira são apresentadas e discutidas as correlações existentes entre os questionários estudados. Já na segunda, é analisado o poder preditivo das respostas de um subconjunto de questionários para prever as respostas dos outros.

4.1 Correlação

Conforme apresentado na Seção 3, as correlações entre os questionários de experiência de usuário e humor foram calculadas. Os resultados foram organizados em tabelas coloridas de forma a facilitar a visualizar das correlações e identificar relações mais significativas entre as perguntas, a fim de verificar algumas das hipóteses iniciais. Nesta subseção, três análises da medida estatística são apresentadas e, para cada uma das tabelas de correlações, há correlações entre as perguntas de um questionário com as do outro, assim como entre perguntas do próprio questionário. Todas as tabelas de correlação seguem o mesmo padrão de cores de fundo, que representam os valores das correlações (C), conforme descrito a seguir:

- preto: correlação perfeita, $C = 1$;
- esverdeado: correlação forte, $0,5 \leq C < 1$ e $-1 < C \leq -0,5$;
- amarelado: correlação média, $0,3 \leq C < 0,5$ e $-0,5 < C \leq -0,3$;
- branco: correlação fraca, $-0,3 < C < 0,3$.

Inicialmente, são analisadas as correlações entre as respostas dos questionários Plex e UEQ-S presente na Tabela 4. Ignorando os resultados duplicados (simétricos), há um total de 253 correlações únicas. Em Plex vs. Plex há 105 correlações totais, destas, seis são fortes (5,71%) e 31 médias (29,52%). Ao se comparar as respostas das questões do questionário Plex em relação aquelas do questionário UEQ-S há 112 correlações resultantes, sendo quatro fortes (3,57%) e 34 médias (30,36%). Já analisando as correlações entre as respostas das perguntas do questionário UEQ-S, há 36 correlações, sendo uma forte (2,78%) e 11 médias (30,56%).

A pergunta Plex8 (*encontrou algo novo ou desconhecido*) foi uma das perguntas com o maior número de correlações fortes sendo elas com a pergunta Plex9 (*expressivo*), Plex13 (*faz parte de uma estrutura maior*) e UEQS8 (*comum/vanguardista*). Com os dados obtidos, pode-se inferir que encontrar algo novo ou desconhecido é algo que os usuários consideram como expressivo, assim como, estar descobrindo algo novo pode contribuir para que com que o usuário se encontre como fazendo parte de algo maior, já que é possível envolver a capacidade de conectar com o novo conhecimento ou experiência. Já a alta correlação desta pergunta do questionário Plex com a pergunta UEQS8 (*comum/vanguardista*) se justifica por serem perguntas que tentam medir, mais ou menos, o mesmo tipo de característica da experiência do usuário.

Adicionalmente, observa-se que as perguntas Plex3 (*competitivo*) e UEQS7 (*convencional/original*) possuem uma relação indireta com Plex8, visto que elas têm cor-

Tabela 4: Correlação dos Questionários Plex vs. UEQ-S

CORRELAÇÃO DOS QUESTIONÁRIOS PLEX VS. UEQ-S																						
y/x	Plex1	Plex2	Plex3	Plex4	Plex5	Plex6	Plex7	Plex8	Plex9	Plex10	Plex11	Plex12	Plex13	Plex14	UEQS1	UEQS2	UEQS3	UEQS4	UEQS5	UEQS6	UEQS7	UEQS8
Plex1	1,000	0,413	0,467	0,567	-0,018	0,092	0,352	0,369	0,484	0,247	0,176	0,404	0,306	0,178	0,082	-0,177	0,204	0,127	0,202	0,565	0,325	0,291
Plex2	0,413	1,000	0,505	0,277	0,214	-0,202	0,143	0,316	0,446	-0,105	-0,019	0,446	0,288	0,221	-0,152	-0,505	0,090	-0,089	-0,135	0,289	0,291	0,224
Plex3	0,467	0,505	1,000	0,326	0,018	-0,046	0,292	0,449	0,622	-0,054	0,143	0,530	0,435	0,282	-0,101	-0,453	0,270	-0,082	-0,103	0,341	0,443	0,406
Plex4	0,567	0,277	0,326	1,000	-0,110	0,235	0,484	0,321	0,398	0,298	0,222	0,385	0,233	0,169	0,175	0,025	0,256	0,170	0,452	0,495	0,254	0,312
Plex5	-0,018	0,214	0,018	-0,110	1,000	-0,231	-0,252	0,022	-0,119	-0,272	-0,101	0,256	0,000	0,032	-0,202	-0,243	-0,466	-0,265	-0,400	-0,145	-0,165	0,007
Plex6	0,092	-0,202	-0,046	0,235	-0,231	1,000	0,449	0,025	0,170	0,395	0,144	-0,005	0,016	0,066	0,452	0,373	0,332	0,371	0,383	0,195	-0,017	-0,015
Plex7	0,352	0,143	0,292	0,484	-0,252	0,449	1,000	0,290	0,471	0,330	0,286	0,319	0,196	0,221	0,235	-0,031	0,362	0,211	0,345	0,277	0,320	0,233
Plex8	0,369	0,316	0,449	0,321	0,022	0,025	0,290	1,000	0,534	0,096	0,288	0,472	0,503	0,313	-0,114	-0,342	0,075	-0,082	-0,043	0,298	0,454	0,541
Plex9	0,484	0,446	0,622	0,398	-0,119	0,170	0,471	0,534	1,000	0,168	0,276	0,449	0,435	0,311	0,037	-0,252	0,410	0,182	0,044	0,448	0,563	0,409
Plex10	0,247	-0,105	-0,054	0,298	-0,272	0,395	0,330	0,096	0,168	1,000	0,411	0,091	0,277	0,172	0,207	0,317	0,103	0,317	0,406	0,169	0,057	0,138
Plex11	0,176	-0,019	0,143	0,222	-0,101	0,144	0,286	0,288	0,276	0,411	1,000	0,180	0,320	0,188	-0,090	-0,081	0,068	0,022	-0,001	0,079	0,228	0,370
Plex12	0,404	0,446	0,530	0,385	0,256	-0,005	0,319	0,472	0,449	0,091	0,180	1,000	0,454	0,284	-0,161	-0,323	0,050	-0,088	-0,045	0,253	0,265	0,370
Plex13	0,306	0,288	0,435	0,233	0,000	0,016	0,196	0,503	0,435	0,277	0,320	0,454	1,000	0,412	-0,087	-0,207	0,205	0,132	0,070	0,283	0,352	0,385
Plex14	0,178	0,221	0,282	0,169	0,032	0,066	0,221	0,313	0,311	0,172	0,188	0,284	0,412	1,000	-0,360	-0,274	0,109	-0,070	0,014	0,244	0,343	0,409
UEQS1	0,082	-0,152	-0,101	0,175	-0,202	0,452	0,235	-0,114	0,037	0,207	-0,090	-0,161	-0,087	-0,360	1,000	0,468	0,242	0,439	0,294	0,139	-0,020	-0,105
UEQS2	-0,177	-0,505	-0,453	0,025	-0,243	0,373	-0,031	-0,342	-0,252	0,317	-0,081	-0,323	-0,207	-0,274	0,468	1,000	0,129	0,389	0,387	-0,038	-0,193	-0,191
UEQS3	0,204	0,090	0,270	0,256	-0,466	0,332	0,362	0,075	0,410	0,103	0,068	0,050	0,205	0,109	0,242	0,129	1,000	0,462	0,369	0,415	0,237	0,052
UEQS4	0,127	-0,089	-0,082	0,170	-0,265	0,371	0,211	-0,082	0,182	0,317	0,022	-0,088	0,132	-0,070	0,439	0,389	0,462	1,000	0,479	0,307	0,012	-0,127
UEQS5	0,202	-0,135	-0,103	0,452	-0,400	0,383	0,345	-0,043	0,044	0,406	-0,001	-0,045	0,070	0,014	0,294	0,387	0,369	0,479	1,000	0,391	0,070	-0,075
UEQS6	0,565	0,289	0,341	0,495	-0,145	0,195	0,277	0,298	0,448	0,169	0,079	0,253	0,283	0,244	0,139	-0,038	0,415	0,307	0,391	1,000	0,432	0,258
UEQS7	0,325	0,291	0,443	0,254	-0,165	-0,017	0,320	0,454	0,563	0,057	0,228	0,265	0,352	0,343	-0,020	-0,193	0,237	0,012	0,070	0,432	1,000	0,634
UEQS8	0,291	0,224	0,406	0,312	0,007	-0,015	0,233	0,541	0,409	0,138	0,370	0,370	0,385	0,409	-0,105	-0,191	0,052	-0,127	-0,075	0,258	0,634	1,000

Legenda: Correlação Perfeita; Correlação Forte; Correlação Média; Correlação Fraca; Questionário UEQ-S; Questionário Plex

relações fortes em comum com Plex9. Nesse sentido, se algo novo/desconhecido é encontrado, resulta em ter o sentimento de fazer parte de uma estrutura maior sendo *comum/vanguardista*, ser expressivo e, consequentemente, em algo *competitivo* e poder ter relação com o sentimento de interagir com algo na dimensão entre o *convencional e original*.

Ao se observar a Tabela 5, referente as correlações dos questionários UEQ-S e Humor Antes (HA), sem os resultados duplicados, há um total de 120 correlações. Em HA vs. HA há um total de 28 correlações, sendo cinco fortes (17,86%) e 13 médias (46,43%). E, em UEQ vs. HA há 56 correlações, mas não há nenhuma correlação forte ou média. Por um lado, infere-se que a grande presença de fortes e médias correlações com as perguntas de HA com ela mesma se dá pelo fato dos sentimentos dos usuários estarem interligados, pois neste caso, todas as correlações fortes estão conectadas. Nesse cenário, se o indivíduo se encontra tranquilo (HA2), ele estará alegre (HA3), animado (HA5) e bem-humorado (HA4) que resulta em estar calmo (HA7). Entretanto, se o usuário se encontra irritado (HA2), é mais provável que ele esteja triste (HA3), desanimado (HA5), mal-humorado (HA4) e, consequentemente, ansioso (HA7).

Por outro lado, exceto as correlações fortes de HA vs. HA e as de UEQ vs. UEQ que já foram citadas, ao correlacionar os dois questionários, não foi possível encontrar nenhuma correlação significativa para a análise. Esta informação é bastante interessante, pois como mencionado nos trabalhos correlatos, questionários de humor não são, normalmente, aplicados juntamente com questionários de experiência de usuário. A falta de correlações médias e fortes poderia servir de base para a não aplicação do questionário de humor. Porém, como será visto na próxima subseção, as respostas aos questionários de humor contribuem significativamente para a predição dos valores das respostas dos questionários de experiência de usuário. Isto indica que apesar de individualmente não haver correlações significativas entre perguntas individuais, o conjunto das perguntas do questionário de humor fornece informações que auxiliam na predição das respostas dos questionários de experiência do usuário.

A última tabela de correlações apresentada (Tabela 6) está associada às correlações dos questionários HA vs. Plex. Sem os resultados duplicados, há um total de 231 correlações únicas. Apesar dessa tabela conter correlações fortes, elas não se referem a HA vs Plex, visto que não há nenhuma correlação forte entre estes questionários e apenas cinco correlações médias (5,10%).

Das cinco correlações médias, três ocorrem com a pergunta Plex3 (*competitivo*) e envolvem as perguntas HA3, HA4 e HA5, respectivamente, *Triste/Alegre*, *Mal-humorado/Bem-humorado* e *Desanimado/Animado*. Estas três perguntas possuem correlações fortes entre si. As demais correlações médias entre Plex e Humor Antes ocorrem entre Plex9 (*Expressivo*) e HA5 (*Desanimado/Animado*), sendo esta uma correlação positiva; e uma correlação negativa entre Plex10 (*Relaxante*) e HA3 (*Triste/Alegre*) indicando que quanto mais triste o usuário estiver, menos relaxante a experiência será (e vice-versa).

Assim como ocorreu para o questionário UEQ-S, para o questionário Plex não foram verificadas correlações fortes entre o Humor Antes e as respostas sobre a experiência do usuário. Porém, conforme será apresentado na próxima seção, o conjunto das respostas de humor será significativo para a predição das respostas sobre a experiência do usuário.

Para as correlações analisadas, conclui-se que, em geral, as correlações mais altas foram entre perguntas do próprio questionário, principalmente as perguntas do questionário de humor com ele mesmo, apesar dessa não apresentar fortes correlações com os outros questionários. No entanto, apesar de não serem muitas, há correlações fortes interessantes entre o questionário Plex e UEQ-S, confirmando

Tabela 5: Correlação dos Questionários UEQ-S vs. HA

CORRELAÇÃO DOS QUESTIONÁRIOS UEQ-S VS. HUMOR ANTES															
y/x	UEQS1	UEQS2	UEQS3	UEQS4	UEQS5	UEQS6	UEQS7	UEQS8	HA1	HA2	HA3	HA4	HA5	HA6	HA7
UEQS1	1,000	0,468	0,242	0,439	0,294	0,139	-0,020	-0,105	-0,175	-0,035	-0,109	0,052	0,012	-0,083	0,005
UEQS2	0,468	1,000	0,129	0,389	0,387	-0,038	-0,193	-0,191	-0,221	-0,064	-0,156	-0,194	-0,133	-0,140	-0,087
UEQS3	0,242	0,129	1,000	0,462	0,369	0,415	0,237	0,052	0,088	0,095	0,139	0,132	0,263	0,073	0,114
UEQS4	0,439	0,389	0,462	1,000	0,479	0,307	0,012	-0,127	-0,131	0,134	-0,084	-0,068	0,031	0,021	0,112
UEQS5	0,294	0,387	0,369	0,479	1,000	0,391	0,070	-0,075	0,014	0,233	-0,036	-0,085	0,090	0,178	0,060
UEQS6	0,139	-0,038	0,415	0,307	0,391	1,000	0,432	0,258	0,063	0,081	-0,001	-0,015	0,011	0,115	0,040
UEQS7	-0,020	-0,193	0,237	0,012	0,070	0,432	1,000	0,634	0,184	-0,023	0,046	0,022	0,167	0,132	0,143
UEQS8	-0,105	-0,191	0,052	-0,127	-0,075	0,258	0,634	1,000	0,058	0,108	0,110	0,068	0,087	0,026	0,171
HA1	-0,175	-0,221	0,088	-0,131	0,014	0,063	0,184	0,058	1,000	0,214	0,428	0,427	0,493	0,445	0,369
HA2	-0,035	-0,064	0,095	0,134	0,233	0,081	-0,023	0,108	0,214	1,000	0,533	0,470	0,276	0,398	0,264
HA3	-0,109	-0,156	0,139	-0,084	-0,036	-0,001	0,046	0,110	0,428	0,533	1,000	0,744	0,574	0,431	0,325
HA4	0,052	-0,194	0,132	-0,068	-0,085	-0,015	0,022	0,068	0,427	0,470	0,744	1,000	0,693	0,407	0,526
HA5	0,012	-0,133	0,263	0,031	0,090	0,011	0,167	0,087	0,493	0,276	0,574	0,693	1,000	0,477	0,483
HA6	-0,083	-0,140	0,073	0,021	0,178	0,115	0,132	0,026	0,445	0,398	0,431	0,407	0,477	1,000	0,474
HA7	0,005	-0,087	0,114	0,112	0,060	0,040	0,143	0,171	0,369	0,264	0,325	0,526	0,483	0,474	1,000

Legenda: Correlação Perfeita; Correlação Forte; Correlação Média; Correlação Fraca; Questionário UEQ-S; Questionário Humor Antes

certa sobreposição entre estes questionários.

4.2 Modelos Preditivos

Apesar do Questionário de Humor não ter apresentado altas correlações com os outros dois questionários, ao estudar os resultados dos modelos preditivos construídos a partir dos diferentes questionários, o humor contribuiu significativamente para a redução do erro na maioria das perguntas tanto para o questionário UEQ-S como para o questionário Plex.

No caso do questionário UEQ-S, considerando os melhores modelos produzidos, somente os modelos para duas das oito perguntas (25%) não fizeram uso do Questionário de Humor. Em outras palavras, 75% dos modelos fizeram uso do Plex + Humor Antes ou Plex + Humor Antes + Humor Depois (*Tudo*). Nesse sentido, o humor estar incluído no melhor modelo produzido para a respectiva pergunta é um forte indicativo da sua importância para a análise. A pergunta UEQS6 (*desinteressante/interessante*), por exemplo, teve uma redução de 48,14% do erro em relação ao *baseline*, esta perguntas apresentou o menor erro absoluto médio (0,598) ao utilizar o questionário de humor em conjunto com o Plex em sua construção. Apesar da previsão para a pergunta UEQS4 (*confuso/evidente*) apresentar o maior erro (1,199), o modelo construído usando todos os atributos conseguiu reduzir o valor do erro em 32,37% em relação ao *baseline*. Destaca-se que mesmo o maior erro, o qual é inferior a 1,2, foi considerado bastante satisfatório, pois muitos usuários têm dúvidas de pelo menos um ponto ao preencher estes questionários e os erros dos melhores modelos para predição da resposta a cada pergunta do questionário UEQ-S indicam que o sistema foi capaz de predizer as respostas dos usuários errando, em média, de, em torno de 0,6 a 1,2 pontos. Todos estes resultados são apresentados na Tabela 7. Em termos de maior redução do percentual do erro em relação ao erro base, destaca-se o modelo produzido para predizer a resposta à pergunta UEQS1.

Ao analisar a coluna "Redução Percentual de Erro em Relação ao Baseline" da Tabela 7, destaca-se o resultado para a pergunta UEQS1 (*obstrutivo/condutor*). Apesar do erro para esta pergunta não ser um dos mais baixos, ela apresenta a melhor redução percentual de erro em relação ao *baseline*, visto que o erro base era de 2,068 pontos e passou para 0,978 ao utilizar todos os atributos em um modelo de Floresta Aleatória (RF). Assim, reduzindo o erro em 52,71%. Por outro lado, a menor redução percentual foi de 31,36%, que foi o caso da pergunta UEQS5 (*aborrecido/excitante*) em que o questionário de Humor não contribuiu para a redução do erro já que o melhor modelo foi produzido utilizando apenas as respostas do questionário Plex.

Para analisar a contribuição média de cada conjunto de atributos (respostas de cada questionário) em relação ao poder preditivo, foram calculadas as médias do erro de todos os algoritmos para cada conjunto de atributos e para algumas de suas combinações. Para a previsão das respostas às perguntas do questionário UEQ-S, é possível afirmar que apesar da média dos questionários de Humor Antes e Humor Depois possuírem uma média de erro maior do que a do uso do questionário Plex, ao serem juntados, a média passa a ser menor que 1, sendo a menor média de erro atingida e, consequentemente, com a maior redução percentual de erro em relação ao *baseline*, conforme apresentado na Tabela 8. Estes resultados confirmam a importância do uso de questionários de humor para a análise da experiência do usuário.

Por mais que o questionário UEQ-S tenha apresentado resultados considerados bastante positivos (com os melhores modelos, na média, produzindo erros abaixo de um ponto), o questionário Plex apresentou resultados ainda melhores.

A Tabela 9 apresenta um resumo dos melhores resultados para a previsão de cada uma das perguntas de Plex. Somente duas das 14 perguntas utilizaram apenas as perguntas do UEQ-S como conjunto de atributos e duas utilizaram UEQ-S mais Humor Antes. As dez perguntas restantes, isto é, 71,42% das perguntas do questionário Plex utilizaram o conjunto de atributos "Tudo" (UEQ-S + HA

Tabela 6: Correlação dos Questionários HA vs. Plex

CORRELAÇÃO DOS QUESTIONÁRIOS PLEX VS. HUMOR ANTES																					
y/x	Plex1	Plex2	Plex3	Plex4	Plex5	Plex6	Plex7	Plex8	Plex9	Plex10	Plex11	Plex12	Plex13	Plex14	HA1	HA2	HA3	HA4	HA5	HA6	HA7
Plex1	1,000	0,413	0,467	0,567	-0,018	0,092	0,352	0,369	0,484	0,247	0,176	0,404	0,306	0,178	0,135	0,022	0,093	0,130	0,204	0,228	0,176
Plex2	0,413	1,000	0,505	0,277	0,214	-0,202	0,143	0,316	0,446	-0,105	-0,019	0,446	0,288	0,221	0,074	-0,003	0,081	0,148	0,224	0,173	0,045
Plex3	0,467	0,505	1,000	0,326	0,018	-0,046	0,292	0,449	0,622	-0,054	0,143	0,530	0,435	0,282	0,177	-0,014	0,332	0,319	0,399	0,288	0,200
Plex4	0,567	0,277	0,326	1,000	-0,110	0,235	0,484	0,321	0,398	0,298	0,222	0,385	0,233	0,169	-0,041	0,080	0,002	0,030	0,176	0,154	0,049
Plex5	-0,018	0,214	0,018	-0,110	1,000	-0,231	-0,252	0,022	-0,119	-0,272	-0,101	0,256	0,000	0,032	-0,109	-0,137	-0,107	0,000	-0,193	-0,157	-0,157
Plex6	0,092	-0,202	-0,046	0,235	-0,231	1,000	0,449	0,025	0,170	0,395	0,144	-0,005	0,016	0,066	-0,103	0,006	-0,174	-0,079	-0,095	0,004	-0,078
Plex7	0,352	0,143	0,292	0,484	-0,252	0,449	1,000	0,290	0,471	0,330	0,286	0,319	0,196	0,221	0,062	-0,039	-0,044	0,053	0,239	0,143	0,147
Plex8	0,369	0,316	0,449	0,321	0,022	0,025	0,290	1,000	0,534	0,096	0,288	0,472	0,503	0,313	0,187	0,168	0,151	0,203	0,235	0,168	0,246
Plex9	0,484	0,446	0,622	0,398	-0,119	0,170	0,471	0,534	1,000	0,168	0,276	0,449	0,435	0,311	0,164	-0,023	0,122	0,133	0,303	0,250	0,196
Plex10	0,247	-0,105	-0,054	0,298	-0,272	0,395	0,330	0,096	0,168	1,000	0,411	0,091	0,277	0,172	-0,093	-0,012	-0,305	-0,228	-0,095	0,144	0,129
Plex11	0,176	-0,019	0,143	0,222	-0,101	0,144	0,286	0,288	0,276	0,411	1,000	0,180	0,320	0,188	0,024	-0,043	-0,130	-0,064	0,056	-0,077	0,095
Plex12	0,404	0,446	0,530	0,385	0,256	-0,005	0,319	0,472	0,449	0,091	0,180	1,000	0,454	0,284	-0,003	0,109	0,061	0,034	0,095	0,231	0,011
Plex13	0,306	0,288	0,435	0,233	0,000	0,016	0,196	0,503	0,435	0,277	0,320	0,454	1,000	0,412	0,071	0,125	0,009	-0,084	-0,002	0,146	0,128
Plex14	0,178	0,221	0,282	0,169	0,032	0,066	0,221	0,313	0,311	0,172	0,188	0,284	0,412	1,000	0,000	0,104	-0,058	-0,166	-0,189	-0,060	-0,090
HA1	0,135	0,074	0,177	-0,041	-0,109	-0,103	0,062	0,187	0,164	-0,093	0,024	-0,003	0,071	0,000	1,000	0,214	0,428	0,427	0,493	0,445	0,369
HA2	0,022	-0,003	-0,014	0,080	-0,137	0,006	-0,039	0,168	-0,023	-0,012	-0,043	0,109	0,125	0,104	0,214	1,000	0,533	0,470	0,276	0,398	0,264
HA3	0,093	0,081	0,332	0,002	-0,107	-0,174	-0,044	0,151	0,122	-0,305	-0,130	0,061	0,009	-0,058	0,428	0,533	1,000	0,744	0,574	0,431	0,325
HA4	0,130	0,148	0,319	0,030	0,000	-0,079	0,053	0,203	0,133	-0,228	-0,064	0,034	-0,084	-0,166	0,427	0,470	0,744	1,000	0,693	0,407	0,526
HA5	0,204	0,224	0,399	0,176	-0,193	-0,095	0,239	0,235	0,303	-0,095	0,056	0,095	-0,002	-0,189	0,493	0,276	0,574	0,693	1,000	0,477	0,483
HA6	0,228	0,173	0,288	0,154	-0,157	0,004	0,143	0,168	0,250	0,144	-0,077	0,231	0,146	-0,060	0,445	0,398	0,431	0,407	0,477	1,000	0,474
HA7	0,176	0,045	0,200	0,049	-0,157	-0,078	0,147	0,246	0,196	0,129	0,095	0,011	0,128	-0,090	0,369	0,264	0,325	0,526	0,483	0,474	1,000

Legenda: Correlação Perfeita; Correlação Forte; Correlação Média; Correlação Fraca; Questionário Plex; Questionário Humor Antes

Tabela 7: Resumo dos Melhores Resultados para a Previsão das Respostas do Questionário UEQ-S

RESUMO DOS MELHORES RESULTADOS PARA A PREVISÃO DAS RESPOSTAS DO UEQ-S					
PERGUNTAS	ATRIBUTOS UTILIZADOS	ERRO	MELHOR MODELO	DIFERENÇA	REDUÇÃO PERCENTUAL DE ERRO EM RELAÇÃO AO BASELINE
UEQS1	Tudo	0,978	RF	1,09	52,71%
UEQS2	Plex	1,132	RegLin	0,673	37,29%
UEQS3	Tudo	0,93	RF	0,97	51,05%
UEQS4	Tudo	1,199	SVM	0,574	32,37%
UEQS5	Plex	1,057	SVM	0,483	31,36%
UEQS6	Plex + Humor Antes	0,598	SVM	0,555	48,14%
UEQS7	Plex + Humor Antes	0,842	RF	0,43	33,81%
UEQS8	Tudo	0,796	RF	0,717	47,39%
Média Total		0,942		0,687	41,76%

Legenda: Baixo — Alto

Tabela 8: Média do Erro para Previsão das Respostas do Questionário UEQ-S

MÉDIAS DO ERRO PARA PREVISÃO DO UEQ-S			
CONJUNTO DE ATRIBUTOS	MÉDIA	DESVIO	REDUÇÃO PERCENTUAL DE ERRO EM RELAÇÃO AO BASELINE
Dummy (baseline)	1,6280	31,47%	0,00%
Humor Antes	1,1786	25,35%	27,60%
Humor Depois	1,1955	24,12%	26,57%
Plex	1,0623	18,66%	34,75%
Plex + Humor Antes	0,9859	21,55%	39,44%
Tudo	0,9464	17,73%	41,87%
Média Total	1,1661	5,04%	28,37%

Legenda: Baixo — Alto

+ HD). Apesar da pergunta Plex2 (*desafiante*) possuir o menor erro e o melhor conjunto de atributos utilizado para a previsão da resposta ter utilizado somente o UEQ-S, esta possui a segunda menor redução de erro percentual. Isto acontece porque, como é possível observar no Boxplot (Fig. 3), ela é uma das que possui a menor variação em termos de valores de resposta, sendo uma pergunta "mais fácil" para os modelos.

A pergunta Plex8 apresentou o maior erro em Plex, sendo ele de 0,857, ainda abaixo de 1,0 que era o nosso parâmetro inicial para uma ótima previsão de resposta (o erro do *baseline* para esta pergunta é de 1,152).

Ainda a partir da Tabela 9 é possível observar que a pergunta Plex11 (*simula algo da vida real*) possui a maior variação das respostas como mostra o Boxplot (Fig. 3) e apresentou a maior redução percentual de erro em relação ao *baseline*. Apesar de ser considerada uma pergunta difícil para previsão, ao utilizar todos os atributos foi possível diminuir o seu erro de 1,67 para 0,519 (redução do erro em quase 70%). No entanto, a pergunta Plex6 (*se sentiu no controle*) apresentou a menor redução de erro (18,55%), utilizando UEQ-S + HA. Apesar de não estar implícito como HA pode estar ajudando a diminuir o erro desta pergunta, é possível inferir que quando o usuário tem um melhor humor, ele sente que tem mais controle no jogo e vice-versa.

As médias de erro para previsão do Plex são bastante satisfatórias, visto que a média total dos conjuntos de atributos, incluindo o *Dummy* é de 0,7749 e, sem o *Dummy* é ainda melhor, sendo uma média de 0,7003. Vale lembrar que o modelo *Dummy* utilizado como *baseline* considera a resposta mais frequente, no conjunto de treinamento, para a respectiva pergunta, não utilizando nenhuma informação adicional. Além disso, HA e HD possuem resultados melhores que aqueles alcançados pelo uso do UEQ-S, como pode ser observado na Tabela 10.

A Tabela 11 consolida os resultados produzidos, colocando em uma única tabela os valores de erro dos melhores modelos para para a previsão das respostas dos questionários UEQ-S e Plex, considerando os diferentes conjuntos de atributos.

Em geral, desconsiderando o *Dummy*, por um lado, as médias de erro dos conjuntos de atributos para a previsão de UEQ-S (Tabela 8) variam de 0,9464 a 1,955. Ao se comparar o uso dos questionários Plex, HA e HD para a

Tabela 9: Resumo dos Melhores Resultados para Previsão das Respostas do Questionário Plex

RESUMO DOS MELHORES RESULTADOS PARA A PREVISÃO DAS RESPOSTAS DO PLEX					
PERGUNTAS	ATRIBUTOS UTILIZADOS	ERRO	MELHOR MODELO	DIFERENÇA	REDUÇÃO PERCENTUAL DE ERRO EM RELAÇÃO AO BASELINE
Plex1	Tudo	0,521	RF	0,198	27,54%
Plex2	UEQ-S	0,458	RF	0,117	20,35%
Plex3	Tudo	0,519	RF	0,72	58,11%
Plex4	UEQ-S	0,528	RF	0,245	31,69%
Plex5	Tudo	0,793	SVM	0,566	41,65%
Plex6	UEQ-S + Humor Antes	0,72	RF	0,164	18,55%
Plex7	Tudo	0,703	KNN	1,065	60,24%
Plex8	Tudo	0,857	RegLin	0,295	25,61%
Plex9	Tudo	0,488	RF	0,409	45,60%
Plex10	Tudo	0,659	SVM	0,784	54,33%
Plex11	Tudo	0,519	RF	1,151	68,92%
Plex12	UEQ-S + Humor Antes	0,696	RF	0,618	47,03%
Plex13	Humor Antes	0,644	RF	0,266	29,23%
Plex14	Humor Antes	0,587	SVM_Poly	0,784	57,18%
Média Total		0,621		0,527	41,86%

Legenda: Baixo Alto

Tabela 10: Média do Erro para Previsão das Respostas do Questionário Plex

MÉDIAS DO ERRO PARA PREVISÃO DO PLEX			
CONJUNTO DE ATRIBUTOS	MÉDIA	DESVIO	REDUÇÃO PERCENTUAL DE ERRO EM RELAÇÃO AO BASELINE
Dummy (baseline)	1,1481	36,34%	0,00%
Humor Antes	0,6989	14,48%	39,13%
Humor Depois	0,6906	12,86%	39,85%
UEQ-S	0,8186	19,91%	28,70%
UEQ-S + Humor Antes	0,6600	13,01%	42,52%
Tudo	0,6332	11,47%	44,85%
Média Total	0,7749	9,45%	32,51%

Legenda: Baixo Alto

predição das respostas do UEQ-S, individualmente, Plex apresenta a menor média de erro (1,0623) e, dentre os três, é o que mais reduz o erro (34,75%). Porém, o uso de todos os conjuntos de atributos, "Tudo" (Plex + HA + HD) apresenta a menor média (0,9464) com a melhor redução percentual de erro em relação ao *baseline* (41,87%) para a previsão do UEQ-S, seguido do Plex + HA (0,9859).

Já para a previsão das respostas do questionário Plex (Tabela 10), as médias de erro variaram de 0,6332 a 0,8186. Ao comparar o uso dos diferentes conjuntos de atributos (UEQ-S, HA e HD), individualmente, o questionário HD apresenta a menor média de erro (0,6906) com redução percentual de erro em relação ao *baseline* de (39,85%). Ao analisar os demais conjuntos de atributos, "Tudo"(UEQ-S + HA + HD) possui a menor média de erro (0,6332) com redução percentual de 44,85%.

Por fim, ao analisar as melhores previsões para Plex e UEQ-S de acordo com o conjunto de atributos, verifica-se que ao comparar a média de cada conjunto de atributos para os dois questionários, Plex apresenta maiores reduções nos erros e médias mais baixas que UEQ-S em todos os resultados, principalmente com o conjunto de atributos "Tudo". A média para esse conjunto em Plex é de 0,6332 enquanto para UEQ-S é de 0,9464. Pode-se inferir que isso acontece uma vez que o modelo consegue fazer um maior aproveitamento do questionário de Humor, o que é evidente em Humor Antes + UEQ-S (para prever Plex) é de 0,66 comparado a Humor Antes + Plex (para prever UEQ-S) é de 0,9859. Adicionalmente, o fato do questionário Plex ter mais perguntas e apenas cinco valores para cada resposta, pode significar que são perguntas mais específicas e por isso mais fáceis de responder pelo usuário, isto justificaria o fato dos valores de erro do *baseline* para Plex serem menores do que para o UEQ-S.

Tabela 11: Melhores Previsões para os Questionários Plex e UEQ-S de acordo com o Conjunto de Atributos

MELHORES PREVISÕES PARA PLEX E UEQ-S DE ACORDO COM O CONJUNTO DE ATRIBUTOS				
CONJUNTO DE ATRIBUTOS	QUESTIONÁRIO	MÉDIA	DESVIO	REDUÇÃO DO ERRO
Dummy (baseline)	PLEX	1,1481	36,34%	0,00%
	UEQ-S	1,628	31,47%	0,00%
Humor Antes	PLEX	0,6989	14,48%	39,13%
Humor Antes	UEQ-S	1,1786	25,35%	27,60%
Humor Depois	PLEX	0,6906	12,86%	39,85%
Humor Depois	UEQ-S	1,1955	24,12%	26,57%
UEQ-S	PLEX	0,8186	19,91%	28,70%
Plex	UEQ-S	1,0623	18,66%	34,75%
Humor Antes + UEQ-S	PLEX	0,6600	13,01%	42,52%
Humor Antes + Plex	UEQ-S	0,9859	21,55%	39,44%
Tudo	PLEX	0,6332	11,47%	44,85%
Tudo	UEQ-S	0,9464	17,73%	41,87%
Média Total	PLEX	0,7749	9,94%	32,51%
	UEQ-S	1,1661	5,04%	28,37%

Legenda: Questionário UEQ-S Questionário Plex

Conforme apresentado, o questionário de Humor pode contribuir significativamente para a previsão de respostas dos questionários que avaliam a experiência do usuário, especialmente para o questionário Plex.

5 Conclusões

A adoção de diferentes sistemas de informação bem como a eficácia e satisfação com seu uso estão diretamente relacionadas com a experiência dos usuários durante a utilização destes sistemas. Conforme destacado na literatura correlata, há diversos questionários padronizadas de UX disponíveis que, apesar de possuírem certa sobreposição, costumam focar em diferentes aspectos da experiência do usuário. Adicionalmente, questionários de humor não costumam ser aplicados em conjunto com esses questionários.

O presente trabalho teve por objetivo analisar e comparar as respostas a dois questionários de UX, UEQ-S e PLEX, bem como de um questionário de humor respondido antes e após a conclusão das interações. Foram analisadas respostas de 146 iterações de 35 usuários com um jogo 3D. Inicialmente, foi analisada a correlação entre as respostas do questionário de humor e dos dois questionários de UX frequentemente na literatura correlata. Diversas correlações médias ou fortes foram encontradas entre as respostas dos questionários UEQ-S e PLEX, evidenciando certa sobreposição conceitual ou de experiência entre suas questões. Não foram identificadas correlações médias ou fortes entre as respostas do questionário de humor e do UEQ-S e apenas algumas correções médias entre o questionário de humor e o PLEX. Foram encontradas ainda algumas correlações fortes internas, entre as perguntas de um mesmo questionário, para os três questionários utilizados mas, analisando a natureza destas perguntas e dos questionários, não se pode interpretar essa correlação como uma

redundância entre as perguntas, mas tão somente como uma indicação de aspectos desta experiência em particular que, nela, estiveram bem correlacionados, o que não necessariamente ocorreria em outras experiências.

Adicionalmente, as respostas dos diferentes questionários foram utilizadas como características preditivas para as respostas dos dois questionários de UX. Observou-se uma boa capacidade preditiva para as respostas do questionário PLEX, com uma média de erro, para os melhores modelos, abaixo de 0,65 (em uma escala de cinco pontos) e uma capacidade satisfatória para a predição ou inferência das respostas do questionário UEQ-S, com uma média de erro inferior a 1 ponto (em uma escala de sete pontos). Esta boa capacidade preditiva responde de forma positiva à pergunta de pesquisa do presente trabalho e mostra que, usando modelos como os discutidos neste trabalho, é possível comparar com precisão razoável medidas de UX obtidas por meio de instrumentos diferentes e até mesmo agregá-las para avaliar a experiência de um maior número de usuários.

Apesar das correlações entre as respostas dos questionários de UX e dos questionários de humor (antes e depois) serem, em geral, fracas, com poucas correlações médias, estas respostas se mostraram bastante importantes como características preditivas dos dois questionários, evidenciando a influência do humor e, em especial, do humor antes da interação, para determinar a experiência do usuário com o sistema. A análise da literatura mostra poucos estudos que considerem esta influência do humor na UX, de forma que esta evidência se mostra como uma linha de investigação interessante para trabalhos futuros.

Como trabalhos futuros adicionais, pretende-se estender o estudo considerando mais usuários e mais jogos ou aplicativos. Adicionalmente, pretende-se explorar características para além dos questionários para analisar automaticamente a experiência do usuário, por exemplo, o uso das emoções extraídas a partir da expressão facial do usuário durante sua experiência.

Referências

Araújo, T., Carneiro, N., Miranda, B., Santos, C. e Meiguins, B. (2015). Aplicações android de realidade aumentada em arquitetura extensível, flexível e adaptável, *Anais do XI Simpósio Brasileiro de Sistemas de Informação*, SBC, Porto Alegre, RS, Brasil, pp. 63–70. https://doi.org/10.5753/sbsi.2015.5802.

Arrasvuori, J., Boberg, M., Holopainen, J., Korhonen, H., Lucero, A. e Montola, M. (2011). Applying the plex framework in designing for playfulness, *Proceedings of the 2011 Conference on Designing Pleasurable Products and Interfaces*, ACM, New York, NY, USA, pp. 1–8. https://doi.org/10.1145/2347504.2347531.

Bertão, R. A. e Joo, J. (2021). Artificial intelligence in UX/UI design: a survey on current adoption and [future] practices, *Blucher Design Proceedings* **9**(5): 404 – 413. http://dx.doi.org/10.5151/ead2021-123.

Bevan, N. (2009). What is the difference between the purpose of usability and user experience evaluation methods, *Proceedings of the Workshop UXEM*, Vol. 9, Springer, Uppsala, Sweden, pp. 1–4. https://api.semanticscholar.org/CorpusID:15182858.

Boscarioli, C., Araújo, R. e Maciel, R. (2017). *I GranDSI-BR–Grand Research Challenges in Information Systems in Brazil 2016-2026*, SBC, Porto Alegre, RS, Brasil. https://doi.org/10.5753/sbc.2884.0.

Choma, J., Zaina, L. A. M., Amaral, A. M. e Oliveira, P. (2014). Identificação das necessidades de interação dos usuários em sistemas erp: proposta de uma metodologia investigativa, *Anais do X Simpósio Brasileiro de Sistemas de Informação*, SBC, Porto Alegre, RS, Brasil, pp. 411–422. https://doi.org/10.5753/sbsi.2014.6132.

da Costa, S. L., Neto, V. V. G., de Oliveira, J. L. e dos Reis Calçado, B. (2014). User interface stereotypes: A model-based approach for information systems user interfaces, *Anais do X Simpósio Brasileiro de Sistemas de Informação*, SBC, Porto Alegre, RS, Brasil, pp. 113–124. https://doi.org/10.5753/sbsi.2014.6106.

Dupont, L., Pallot, M., Christmann, O. e Richir, S. (2018). A universal framework for systemizing the evaluation of immersive and collaborative performance, *Proceedings of the Virtual Reality International Conference - Laval Virtual*, VRIC '18, Association for Computing Machinery, New York, NY, USA. http://doi.org/10.1145/3234253.3234306.

Ferreira, T. M., Costella, F. L., Zanetti, A. B., da Silva, S. E., Zanatta, A. L. e De Marchi, A. C. B. (2019). Crowdrec: A prototype recomendation system for crowdsourcing platforms using google venture design: Google venture design sprint, *Proceedings of the XV Brazilian Symposium on Information Systems*, SBC, Aracaju, Brazil, pp. 26:1–26:8. https://doi.org/10.1145/3330204.3330235.

Goodhue, D. L. e Thompson, R. L. (1995). Task-technology fit and individual performance, *MIS quarterly* **19**(2): 213–236. https://doi.org/10.2307/249689.

Laugwitz, B., Held, T. e Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire, *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society*, Springer, Graz, Austria, pp. 63–76. https://doi.org/10.1007/978-3-540-89350-9_6.

LaViola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D. e Poupyrev, I. P. (2017). *3D user interfaces: theory and practice*, Addison-Wesley Professional, Boston, MA, USA. =https://books.google.com.br/books?id=ilUyjwEACAAJ.

Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P. e Kort, J. (2009). Understanding, scoping and defining user experience: a survey approach, *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, Boston, MA, USA, pp. 719–728. https://doi.org/10.1145/1518701.1518813.

Lucero, A., Holopainen, J., Ollila, E., Suomela, R. e Karapanos, E. (2013). The playful experiences (plex) framework as a guide for expert evaluation, *Proceedings of the 6th international conference on designing pleasurable products*

and interfaces, ACM, Newcastle upon Tyne, United Kingdom, pp. 221–230. https://doi.org/10.1145/2513506.2513530.

Lyytinen, K. e Grover, V. (2017). Management misinformation systems: A time to revisit?, *Journal of the Association for Information Systems* **18**(3): 2. https://doi.org/10.17705/1jais.00453.

Quek, A. e See, J. (2015). Obscura: A mobile game with camera based mechanics, *2015 Game Physics and Mechanics International Conference (GAMEPEC)*, IEEE, Langkawi, Malaysia, pp. 21–25. https://doi.org/10.1109/GAMEPEC.2015.7331850.

Schroeder, W., Martin, K., Lorensen, B. e Kitware, I. (2006). *The Visualization Toolkit: An Object-oriented Approach to 3D Graphics*, Kitware, Hoboken, New Jersey, USA. https://books.google.com.br/books?id=rx4vPwAACAAJ.

Souza, K. E. S., Seruffo, M. C. R., De Mello, H. D., Souza, D. D. S. e Vellasco, M. M. B. R. (2019). User experience evaluation using mouse tracking and artificial intelligence, *IEEE Access* **7**: 96506–96515. http://dx.doi.org/10.1109/ACCESS.2019.2927860.

Szajna, B. (1996). Empirical evaluation of the revised technology acceptance model, *Management science* **42**(1): 85–92. https://www.jstor.org/stable/2633017.

Veriscimo, E. d. S., Bernardes Junior, J. L. e Digiampietri, L. A. (2020). Evaluating user experience in 3d interaction: a systematic review, *XVI Brazilian Symposium on Information Systems*, SBC, São Bernardo do Campo, Brasil, pp. 1–8. https://doi.org/10.1145/3411564.341164.

Vessey, I. e Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition, *Information Systems Research* **2**(1): 63–84. https://doi.org/10.1287/isre.2.1.63.

Viana, M. F., Almeida, P. e Santos, R. C. (2001). Adaptação portuguesa da versão reduzida do perfil de estados de humor–poms, *Análise Psicológica* **19**(1): 77–92. http://hdl.handle.net/10400.12/1874.

Yangguang, L., Yue, L. e Xiaodong, W. (2014). Multiplayer collaborative training system based on mobile ar innovative interaction technology, *2014 International Conference on Virtual Reality and Visualization*, IEEE, Shenyang, China, pp. 81–85. https://doi.org/10.1109/ICVRV.2014.66.