

Propostas de Trabalho Final

Oi Irene,

Muito legal a proposta A, pode começar a trabalhar nela. Qq problema estaremos à disposição.

— [Alexandre Adalardo de Oliveira](#) 2017/06/03 13:38

Muito obrigada Alê!! Já estou trabalhando nela! Abraços. — [Irene Delval](#) 2017/06/05 08:58

PROPOSTA A: Ontogenizar

Função: ontogenizar () = Organiza pastas com dados longitudinais cronologicamente.

Descrição: A função “ontogenizar” é uma função simples para a classificação de informação coletada e armazenada em diversos formatos de vídeo digital.

Justificativa: Quando podemos dizer que um comportamento é geneticamente determinado? Existem comportamentos puramente inatos ou puramente aprendidos? Biólogos, etólogos e estudantes do comportamento em geral, sabemos que não há uma resposta clara para essas perguntas pois existe uma interação genes x ambiente. Contudo, determinar o momento da aparição de um comportamento qualquer (i.e. brincadeira, alimentação independente da mãe, agressividade, canto [nas espécies de aves cantoras], migração de grupo, comportamento sexual) ao longo da ontogenia pode nos ajudar à obter respostas para estas questões. É por isso que, pese ao alto custo e esforço requerido, existem cada vez mais pesquisas longitudinais. Ainda, a democratização e barateamento das filmadoras digitais fez com que seja cada vez mais comum que os pesquisadores interessados no comportamento (animal ou humano), filmem esse para uma análise posterior. Assim, estão sendo geradas enormes quantidades de dados em formato de vídeo digital que contém valiosa informação, e que poderiam, inclusive, ser reaproveitados para outras pesquisas desde que estejam corretamente classificados. Portanto, a função “ontogenizar” irá organizar os dados de vídeo introduzidos pelo usuário em pastas classificadas temporalmente na escala escolhida por ele (dias, semanas, meses, ou anos) e fornecer como resultado, ademais das pastas organizadas, um data.frame contendo toda essa informação.

Argumentos:

- ind= nome do sujeito, vetor carater
- db = date_of_birth, em fomato YYYY-MM-DD, vetor data
- in_folder=pasta contendo os dados do sujeito a ser classificado
- out_folder= endereço da pasta que será criada
- unit= “days”, “weeks”, “months”, “year”
- period= fração da unidade (por exemplo, 8 semana, 5 mês, 1 ano)

Resultado: Pastas organizadas conforme o solicitado e um data.frame com os dados agrupados.

Uso: ontogenizar = function (ind= "RITA", db=2012-12-07, in_folder="C:/Users/Irene/Desktop/Videos", out_folder="C:/Users/Irene/Desktop/Output ", unit="weeks", period= 8)

Pacotes necessários: "zoo", "media.info", "ffmpeg" (ainda tenho dúvidas ao respeito)

Oi Irene, tudo bem?

Achei sua proposta A muito interessante, parece bem generalizável para pesquisadores de comportamento animal e muito útil, mas não entendi muito bem como são os dados! Como não sou da área, você poderia me esclarecer melhor como seria a entrada da função? Uma pasta contendo vários arquivos em formato .mp4, por exemplo? Ela teria vários arquivos, todos do mesmo tipo, ou com vários formatos diferentes de vídeo? Assim que eu entender melhor, posso opinar mais.

Qualquer dúvida, estou à disposição. Poste sempre por aqui e me avise por email que postou para eu voltar aqui, ok?

Abraços! — [Rodolfo Liporoni Dias](#) 2017/06/02 10:07

Oi Rodolfo, muito obrigada pela rápida resposta! Eu tb acho que a função vai ser útil para qualquer pesquisador com dados longitudinais (mas para isso tem que funcionar, kkkk). Os dados de entrada seriam pastas, cada uma de um mesmo indivíduo conhecido, com arquivos de vídeo:



Acho que também iria funcionar com arquivos de imagem, ou seja, com fotos, mas não sei se tudo formato, ainda não testei. Por enquanto provei quatro formatos de vídeo (mts, mpg, mod e mp4) e todos eles foram lidos pelo comando file.info, que é o que vou usar para extrair os metadados com a data da filmagem no R, e assim poder calcular a idade do indivíduo. A saída deve ser uma coisa assim:



Por enquanto estou investigando como fazer para os arquivos escolhidos serem copiados em outro diretório... Isso deve ser simples para usuários habituais do R, mas não para usuárias "babies" que nem eu, kkkkk. Tenho outras ideias de aprimoramento como opções que apaguem os arquivos nas pastas de origem se o usuário quer resgatar espaço no HD, ou como mudar o nome dos novos

arquivos copiados nas pastas organizadas (i.e. em vez de o arquivo se chamar 000060.MTS ele seria Sofia8month_1, seguindo a ordem consecutiva dos vídeos para essa idade). Porém, antes de tudo isso, queria fazer a função funcionar. 😊



Muito obrigada pelos comentários! Qualquer sugestão é bem-vinda!

Abraços! — [Irene Delval](#) 2017/06/03 10:37

Oi Irene!

Muito obrigado pelos esclarecimentos, já entendi melhor e sua proposta A me parece mais interessante ainda! Bom, o Alê já avaliou e já disse para fazer esta né, mas vou colocar aqui algumas sugestões que pensei.

Primeiro, o argumento “ind” talvez não seja obrigatório né? A pasta original já contém o nome do indivíduo, certo? Se sim, pense num jeito da função ler o nome da pasta original e entender o nome que o usuário já tinha dado (para ele não ter que repetir dentro desse primeiro argumento). Caso a pasta original não tenha o nome do indivíduo, aí você precisa dessa informação e poderia vir no argumento “ind” mesmo, mas se ele fosse deixado em branco seria legal se a função entendesse isso e rodasse com o nome da pasta original (qualquer que fosse). O que acha?

Segundo, pelo visto você terá que trabalhar muito com formatos de data e hora no R (vulgo, objetos das classes “POSIXlt” e “POSIXct”) para fazer os cálculos de idade. Um bom pacote para lidar com essas classes de objeto e que eu uso muito é o pacote “anytime”. Caso esteja apanhando com as funções básicas do R para manipular esses objetos, recomendo explorar se fica mais fácil com as funções do anytime! 😊

Terceiro, para deixar a função mais turbinada, acho que seria legal deixar o usuário com o máximo de liberdade possível para escolher o destino das pastas originais e novas. Você poderia colocar mais argumentos que deixassem o usuário escolher se ele quer: 1) apenas criar pastas novas organizadas sem mexer nas pastas originais, ou 2) se ele quer criar essas pastas novas e apagar as originais, ou 3) se ele quer criar essas pastas novas em outro local, e mover as antigas para um outro local ou manter no local original, coisas assim, entende?

E que tal um argumento que permite criar uma cópia do output da função também num serviço de nuvem (como Dropbox, GoogleDrive, etc.) e não só dentro do computador

do usuário? 😊

Como disse, são só sugestões. Manda bala na proposta A, e qualquer dúvida ou dificuldade, só falar!

Abraços! — [Rodolfo Liporoni Dias](#) 2017/06/05 14:46

Oi Rodolfo,

Adorei todas as suas sugestões! Muito obrigada por dedicar um tempo para pensar aprimoramentos para minha função. Acho que todas as sugestões são muito boas. Vamos ver uma por uma:

1) **Argumento “ind” opcional:** nos meus arquivos brutos a pasta de origem tem a data de nascimento, nome do indivíduo e nome da mãe (p.e. 2015-03-12_SOFIA-Samantha), então vou tentar descobrir como se faz para extrair essa informação de um jeito seletivo, mas acho um pouco difícil... Porém, como vc mesmo disse, nem todos os usuários terão as pastas brutas organizadas do mesmo jeito, então é bom ter esse “ind” entre os argumentos.

2) **Pacote “anytime”:** na verdade eu já estava apanhando com o R básico, por isso que teve que abaixar o pacote “zoo”, pois com ele consegui calcular a idade em meses e anos (no R só calculava em dias e semanas). Acabei de abaixar esse pacote “anytime”. Vou testar ele. Valeu!

3) **Opções de copiado:** gostei muito das 3 opções. Veremos se eu consigo ir em todas elas mas vou tentar. Contudo, no meu trabalho, eu nunca apago nada, tenho medo de perder informação, kkkk. Mas é bom ter a opção de apagar.

4) **Copiar na nuvem:** achei muuuuito massa também, vou pesquisar e ver se sei fazer. Se não, vou ter que te pedir ajuda ;)

Depois, mais uma dúvida: queria também poder acessar à duração da filmagem, mas o R não faz isso “nativamente” (assim falamos no stackoverflow). Tentei instalar dois programas (media.info e ffmpeg), mas nenhum dos dois deu certo para acessar essa informação e poder criar no dataframe de saída, uma coluna com a duração do vídeo. São programas mesmo, não pacotes, pois não existem pacotes do R para fazer isso, o que eu acho muito esquisito. Se vc souber como fazer isso me diga, por favor.

A sua ajuda está sendo ótima!

Obrigada! Abraços! — [Irene Delval](#) 2017/06/07 20:23

Olá Irene, tudo bem?

Que bom que as sugestões foram úteis! Acho que você não terá muitas dificuldades para implementar não.

Quanto a sua última dúvida, confesso que eu nunca trabalhei com isso no R, então tem coisas que você terá que descobrir sozinha mesmo. Podemos sempre perguntar para outros monitores também, mas por hora acho que você precisa investigar mais se não consegue resolver sozinha. Como isso é mais um bônus da função, podemos deixar para resolver mais para frente, ok?

E provavelmente não existem pacotes no R para fazer isso porque é uma tarefa aparentemente bem complexa. Dei uma pesquisada aqui e o pessoal geralmente não sabe dizer porque o Windows às vezes não coloca a duração do vídeo quando você pede mais informações do arquivo com o botão direito. No Mac acontece a mesma coisa, tem arquivos que às vezes você não acessa essa informação facilmente, precisa abrir o vídeo e ver mesmo. Então taaaalvez não seja um problema fácil com ou sem R. Ah, e fique atenta se sua função vai funcionar igual independente do sistema operacional do usuário! Só um lembrete, deixe para se preocupar com isso mais para a frente.

No momento, acho que você já está bem encaminhada. Continue usando o StackOverFlow para essas dúvidas mais específicas. Você já viu [este](#) post? Parece que resolveram o problema. Investigue o passo a passo e se consegue resolver pelo mesmo caminho. Se não conseguir, me avise quando a função já estiver mais encorpada e vamos investigar.

Qualquer outra dúvida, só perguntar.

Abraços! — [Rodolfo Liporoni Dias](#) 2017/06/08 11:33

Bom dia Rodolfo!

Foi esse post exatamente o que eu segui, e não funcionou... Não sei se eu não fiz direto ou se é que esse jeito não funciona mesmo.

De qualquer jeito, estou enfrentando outros problemas que tem a ver com trabalhar com datas: não consigo obter a diferença exata em meses de duas listas de datas. Nenhuma das coisas que tentei funcionaaaa, socorro!! kkkkk

Tudo isso faz com que esteja mais focada em isso que em obter a duração.

Será que podemos marcar um encontro?

Abraços! — [Irene](#) 2017/06/13 10:25

PROPOSTA B: PhD Nobel Prize

Função: phdnobel () = calculador de chances de passar na defesa do seu PhD ou MSc.

Descrição: A função “phdnobel” irá calcular as chances de ser aprovado na defesa, em base às informações fornecidas pelo usuário. Melhor não malgastar seu tempo se a função dá um resultado perto de 0.

Justificativa: a vida é muito curta para andar gastando o tempo. Essa função ajudará aos estudantes de pós-graduação a determinar se vale a pena continuar investindo o tempo na tese, ou se pelo contrário, seria melhor parar e pensar em outra coisa. Ainda, a interpretação do resultado pode servir para relaxar frente a ansiedade da defesa, ou como um pequeno “tapa” a tempo para acordar antes de que seja tarde. O usuário irá a introduzir variáveis referidas às características da sua pós (p.e. horas de campo, número de artigos publicado, horas de análise de dados, área de concentração, nota CAPES, participação em congressos, fez sanduiche?) e obterá um resultado de 0 a 10 sobre as possibilidades de passar na defesa. Curiosamente, nem todas variáveis aumentam as possibilidades de passar.

Argumentos:

- pos= vetor caráter. Título que o usuário pretende (MSc ou PhD)
- lab= vetor lógico, TRUE, FALSE. Teve laboratório?
- field= vetor lógico, TRUE, FALSE. Houve trabalho de campo?
- p_paper= vetor numeric. Número de papers publicados, pode ser NULL
- CAPES= vetor numérico. Coinceito CAPES do curso.
- deadline= vetor data em formato YYYY-MM-DD. Data de entrega.

Resultado: um valor numérico entre 0 e 10.

Uso: phdnobel = (pos= phd, lab=T, field=F, p.papers= 2, capes= 5, deadline= 2018-08-01)

Pacotes necessários: ainda não sei

Bibliografia:

Wright, T., & Cochrane, R. (2000). Factors Influencing Successful Submission of PhD Theses. *Studies in Higher Education*, 25(2), 181-195. <https://doi.org/10.1080/713696139>

Philips, E., & Pugh, D. (2010). *How to get a PhD: A handbook for students and their supervisors*. McGraw-Hill Education (UK)

Oi de novo!

A proposta B é divertida, mas não deu para entender como você irá calcular esse índice de 0 a 10! Poderia explicar

melhor (em texto mesmo) quais seriam as etapas para se calculá-lo? Assim eu posso avaliar melhor se é uma proposta viável ou não. Eu dei uma olhada rápida nas referências que você colocou, mas não achei nada explicitamente calculando algo do tipo. Também fiquei em dúvida (e curioso!) em como você vai ponderar cada critério para definir o valor final do índice...

E só para avisar, não esperamos que os alunos usem nenhum pacote para fazer suas funções (pode até usar, mas não é obrigatório, ok? Só o pacote básico do R já serve para fazer muita coisa).

Aguardo mais informações.

Abraços! — [Rodolfo Liporoni Dias](#) 2017/06/02 10:22

Oi, Rodolfo!! Então, usei as referências somente como guia para ver quais das variáveis usar. Por exemplo, na pesquisa de Wright e Cochrane (2000) não foi significativo o efeito do sexo para o sucesso de apresentar a tese em quatro anos, assim, eu decidi não colocar essa variável dentre os argumentos. Depois, o que eu pensei, (muito rapidamente, sendo sincera) é inventar os pesos para as categorias de forma que no final das contas some 100, e depois posso dividir entre 10, para obter esse índice. Desse jeito, o argumento "pós" fornece 50 pontos ao estudante de MSc e 40 ao de PhD (pois pelo menos o 50% dos alunos de mestrado passam na defesa). O fato de ter feito campo ou laboratório também aumenta o valor em 10 pontos, pois, embora aumenta a dificuldade do curso e diminui o tempo da análise e escrita, consideramos que o aluno tem que ser um bom gestor do tempo. Cada artigo publicado soma 10 pontos no doutorado e 15 no mestrado, já que é bem mais raro publicar no mestrado, enquanto no doutorado é obrigatória na maioria dos cursos. O argumento CAPES é avaliado do jeito seguinte: um conceito CAPES de 4 pontos ou menos não soma nenhum valor à pontuação total; nota 5 e 6 acrescentam esse valor nas suas chances de passar enquanto nota 7 diminui a chance pela dificuldade. Por último, o argumento do tempo para o deadline é mais complicado: pode multiplicar o total obtido por uma fração maior quanto mais baixo esse total, e menos dias para o depósito (i.e., se você obteve entre 100 e 90 pontos e faltam 3 meses ou menos suas chances de passar são multiplicadas por 1, de modo que a chance de aprovar não diminui. Porém, se vc obteve entre 80 e 70 pontos e faltam 3 meses ou menos sua pontuação é multiplicada por 0.9, é assim em diante). Acho que assim fica meio confuso, mas vejamos a tabela:



Tenho que dizer que são valores "made in Irene", pensados sem rigor científico nenhum, mas com o objetivo de ter medidas para podermos criar e testar uma função.

De qualquer jeito, eu também prefiro o plano A, pois, embora desafiador, vejo ele bem mais útil e com muuuitas mais chances de ser usada no futuro por qualquer pessoa, não só por mim.



Muito obrigada! — *Irene Delval* 2017/06/03 10:40

Oi de novo!

Entendi como seria o cálculo do índice. Divertido, mas um tanto especulativo né? Fiquei curioso para saber qual seria o meu resultado! 😊

Só para constar: a proposta B também está muito boa, parabéns! Mas a A está muito melhor!

Boa sorte com o plano A! — *Rodolfo Liporoni Dias* 2017/06/05 15:04


Como você mesmo disse, não ia dar para confiar nesse índice, então melhor nem pensar nos possíveis resultados. Eu tb não quis testar, kkkk. Vou com tudo na proposta A!!



Obrigada pelas dicas e os ânimos, isso já é uma ajuda!

Abraços

From:
<http://ecor.ib.usp.br/> - **ecoR**

Permanent link:
http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2017:alunos:trabalho_final:irenedelval:trabalho_final 

Last update: **2020/08/12 06:04**