

Exercício de Reprodução de Experimento

Artigo Original: A Statistical Comparative Study of Different Similarity Measures of Consensus in Group Decision Making

Tiago Lucas Pereira Clementino

14 de julho de 2018

Resumo

Um dos objetivos essenciais em problemas de decisão entre especialistas (do inglês Group Decision Models, GDM) é a obtenção de níveis aceitáveis de consenso. Uma solução não consensual pode não ter validade prática. Medir a similaridade entre as soluções apontadas por cada especialista é essencial para trilhar o caminho até o consenso. Decidir como fazer isto é importante para o sucesso neste processo. Em (Chiclana et al. 2013) o autor realizou uma comparação entre cinco dos principais (mais usados) modelos de medição da distância aplicados ao consenso (Manhattan, Euclidian, Dice, Cosine e Jacard). Pretendo questionar/validar este estudo, o reproduzindo e analisando utilizando um modelo de regressão linear aliado a uma análise estatística por intervalos de confiança. Metodologia diferente daquela adotada pelo autor.

1 Introdução

Soluções voltadas ao suporte à decisão em grupo trabalham para ajudar a escolher a melhor opção a partir de um conjunto de opções viáveis. Decidir qual a melhor opção frequentemente representa um problema complexo, sobretudo quando envolve o consenso entre um grupo de indivíduos. Com isto, medir com precisão a diferença entre múltiplas opções de solução é uma importante questão projetual. Para tanto, é preciso escolher um modelo adequado para mensurar tal distância. O que se segue é a reprodução de um estudo comparativo que busca evidências que apontem os melhores modelos de distância para problemas de decisão em consenso.

GDM (do inglês, *Group Decision Making*) são

modelos onde um grupo de indivíduos mobiliza seus conhecimentos a certa e um problema em busca da melhor solução dentre um conjunto de critérios e alternativas possíveis. Neste contexto, CRP (*Consensus Reaching Process*) descreve um processo onde indivíduos engajados na argumentação alteram suas opiniões visando elevar indicadores de consenso e satisfação coletivos (Cabrerizo, Alonso, and Herrera-Viedma 2009; Cabrerizo et al. 2010). Para se atingir indicativos aceitáveis de consenso é preciso quantificar e comparar as preferências de cada indivíduo dentre pares de alternativas viáveis. Em tal comparação são aplicadas estratégias matemáticas de medição de distância. Diante disto, as cinco estratégias mais aplicadas na literatura (segundo o autor original) são as distâncias de Manhattan, Euclidiana, Dice, Jacard e Cosseno (Cha 2007; Chen et al. 1995; M. M. Deza and Deza 2009; Wang, De Baets, and Kerre 1995).

Quando mencionamos a palavra consenso nos remetemos imediatamente ao conceito de unanimidade. Porém, nas últimas décadas tem crescido o entendimento de que uma interpretação mais flexível da ideia de consenso refletiria melhor as relações entre os seres humanos. Daí o conceito de Soft Consensus, onde o consenso é representado por um índice entre zero e um.

Soft Consensus pode ser aplicado representando o consenso por quantificadores Fuzzy tais como “most”, “leasthalf” ou “asmany” (margens de

maioria), ao invés de apenas “todos ou nenhum”(unanimidade). Tal consenso ainda pode ser mensurado usando três estratégias (níveis): comparação entre pares de alternativas, comparação entre todas as relações ligadas à uma alternativa específica ou entre todas as relações globalmente. Todos os conceitos apresentados até aqui serão tratados com mais detalhes na seção seguinte, caso esta introdução já tenha sido suficiente ou o leitor tenha familiaridade com o assunto, pode avançar da introdução para a seção 3.

Aqui foi feito a reprodução de um estudo comparativo originalmente reportado em (Chiclana et al. 2013). Os autores do estudo original fizeram uma comparação estatística entre os cinco principais modelos de medição de distância utilizadas no suporte ao consenso. Testes estatísticos pareados utilizando o método não-paramétrico de Wilcoxon foram reportados. Os resultados visavam responder se a aplicação de um ou outro modelo de distância exercia alguma influência no resultado do processo até o consenso. Dentre os fatores do experimento estão o quantificador Fuzzy aplicado, o nível de consenso aplicado, o total de indivíduos, o total de alternativas e o modelo de distância (variável de controle). O experimento original será descrito em detalhes na seção 3.

Para esta reprodução, procurei seguir detalhadamente todos os passos do estudo original, porém aplicando o meu próprio modelo operacional (disponível em github.com/tiagoclementino/GDMR/tree/ref18¹) e métodos es-

tatísticos diferentes. A análise procedeu com base em duas abordagens. Uma utilizando intervalos de confiança com *bootstrap* (por observar a distribuição gaussiana nos dados) para reproduzir exatamente o mesmo quadro do artigo original (que utilizou o teste de Wilcoxon) e outra utilizou regressão linear para criar um modelo de correlação entre todos os fatores. Os detalhes desta reprodução estão descritos na seção 3, na seção 4 estão os resultados comparados e a seção 5 traz uma sumarização e algumas considerações finais.

2 Contexto

Para contextualizar e tornar este documento alto explicativo, vamos esclarecer alguns conceitos que nortearam o experimento original e a reprodução.

2.1 GMD (*Group Decision Making*)

GMD é o processo que visa atingir um julgamento comum a um conjunto de múltiplos participantes para um problema de decisão, o qual é definido na forma de um conjunto de possíveis soluções. De acordo com (Kacprzyk 1986), GMD pode ser formalmente definido do seguinte modo:

- Um grupo de participantes, $E = \{e_1, e_2, \dots, e_m\}$, cada um deles com um certo grau de conhecimento referente ao domínio do problema.
- Um problema de decisão definido por n possíveis alternativas, o qual é denotado por

completo

¹Neste repositório você encontrará também com os scripts da análise, um guia de reprodução e até o texto deste relatório

$$X = \{x_1, x_2, \dots, x_n\}.$$

- O processo onde os participantes tentam chegar a uma solução comum.

Para cada participante $e_i \in E$, construa a função $\mu_{p_i} X \times X \implies D$, onde D é o domínio de representação da informação (em $[0, 1]$) e $\mu_{p_i}(x_l, x_k) = (p_{lk})^i(l, k \in \{1, 2, \dots, n\})$ denota o grau de preferência da alternativa x_l frente à x_k em D e $p_{ij} + p_{ji} = 1, \forall i, j \in \{1, \dots, n\}$. Assim, as preferências dos participantes para todas as alternativas em X podem ser descritas como uma matriz $P^i = (p_{lk}^i)_{(n \times n)}$ (Kacprzyk 1986).

2.2 CRP (*Consensus Reaching Process*)

O consenso nada mais é do que o resultado da discussão em si, fruto de um processo recorrente e iterativo onde um grau de consenso é checado a cada “rodada”.

A cada rodada, o processo de suporte ao consenso calcula dois parâmetros (Cabrerizo, Pérez, and Herrera-Viedma 2010; Herrera, Herrera-Viedma, and others 1996; Herrera, Herrera-Viedma, and Verdegay 1997): (i) o grau de consenso, que vai definir o quão próximo estamos de um limiar estipulado como aceitável (ii) uma medida de proximidade entre as preferências dos participantes, o que dará suporte ao processo rumo ao consenso. Estas medidas são utilizadas para mensurar o estágio do consenso em três diferentes níveis de relações de preferência Fuzzy: preferências entre pares de alternativas (*pairs of alternatives* no artigo original), preferências envolvendo todas as relações ligadas a uma alternativa específica (*alternatives*

level no artigo original) e uma medida geral (*relation level* no artigo original). Isto nos permite identificar quais indivíduos estão mais próximos da solução consensual, ou a respeito de quais alternativas o consenso será mais difícil.

O cálculo do índice de consenso envolve, necessariamente, alguma métrica de similaridade. Uma função de similaridade deve ser utilizada para mensurar ambos, o índice de consenso e uma matriz de proximidade. Proximidade é calculada mensurando a similaridade entre as preferências (pares de alternativas), para cada participante r , e uma agregação geral em uma matriz de similaridade de acordo com $SM^r = (sm_{ij}^r)$, onde

$$sm_{ij}^r = s(p_{ij}^r, p_{ij})$$

Sendo p_{ij}^r a medida da preferência do participante r entre as alternativas $i, j \in X$, e $p_{ij} = \phi(p_{ij}^1, \dots, p_{ij}^{(r-1)}, p_{ij}^{(r+1)}, \dots, p_{ij}^m)$, onde ϕ é uma função que descreve a maioria segundo um quantificador Fuzzy (no experimento foi usado “most”, “leathalf” e “asmany”), e $s : [0, 1]^{(m-1)} \times [0, 1]^{(m-1)} \implies [0, 1]$ é a função de similaridade. Esta matriz de similaridade é útil tanto no cálculo do consenso, quanto em prover o feedback necessário para que cada participante possa modificar sua opinião corretamente rumo ao consenso.

3 Reprodução do Experimento

Muitos dos elementos descritivos foram omitidos aqui na reprodução por serem idênticos aos do experimento original.

3.1 Objetivos

Formalmente, o objetivo desta comparação pode ser definido no formato GQM como analisar as funções de distância aplicadas ao suporte ao consenso com a intenção de compará-las a respeito de seus graus de consenso finais do ponto de vista dos projetistas de sistemas de suporte ao consenso no contexto dos resultados de simulações envolvendo múltiplos fatores.

3.1.1 Design do experimento

Trata-se de uma reprodução do um experimento. A maior parte deste trabalho replica o experimento original.

Para testar a hipótese nula (a mesma do experimento original), doze (12) conjuntos de relações de preferência (pares de alternativas) foram gerados aleatoriamente para cada combinação possível de participantes ($E_m : m \in M = \{4, 6, 8, 10, 12\}$) e alternativas ($X_n : n \in N = \{4, 6, 8\}$). As diferentes funções de distância foram aplicadas, por sua vez, para mensurar consenso nos três níveis possíveis (pares de alternativas, alternativas e relação geral), usando os três diferentes operadores quantificadores guiados OWA, tal como no experimento sob reprodução. Com isto, um modelo operacional de um sistema de CRP foi aplicado à este ambiente simulado e executado um total de $12 * |M| * |N| * 3 * 3 * 5 = 8100$ vezes.

No gráfico 1 abaixo, temos uma visão da dispersão dos dados gerados a partir de todas as simulações. Estes dados estão dispostos em um plano

que coloca a distância utilizada em função do grau de consenso. Note que já é possível perceber uma nítida diferença no posicionamento dos percentis.

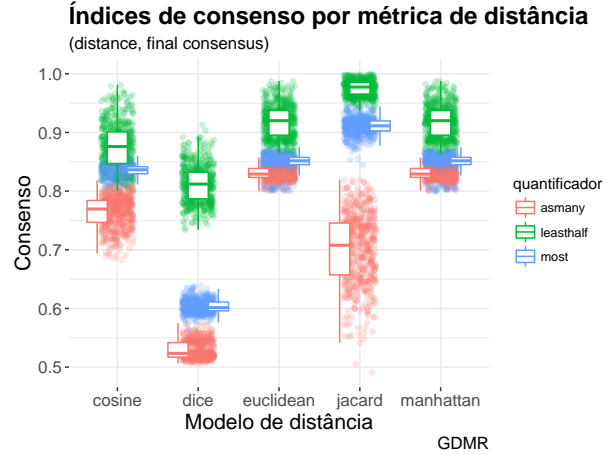


Gráfico 1: Gráfico de dispersão que nos dá uma visão geral dos dados tomando a distância em função do grau de consenso.

Todas as funções de distância foram analisadas aos pares, $d_i \times d_j : i = \{1, 2, 3, 4\}, j = \{1, 2, 3, 4, 5\}$ ($\frac{5*5-5}{2} = 10$ testes). Assim, foram realizadas um total de $|M| * |N| * 3 * 3 * 10 = 1350$ análises.

3.2 Execução da reprodução

3.2.1 Questões de pesquisa

“The application of the Manhattan, Euclidean, Cosine, Dice and Jaccard distance functions in GDM problems do not produce significant differences in the measurement of consensus” (Chiclana et al. 2013), traduzindo: “A aplicação das funções de distância de Manhattan, Euclidiana, Cosseno, Dice e Jaccard em problemas GDM não produzem diferenças significativas ao mensurar consenso.”

3.2.2 Fatores e variáveis de resposta

As variáveis independentes (e também fatores) utilizados no experimento original são repetidas aqui, além de mais duas utilizadas apenas nos questionamentos finais (em função das restrições de prazo):

- **Função de distância utilizada:** Uma dentre cinco funções tomadas como predominantes na literatura (variável de controle).
- **Total de participantes:** Número de participantes utilizados no processo de consenso (simulação) $E_m : m = \{4, 6, 8, 10, 12\}$.
- **Total de alternativas viáveis:** Número de alternativas de solução viáveis $X_n : n = \{4, 6, 8\}$.
- **Quantificadores Fuzzy:** Além das funções de distância, quantificadores de maioria Fuzzy podem ter influência no resultado. Aqui foram usados "most", "asmany" e "least-half".
- **Nível de consenso:** São relações de preferência entre alternativas associadas a maiores Fuzzy. São eles: (1) nível de pares de alternativas (compara cada alternativa entre si), (2) nível de alternativas (compara cada alternativa com as demais) e (3) nível da relação (comparação geral).
- **Completeness*:** Grandeza $\alpha \in [0, 1]$ que vai de 0, onde a chance de um participante apontar sua preferência é aleatória, a 1, onde todos os participantes sempre apontam todas as suas preferências.
- **Coerência*:** Grandeza $\delta \in [0, 1]$ que de-

fine coerência nas preferências dos participantes com base em simetria e comutatividade. Como o autor original não trata desta variável, tomamos como 0 (sem verificação de coerência) para nossa reprodução.

Para as variáveis de resposta, incluí apenas mais uma, também apenas nos questionamentos finais (em função das restrições de prazo):

- **Grau de consenso:** Saldo final do processo de consenso que descreve o quão de acordo os participantes estão com o resulta: $cr \in [0, 1]$
- **Diferença entre índice inicial e final de consenso*:** Necessário para medir se o modelo de distância é excessivamente otimista ou pessimista.

*Variáveis de resposta e fatores que julgo necessários, mas que não são tratados no experimento original. Em função das restrições de prazo do curso de FPCC (fundamento de Pesquisa em Ciência da Computação), apenas reproduzi o experimento do autor original, deixando estes ajustes para trabalhos futuros.

3.2.3 Níveis dos fatores

Os níveis dos fatores e das variáveis de resposta estão definidos de acordo com as tabelas 2 e 3.

Fator	Níveis
Função de distância	{Manhatan, Euclidean, Cosine, Dice, Jaccard}
Participantes	{4, 6, 8, 10, 12}
Alternativas	{4, 6, 8}
Quantificadores	{most, leasthalf, as-many}
Níveis	{1, 2, 3}
Compleitude	[0,1]
Coerência	[0,1]

Tabela 2: Níveis dos fatores utilizados na reprodução

Variável	Níveis
Grau de consenso	[0,1]
Diferença entre consenso inicial e final	[0,1]

Tabela 3: Níveis das variáveis de resposta utilizadas (ou apenas propostas) na reprodução

3.2.4 Unidade experimentais

As unidades experimentais são os resultados da comparação entre cada duas funções de distância distintas dadas cada combinação diferente dos fatores.

3.2.5 Execução

Tal como no experimento original, execução do experimento envolve apenas dois passos:

1. Execução de simulações de consenso. Uma para cada combinação de fatores.
2. Análise estatística dos resultados.

3.2.5.1 Simulações

Para a simulação, utilizei a plataforma R e a IDE RStudio para criar meu modelo operacional. O código fonte utilizou como base um software livre de apoio ao consenso chamado GDMR(Group Decision Making R)(Ureña et al. 2016). Daí um simulador simples foi desenvolvido embutido no próprio código do modelo.

Para validação, me ative a apenas ao modelo operacional, por utilizar modelos conceituais já amplamente utilizados e validados na literatura (Cha 2007; Chen et al. 1995; M. M. Deza and Deza 2009; Wang, De Baets, and Kerre 1995; Chiclana et al. 2004; Chiclana et al. 2007; Dubois and Prade 1985; Fodor and Roubens, n.d.; Herrera, Herrera-Viedma, and Chiclana 2003; Klir and Folger 1988; Xu and Da 2003; Yager 1988; Zhou et al. 2008; Zhou et al. 2011; Kacprzyk 1986). Neste contexto, o gráfico abaixo (gráfico 2), onde o consenso inicial (aleatório) é representado no eixo “Y” e o consenso final no eixo “X”, demonstra o sucesso do meu modelo pela checagem $X - Y < 0$ sendo igual a “FALSE” em todos os casos. Observando também que todos os pontos estão abaixo da diagonal (indica que $x \geq y$, sempre). Com isto temos uma evidência da validade do modelo. A eficiência poderia ser validada por comparação a outro modelo já comprovadamente eficiente, porém isto não é viável (pelo menos não nos prazos da disciplina de FPCC). Em alternativa, utilizamos novamente a validação visual. Perceba no gráfico 3 que a maioria das soluções alcançam um nível aceitável de consenso de 0.8.

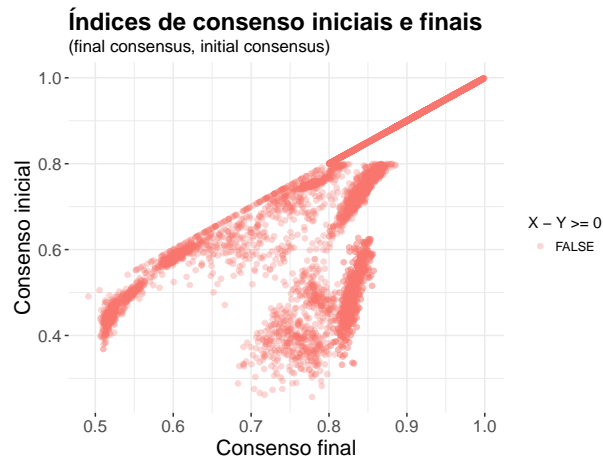


Gráfico 2: Demonstração de que o consenso inicial é sempre inferior ao final.

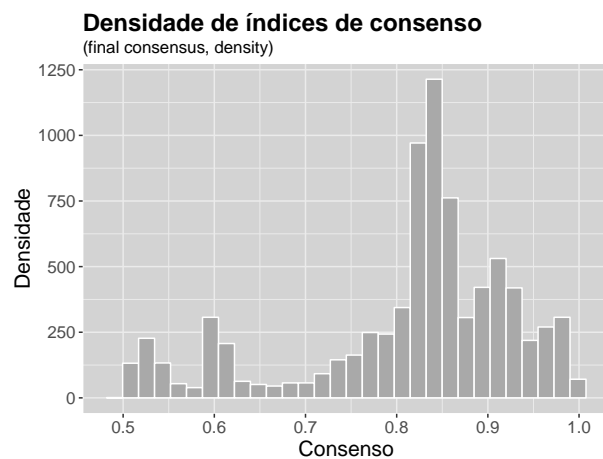


Gráfico 3: Densidade da variável "consenso final".

3.2.5.2 Os dados

A execução das simulações proporcionou um conjunto de dados de 8100 registros, sendo estes compostos pelos seguintes campos:

Campo	Tipo	valores
initial consensus	real	O grau de consenso inicial aleatório
final consensus	real	O grau de consenso final propriamente
iterations	inteiro	total de iterações até o consenso
distance	caracteres	o nome da função de distância
quantifier	caracteres	o nome do quantificador
experts	inteiro	o total de participantes
alternatives	inteiro	o total de alternativas
level	inteiro	o nível de consenso {1,2,3}
sample	inteiro	indica qual amostra aleatória foi utilizada (cada combinação de parâmetros tem 12 amostras)

Tabela 4: Campos no conjunto de dados

3.2.5.3 Análise

Para cada par de funções de distância em com-

paração foram analisadas duas amostras relacionadas, tal como no experimento original. Utilizando o método de reamostragem sem reposição *bootstrap* (Efron and Tibshirani 1997) podemos identificar a distribuição dos dados de consenso (pela mediana) para cada métrica de distância e para a diferença entre cada duas métricas. Observe que é possível visualizar uma distribuição gaussiana para a diferença entre as medianas de consenso para cada duas métricas, conforme o gráfico 4a abaixo, e para as medianas de consenso de cada métrica diretamente, conforme gráfico 4b.

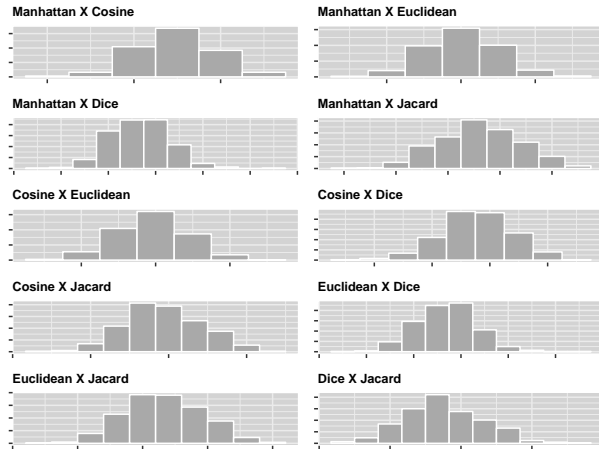


Gráfico 4a: Distribuição da diferença de consenso entre cada par de funções de distância.

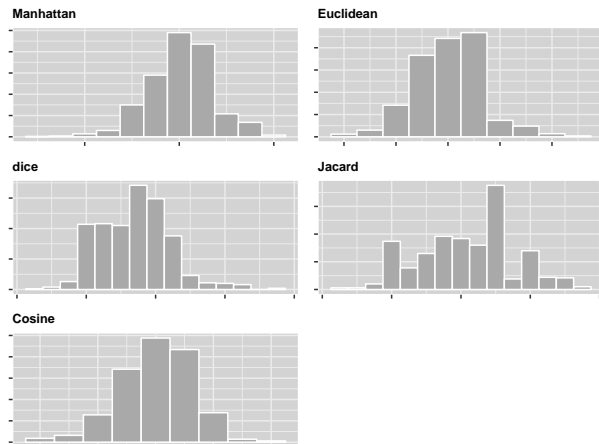


Gráfico 4b: Distribuição de consenso para cada função de distância.

O autor utilizou o teste não-paramétrico de Wilcoxon, porém, como já mencionado, pretendendo utilizar Intervalos de Confiança (Gardner and Altman 1986) com *Bootstrap* e Regressão Linear (Neter, Wasserman, and Kutner 1989). Intervalos de confiança para reproduzir o mesmo cenário do experimento original (substituindo o teste de Wilcoxon) e por ter a comprovação gráfica de que os dados seguem a distribuição normal. Regressão linear para obter uma informação definitiva do quanto a variação da função de distância descreve o comportamento dos dados, sem ignorar os demais fatores.

3.2.6 Ameaças à validade

Não levar em consideração o índice de consenso inicial (aleatório) pode ser uma importante ameaça ao constructo. Como é possível observar no gráfico 5 abaixo, para o quantificador “*leathalf*”, o consenso inicial e final são sempre iguais. Isto implica em uma função de distância demasiadamente otimista, mas que parece eficiente nos resultados. Esta ameaça só poderia, de fato, ser eliminada se mudássemos a variável de resposta do experimento de grau de consenso final para a diferença entre grau inicial e final de consenso.

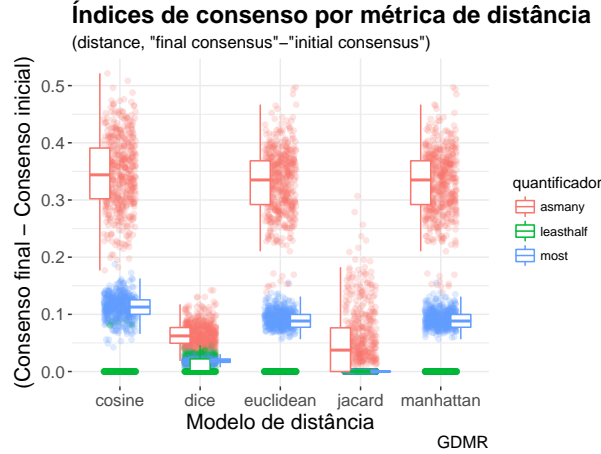


Gráfico 5: Gráfico de dispersão que nos dá uma visão geral do ganho no grau de consenso após o processo, tomando a diferença entre consenso inicial e final em função da distância.

Além da ameaça acima, o autor original (e este autor também, em função do prazo) negligenciou a informação “número de iterações necessárias ao consenso”, o que pode comprometer o resultado. Poucas iterações pode ser sinal de eficiência ou outro indício de excesso de otimismo. Diferenciar estas duas indicações poderia ser feita, novamente, tomando o consenso inicial como medida de otimismo (solucionando duas ameaças ao mesmo tempo).

4 Resultados

Aqui trataremos dos resultados obtidos no nosso experimento fazendo um paralelo com os resultados do experimento original. Como dito anteriormente, utilizamos intervalos de confiança com *bootstrap* e regressão linear para analisar os dados.

4.1 Convergência

Em um primeiro momento, vamos analisar a magnitude ou convergência de consenso para cada função de distância, ainda sem comparações. O gráfico 6 abaixo exibe o centro do intervalo de confiança da mediana à 95% de confiança para cada conjunto de parâmetros {quantificador, função de distância, alternativas e participantes}.

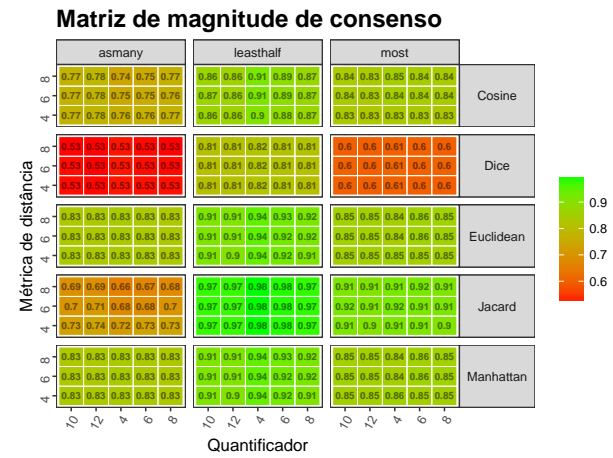


Gráfico 6: Quadro de magnitudes de consenso com base em IC à 95% de significância e envolvendo quatro dos fatores do experimento (total de alternativas, total de participantes, quantificadores e distâncias).

Observe no gráfico que as distâncias de Manhattan, Euclidiana, Cosseno e Jaccard parecer boas medianas no geral. Porém, Dice tem valores muito baixos para o quantificador “asmany” e “most”. As distâncias de Euclides e de Manhattan tem valores praticamente iguais. Note ainda que não existem consensos piores que 0.5, o aleatório. Constatação que ser de evidência à validade do modelo.

Para finalizar, perceba que “leasthalf” apre-

senta os melhores índices de consenso. Lembrando que este quantificador apresenta uma importante ameaça a sua validade neste experimento (e por consequência no experimento original também), por demonstrar ser um modelo de distância muito otimista, atingindo o consenso algumas vezes sem nenhuma iteração.

4.2 Comparações

Comparações pareadas, tal como no experimento original, são o ponto de partida desta análise. Os gráficos 7a e 7b apresentam uma análise por intervalo de confiança e *bootstrap* para cada combinação de fatores tomados todos os pares de combinações. Aqui as métricas são comparadas em pares e, à 95% de confiança, caso o intervalo de confiança da diferença entre as medianas do grau de consenso medido não inclua zero, são classificadas como significativamente diferentes.

Para o gráfico 7a, no eixo x está cada valor utilizado para o parâmetro “total de participantes”, agrupado por quantificador. No eixo y estão os totais de alternativas utilizados, agrupados por nível de consenso. A escala e cores representa cada uma das 10 comparações possíveis, sendo o vermelho igual à 7 comparações e o verde igual à 9. Estas cores representam o total de comparações que apresentaram diferença à 95% de significância (o mínimo foi 7 e o máximo foi 9).

Para o gráfico 7b, o eixo x também mostra cada valor utilizado para o parâmetro “total de participantes”, agora agrupados por comparação entre distâncias. No eixo y estão os totais de

alternativas utilizados, agrupados por nível de consenso. A escala e cores representa cada um dos 3 níveis de consenso, sendo o vermelho igual à 0 níveis e o verde igual à 3 níveis. Estas cores representam em que níveis tais comparações apresentam diferença à 95% de confiança.

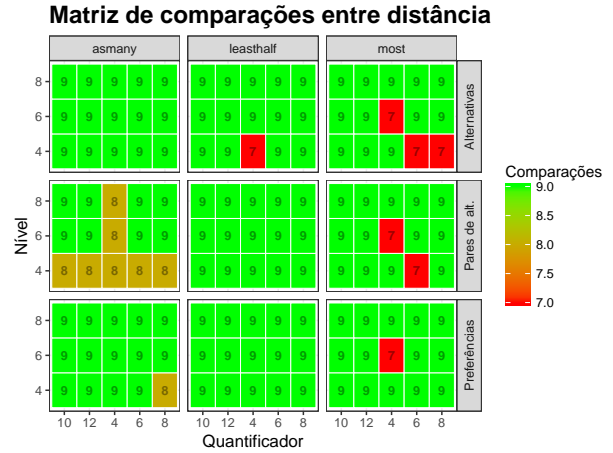


Gráfico 7a: Quadro de comparações com base em IC à 95% de significância e envolvendo quatro dos fatores do experimento (total de alternativas, total de participantes, quantificadores e níveis de consenso). A cor verde indica que 9 comparações entre distâncias apresentaram diferença. Já a cor vermelha, apenas 7.

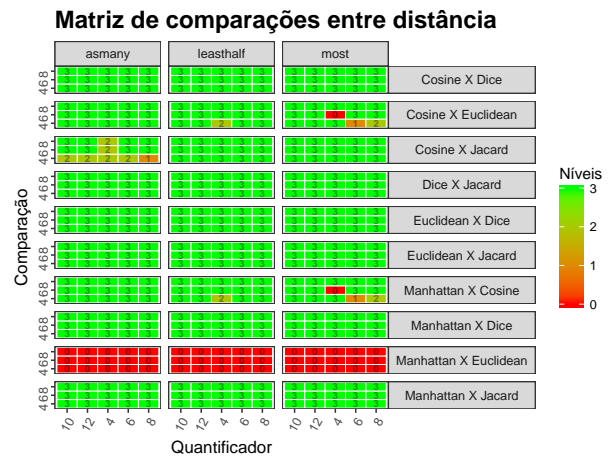


Gráfico 7b: Quadro de comparações com base em IC à 95% de sig-

nificância e envolvendo quatro dos fatores do experimento (total de alternativas, total de participantes, quantificadores e comparações pareadas entre distâncias). A cor verde indica que 3 níveis e consenso apresentaram diferença. Já a cor vermelha, zero.

Observando agora o gráfico 7b, tal como o autor original, não podemos afirmar nenhuma diferença entre as distâncias euclidiana e de Manhattan em nenhum dos três níveis. Tal resultado é interessante, visto que, de acordo com nossa análise de magnitude, a distância euclidiana e a distância de Manhattan, além da distância cosseno, apresentam os maiores graus de consenso no geral.

A distância cosseno também apresenta semelhanças pontuais com a distância euclidiana (para o quantificador “most” e 4 ou 6 participantes), com Jaccard em boa parte do quantificador “as-many” e com Manhattan em um padrão muito semelhante ao da distância euclidiana.

Com isto podemos concluir que podemos aplicar a distância Euclidiana ou a distância de Manhattan e ganharemos tanto na magnitude quanto com relação as demais funções de distância (mesmo resultado do autor original). Não há distinção significativa entre estas duas distâncias como pode ser comprovado no gráfico 7b, onde não foi identificado nenhuma diferença significativa em nenhum nível de consenso. Isto é coerente com o resultado do gráfico 7a, onde todas as células guardam pelo menos uma comparação diferente (a escala só vai até 9.0).

4.3 Regressão Linear

O modelo de regressão que propus seguirá a mesma lógica comparativa pareada do experimento original, analisando as distâncias duas à duas. Utilizamos duas estratégias associadas a regressão, o cálculo do R^2 e, novamente, o intervalo de confiança para cada coeficiente da equação linear resultante. Com isto, considerando sempre um $p \leq 0.05$, comprovamos as afirmações apontadas anteriormente e chegamos a algumas novas, conforme veremos em seguida.

4.3.1 Regressão e R^2

O R^2 mede o quanto uma variável é capaz de descrever o comportamento da outra. No gráfico 8 apresento um quadro comparativo que coloca lado a lado o R^2 de cada um dos fatores do experimento para cada uma das dez comparações possíveis mais um cenário geral, sempre com um $p \leq 0.05$. Perceba no gráfico que o fator que melhor descreve o grau de consenso é o quantificador, chegando a um R^2 superior a 0.8 (o que significa que a variável quantificador descreve 80% do comportamento da variável consenso) para a comparação Manhattan X Euclidean (a mais promissora, de acordo com a análise anterior), além de descrever 33% do comportamento no geral.

Tratando exclusivamente da relação entre consenso e distância, chegamos a um resultado um pouco diferente do que chegamos para Intervalos de confiança. Como antes, as distâncias de Manhattan e Euclidiana não apresentam diferença considerável à pelo menos 95% de significân-

cia. Já Jaccard, que na análise anterior parecia bem diferente das outras duas, agora interfere no consenso em apenas 0.034%, com relação a escolher as distâncias Euclidiana ou de Manhattan, à um $p \leq 0.05$. Resultado que distoa das análises anteriores e até do experimento original, onde utilizou-se teste de hipóteses. Perceba ainda que todas estas comparações incluem a distância de Manhattan e a distância Euclidiana, as mais promissoras em magnitude de consenso (de acordo com a análise anterior).

Entenda que na análise anterior, tal como fez o autor, não comparei magnitude de diferença, apenas as identifiquei com 95% de significância. Com base na natureza da análise feita, qualquer diferença significativa apareceria nos dados, ainda que de baixa magnitude. Por isto não percebemos semelhança entre Jaccard, Euclidean e Manhattan pela análise anterior (já que temos apenas um $R^2 \simeq 0.00034$).

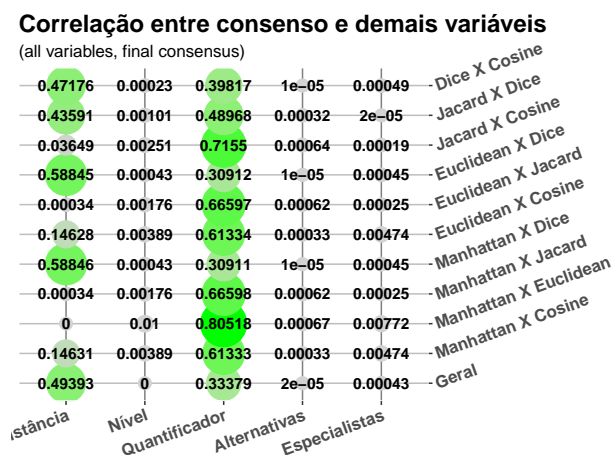


Gráfico 8: R^2 de cada fator do experimento para cada comparação separadamente e para um quadro de consenso geral.

4.3.2 Regressão Linear e Intervalos de Confiança

Para finalizarmos nossas análises por Regressão Linear e Intervalos de Confiança, vamos testar um modelo envolvendo as duas técnicas e todos os fatores ao mesmo tempo. Regressão linear definindo uma equação que descreve o comportamento dos dados e intervalos de confiança para cada coeficiente desta equação. Cada coeficiente, com seu respectivo intervalo de confiança, é considerado válido à pelo menos 95% de significância ($p \geq 0.05$).

O gráfico 9 apresenta um quadro que coloca lado a lado cada um dos valores descritos acima. Na coluna Termo, temos a variável proposta. Perceba que as variáveis categóricas (distância e quantificador) foram distribuídas em várias variáveis booleanas (cada uma pertencente ao domínio $\{0,1\}$), tomando sempre 1 como sendo um valor específico (uma distância, por exemplo) e zero como sendo os demais valores. A coluna Coeficiente descreve o grau de influência linear daquele fator para a variável consenso. Limite inf. e Limite sup. representam os limites inferior e superior do intervalo de confiança de Coeficiente, respectivamente. Observando a coluna P-valor, vemos que nem todos os intervalos contam com mais de 95% de confiança.

Os intervalos de confiança das variáveis level e alternatives (dos dados, vide tabela 4), incluem zero. Porém, levando em consideração seus P-valor elevados (0.86 e 0.31 respectivamente), não podemos tirar conclusões sobre os resultados. Já a variável experts, cujo intervalo de confiança também inclui zero, conta com um

P-valor < 0.05 (significância muito maior que 95%) e, por isso, podemos concluir que esta variável não exerce influência significativa na variável consenso.

Como era de se esperar, com base na análise anterior, os quantificadores exercem influência significativa no consenso. Já as distâncias tem comportamentos variados. Todas exercem alguma influência significativa junto ao consenso, porém para Dice esta influência é negativa. Como já jávamos constatado, a distância de Euclides e de Manhattan tem exatamente a mesma influência para o consenso. Jaccard tem uma influência muito similar às distâncias de Manhattan e de Euclides (coerente com resultado anterior). Como os intervalos de confiança de Manhattan, Euclides e Jaccard se interceptam, não podemos concluir se estas são de fato diferentes com base neste modelo.

Termo	Coefficiente	P-valor	Limite inf.	Limite sup.
distance dice	-0.1777	0.000000e+00	-0.1812	-0.1743
distance euclidean	0.0402	1.209516e-113	0.0368	0.0436
distance jaccard	0.0368	4.328609e-96	0.0334	0.0402
distance manhattan	0.0402	1.145792e-113	0.0368	0.0436
quantifier least half	0.1694	0.000000e+00	0.1667	0.1720
quantifier most	0.0780	0.000000e+00	0.0754	0.0807
level	0.0001	8.576736e-01	-0.0012	0.0014
alternatives	-0.0003	3.118017e-01	-0.0010	0.0003
experts	-0.0009	7.132424e-06	-0.0013	-0.0005

Gráfico 9: Quadro de coeficientes lineares para cada fator em um modelo conjunto envolvendo todas as variáveis e as duas técnicas de análise utilizadas.

5 Considerações finais

Sumarizando, verificamos que as distâncias de Manhattan e de Euclides são as mais promissoras em termos de magnitude de consenso e são equivalentes à 95% ou mais de confiança. A distância de Jaccard é apenas ligeiramente inferior as duas primeiras. Dice é o modelo de distância menos recomendável para consenso e a distância Cosseno é o meio termo. Além disto, o modelo de Regressão Linear nos mostrou que a escolha dos quantificadores é uma decisão projetual muito mais crítica para sistemas de suporte ao consenso que a escolha da métrica de distância.

Para além deste experimento, poderia propor incluir as variáveis coerência e completude, porpostos na nossa reprodução mas não aplicados. A verificação de coerência exigiria verificações e correções que, eventualmente, poderiam fazer o consenso retroceder. Uma completude inferior a 1.0 levaria o processo a lidar com incerteza. Um novo experimento para avaliar como estes modelos de distância se comportariam nestes casos poderia chegar a resultados diferentes.

Referências

Cabrerizo, Francisco Javier, Sergio Alonso, and Enrique Herrera-Viedma. 2009. “A Consensus Model for Group Decision Making Problems with Unbalanced Fuzzy Linguistic Information.” *International Journal of Information Technology & Decision Making* 8 (01). World Scientific:

109–31.

Cabrerizo, Francisco Javier, Juan Manuel Moreno, Ignacio J Pérez, and Enrique Herrera-Viedma. 2010. “Analyzing Consensus Approaches in Fuzzy Group Decision Making: Advantages and Drawbacks.” *Soft Computing* 14 (5). Springer: 451–63.

Cabrerizo, Francisco Javier, Ignacio J Pérez, and Enrique Herrera-Viedma. 2010. “Managing the Consensus in Group Decision Making in an Unbalanced Fuzzy Linguistic Context with Incomplete Information.” *Knowledge-Based Systems* 23 (2). Elsevier: 169–81.

Cha, Sung-Hyuk. 2007. “Comprehensive Survey on Distance/Similarity Measures Between Probability Density Functions.” *City* 1 (2): 1.

Chen, Shyi-Ming, Ming-Shiow Yeh, Pei-Yung Hsiao, and others. 1995. “A Comparison of Similarity Measures of Fuzzy Values.” *Fuzzy Sets and Systems* 72 (1): 79–89.

Chiclana, Francisco, JM Tapia García, Maria Jose del Moral, and Enrique Herrera-Viedma. 2013. “A Statistical Comparative Study of Different Similarity Measures of Consensus in Group Decision Making.” *Information Sciences* 221. Elsevier: 110–23.

Chiclana, Francisco, Enrique Herrera-Viedma, Francisco Herrera, and Sergio Alonso. 2004. “Induced Ordered Weighted Geometric Operators and Their Use in the Aggregation of Multiplicative Preference Relations.” *International Journal of Intelligent Systems* 19 (3). Wiley Online Library: 233–55.

———. 2007. “Some Induced Ordered

Weighted Averaging Operators and Their Use for Solving Group Decision-Making Problems Based on Fuzzy Preference Relations.” *European Journal of Operational Research* 182 (1). Elsevier: 383–99.

Deza, Michel Marie, and Elena Deza. 2009. “Encyclopedia of Distances.” In *Encyclopedia of Distances*, 1–583. Springer.

Dubois, Didier, and Henri Prade. 1985. “A Review of Fuzzy Set Aggregation Connectives.” *Information Sciences* 36 (1-2). Elsevier: 85–121.

Efron, Bradley, and Robert Tibshirani. 1997. “Improvements on Cross-Validation: The 632+ Bootstrap Method.” *Journal of the American Statistical Association* 92 (438). Taylor & Francis: 548–60.

Fodor, János, and Marc Roubens. n.d. “Fuzzy Preference Modelling and Multicriteria Decision Support. 1994.” Kluwer Academic Publishers, Dordrecht.

Gardner, Martin J, and Douglas G Altman. 1986. “Confidence Intervals Rather Than P Values: Estimation Rather Than Hypothesis Testing.” *Br Med J (Clin Res Ed)* 292 (6522). British Medical Journal Publishing Group: 746–50.

Herrera, Francisco, Enrique Herrera-Viedma, and Francisco Chiclana. 2003. “A Study of the Origin and Uses of the Ordered Weighted Geometric Operator in Multicriteria Decision Making.” *International Journal of Intelligent Systems* 18 (6). Wiley Online Library: 689–707.

Herrera, Francisco, Enrique Herrera-Viedma, and others. 1996. “A Model of Consensus in

- Group Decision Making Under Linguistic Assessments.” *Fuzzy Sets and Systems* 78 (1). Elsevier: 73–87.
- Herrera, Francisco, Enrique Herrera-Viedma, and José L Verdegay. 1997. “Linguistic Measures Based on Fuzzy Coincidence for Reaching Consensus in Group Decision Making.” *International Journal of Approximate Reasoning* 16 (3). [New York, NY]: North-Holland, c1987-: 309–34.
- Kacprzyk, Janusz. 1986. “Group Decision Making with a Fuzzy Linguistic Majority.” *Fuzzy Sets and Systems* 18 (2). Elsevier: 105–18.
- Klir, George J, and Tina A Folger. 1988. “Fuzzy Sets, Uncertainty, and Information.” Prentice Hall.
- Neter, John, William Wasserman, and Michael H Kutner. 1989. “Applied Linear Regression Models.” Irwin Homewood, IL.
- Ureña, Raquel, Francisco Javier Cabrerizo, Juan Antonio Morente-Molinera, and Enrique Herrera-Viedma. 2016. “GDM-R: A New Framework in R to Support Fuzzy Group Decision Making Processes.” *Information Sciences* 357. Elsevier: 161–81.
- Wang, Xuzhu, Bernard De Baets, and Etienne Kerre. 1995. “A Comparative Study of Similarity Measures.” *Fuzzy Sets and Systems* 73 (2). Elsevier: 259–68.
- Xu, Zeshui, and Qing-Li Da. 2003. “An Overview of Operators for Aggregating Information.” *International Journal of Intelligent Systems* 18 (9). Wiley Online Library: 953–69.
- Yager, Ronald R. 1988. “On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decisionmaking.” *IEEE Transactions on Systems, Man, and Cybernetics* 18 (1). IEEE: 183–90.
- Zhou, Shang-Ming, Francisco Chiclana, Robert I John, and Jonathan M Garibaldi. 2008. “Type-1 Owa Operators for Aggregating Uncertain Information with Uncertain Weights Induced by Type-2 Linguistic Quantifiers.” *Fuzzy Sets and Systems* 159 (24). Elsevier: 3281–96.
- . 2011. “Alpha-Level Aggregation: A Practical Approach to Type-1 Owa Operation for Aggregating Uncertain Information with Applications to Breast Cancer Treatments.” *IEEE Transactions on Knowledge and Data Engineering* 23 (10). IEEE: 1455–68.