



BIG DATA E FILOSOFIA: NOVOS DESAFIOS PARA OS FILÓSOFOS

Danúbio Viana Nogueira

Possui bacharelado em Administração pelo Centro de Ensino Superior de Dracena. Atualmente, graduando em Filosofia pela FFC – UNESP/ Marília. E-mail: danubionogueira@gmail.com

Mariana Cláudia Broens

Possui graduação em Filosofia pela Universidade Federal do Paraná (1985), graduação em Direito pela Pontifícia Universidade Católica do Paraná (1983), Diplôme d'Études Approfondies en Logique et Philosophie anglo-saxonne pela Université de Nantes - França (1989) e doutorado em Filosofia pela Universidade de São Paulo (1996). Atualmente é professora adjunta (Livre Docente) da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista Júlio de Mesquita Filho. E-mail: mbroens@marilia.unesp.br

RESUMO

Este artigo apresenta uma breve reflexão a respeito dos *Big Data* e de algumas das suas relações possíveis com a filosofia, mais especificamente com a epistemologia e com a filosofia das ciências. É abordado o problema filosófico da indução enquanto método investigativo, tal como apresentado por David Hume, a aplicação deste método nas ciências empíricas, e de que forma os *Big Data* estão sendo propostos como um novo método investigativo para estas ciências. Por fim, é apresentada a razão pela qual acreditamos que os *Big Data* representam novos desafios para os filósofos atuais.

PALAVRAS-CHAVE: Big Data. Filosofia. Ciência. Método Indutivo. Estatística Inferencial.

BIG DATA AND PHILOSOPHY: NEW CHALLENGES FOR THE PHILOSOPHERS

ABSTRACT

This article presents a brief reflection about the Big Data and some of its possible relations with philosophy, more specifically with epistemology and with the philosophy of sciences. It addresses the philosophical problem of induction as

investigative method, as it was presented by David Hume, the application of this method in the empirical sciences, and how the Big Data are being proposed as a new investigative method for these sciences. Finally, are presented the reasons for we to believe that Big Data represent new challenges for the current philosophers.

KEYWORDS: Big Data. Philosophy. Science. Inductive Method. Inferential Statistics.

1 INTRODUÇÃO

Ao longo do texto procuramos mostrar, de maneira introdutória, algumas das relações possíveis entre a filosofia e os *Big Data*, e de que forma estes trazem alguns questionamentos novos aos debates filosóficos em andamento, exigindo que os filósofos atuais estejam cientes quanto à relevância alcançada por esta tecnologia nos dias atuais. O artigo se concentra no debate epistemológico, muito embora os questionamentos éticos sejam, no mínimo, tão importantes quanto os epistemológicos.

Um exemplo bastante claro de implicação epistemológica dos *Big Data* foi apresentado por Chris Anderson (2008), num artigo onde ele afirma que uma *nova era* das ciências por meio dos *Big Data* estava surgindo, tornando o método científico tradicional obsoleto, e modificando completamente a maneira de se fazer ciência. Pelas palavras de Anderson, podemos perceber o alto grau de confiança de que os *Big Data* substituirão os métodos científicos tradicionais:

Os petabytes nos permitem dizer: “Correlação é suficiente”. Podemos parar de procurar por modelos. Podemos analisar os dados sem hipóteses sobre o que isso possa mostrar. Podemos jogar os números nos maiores clusters¹ de computadores que o mundo já viu e deixar os algoritmos estatísticos encontrarem padrões onde a ciência não pode. (Anderson, 2008, tradução e grifo nossos).

Mas por que os *Big Data* representam uma mudança tão grande no mundo da tecnologia, a ponto de haver opiniões de que até mesmo a ciência será revolucionada por eles? Que tipo de tecnologia seria capaz de realizar uma mudança tão radical?

2 COMPUTAÇÃO

A evolução do hardware e de seus softwares de controle e o desenvolvimento de novas tecnologias computacionais permitiram o surgimento de três paradigmas tecnológicos relacionados, e que são de importância fundamental para o surgimento dos *Big Data*: a *computação paralela e distribuída*, a *computação em nuvem* e a *computação ubíqua*.

A computação paralela e distribuída permite que processamentos de volumes muito grandes possam ser divididos em partes menores, para serem processados simultaneamente em diferentes sistemas computacionais interligados por meio de redes, possibilitando, desta forma, que o tempo total de processamento seja drasticamente reduzido.

A computação em nuvem permite que, não apenas o processamento, mas também o armazenamento dos dados seja realizado em paralelo, proporcionando maior capacidade de armazenamento e maior confiabilidade na integridade dos dados, por custos relativamente pequenos. A computação em nuvem permite também que a localização física dos computadores tenha pouca relevância para o trabalho a ser realizado.

A transmissão de dados por fibra ótica e por redes sem fio permite transferências com alta velocidade e qualidade e possibilita o tráfego de enormes quantidades de dados de maneira quase instantânea, permitindo assim a existência não apenas da computação paralela, distribuída e em nuvem, mas também a criação de enormes redes de dispositivos que se distribuem de forma descentralizada, dando a impressão de que podem estar em todo e qualquer lugar e que sua presença pode ser considerada *ubíqua*²

Com o advento dos softwares inferenciais³, dentre os quais estão os *Big Data*, a tecnologia da informação parece ter alcançado um novo patamar, onde o *manuseio* dos dados por agentes humanos é não apenas desnecessário, mas até mesmo contraproducente e indesejável.

3 TECNOLOGIA DA INFORMAÇÃO

Os *Big Data* possuem alguns aspectos que formam as suas características distintivas, conhecidos como os “*Vs*” dos *Big Data*. A característica do *volume* diz respeito à quantidade de dados, que só podem ser processados por organizações que possam arcar com os custos de grandes parques computacionais; resumida esta característica em algumas palavras, “volume” significa: poder de processamento. A característica da *variedade* diz respeito às diversas origens dos dados que servem de entrada para os processamentos; resumindo em poucas palavras, “variedade” significa: computação ubíqua. A característica da *velocidade* se refere ao tempo em que os *Big Data* processam os grandes volumes de dados, visando apresentar os resultados das análises tão rapidamente quanto possível; resumindo em poucas palavras, “velocidade” significa: processamento instantâneo.

A característica da *veracidade* se refere aos critérios utilizados para a seleção dos dados

que serão analisados, visando selecionar apenas os dados que tenham significado para a análise pretendida; resumindo, “veracidade” significa: relevância. A característica do *valor* se refere à finalidade com que são realizadas as análises por meio dos *Big Data*, denotando que de alguma forma estas têm o propósito de criar valor para as organizações, seja em sentido de valor financeiro, valor estratégico, valor social, etc.; resumindo esta característica, “valor” significa: obtenção de vantagens e benefícios.

Diferentes enfoques sobre os *Big Data* podem abordar estas características de maneiras distintas. Para a computação o volume, a variedade e a velocidade podem ser suficientes, enquanto para o mundo dos negócios o aspecto mais relevante pode ser o valor, e para a filosofia, certamente, a veracidade é o aspecto fundamental.

Analisando as propostas apresentadas por alguns proponentes dos *Big Data* (BOLLIER, 2010), chegamos à conclusão de que a utilização dos *Big Data* poderia ser resumida nas seguintes etapas: a) coletar e *processar* o maior volume e variedade de dados possível, *na maior velocidade possível*; b) *analisar* os dados *de forma imparcial*, de maneira que os resultados surjam espontaneamente, sem que haja influência por parte dos analistas; c) *identificar* padrões e *correlações* que sejam *significativas* para a análise em questão, permitindo, assim, que seja realizado o aspecto da veracidade; d) *inferir* situações presentes e *tendências futuras*, no intuito de permitir a antecipação de fenômenos de larga escala, que de outra maneira seriam muito difíceis de identificar, permitindo assim, a realização do aspecto do valor. Para sabermos se os *Big Data* podem, de fato, alcançar seus objetivos e cumprir suas metas, a análise filosófica e epistemológica é de especial auxílio.

4 FILOSOFIA

Antes que os *Big Data* possam ser tomados como uma nova forma válida de se obter conhecimento, é preciso que estes sejam analisados sob o ponto de vista da epistemologia, pois caso se constate que os *Big Data* não conduzem a uma forma válida de conhecimento, a discussão filosófica neste sentido perderá seu propósito. Parece-nos, pois, que antes de os *Big Data* representarem um desafio à filosofia, a própria filosofia representa um desafio aos *Big Data*.

Sob o ponto de vista filosófico o conceito de *Big Data* é complexo, sendo necessário para a sua compreensão, passar por diversas áreas das ciências e da própria filosofia. Muito além das questões envolvendo a lógica, as ontologias e as taxonomias utilizadas em

desenvolvimento de software, a análise filosófica dos *Big Data* precisa lidar com questões que apresentam abrangências bem maiores.

Dentre as questões filosóficas concernentes aos *Big Data*, parece-nos que as de maior relevância estão relacionadas a outras questões filosóficas já pertencentes a debates filosóficos anteriores, mas que sob a ótica da problemática dos *Big Data*, aparecem como que estendendo estas questões já existentes.

5 EMPIRISMO E INDUTIVISMO

Uma das diretrizes da utilização dos *Big Data*, é a imparcialidade na coleta e análise dos dados, evitando-se que o processo seja influenciado pelas ideias dos analistas, pois apenas por meio deste processo imparcial seria possível que a realidade se revelasse espontaneamente através dos dados. Esta confiança de que apenas através dos dados empíricos é possível se obter conhecimento do mundo é conhecida na filosofia como “empirismo” (EMPIRICISM, 2015). No caso dos proponentes dos *Big Data*, este empirismo parece revelar, ainda, a crença de que quanto mais dados são coletados, mais conhecimento sobre a realidade será obtido.

Além do amplo debate que o empirismo já suscita dentro da filosofia, podemos fazer um questionamento ainda mais básico: é possível para um sistema computacional coletar todos os dados relevantes para uma análise acurada das diversas e complexas situações do mundo real? Outra questão que a filosofia deve analisar é a da possibilidade da realização de observações empíricas de maneira completamente imparcial, pois há filósofos que afirmam que as observações não são neutras, mas *carregadas* pelas ideias do observador.

A etapa final da utilização dos *Big Data* é a realização de inferências a respeito tanto das situações atuais, quanto das situações futuras. A inferência neste caso é do tipo indutiva, pois, a partir de observações particulares, procura-se inferir generalizações. As análises realizadas pelos *Big Data* se baseiam principalmente na observação das correlações entre metadados⁴, a fim de encontrar quais destas correlações são significativas, ou seja, necessárias, e quais delas são contingenciais, ou seja, apenas coincidências.

Este aspecto da utilização dos *Big Data* parece revelar uma forte crença de que basta encontrar correlações significativas para que sejamos capazes de fazer previsões. Quanto a este assunto, há uma discussão filosófica profunda a respeito do chamado “problema da indução”, que gira basicamente em torno do conceito de causalidade. Conforme exposto por David Hume, a simples observação de que duas coisas ocorrem de maneira conjunta não é suficiente para

afirmarmos que uma coisa causa a outra, ou que ambas se causem mutuamente, não sendo possível, desta forma, assumir uma relação universal de causalidade entre as duas coisas (HENDERSON, 2018).

Sob este aspecto, podemos nos perguntar se seria possível que um programa de computador recriasse por meio de modelos matemáticos as complexidades das situações existentes no mundo. Ou ainda, se encontrar correlações significativas com poder preditivo por meio dos *Big Data* conseguirá resolver o problema da indução apresentado por Hume?

6 FILOSOFIA DA ESTATÍSTICA E FILOSOFIA DA CIÊNCIA

Talvez a principal razão para os *Big Data* terem alcançado a fama de tecnologia revolucionária, seja o fato de eles possuírem a capacidade de processar dados estatísticos em larga escala. Os processamentos dos dados nos *Big Data* seguem os padrões e modelos da estatística inferencial (ou, estatística indutiva) e portanto, acabam por *herdar* as discussões desta área específica. A Estatística Inferencial utiliza modelos matemáticos de análises de dados, na tentativa de compreender, explicar e prever a ocorrência de fenômenos, sendo uma área da matemática cujos resultados se expressam não de forma determinista (como certezas), mas de forma probabilística. Neste ponto, deparamo-nos com mais uma grande discussão dentro da filosofia, a chamada “interpretação da probabilidade”, que se refere às diversas formas pelas quais os filósofos e pensadores deram explicações para o que é a probabilidade (HÁJEK, 2012).

Sob este ponto de vista, parece-nos que uma questão fundamental a ser feita, é sobre como justificar o conhecimento em termos de probabilidade. Pois, uma vez que o cálculo probabilístico não nos permite ter completa certeza a respeito do que ocorrerá, como podemos saber se se trata realmente de um conhecimento ou apenas de uma opinião fundamentada em cálculos?

O questionamento dos *Big Data* quanto ao método científico nos conduz até a questão da utilização da estatística como método de comprovação para a investigação empírica. A ciência depende da estatística de tal maneira que quando falamos em ciência empírica, ou seja, em ciência através da experiência, a própria noção de experiência é colocada em termos estatísticos, tornando a estatística uma técnica fundamental para a pesquisa científica. De modo geral:

empírica, servindo para apoiar e comunicar as descobertas científicas, a filosofia da estatística é de importância fundamental para a filosofia da ciência. Ela tem um impacto na avaliação filosófica do método científico, e no debate sobre o estatuto epistêmico e ontológico da teoria científica. (ROMEIJN, 2017, tradução nossa).

A pergunta que os filósofos podem fazer quanto a este aspecto é se os *Big Data* constituem realmente uma maneira *melhor* de se fazer estatística, e portanto de se fazer ciência, ou se são apenas uma maneira *maior* de se fazer estatística. Em outras palavras: os *Big Data* podem realmente melhorar o método científico de maneira qualitativa, ou podem apenas aumentar a quantidade de dados que são processados, sem que isto tenha impacto significativo na qualidade?

7 FILOSOFIA DA LINGUAGEM

O aspecto qualitativo da análise realizada pelos softwares inferenciais é o que torna importante a característica da variedade dos dados (NEGASH, 2004). Estes dados são constituídos não apenas por dados numéricos, mas também por palavras, frases, comunicações, documentos, etc., envolvendo, desta forma, algum nível de análise linguística. A compreensão dos limites linguísticos dos *Big Data* pode ser de grande importância para a compreensão dos limites da própria utilização dos *Big Data*. Sem um entendimento suficiente da linguagem humana, um software dificilmente será capaz de realizar operações linguísticas além daquelas para as quais foi programado.

A dinâmica dos sistemas humanos pode exigir um nível de adaptabilidade maior do que o apresentado por programas que se baseiem apenas em análises quantitativas, pois, quando se trata da linguagem humana, as análises estão relacionadas a valores qualitativos e subjetivos (LANGUAGE, 2015).

Com relação a este aspecto, um dos questionamentos que podemos fazer é a respeito da consistência e amplitude das análises linguísticas feitas pelos *Big Data*. Qual seria a extensão possível para a compreensão linguística dos *Big Data*? É possível que eles alcancem uma compreensão semântica dos dados que analisam, ou ficarão restritos apenas ao aspecto sintático? Podem os *Big Data* ter inteligência suficiente para compreender as nuances e sutilezas da linguagem e da mente humana?

8 INTELIGÊNCIA ARTIFICIAL E FILOSOFIA DA MENTE

algoritmos formulados previamente, mas também envolve inteligência artificial e aprendizado de máquinas. Neste aspecto, já há um amplo debate filosófico a respeito da capacidade das máquinas de aprenderem ou de apenas imitarem o comportamento humano. A discussão ocorre principalmente a respeito do conceito de inteligência e da extensão do aprendizado possível pelas máquinas, se seria possível para as máquinas aprenderem como os seres humanos aprendem, apresentando características criativas, ou se seria possível apenas que as máquinas reproduzissem padrões inseridos previamente.

Grande parte do debate sobre a inteligência artificial tem sido conduzido pelas visões contrárias apresentadas por Alan Turing e por John Searle, a respeito do que é a inteligência e de como pode-se verificar ou falsificar a sua existência em máquinas. Turing apresenta uma visão sobre a inteligência que na filosofia da mente é chamada de “funcionalismo”, e que poderia ser resumida da seguinte maneira: assumindo-se que o ser humano possui uma mente inteligente, para sabermos se uma máquina possui também inteligência, façamos com que ela interaja com a inteligência de uma pessoa, e se a pessoa não for capaz de distinguir se está interagindo com uma máquina ou com um outro ser humano, verificamos que esta máquina apresenta inteligência comparável à inteligência humana (OPPY, 2018). Searle, no entanto, se opôs diretamente à tese de Turing, argumentando que mesmo que as máquinas consigam imitar o ser humano de maneira satisfatória através da manipulação de símbolos, isto não significa que sejam capazes de compreender os símbolos que manipulam, razão pela qual, não podem ser chamadas de inteligentes em sentido forte, ou seja, no mesmo sentido que utilizamos para falar da inteligência das pessoas (COLE, 2015).

Neste aspecto, a questão filosófica principal parece ser bastante simples e direta (embora sua investigação se direcione em sentido oposto): podem as máquinas ter uma mente ou inteligência, no mesmo sentido que assumimos que os seres humanos têm?

9 CONSIDERAÇÕES FINAIS

Pelo que foi apresentado até aqui, chegamos à conclusão de que, sob a ótica da epistemologia, as inovações trazidas pelos *Big Data* se inserem dentro de questionamentos filosóficos já existentes, sem, no entanto, revolucionar completamente a filosofia e as ciências.

Os *Big Data* modificam, de certa forma, o debate filosófico, porém ainda não representam um novo paradigma epistemológico. No entanto, isto não significa que eles não possam representar algo novo para outras perspectivas filosóficas ou para outras disciplinas,

como a ética, a ciência da informação, administração de empresas, psicologia, ciências sociais, tecnologia da informação, etc.

É tarefa dos filósofos investigar e compreender os limites da utilização dos *Big Data* como um método auxiliar à ciência empírica, e os *Big Data* podem contribuir muito quando se trata de cálculos e análises estatísticas. Porém, a afirmação de que eles tornarão obsoleto o método científico tradicional, parece-nos, na melhor das hipóteses, uma grande precipitação.

Não apenas a utilização dos *Big Data* apresenta novos desafios aos filósofos, mas também a filosofia apresenta vários desafios aos *Big Data*, antes que a validade destes como parte do método de investigação científica possa ser aceita. O debate acerca dos *Big Data* envolve aspectos epistemológicos que ainda não foram completamente resolvidos dentro da filosofia. Parece-nos que para a filosofia solucionar satisfatoriamente os desafios apresentados pelos *Big Data*, é necessário que a filosofia solucione antes os desafios epistemológicos que lhes são antecedentes.

Diante dos desafios aqui apresentados, acreditamos que a investigação filosófica a respeito dos *Big Data* ainda está longe de chegar à sua conclusão.

NOTAS

¹ Os *clusters* são uma das formas de realizar a computação paralela e distribuída, conforme exposto na seção seguinte.

² No mundo dos negócios, a computação ubíqua é chamada de “internet das coisas”.

³ Softwares inferenciais são softwares que auxiliam na realização de inferências, visando uma melhor capacidade de compreensão das situações presentes ou de previsão das situações futuras. Outros exemplos deste tipo de software são: *data warehouse* e *business intelligence*.

⁴ Metadados são dados gerados a respeito dos dados coletados no mundo real. Em geral expressam agrupamentos, contagens, somas, produtos, médias, etc.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSON, Chris. The end of theory: the data deluge makes the scientific method obsolete. *Wired*, 2008. Disponível em: <<https://www.wired.com/2008/06/pb-theory/>>. Acesso em: 25/03/2018.

BOLLIER, David (Relator). *The Promise and Peril of Big Data*. Washington (EUA): The Aspen Institute, 2010. Disponível em: <https://assets.aspeninstitute.org/content/uploads/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf>. Acesso em: 25/03/2018.

COLE, David. The Chinese Room Argument. In: *The Stanford Encyclopedia of Philosophy*. Winter 2015 Edition. Disponível em: <<https://plato.stanford.edu/archives/win2015/entries/chinese-room/>>. Acesso em: 22/07/2018.

EMPIRICISM. In: *Encyclopædia Britannica Ultimate Reference Suite*. Chicago: Encyclopædia Britannica, 2015.

HÁJEK, Alan. Interpretations of Probability. In: *The Stanford Encyclopedia of Philosophy*. Winter

2012 Edition. Disponível em:
<<https://plato.stanford.edu/archives/win2012/entries/probability-interpret/>>.
Acesso em: 28/07/2018.

HENDERSON, Leah. The Problem of Induction. In: *The Stanford Encyclopedia of Philosophy*. Summer 2018 Edition. Disponível em:
<<https://plato.stanford.edu/archives/sum2018/entries/induction-problem/>>.
Acesso em: 21/07/2018.

LANGUAGE, Philosophy of. In: *Encyclopædia Britannica Ultimate Reference Suite*. Chicago: Encyclopædia Britannica, 2015.

NEGASH, Solomon. *Business Intelligence*. 2004. Disponível em:

<<http://site.xavier.edu/sena/info600/businessintelligence.pdf>>. Acesso em: 15/06/2018.

OPPY, Graham; DOWE, David. The Turing Test. In: *The Stanford Encyclopedia of Philosophy*. Spring 2018 Edition. Disponível em:
<<https://plato.stanford.edu/archives/spr2018/entries/turing-test/>>. Acesso em: 29/07/2018.

ROMEIJN, Jan-Willem. Philosophy of Statistics. In: *The Stanford Encyclopedia of Philosophy*. Spring 2017 Edition. Disponível em:
<<https://plato.stanford.edu/archives/spr2017/entries/statistics/>>. Acesso em: 28/07/2018.

NOGUEIRA, D. V.; BROENS; M. C. Big Data e Filosofia: Novos Desafios Para os Filósofos. *Complexitas - Rev. Fil. Tem.* Belém, v. 3, n. 1, p. 40-49, jan./jun. 2018. Disponível em:<<http://www.periodicos.ufpa.br/index.php/complexitas/article/view/6644>>. Acesso em: 20 de fevereiro de 2019.
