The use of 'Big Data' in a modern swine breeding program now and in the future

Jeremy T. Howard, Ph.D.
Geneticist, Smithfield Premium Genetics

Introduction

Considerable progress has been made in the use of tools to routinely monitor and collect data on animals without the need for humans to manually capture that data. Furthermore, the computing power continues to increase, along with a decrease in the amount of space required for its components, such that smartphones now have more processing power than a supercomputer 20 years ago. As a result, innovative high-throughput data recording and phenotyping platforms via the use of pictures, sensor (e.g., temperature, GPS position, accelerometer data, RFID, etc.) and sound data have begun to be prototyped and/or used within commercial swine companies (Brünger et al. 2019; Fernandes et al. 2019). 3D Camera technology has improved rapidly and with it, innovative uses have been investigated and include monitoring behavior in sows and pigs, predicting pig weights, reading ear identification tags, estimating the lean meat percentage (Lohumi et al. 2018) and more recently facial recognition to identify individual pigs (Hansen et al. 2018). Furthermore, a system that continuously monitors a group of animals has been developed to detect abnormal behavior in the form of how much an animal is eating and drinking, along with animal movement (Psota et al. 2019).

The development of innovating high-throughput data recording platforms will continue, but another challenge that is actively being researched, is reliable and in real-time extraction of important conclusions from the data generated. The large amount of data generated can be defined as 'big data', although the definition varies considerably across disciplines (Morota et al. 2018). The large amount of data that is generated from high-throughput platforms often contains a greater amount of errors (e.g., missing data, outliers, etc.) compared to traditional data collected via humans (e.g., body weights, litters size). Furthermore, visually inspecting the data is much more time consuming and, in some cases, no longer possible to effectively visualize all the data. As a result, effective ways to manage the data being collected along with diagnosing data issues in real-time is important to ensure the data generated is useful and accurately portrays what is actually occurring for any downstream analysis.

Historical Use of Big Data

Historically, the swine industry has been using big data in the form of high throughput feed intake data beginning around the early 1990s. Electronic feeders utilize radio frequency identification (RFID) tags to determine which pig is at the feeder, the amount of feed consumed and the weight of the animal. Electronic feeder systems have been traditionally utilized to improve the genetics of feed efficiency. Several other metrics are generated from electronic feeding systems and include feeding duration, amount of feed consumed for a visit and time at which an animal eats. Recently, the use of electronic feeders has been used to extract resilience phenotypes based on the variability of feed intake or feed intake duration within an animal across time during the growth period (Putz et al. 2019). Animals with higher resilience are less affected by environmental changes such as disease or weather and, as a result, show fewer fluctuations in their daily feed intake. Although electronic feeders provide a wealth of data, the cost for each station is high, which usually limits the number of pigs that can be evaluated at one time. Furthermore, in a swine breeding program, the breeding goal is to maximize crossbred performance (e.g., commercial market animals and commercial dam), which can be accomplished via selection at the purebred level when the genetic correlation between purebred and crossbred performance is close to unity (Bijma et al. 1998). Literature estimates of the genetic correlation between purebred and crossbred performance for feed intake traits have ranged from 0.62 to 0.67 (Godinho et al. 2018). Traditionally electronic feeders have been placed on disease-free nucleus farms and measured on purebred animals. As a result, the feed efficiency response achieved at the purebred level is not being fully realized at the commercial level as a result of environmental differences between the two environments.

Future of Big Data In An Integrated Swine Company

Use of big data allows for one to more effectively focus on how the biological system can be managed at the individual animal level, in order to reduce the phenotypic variability and minimize the impact of environmental, disease and/or technological issues (e.g. ventilation malfunctioning) when they occur. Integrated companies that have genetic data from the commercial level flowing back to selection candidates at the nucleus can also leverage the data to not only select animals that excel in economically important traits, but can be used to more effectively manage a group of pigs at the commercial sector to achieve the maximum productivity and uniformity. As a result, integrated companies can leverage big data across multiple sectors (genetics, animal production, packing plant) in order to improve the profitability of the whole system. Furthermore, the highest value of any given high-throughput recording platform is not realized one technology, at a time in isolation, rather, through the broad adoption of multiple platforms. As an example, electronic feeders provide a wealth of data regarding how much an animal is consuming, but it doesn’t provide any information on what an animal is doing when it is not eating or environmental stressors that may have caused an animal to reduce its feed intake. For example, a finisher animal on average spends a little over an hour (e.g. 76.7 minutes; Brown-Brandl et al. 2013) a day eating, which provides only 5 % of the activity information for an animal in a given day. As a result, behavior data in the form of animal activity and barn temperature along with a feed intake recording system could potentially provide more information than any one recording platform. Lastly, machine vision technology offers the potential to realize a low cost and non-intrusive method to identify individual animals, which when scaling data capture and animal traceability at the commercial level, could greatly reduce the complexity of tracking a large number of commercial animals from birth to slaughter. The use of high-throughput data recording, animal identification, and phenotyping platforms has the potential to revolutionize the way pigs are managed to achieve maximum production along with phenotype collection that is less labor intensive. The novel phenotypes collected via automated recording platforms could provide a more comprehensive overview of the genetic potential of an animal in regards to how its behavior and response to environmental stressors interact with routinely collected weight, reproductive and carcass information.

References:

Bijma, P., and J. A. M. van Arendonk. 1998. Maximizing genetic gain for the sire line of a crossbreeding scheme utilizing both purebred and crossbred information. Animal Sci. 66: 529-542.

Brünger, J., S. Dippel, R. Koch, and C. Veit. 2019. 'Tailception': Using neural networks for assessing tail lesions on pictures of pig carcasses. Animal 13(5):1030-1036.

Fernandes, A. F. A., J. R. R. Dórea, R. Fitzgerald, W. Herring and G. J. M. Rosa. 2019. A novel automated system to acquire biometric and morphological measurements and predict body weight of pigs via 3D computer vision. J Anim. Sci., 97(1):496-508.

Godinho, R. M., R. Bergsma, F. F. Silva, C. A. Sevillano, E. F. Knol, M. S. Lopes, P. S. Lopes, J. W. M. Bastiaansen, and S. E. F. Guimaraes. 2018. Genetic correlations between feed efficiency traits, and growth performance and carcass traits in purebred and crossbred pigs. J Anim. Sci., 96(3):817-829.

Hansen, M. F., M. L. Smith, L. N. Smith, M. G. Salter, E. M. Baxter, M. Farish, and B. Grieved. 2018. Towards on-farm pig face recognition using convolutional neural networks. Comput. Ind., 98:145-152.

Lohumi, S., C. Wakholi, J. H. Baek, B. D. Kim, S. J. Kang, H. S. Kim, Y. K. Yun, W. Y Lee, S. H. Yoon, B. K. Cho. (2018). Nondestructive Estimation of Lean Meat Yield of South Korean Pig Carcasses Using Machine Vision Technique. Korean journal for food science of animal resources, 38(5), 1109-1119.

Morota, G., R. V. Ventura, F. F. Silva, M. Koyama, S. C. Fernando. 2018. BIG DATA ANALYTICS AND PRECISION ANIMAL AGRICULTURE SYMPOSIUM: Machine learning and data mining advance predictive big data analysis in precision animal agriculture. J Anim. Sci., 96(4):1540-1550.

Putz, A. M., J. C. S. Harding, M. K. Dyck, F. Fortin, G. S. Plastow, J. C. M. Dekkers and PigGen Canada. 2019. Novel Resilience Phenotypes Using Feed Intake Data From a Natural Disease Challenge Model in Wean-to-Finish Pigs. Front Genet, 9:660.

Psota, E. T., M. Mittek, L. C. Pérez, T. Schmidt & B. Mote. 2019. Multi-Pig Part Detection and Association with a Fully-Convolutional Network. Sensors, 19(4):852.