DOI: 10.26650/forestist.2018.340634
Forestist 2018, 68(2): 136-140

İSTANBUL UNIVERSITY PRESS

Short Note

Comparing Shannon entropy with Deng entropy and improved Deng entropy for measuring biodiversity when a priori data is not clear

Öncü verinin belirsizliği durumunda biyoçeşitliğin belirlenmesinde Shannon entropisinin Deng entropisi ve geliştirilmiş Deng Entropisi ile karşılaştırılması

Kürşad Özkan

Department of Soil Science and Ecology, Süleyman Demirel University, Faculty of Forestry, 32200, Isparta, Turkey

ABSTRACT

The various diversity measures used to measure biodiversity include the Margalef index, McIntosh index, Simpson index, Brillouin index, and Shannon entropy. Of these measures, the most popular is Shannon entropy (H). In this study, with respect to measuring biodiversity, we compare Shannon entropy-the essential aspect of information theory-with the Deng and improved Deng entropies, as proposed within the framework of the Dempster–Shafer evidential theory. To do so, we used a hypothetical dataset of three complexes. Based on this hypothetic data, ecologically speaking, we obtained the most reasonable result from the improved Deng entropy. There are two reasons for this result: 1) Mass functions cannot be used when computing the Shannon entropy, and 2) Deng entropy does not take into consideration the scale of the frame of discernment.

Keywords: Improved belief entropy, information theory, uncertainly, mass function, basic probability assignment, frame of discernment, alpha diversity

ÖZ

Biyolojik çeşitliliğin belirlenmesinde Margalef indeksi, McIntosh indeksi, Simpson indeksi, Brillouin indeksi ve Shannon entropisi gibi birçok çeşitlilik indisi kullanılmaktadırlar. Bu indisler arasındaki en popular olanı Shannon entropisidir. Bu çalışma biyolojik çeşitliğin ölçümüne yönelik olarak bilgi teorisinin temel eşitliği olan Shannon entropis ile Demster-Shafer Delil Teorisi'nin ölçümlerinden olan Deng entropisi ve Geliştirilmiş Deng entropisini karşılaştırmak için gerçekleştirilmiştir. Çalışmada 3 kompleksten oluşan hipotetik bir veri kullanılmıştır. Kullanılan hipotetik veri ile gerçekleştirilen hesaplamaların sonucunda, ekolojik açıdan en makul sonuçlar Geliştirilmiş Deng entropisi ile elde edilmiştir. Bu sonucun iki sebebi bulunmaktadır. Birincisi Shannon entropisi hesaplanırken kütle fonksiyonları kullanılamamaktadır. İkincisi ise Deng entropisinin sezgisel yapı ölçeğini dikkate almamasıdır.

Anahtar Kelimeler: Geliştirilmiş kanaat entropisi, bilgi teorisi, belirsizlik, kütle fonksiyonu, temel olasılık ataması, sezgisel yapı, alfa çeşitliliği

Cite this paper as:
Özkan, K. 2018. Comparing Shannon entropy with Deng entropy and improved Deng entropy for measuring biodiversity when a priori data is not clear. *Forestist* 68(2): 136-140.

Corresponding author:
Kürşad Özkan
e-mail:
kursadozkan@sdu.edu.tr

Received Date:
29.09.2017
Accepted Date:
28.06.2018

INTRODUCTION

Biodiversity is one of the most central topics in conservation biology, community ecology, and environmental geography. There is a wide variety of indices to measure biodiversity**.** In this context, Shannon entropy, a theory for uncertainty measurement first introduced by Claude Shannon (Shannon, 1948), is the most well-known measure (Gorelick, 2006).

Even though Shannon entropy is the most popular theory for uncertainty measurement, it cannot be used directly in the framework of Dempster-Shafer Evidential Theory (DSET) which is effective in uncertain information processing (Zhou et al., 2017). This is because, unlike Shannon entropy, DSET provides the frame of discernment (FOD) and the basic probability assignment (BPA). It has, therefore, been frequently used in many fields such as pattern recognition (Liu et al., 2013; Liu et al., 2016), fault diagnosis (Su et al., 2012; Jiang et al., 2016c; Jiang et al., 2016d; Yuan et al., 2016), multiple attribute decision-making (Chin et al., 2015; Fu et al., 2015), risk evaluation (Wang and Elhag, 2007; Su et al., 2012; Chin et al., 2015; Fu and Wang, 2015; Du and Hu, 2016; Jiang et al., 2016a; Jiang et al., 2016c; Jiang et al., 2016d; Yuan et al., 2016), controller design (Yager and Filev, 1995; Tang et al., 2016), and so on (Wang et al., 2009; Ma et al., 2015; Zhou et al., 2015).

In the Dempster Shafer framework, many methods have been proposed to measure the uncertain degree of evidence, such as discord measurement (Klir and Ramer, 1996), weighted Hartley entropy (Dubois and Prade, 1985), dissonance measurement (Yager, 1983), total conflict measurement (George and Pal, 1996), distance-based total uncertainty measurement (Yang and Han, 2016), Deng entropy (Deng, 2016), Improved Deng entropy (Zhou et al., 2017) and so on (Song et al., 2015; Song et al., 2016).

Deng entropy was first introduced by Deng (Deng, 2016) and has started to be used in many real applications. Deng entropy is the generalization of Shannon entropy. When the BPA is degenerated as a probability distribution, it is degenerated as Shannon entropy (Deng, 2016). Deng entropy may therefore be considered for use in measuring biodiversity. However, Deng entropy does not take the scale of the FOD into consideration, which means a loss of information while processing information. Improved Deng entropy proposed by Zhou et al. (2017) overcomes this limitation.

This paper was organized to compute Shannon entropy, Deng entropy and Improved Deng entropy using an unclear priori hypothetical data. The results of these entropic measures were then compared and discussed from an ecological perspective.

Shannon Entropy

In information theory, Shannon entropy is often used to measure the information volume of a process or a system, and quantify the expected value of the information contained in a message. Information theory denoted as H (Shannon, 1948), is defined as:

$$H = -\sum_{i=1}^{N} p_i \log_b p_i$$

Where N is the number of basic states, p_i is the probability of state i and p_i satisfies and b is the basis of the logarithm which accounts for the scaling of H. Although b is arbitrary, b is usually chosen to be 2, and the unit of information entropy is bit. If b is the nature base, then the unit of information entropy will be Nat.

Deng Entropy

Deng proposed a new belief entropy called Deng entropy (Deng, 2016). It is presented to measure the uncertainty degree of basic probability assignment as a generalized Shannon entropy in Dempster-Shafer evidence theory. Deng entropy is given by:

$$E_d = -\sum_i m(F_i) \log \frac{m(F_i)}{2^{F_i} - 1}$$

Where F_i is a proposition in mass function m, and $|F_i|$ is the cardinality of F_i. Deng entropy is similar to Shannon entropy in form. The difference is that the belief for each proposition F_i is divided by a term (2^{F_i}-1) which represents the potential number of states in F_i (The empty set is not included). So Deng entropy is the generalization of Shannon entropy, which is used to measure the uncertainty degree of BPA (Deng, 2016).

Deng entropy can definitely degenerate to the Shannon entropy if the belief is only assigned to single elements. The process is shown as follows.

$$E_d = -\sum_i m(\theta_i) \log \frac{m(\theta_i)}{2^{|\theta_i|} - 1} = -\sum_i m(\theta_i) \log m(\theta_i)$$

Improved Deng Entropy

In Dempster-Shafer framework, the Improved Deng Entropy (Zhou et al., 2017) is proposed as follows:

$$E_{Id}(m) = -\sum_{A \subseteq X} m(A) \log_2 \left(\frac{m(A)}{2^{|A|-1}} e^{\frac{|A|-1}{|X|}} \right)$$

Where X is the FOD, $|A|$ denotes the cardinality of the focal element A, and $|X|$ is the number of elements in the FOD. Compared with some other uncertainty measures in Yager (1983), Dubois (1985), Klir and Ramer (1990), George and Pal and (1996), Song et al. (2015), Improved Deng Entropy addresses more information in a BOE. The uncertain information addressed by the new belief entropy includes the information represented by the mass function, the cardinality of each proposition, the scale of FOD (denotes as $|X|$, and the relative scale of a focal element with respect to the FOD (denoted as $((|A|\text{-}1)/|X|)$).

Numerical example

Assume that the data is taken from 3 different sites or complexes (C_1, C_2 and C_3) of a given ecosystem. In this hypothetic data, each complex is divided into 9 subsamples and plant species (S) are recorded in each subsample. C_1 and C_2 include 15 species whereas C_3 has 6 species (Table 1).

If we decide to use Shannon entropy, we have to use proportional values for each species (p_i). Proportional values (p_i) of the species (S_i) from S_1 to S_{15} in C_1 are 0.0625; 0.0625; 0.0625; 0.0625; 0.0625; 0.0625; 0.125; 0.0625; 0.0625; 0.0625; 0.0625; 0.0625; 0.0625; 0.0625 and 0.0625 respectively. Proportional values (p_i) from S_1 to S_{15} in C_2 are 0.04348; 0.04348; 0.08696; 0.04348; 0.21739; 0.04348; 0.08696; 0.04348; 0.13043; 0.04348; 0.04348; 0.04348; 0.04348; 0.04348 and 0,04348 respectively. With regards to C_3, proportional values (p_i) are 0,25; 0,25; 0,19; 0,13; 0,06; 0,13 from S_1 to S_6 respectively. When the Shannon entropy values of the complexes are computed using p_i values, those values (H values) are found to be 2.686; 2.522 and 1.7 for C_1, C_2 and C_3 (Figure 1).

Table 1. A hypothetical data composed of 3 complexes

C1	S1, S2	S3, S4, S5	S6
	S7	S7	S8, S9
	S10, S11	S12, S13, S14	S15
C2	S1, S2, S3	S3,S4, S5	S5, S6
	S5, S7	S5, S7	S5, S8, S9
	S9, S10, S11	S9, S12, S13, S14	S15
C3	S1, S2	S1,S2, S3	S1
	S2	S2	S1, S3
	S3, S4	S4, S5, S6	S6

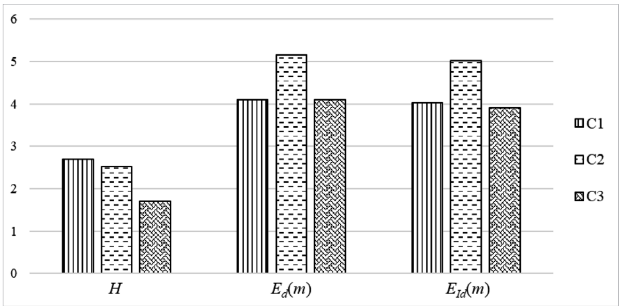

Figure 1. H, E_d and E_{Id} values of the complexes

If we prefer to use Deng entropy and/or Improved Deng entropy, then we will use mass function, m. In this case, the mass functions of C_1 are $m_1(\{S_1, S_2\})=0.11111$; $m_2(\{S_3, S_4, S_5\})=0.11111$; $m_3(\{S_6\})=0.11111$; $m_4(\{S_7\})=0.22222$; $m_5(\{S_8, S_9\})=0.11111$; $m_6(\{S_{10}, S_{11}\})=0.11111$; $m_7(\{S_{12}, S_{13}, S_{14}\})=0.11111$ and $m_8(\{S_{15}\})=0.11111$. The mass functions of C_2 are $m_1(\{S_1, S_2, S_3\})=0.11111$;$m_2(\{S_3,S_4,S_5\})=0.11111$;$m_3(\{S_5,S_6\})=0.11111$;$m_4(\{S_5, S_7\})=0.22222$; $m_5(\{S_5, S_8, S_9\})=0.11111$; $m_6(\{S_9, S_{10}, S_{11}\})=0.11111$; $m_7(\{S_9, S_{12}, S_{13}, S_{14}\})=0.11111$ and $m_8(\{S_{15}\})=0.11111$. Lastly, the mass functions of C_3 are $m_1(\{S_1, S_2\})=0.11111$; $m_2(\{S_1, S_2, S_3\})=0.11111$; $m_3(\{S_1\})=0.11111$; $m_4(\{S_2\})=0.22222$; $m_5(\{S_1, S_3\})=0.11111$; $m_6(\{S_3, S_4\})=0.11111$; $m_7(\{S_4, S_5, S_6\})=0.11111$ and $m_8(\{S_6\})=0.11111$. According to mass function values, Deng entropy (*Ed*) values of C_1, C_2 and C_3 are found to be 4.09988; 5.157836 and 4.09988 whereas Improved Deng entropy (*EId*) values are 4.025074; 5.008223 and 3.912864 for C_1, C_2 and C_3 respectively (Figure 1).

Comparisons and Interpretations

H value is the maximum in C_1. This result is not confirmed by the results of E_d and E_{Id}. Because the maximum values of E_d and E_{Id} are found in C_2. In addition to this, it seems that C_3 has minimum entropic value in accordance with the results of H and E_{Id}. However, C_1 and C_3 have the same entropic value when using E_d (Figure 1).

It is clear that the computed results of the entropic measures include disagreements in terms of grading by considering entropic values of the complexes. It should be explained why the differences of the results among the entropic measures has occurred. More importantly, it should be decided which entropic measure gives the most reasonable grading ecologically speaking.

C_1 and C_2 have the same number of species. Namely, 15 plants are found in each of C_1 and C_2. However, the total number of the individuals found in C_1 is 16 compared with 23 in C_2 (Table 1). In this case, we conclude that C_2 should have a higher entropic value compared to C_1. This result could be provided by E_d and E_{Id} but H. As explained before, the reason of the incomplete result of H compared to E_d and E_{Id} is due to the fact that proportional values (p_i) are only used to compute the Shannon entropy (H) value while Deng entropy (E_d) and Improved Deng entropy (E_{Id}) are computed by using mass function value, m. In other words, since the information is not clear in the hypothetic data (Table 1), as usual, the values of Deng entropy and Improved Deng entropy show differences from the values of the Shannon entropy.

With regard to the grading difference between C_1 and C_3 considering the computed values of E_d and E_{Id}, as explained by Zhou et al. (2017), E_d does not take into consideration the scale of the FOD, which means a loss of information while processing information. However, E_{Id} overcomes this limitation. Unlike E_d, the entropic value differences can therefore be detected between C_1 and C_3 when using E_{Id}. In other words, even though the number of the elements found in the mass functions of C_1 and C_2 includes the same values, these mass functions of C_1 and C_2 do not include the same species. When E_d is computed, this difference is ignored. When E_{Id} is computed, this difference is taken into consideration.

CONCLUSION

Biodiversity plays a very important role in maintaining the balance and protecting the health of ecosystems and has attracted increasing interest in recent years. This topic was stressed specifically at the Rio Declaration and again at the Lisbon Conference in 1988. Biodiversity should always be defined using quantities (Özkan, 2016a).

There are a wide variety of quantities available for computing biodiversity such as the Margalef index, McIntosh index, Simpson index, Fisher alpha, Brillouin index, Shannon entropy and so on (Özkan, 2016b). Among these measures, the most popular metric of biodiversity, derived from information theory, is the Shannon entropy (Shannon, 1948). In fact, Shannon entropy originating in physics and engineering has been frequently used not only to measure biodiversity but also to process data in many areas of science such as chemistry, genetic, music, architecture, urban planning, computer languages and human languages (Robinson, 2008; Doyle, 2009).

Even if Shannon entropy is the most popular measure, as explained by Jost (2006), it cannot be relied upon to measure biodiversity in all conditions. That is particularly valid when a priori information is not clear. In this case, the application of the various forms of Shannon entropy is reasonable. Deng entropy and Improved Deng entropy in the Dempster-Shafer framework are the alternative measures to Shannon entropy (Jiang et al, 2016b). Because Deng entropy is the generation form of Shannon entropy (Deng, 2016) and Improved Deng entropy is the entropy-based Deng entropy (Zhou et al., 2017).

According to the entropic measure values obtained using the hypothetical data given in this study, the most reasonable result was obtained using Improved Deng entropy from an ecological point of view. The reason for this is due to the fact that Shannon entropy merely uses proportional values of the species, Deng entropy ignores the scale of FOD, but Improved Deng entropy takes into consideration not only BPA but also FOD.

Although this study indicated that Improved Deng entropy is the best option for measuring biodiversity compared to Shannon entropy and Deng entropy when a priori information is not clear, further studies should be generated to confirm the inference obtained from this study using various types of real ecological data.

Ethics Committee Approval: N/A.

Informed Consent: N/A.

Peer-review: Externally peer-reviewed.

Acknowledgements: I am thankful for the constructive comments and suggestions by the Editor and the reviewers which further improved the earlier manuscript. Additionally I thank Yong Deng for giving the answers to my questions about Deng entropy by email.

Conflict of Interest: The author have no conflicts of interest to declare.

Financial Disclosure: The author declared that this study has received no financial support.

REFERENCES

- Chin, K.S., Fu, C., Wang, Y., 2015. A method of determining attribute weights in evidential reasoning approach based on incompatibility among attributes. *Computers and Industrial Engineering* vol. 87, pp. 150-162. [CrossRef]
- Deng, Y., 2016. Deng entropy. *Chaos, Solitions and Fractals* vol. 91, pp. 549-553. [CrossRef]
- Doyle, L.R., 2009. Quantification of Information in a One-Way Plant to Animal Communication System. *Entropy* 11, pp. 431-442. [CrossRef]
- Du, W.S., Hu, B.Q., 2016. Attribute reduction in ordered decision tables via evidence theory. *Information Science* vol. 364-365, pp. 91-110. [CrossRef]
- Dubois, D., Prade, H., 1985. A note on measures of specificity for fuzzy sets. *International Journal of General Systems vol.* 10, no. 4, pp. 279-283. [CrossRef]
- Fu, C., Wang, Y., 2015. An interval difference based evidential reasoning approach with unknown attribute weights and utilities of assessment grades. *Computers and Industrial Engineering* vol.81, pp. 109-117. [CrossRef]
- George, T., Pal, N.R., 1996. Quantification of conflict in Dempster-Shafer framework: a new approach. *International Journal of General Systems* vol. 24, no. 4, pp. 407-423. [CrossRef]
- Gorelick, R., 2006. Combining richness and abundance into a single diversity index using matrix analogues of Shannon's and Simpson's indices. *Ecography* 29, 525-530. [CrossRef]
- Jiang, W., Xie, C., Wei, B., Zhou, D., 2016a. A Improved method for risk evaluation in failure models and effects analysis of aircraft turbine rotor blades. *Advances in Mechanical Enginnering* vol. 8, no. 4, pp. 1-16. [CrossRef]
- Jiang, W, Wei, B, Xie, C, Zhou, D., 2016b. An evidential sensor fusion method in fault diagnosis. *Advances in Mechanic Engineering* vol 8(3), pp. 1-7. [CrossRef]
- Jiang, W., Wei, B., Qin, X., Zhan, J., Tang, Y., 2016c. Sensor data fusion based on a new conflict measure. *Mathematical Problems in Engineering* vol. 2016, p.11. [CrossRef]
- Jiang, W., Xie, C., Zhuang, M., Shou, Y., Tang, Y., 2016d. Sensor data fusion with z-numbers and its application in fault diagnosis. *Sensor* vol. 16, no. 9, p.1509. [CrossRef]
- Jost, L., 2006. Entropy and diversity. *Oikos* 113, 363-375. [CrossRef]
- Klir, G.J., Ramer, A., 1990. Uncertainly in the Dempster-Shafer theory: a critical re-examination. *International Journal of General Systems* vol. 18, no.2, pp. 155-166. [CrossRef]
- Liu, Z., Pan, Q., Dezert, J., 2013. A new belief-based K-nearest neighbor classification method. *Pattern Recognition* vol. 46, no.3, pp. 834-844. [CrossRef]
- Liu, Z.G., Pan, Q., Dezert, J., Martin, A. 2016. Adaptive imputation of missing values for incomplete pattern classification. *Pattern Recognition* vol. 52, pp. 85-95. [CrossRef]
- Ma, J., Liu, W., Benferhat, S., 2015. A belief revision framework for revising epistemic states with partial epistemic states. *International Journal of Approximate Reasoning* 59; 20-40. [CrossRef]
- Özkan, K., 2016a. Application of Information Theory for an Entropic Gradient of Ecological Sites. *Entropy* 18: 340. [CrossRef]
- Özkan, K., 2016b. Biyolojik Çeşitlilik Bileşenleri (α, ß, γ) Nasıl Ölçülür. Süleyman Demirel Üniversitesi, Orman Fakültesi Yayın No: 98, ISBN: 976-9944-452-89-2, Isparta, 142 s.
- Robinson, D.W., 2008. Entropy and uncertainty. *Entropy* 10(4): 493-506. [CrossRef]
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell Syst Tech J* 27: 379-423. [CrossRef]
- Song, Y., Wang, X., Lei, L., Yue, S., 2016. Uncertainly measure for interval-valued belief structures. *Measurement* vol. 80, pp. 241-250. [CrossRef]
- Song, Y., Wang, X., Zhang, H., 2015. A distance measure between intuitionistic fuzzy belief functions. *Knowlegde-Based Systems* vol. 86, pp. 288-298. [CrossRef]
- Su, X., Deng, Y., Mahadevan, S., Bao, Q., 2012. An improved method for risk evaluation in failure models and effects analysis of aircraft engine rotor blades. *Engineering Failure Analysis* vol.26, pp.164-174. [CrossRef]
- Tang, Y., Zhou, D., Jiang, W., 2016. A new fuzzy-evidential controller for stabilization of the planar inverted pendulum systems. *PLos ONE* vol. 11, no. 8. [CrossRef]
- Wang, Y.M., Elhag, T.M.S., 2007. A comparison of neural network, evidential reasoning and multiple regression analysis in modelling bridge risks. *Expert Systems with Applications* vol. 32, no. 2, pp.336-348. [CrossRef]

- Wang, Y.M., Yang, J.B., Xu, D.L., Chin, K.S., 2009. Consumer preference prediction by using a hybrid evidential reasoning and belief rule-based methodology. *Expert Systems with Applications* vol. 36, no. 4, pp. 8421-8430. [CrossRef]
- Yager, R.R., 1983. Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems* vol. 9, no. 4, pp. 249-260. [CrossRef]
- Yager, R.R., Filev, D.P., 1995. Including probabilistic uncertainly in fuzzy logic controller modeling using dempster-shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics* vol. 25, no. 8, pp.1221-1230. [CrossRef]
- Yang, Y., Han, D., 2016. A new distance-based total uncertainly measure in the theory of belief functions. *Knowledge-Based Systems* vol. 94, pp.114-123. [CrossRef]
- Yuan, K., Xiao, F., Fei, L., Kang, B., Deng, Y., 2016. Modelling sensor reliability in fault diagnosis based on evidence theory. *Sensor* vol. 16, no. 1, article 113.
- Zhou, K., Martin, A., Pan, Q., Liu, Z.G., 2015. Medial evindential c-means algorithm and its application to community detection. *Knowlegde-Based Systems* vol. 74, pp. 69-88. [CrossRef]
- Zhou, D., Tang, Y., Jiang, W., 2017. A Improved belief entropy in Dempster-Shafer framework. *PLoS ONE* 12(5): e0176832. [CrossRef]