Information Entropy: Illustrating Example

Andrew Kusiak
Intelligent Systems Laboratory
2139 Seamans Center
The University of Iowa
Iowa City, Iowa 52242 - 1527

andrew-kusiak@uiowa.edu
http://www.icaen.uiowa.edu/~ankusiak
Tel: 319 - 335 5934 Fax: 319-335 5669

Etymology of "Entropy"

Origin in chemistry

- Entropy = randomness

Information theory

- Amount of uncertainty

Shannon Entropy

- S = final probability space composed of two disjoint events E_1 and E_2 with probability $p_1 = p$ and $p_2 = 1 - p$, respectively.
- The Shannon entropy is defined as

$$H(S) = H(p_1, p_2) = -p\log p - (1-p)\log(1-p)$$

Single term only

Definitions

Information content

$$I(s_1, s_2, \ldots, s_m) = -\sum_{i=1}^{m} \frac{s_i}{s} \log_2 \frac{s_i}{s}$$

Entropy

$$E(A) = \sum_{j=1}^{v} \frac{s_{1j} + \ldots + s_{mj}}{s} I(s_{1j}, \ldots, s_{mj})$$

Weighted Shannon entropy

Information gain

$$Gain(A) = I(s_1, s_2, \ldots, s_m) - E(A)$$

Example

Case 1 **2 classes; 2 nodes**

No.	F1	D
1	Blue	1
2	Blue	1
3	Blue	1
4	Blue	1
5	Red	2
6	Red	2
7	Red	2
8	Red	2

Info content

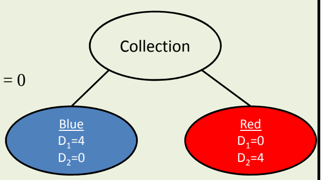
$I(D_1, D_2) = -4/8*\log_2 (4/8) - 4/8*\log_2 (4/8) = 1$

For Blue $D_{11} = 4, D_{21} = 0$
$I(D_{11}, D_{21}) = -4/4* \log_2 (4/4) = 0$

For Red $D_{12} = 0, D_{22} = 4$
$I(D_{12}, D_{22}) = -4/4* \log_2 (4/4) = 0$

$E(F1) = 4/8\, I(D_{11}, D_{21}) + 4/8\, I(D_{12}, D_{22}) = 0$

$\text{Gain}(F1) = I(D_1, D_2) - E(F1) = 1$



Gain(A) = D_1= No of examples in class 1
D_2= No of examples in class 2

$I(s_1,s_2,...,s_m) = -\sum_{i=1}^{m}\frac{s_i}{s}\log_2\frac{s_i}{s}$ $E(A) = \sum_{j=1}^{v}\frac{s_{1j}+...+s_{mj}}{s}I(s_{1j},...,s_{mj})$ $\text{Gain}(A) = I(s_1,s_2,...,s_m) - E(A)$

Example

Case 2 **3 classes; 2 nodes**

No.	F1	D
1	Blue	1
2	Blue	1
3	Blue	2
4	Blue	2
5	Red	2
6	Red	3
7	Red	3
8	Red	3

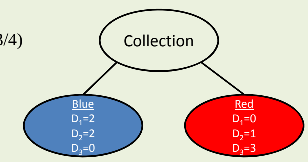
$I(D_1, D_2, D_3) = -2/8*\log_2(2/8) - 3/8*\log_2(3/8) - 3/8*\log_2(3/8) = 1.56$

For Blue $D_{11} = 2, D_{21} = 2, D_{31} = 0$
$I(D_{11}, D_{21}) = -2/4* \log_2 (2/4) -2/4* \log_2 (2/4) = 1$

For Red $D_{12} = 0, D_{22} = 1, D_{32} = 3$
$I(D_{22}, D_{32}) = -1/4* \log_2 (1/4) -3/4* \log_2 (3/4) = 0.81$

$E(F1) = 4/8\, I(D_{11}, D_{21}) + 4/8\, I(D_{22}, D_{32}) = 0.905$

$\text{Gain}(F1) = I(D_1, D_2) - E(F1) = 0.655$



$I(s_1,s_2,...,s_m) = -\sum_{i=1}^{m}\frac{s_i}{s}\log_2\frac{s_i}{s}$ $E(A) = \sum_{j=1}^{v}\frac{s_{1j}+...+s_{mj}}{s}I(s_{1j},...,s_{mj})$ $\text{Gain}(A) = I(s_1,s_2,...,s_m) - E(A)$

Example

Case 3 **3 classes; 2 nodes**

No.	F1	D
1	Blue	1
2	Blue	2
3	Blue	2
4	Blue	2
5	Red	3
6	Red	3
7	Red	3
8	Red	3

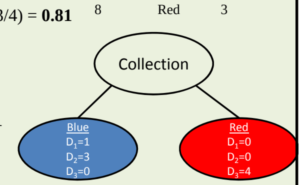
$I(D_1, D_2, D_3) = -1/8*\log_2(1/8) - 3/8*\log_2(3/8) - 4/8*\log_2(4/8) = \mathbf{1.41}$

For Blue $D_{11} = 1, D_{21} = 3, D_{31} = 0$
$I(D_{11}, D_{21}) = -1/4* \log_2 (1/4) -3/4* \log_2 (3/4) = \mathbf{0.81}$

For Red $D_{12} = 0, D_{22} = 0, D_{32} = 4$
$I(D_{32}) = -4/4* \log_2 (4/4) = \mathbf{0}$

$E(F1) = 4/8\, I(D_{11}, D_{21}) + 4/8\, I(D_{32}) = \mathbf{0.41}$

$\text{Gain}(F1) = I(D_1, D_2) - E(F1) = \mathbf{1}$



$I(s_1,s_2,...,s_m) = -\sum_{i=1}^{m}\frac{s_i}{s}\log_2\frac{s_i}{s}$ $E(A) = \sum_{j=1}^{v}\frac{s_{1j}+...+s_{mj}}{s}I(s_{1j},...,s_{mj})$ $\text{Gain}(A) = I(s_1,s_2,...,s_m) - E(A)$

Example

Case 4 **3 classes; 3 nodes**

No.	F1	D
1	Blue	1
2	Blue	1
3	Green	3
4	Green	3
5	Green	3
6	Red	2
7	Red	2
8	Red	2

$I(D_1, D_2, D_3) = -2/8*\log_2(2/8) -3/8*\log_2(3/8) -3/8*\log_2(3/8) = 1.56$

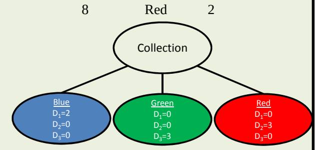
For Blue $D_{11} = 2, D_{21} = 0, D_{31} = 0$
$I(D_{11}) = -2/2* \log_2 (2/2) = 0$

For Red $D_{12} = 0, D_{22} = 3, D_{32} = 0$
$I(D_{32}) = -3/3* \log_2 (3/3) = 0$

For Green $D_{13} = 0, D_{23} = 0, D_{33} = 3$
$I(D_{33}) = -3/3* \log_2 (3/3) = 0$

$E(F1) = 2/8\, I(D_{11}) + 3/8\, I(D_{32}) + 3/8\, I(D_{32}) = 0$

$\text{Gain}(F1) = I(D_1, D_2) - E(F1) = 1.56$



$I(s_1,s_2,...,s_m) = -\sum_{i=1}^{m}\frac{s_i}{s}\log_2\frac{s_i}{s}$ $E(A) = \sum_{j=1}^{v}\frac{s_{1j}+...+s_{mj}}{s}I(s_{1j},...,s_{mj})$ $\text{Gain}(A) = I(s_1,s_2,...,s_m) - E(A)$

Example

Case 5

$I(D_1, D_2, D_3, D_4) = -2/8*\log_2(2/8) -2/8*\log_2(2/8) -2/8*\log_2(2/8) -2/8*\log_2(2/8) = 2$

For Blue $D_{11} = 1, D_{21} = 0, D_{31} = 0, D_{41} = 0$
$I(D_{11}) = -1/1* \log_2 (1/1) = 0$

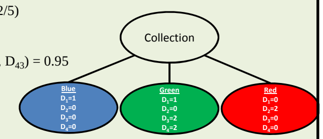
For Red $D_{12} = 0, D_{22} = 2, D_{32} = 0, D_{42} = 0$
$I(D_{22}) = -2/2* \log_2 (2/2) = 0$

For Green $D_{13} = 1, D_{23} = 0, D_{33} = 2, D_{43} = 2$
$I(D_{13}, D_{33}, D_{43}) = -1/5* \log_2 (1/5) - 2/5* \log_2 (2/5) - 2/5* \log_2 (2/5) = 1.52$

$E(F1)= 1/8\ I(D_{11}) + 2/8\ I(D_{22}) + 5/8\ I(D_{13}, D_{33}, D_{43}) = 0.95$

$Gain\ (F1)= I(D_1, D_2) - E\ (F1) = 1.05$

No.	*F1*	*D*
1	Blue	1
2	Green	1
3	Green	3
4	Green	3
5	Green	4
6	Green	4
7	Red	2
8	Red	2



$$I(s_1,s_2,...,s_m) = -\sum_{i=1}^{m}\frac{s_i}{s}\log_2\frac{s_i}{s}$$

$$E(A) = \sum_{j=1}^{v}\frac{s_{1j}+...+s_{mj}}{s}I(s_{1j},...,s_{mj})$$

$$Gain(A) = I(s_1, s_2,..., s_m) - E(A)$$

Summary

Case 1 No.	Case 1 F1	Case 1 D	Case 2 No.	Case 2 F1	Case 2 D	Case 3 No.	Case 3 F1	Case 3 D	Case 4 No.	Case 4 F1	Case 4 D	Case 5 No.	Case 5 F1	Case 5 D
1	Blue	1	1	Blue	1	1	Blue	1	1	Blue	1	1	Blue	1
2	Blue	1	2	Blue	1	2	Blue	2	2	Blue	1	2	Green	1
3	Blue	1	3	Blue	2	3	Blue	2	3	Green	3	3	Green	3
4	Blue	1	4	Blue	2	4	Blue	2	4	Green	3	4	Green	3
5	Red	2	5	Red	2	5	Red	3	5	Green	3	5	Green	4
6	Red	2	6	Red	3	6	Red	3	6	Red	2	6	Green	4
7	Red	2	7	Red	3	7	Red	3	7	Red	2	7	Red	2
8	Red	2	8	Red	3	8	Red	3	8	Red	2	8	Red	2

E(F1) = 0 | E(F1) = 0.905 | E(F1) = 0.41 | E(F1) = 0 | E(F1) = 0.95

Gain (F1) = 1 | Gain (F1) = 0.655 | Gain (F1) = 1 | Gain (F1) = 1.56 | Gain (F1) = 1.05

Continuous values → a split

http://www.icaen.uiowa.edu/%7Ecomp/Public/Kantardzic.pdf

Play Tennis:Training Data Set

Example

Outlook	Temperature	Humidity	Wind	Play tennis
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rain	mild	high	weak	yes
rain	cool	normal	weak	yes
rain	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rain	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rain	mild	high	strong	no

Decision

Feature (Attribute)

Feature values

Feature Selection

feature wind
Gain(S, wind) = 0.048

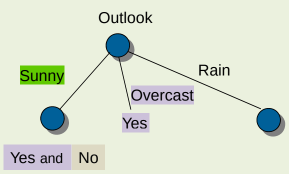
feature outlook
Gain(S, outlook) = 0.246

feature humidity
Gain(S, humidity) = 0.151

feature temperature
Gain(S, temperature) = 0.029

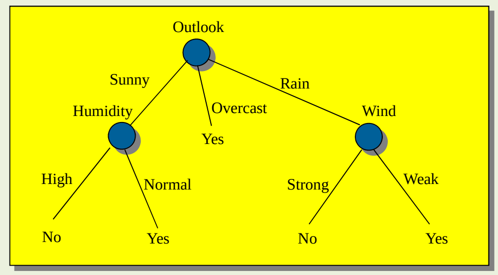
Outlook	Temperature	Humidity	Wind	Play tennis
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rain	mild	high	weak	yes
rain	cool	normal	weak	yes
rain	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rain	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rain	mild	high	strong	no

Constructing Decision Tree

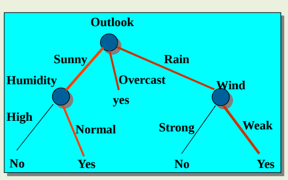


Outlook	Temp	Humidity	Wind	Play tennis
sunny	hot	high	weak	no
sunny	hot	high	strong	no
overcast	hot	high	weak	yes
rain	mild	high	weak	yes
rain	cool	normal	weak	yes
rain	cool	normal	strong	no
overcast	cool	normal	strong	yes
sunny	mild	high	weak	no
sunny	cool	normal	weak	yes
rain	mild	normal	weak	yes
sunny	mild	normal	strong	yes
overcast	mild	high	strong	yes
overcast	hot	normal	weak	yes
rain	mild	high	strong	no

Complete Decision Tree



From Decision Trees to Rules



If Outlook = Overcast
OR
Outlook = Sunny *AND* **Humidity = Normal**
OR
Outlook = Rain *AND* **Wind = Weak**
THEN Play tennis

Decision Trees: Key Characteristics

- **Complete space of finite discrete-valued functions**
- **Maintaining a single hypothesis**
- **No backtracking in search**
- **All training examples used at each step**

Avoiding Overfitting the Data

Reference

J. R. Quinlan, Induction of decision trees, *Machine Learning*, 1, 1986, 81-106.