Solving Wordle using Information Theory: Tasks 1-6

Andy Arrigony Pérez, Maryam Abuissa, Daniel Flores García

March 2023

1 Task 1

Consider $I[T;G] = H(T) - H(T|G)$. But note that $H(T) = H(T|G)$ because T and G are independent. So $I[T;G] = 0$.

2 Task 2

Consider $I[T;G|P] = H(T|P) - H(T|G,P)$. But note that $H(T|P) > H(T|G,P)$, intuitively because knowing both the guess and the pattern leads to less uncertainty about T than knowing just the pattern.

3 Task 3

If we know both the guess and the target, we can deduce the pattern. Hence, $H[P|G,T] = 0$.

4 Task 4

Consider: $I[T;P|G] = H(T|G) - H(T|P,G)$. But note that $H(T|G) = H(T) - I[T;G] = H(T)$, which does not depend on π. Thus, we need to minimize the subtractive term $H(T|P,G)$ (since it *does* depend on π) in order to maximize $I[T;P|G]$.

5 Task 5

Let R be the alphabet of P. Consider:

$$H[T|G,P] = \sum_{t\in W}\sum_{g\in W}\sum_{r\in R} p(t,g,r)\log(\frac{1}{p(t|g,r)}) = \sum_{t\in W}\sum_{g\in W}\sum_{r\in R} p(r|t,g)p(t,g)\log(\frac{1}{p(t|g,r)})$$

But note that $\exists r' \in R$ s.t. $p(r'|t,g) = 1$ and therefore $p(r^*|t,g) = 0\ \forall r^* \in R$ s.t. $r* \neq r'$. Therefore:

$$= \sum_{t\in W}\sum_{g\in W} p(t,g)\log(\frac{1}{p(t|g,r')})$$

Moreover, by uniformity of T, $p(t|g,r') = \frac{1}{|W'(g,t)|}$, where $W'(g,t)$ is a pruned alphabet. Moreover, changing the summation order:

$$= \sum_{g\in W}\sum_{t\in W} p(t,g)\log(|W'(g,t)|) = \sum_{g\in W}\sum_{t\in W} p(t)p(g)\log(|W'(g,t)|) = \frac{1}{|W|}\sum_{g\in W}\pi(g)\sum_{t\in W}\log(|W'(g,t)|)$$

And letting $\alpha_g = \sum_{t\in W}\log(|W'(g,t)|)$:

$$= \frac{1}{|W|}\sum_{g\in W}\pi(g)\alpha_g$$

6 Task 6

Suppose there are multiple guesses $g_1, g_2, ..., g_k \in W$ with the same minimum value for $\alpha_{g_1} = \alpha_{g_2} = ... = \alpha_{g_k}$. Then, let π be a distribution s.t. $\sum_{i=1}^{k}\pi(g_i) = 1$ and $\pi(g_i) \geq 0$. Note that $\pi(g_j) = 0 \ \forall g_j \in W$ s.t. $g_j \neq g_i$. Additionally, let $\alpha_\phi = \alpha_{g_i}$ for any $i \in [1,..,k]$. Consider:

$$\frac{1}{|W|}\sum_{g\in W}\pi(g)\alpha_g = \frac{1}{|W|}\sum_{i=1}^{k}\pi(g_i)\alpha_\phi = \frac{\alpha_\phi}{|W|}\sum_{i=1}^{k}\pi(g_i)$$

$$= \frac{\alpha_\phi}{|W|}$$

Now, to show that π is optimal, we show that for an arbitrary distribution π^*, $\frac{1}{|W|}\sum_{g\in W}\pi^*(g)\alpha_g \geq \frac{1}{|W|}\sum_{g\in W}\pi(g)\alpha_g$. Let g_j be a guess such that $g_j \neq g_i \ \ \forall i \in [1,...,k], \forall j \in [1,...,t]$. Consider:

$$\frac{1}{|W|}\sum_{g\in W}\pi^*(g) = \frac{1}{|W|}(\sum_{i=1}^{k}\pi^*(g_i)\alpha_\phi + \sum_{j=1}^{t}\pi^*(g_j)\alpha_j)$$

But note that:

$$\sum_{j=1}^{t}\pi^*(g_j)\alpha_j = \sum_{j=1}^{t}\pi^*(g_j)(\alpha_\phi + \beta_j)$$

for some $\beta_j \geq 0$, because α_ϕ is minimum and ≥ 0. So:

$$= \sum_{j=1}^{t}\pi^*(g_j)\alpha_\phi + \sum_{j=1}^{t}\pi^*(g_j)\beta_j$$

Plugging this into the earlier expression:

$$\frac{1}{|W|}(\sum_{i=1}^{k}\pi^*(g_i)\alpha_\phi + \sum_{j=1}^{t}\pi^*(g_j)\alpha_j) = \frac{1}{|W|}(\sum_{i=1}^{k}\pi^*(g_i)\alpha_\phi + \sum_{j=1}^{t}\pi^*(g_j)\alpha_\phi + \sum_{j=1}^{t}\pi^*(g_j)\beta_j)$$

$$= \frac{1}{|W|}(\alpha_\phi(\sum_{i=1}^{k}\pi^*(g_i) + \sum_{j=1}^{t}\pi^*(g_j)) + \sum_{j=1}^{t}\pi^*(g_j)\beta_j) = \frac{1}{|W|}(\alpha_\phi + \sum_{j=1}^{t}\pi^*(g_j)\beta_j)$$

Because $\sum_{j=1}^{t}\pi^*(g_j) \geq 0$:

$$\geq \frac{\alpha_\phi}{|W|}$$