System wspomagania podejmowania decyzji

Piotr Fulmański

16 czerwca 2019

Rozdział 1

System wspomagania podejmowania decyzji

1.1 Tytułem wstępu

W promieniu kilkuminutowego spaceru od miejsca gdzie mieszkam jest kilka cukierni. I choć sam zwykle ciastek nie kupuję, z wyjątkiem jednego rodzaju - bajaderki - od czasu do czasu, to przyszedł mi kiedyś, za sprawą mojego syna, do głowy pewien pomysł. Syn mój, zapewne za sprawą dążenia do doskonałości, ma kłopoty z podejmowaniem decyzji. Cały problem polega jak sądzę na tym, w jaki sposób wybrać najlepszą opcję biorąc pod uwagę różne czynniki. I dotyczy to wielu aspektów jego życia, od menu na śniadanie zaczynając, przez ubranie jakie warto danego dnia założyć, ciastka jakie warto zjeść na bajkach jakie warto obejrzeć kończąc. Sam nawet zamiast *warto*, zaczął mówić *opłaca się*.

1.2 System rekomendacji ciastka

1.2.1 Idea

Choć oczywiście najlepszym rozwiązaniem będzie, gdy mój syn nauczy się samodzielnie podejmować decyzje, to przyszedł mi kiedyś pomysł na *system rekomendacji ciastka*. W tym miejscu uwaga, którą warto zrozumieć. Uwaga ta dotyczy przede wszystkim tych, którzy myślą, że taki system to „czarna skrzynka", która „z niczego" odgadnie nasze pragnienia; włączymy i natychmiast będziemy otrzymywać fantastyczne „porady". Otóż, żaden system rekomendacji nie powstaje z niczego. Aby system mógł rekomendować nam ciastko zgodnie z naszym gustem, musi się tego nauczyć. Sam z siebie nie będzie tego wiedział, tak jak my nie wiemy jakie ciastka lubią ludzie mijani przez nas na ulicy. Sposoby nauki różnią się w zależności od systemu, ale niezależnie od tego na jaki się zdecydujemy, to dane musimy pozyskać. Na potrzeby naszego przykładu, przyjmiemy następujący sposób pozyskiwania danych, które w dalszej części posłużą nam do nauki.

1.2.2 Pozyskiwanie danych

Za każdym razem gdy (w końcu) uda się dokonać zakupu ciastka, zapiszemy sobie pewne informacje, np. czas na podjęcie decyzji (`czasDecyzji`), czy pogoda jest na tyle dobra aby można pokusić się o odwiedzenie innej cukierni (`dobraPogoda`), czy w cukierni są inne ciastka z kręgu zainteresowania (`saInne`) włączając w to także ostatecznie podjętą decyzje (`kupic?`). W ten sposób utworzymy daną nazywaną w dalszej części *rekordem* składającą się z zestawu *cech* (czynników, *atrybutów*) oraz oczekiwanej *odpowiedzi* (podjęta *decyzja*, *cel*). Załóżmy dla ustalenia uwagi, że udało nam się zebrać następujące dane, nazywane w dalszej części *zbiorem uczącym*

```
numer czasDecyzji cena       dobraPogoda rodzaj     wielkosc saInne kupic?

    1 malo        drogie     tak         bajaderka  srednie  tak    tak
    2 srednio     drogie     tak         bajaderka  male     nie    nie
    3 malo        przystepne nie         drozdzowka duze     nie    tak
    4 srednio     tanie      tak         drozdzowka duze     tak    tak
    5 srednio     tanie      tak         drozdzowka srednie  tak    nie
```

```
    6 malo        drogie     tak         bajaderka  male     nie    tak
    7 srednio     przystepne nie         paczek     duze     tak    tak
    8 duzo        tanie      tak         drozdzowka male     tak    nie
    9 duzo        przystepne nie         pierniczki duze     nie    nie
   10 srednio     przystepne nie         paczek     srednie  tak    nie
   11 srednio     drogie     nie         pierniczki srednie  nie    tak
   12 duzo        tanie      nie         bajaderka  male     nie    nie
```

1.2.3 Co z danych można wyczytać

Przyjrzyjmy się teraz pozyskanym danym. W przypadku atrybutu `rodzaj` otrzymaliśmy

```
           tak    nie
bajaderka    2      2
drozdzowka   2      2
paczek       1      1
pierniczki   1      1
```

Intuicyjnie czujemy, że atrybut `rodzaj` nie jest czynnikiem decydującym w pierwszym rzędzie. Patrząc tylko na tę cechę, tyle samo razy podejmowaliśmy decyzję `tak` co decyzję `nie`. Znajomość wartości tej cechy zupełnie nic nam nie daje – równie dobrze, podejmując decyzję końcową w oparciu o wartość tej cechy, moglibyśmy rzucać monetą.

Co innego cecha `czasDecyzji`

```
         tak    nie
malo       3      0
srednio    3      3
duzo       0      3
```

W tym przypadku nasze decyzje były dużo bardziej jednoznaczne

- Gdy wartość atrybutu `czasDecyzji` równa była `malo`, wówczas we wszystkich przypadkach odpowiedzią było `tak`.
- Gdy wartość atrybutu `czasDecyzji` równa była `duzo`, wówczas we wszystkich przypadkach odpowiedzią było `nie`.
- Gdy wartość atrybutu `czasDecyzji` równa była `srednio`, wówczas we połowie przypadków odpowiedzią było `tak`.

Jak więc widać, w dwóch najbardziej skrajnych przypadkach, tj. gdy dysponowaliśmy małą albo dużą ilością czasu na podjęcie decyzji, odpowiedzi były jednoznaczne.

1.2.4 Budowanie hierarchii pytań

W systemie wspomagania podejmowania decyzji (systemie rekomendacji) dążymy do tego, aby zadać użytkownikowi minimalną liczbę pytań pozwalającą wskazać, *na podstawie dotychczas zgromadzonych informacji*, jaką decyzję należy rozważyć jako najbardziej adekwatną do sytuacji.

Zdroworozsądkowe podejście do problemu nakłada na system warunek testowania najważniejszych cech (zadawania pytań o najważniejsze cechy) – najważniejszych, czyli takich, gdzie otrzymujemy najbardziej jednoznaczne odpowiedzi w ramach zbioru z przykładami (zbioru uczącego). Dodatkowo, aby ułatwić użytkownikowi udzielanie odpowiedzi, zawsze będziemy zadawać pytanie tylko o jedną cechę – o kolejną możemy zapytać dopiero po udzieleniu odpowiedzi na wcześniejsze pytanie. Postępując w ten sposób, dzielimy problem na podproblemy prostsze a każda udzielona odpowiedź powinna nas przybliżać do uzyskania odpowiedzi końcowej, zmniejszając stopień niepewności podejmowanej decyzji.

Taką metodologię postępowania nazwiemy *zachłanną strategię dziel i zwyciężaj*.

Zachłanną, gdyż poszukujemy takiego pytania (ciągu pytań), które najszybciej zmniejszy naszą niepewność odnośnie decyzji jaką należy podjąć.

Dziel, gdyż duży i złożony problem, w przypadku którego trudno nam zebrać myśli (wszak mamy wiele pytań prowadzących do końcowej decyzji), dzięki zadawanym kolejno (a nie jednocześnie) pytaniom, rozkładamy na podproblemy prostsze tego samego typu (każde udzielenie odpowiedzi na zadane pytanie prowadzi do otrzymania pomniejszonego zbioru pytań na jakie jeszcze należy odpowiedzieć). Do każdego problemu prostszego możemy następnie zastosować to samo podejście, to znaczy podzielić go na jeszcze prostsze podproblemy. Postępowanie to kontynuujemy dopóki nie dojdziemy do podproblemu, którego rozwiązanie jest oczywiste.

Zwyciężaj, gdyż systematyczne upraszczanie problemu powinno pozwolić na szybkie rozwiązanie podproblemów a znając ich rozwiązania łatwo znajdziemy rozwiązanie zadania głównego. Zwykle następuje to szybciej niż miałoby to miejsce przy rozważaniu tylko i wyłącznie problemu w całości.

Załóżmy teraz, że w jakiś sposób wybraliśmy najważniejszą (najbardziej istotną) cechę. I co dalej? Teraz nasz zbiór przykładów dzielimy na tyle podzbiorów, ile możliwych wartości przyjmuje ta wybrana cecha (mówiąc inaczej: na tyle podzbiorów, ile możemy udzielić różnych odpowiedzi pytani o tę cechę). Każdy taki podzbiór zawiera te dane, w przypadku których rozważana cecha przyjmuje identyczną wartość. W przypadku naszego przykładowego zbioru danych i cechy `czasDecyzji` otrzymamy

`czasDecyzji = malo`

numer	czasDecyzji	cena	dobraPogoda	rodzaj	wielkosc	saInne	kupic?
1	malo	drogie	tak	bajaderka	srednie	tak	tak
3	malo	przystepne	nie	drozdzowka	duze	nie	tak
6	malo	drogie	tak	bajaderka	male	nie	tak

`czasDecyzji = srednio`

numer	czasDecyzji	cena	dobraPogoda	rodzaj	wielkosc	saInne	kupic?
2	srednio	drogie	tak	bajaderka	male	nie	nie
4	srednio	tanie	tak	drozdzowka	duze	tak	tak
5	srednio	tanie	tak	drozdzowka	srednie	tak	nie
7	srednio	przystepne	nie	paczek	duze	tak	tak
10	srednio	przystepne	nie	paczek	srednie	tak	nie
11	srednio	drogie	nie	pierniczki	srednie	nie	tak

`czasDecyzji = duzo`

numer	czasDecyzji	cena	dobraPogoda	rodzaj	wielkosc	saInne	kupic?
8	duzo	tanie	tak	drozdzowka	male	tak	nie
9	duzo	przystepne	nie	pierniczki	duze	nie	nie
12	duzo	tanie	nie	bajaderka	male	nie	nie

Ponieważ w ramach każdego podzbioru wybrana cecha ma taką samą wartość, więc może ona zostać ze zbioru usunięta (ponieważ ma taką samą wartość, więc w ramach podzbioru nie dostarcza ona już żadnej nowej informacji). W ten sposób upraszczamy problem: w podzbiorze mamy mniej danych niż w zbiorze a poza tym rozważamy o jedną cechę mniej. Ogólnie rzecz biorąc, po każdym wyborze cechy dane podzielone zostają na podzbiory z mniejszą liczbą przykładów i mniejszą liczbą atrybutów. Każdy taki podzbiór, w pewnym sensie, staje się nowym problemem decyzyjnym. Należy rozważyć cztery przypadki.

- Jeśli wszystkie pozostałe w podzbiorze przykłady dają taką samą odpowiedź (odpowiedź jest jednoznaczna), to odpowiedź ta staje się końcową odpowiedzią systemu.
- Jeśli odpowiedź nie jest jednoznaczna, to poszukujemy w podzbiorze najbardziej istotnej cechy aby na jej podstawie dokonać kolejnego podziału.
- Jeśli nie ma żadnych przykładów dla tego przypadku, oznacza to, że nie zaobserwowano żadnego przykładu dla tej kombinacji wartości cech. Co możemy w takiej sytuacji zrobić? Przede wszystkim powinniśmy bardzo wyraźnie zasygnalizować użytkownikowi, że żaden przykład o takiej kombinacji cech w zebranych danych nie występuje. W związku z tym, odpowiedź systemu może znacząco rozmijać się z oczekiwaniami.

Najczęściej w takiej sytuacji system zwraca odpowiedź opartą na częstotliwości podejmowania poszczególnych wariantów decyzji w innych przypadkach, co oczywiście wcale nie musi być adekwatne do tej sytuacji. Dlatego tak ważne jest aby człowiek uzyskał w tym przypadku wyraźne ostrzeżenie.

- Jeśli odpowiedź nie jest jednoznaczna, ale nie ma już żadnej cechy o którą moglibyśmy zapytać, oznacza to, że pozostałe przykłady mają dokładnie ten sam opis, ale różne klasyfikacje. Może się tak zdarzyć, na przykład w przypadku gdy w danych występuje błąd. Podobnie jak poprzednio, najczęściej w takiej sytuacji system zwraca odpowiedź opartą na częstotliwości podejmowania poszczególnych wariantów decyzji w innych przypadkach. Także i tym razem ważne jest aby człowiek uzyskał wyraźne ostrzeżenie o zaistniałej sytuacji.

Cała procedura prowadzi do uzyskania pewnej hierarchicznej zależności pomiędzy kolejnymi pytaniami jakie należy zadać. Najpierw zadajemy pytanie pierwsze (początek hierarchii). W zależności od udzielonej odpowiedzi, zadajmy kolejne, konkretnie wskazane, pytania (są one dziećmi pytania pierwszego). Następnie kolejne i kolejne aż do otrzymania odpowiedzi z systemu. W naszym przypadku taka hierarchia pytań wyglądałaby następująco

```
czasDecyzji = malo: tak (3.0)
czasDecyzji = duzo: nie (3.0)
czasDecyzji = srednio
          |
          wielkosc = male: nie (1.0)
          wielkosc = duze: tak (2.0)
          wielkosc = srednie
                 |
                 inne = tak: nie (2.0)
                 inne = nie: tak (1.0)
```

Widzimy z niej m.in., iż

- W przypadku, gdy mam mało czasu na podjęcie decyzji, to kupuję to co jest.
- W przypadku, gdy mam dużo czasu na podjęcie decyzji, to nie kupuję tego co mogę w danej chwili (najprawdopodobniej chcę się dalej rozejrzeć w dostępnym asortymencie).
- W przypadku, gdy mogę się trochę (ale nie za długo) zastanawiać, to biorę pod uwagę wielkość ciastka. W tym przypadku wartości skrajne (ciastko małe albo duże) natychmiast prowadzą mnie do ostatecznej decyzji.
- W przypadku, gdy mogę się trochę (ale nie za długo) zastanawiać, to biorę pod uwagę wielkość ciastka. Gdy okazuje się ono mieć średnią wielkość to moje dalsze decyzje zależą od dostępności innych ciastek – jeśli nie ma innych to kupuje to co mam w danej chwili, w przeciwnym razie odkładam zakup i poszukuję alternatyw.
- W świetle zebranych danych, żadna inna cecha nie jest istotna przy podejmowaniu decyzji. Oczywiście może to ulec drastycznej zmianie wraz ze zdobycie większej liczby przykładów.

Zauważmy, że tak skonstruowany system rekomendacji doskonale „tłumaczy" nam nasz wewnętrzny mechanizm podejmowania decyzji. Dzięki temu, na podstawie surowych *danych* (rekordy uczące), które poddane pewnej interpretacji stają się *informacją* (możemy stwierdzić na przykład, że w przypadku gdy dysponowałem małą ilością czasu na podjęcie decyzji, zdecydowałem się na zakup), możemy uzyskać *wiedzę* (wiem, które cechy są dla mnie najistotniejsze przy podejmowaniu decyzji – może dzięki temu będzie mi teraz łatwiej je podejmować).

1.2.5 Obliczanie istotności atrybutów

Jak do tej pory prześlizgnęliśmy się nad kluczowym dla działania systemu tematem, to znaczy nad zagadnieniem wyboru najważniejszej cechy (najbardziej istotnego atrybutu).

W tym celu wykorzystujemy *entropię*. Entropię można interpretować jako niepewność wystąpienia pewnego zdarzenia. Jeżeli jakieś zdarzenie występuje z prawdopodobieństwem równym 1, to entropia układu wynosi wówczas 0, gdyż z góry wiadomo, co się stanie – nie ma niepewności. Pozyskiwanie nowych danych (informacji) powinno zmniejszać niepewność a więc także zmniejszać entropię.

Zmienna losowa z tylko jedną wartością (na przykład moneta, która rzucona zawsze upada na tą samą stronę) nie wprowadza żadnej niepewności; jej entropia wynosi 0. Obserwując taką monetę nie uzyskujemy żadnej nowej informacji, przez co mamy zerową zmianę naszego stanu wiedzy (zerowy przyrost informacji). Z drugiej strony „sprawiedliwa" moneta z takim samym prawdopodobieństwem upada na obie strony. W tym przypadku mamy maksymalną entropię wynoszącą 1. Nasza niepewność co do rezultatu jest w tym przypadku największa i największy jest przyrost naszej wiedzy, gdy znamy już wynik upadku monety.

W ogólności entropię H zmiennej losowej V przyjmującej k różnych wartości v_i, $i = 1, \ldots, k$, każdą z prawdopodobieństwem $P(v_i)$, definiujemy jako

$$H(V) = \sum_k P(v_i) \log_2 \frac{1}{P(v_i)} = -\sum_i P(v_i) \log_2 P(v_i).$$

W przypadku gdy $P(v_i) = 0$ dla pewnego i, wartość składnika $0 \log_2 0$ przyjmujemy jako 0, co jest zgodne z granicą wyrażenia $\lim_{p \to 0+} p \log(p) = 0$.

Korzystając z tego wzoru

- w przypadku „oszukanej" monety otrzymujemy (załóżmy, że moneta upada zawsze na stronę oznaczoną umownie przez 0, druga strona oznaczona jest przez 1)
  - możliwe wartości: $v_0 = 0$ i $v_1 = 1$,
  - prawdopodobieństwo upadku na stronę 0: $P(v_0) = 1$,
  - prawdopodobieństwo upadku na stronę 1: $P(v_1) = 0$,

$$H = -(P(v_0) \log_2 P(v_0) + P(v_1) \log_2 P(v_1)) = -(1 \log_2 1 + 0 \log_2 0) = 0 + 0 = 0.$$

- w przypadku „sprawiedliwej" monety otrzymujemy
  - możliwe wartości: $v_0 = 0$ i $v_1 = 1$,
  - prawdopodobieństwo upadku na stronę 0: $P(v_0) = 0,5$,
  - prawdopodobieństwo upadku na stronę 1: $P(v_1) = 0,5$,

$$H = -(P(v_0) \log_2 P(v_0) + P(v_1) \log_2 P(v_1)) = -(0,5 \log_2 0,5 + 0,5 \log_2 0,5) = -(-0,5 - 0,5) = 1.$$

W ogólności entropię $B(q)$ zmiennej przyjmującej dwie wartości 0 i 1 z prawdopodobieństwem odpowiednio q i $1 - q$ zdefiniować możemy jako

$$B(q) = -(q log_2 q + (1 - q) log_2 (1 - q)).$$

Wracając na chwilę do naszego przykładu, jeśli w całym zbiorze uczącym L mamy $p = 6$ przykładów pozytywnych (odpowiedź `tak`) oraz $n = 6$ negatywnch (odpowiedź `nie`) wówczas entropia zbioru uczącego wynosi 1 – mamy maksymalną niepewność co do tego jaką podjąć decyzję, gdyż tyle samo przykładów mamy na `tak` co na `nie`. Potwierdzają to także obliczenia:

$$H(L) = B\left(\frac{p}{p+n}\right) = B\left(\frac{6}{6+6}\right) = B(0,5) = 1$$

Możemy teraz spróbować określić jak zmniejszyć tę niepewność. Każdorazowo wybierając konkretną wartość A dla konkretnego atrybutu i dzieląc zbiór uczący właśnie względem tej wartości tego atrybutu (to znaczy wybierając tylko te dane uczące, w których ten atrybut ma taką właśnie wartość) możemy otrzymywać inną liczbę przykładów na `tak` oraz na `nie`. Na przykład

```
czasDecyzji = malo

numer czasDecyzji cena       dobraPogoda rodzaj     wielkosc saInne kupic?
    1 malo        drogie     tak         bajaderka  srednie  tak    tak
    3 malo        przystepne nie         drozdzowka duze     nie    tak
    6 malo        drogie     tak         bajaderka  male     nie    tak
```

W ten sposób możemy poszukiwać entropii jaka pozostanie po zbadaniu *wszystkich* możliwych wartości wybranego atrybutu. Jeśli przyjmiemy, że atrybut A posiada d różnych możliwych wartości to pozwala nam to podzielić zbiór uczący względem tego atrybutu na L_i^A, $i = 1, \ldots, d$ podzbiorów a w każdym z nich będzie p_i^A przykładów na `tak` oraz n_i^A przykładów na `nie`. Licząc, na analogicznej zasadzie jak entropię całego zbioru uczącego, entropię wybranego zbioru L_i^A otrzymujemy ogólny wzór

$$H(L_i^A) = B\left(\frac{p_i^A}{p_i^A + n_i^A}\right).$$

Ponieważ losowo wybrany element ze zbioru uczącego L będzie miał wartość atrybutu A równą v_i^A z prawdopodobieństwem $\frac{p_i^A + n_i^A}{p+n}$, więc ostatecznie entropia po wykonaniu testu na atrybucie A daje się opisać wzorem

$$H(L^A) = \sum_{i=1}^{d} \frac{p_i^A + n_i^A}{p+n} B\left(\frac{p_i^A}{p_i^A + n_i^A}\right)$$

Oczywiście najlepszy atrybut to taki, dla którego entropia wynosi 0 – nie mamy wówczas żadnej niepewności i znajomość wartości tego atrybutu natychmiast pozwala nam podjąć decyzję. Najgorszy atrybut to taki, dla którego entropia wynosi 1 – mamy wówczas maksymalną niepewność i znajomość wartości tego atrybutu zupełnie nic nam nie daje. Tak więc mniejsza wartość oznacza lepszy atrybut co kłóci się z naszą wewnętrzną intuicją – zwykle wolimy aby większe wartości odpowiadały lepszym wynikom. Dlatego czasami definiuje się *zysk informacji* jako różnicę pomiędzy wartością początkową entropii (przed podziałem zbioru za pomocą wybranego atrybutu) a wartością entropii jaka pozostanie po zbadaniu wszystkich możliwych wartości wybranego atrybutu

$$G(A) = B\left(\frac{p}{p+n}\right) - H(L^A)$$

Znajomość wartości G dla wszystkich atrybutów pozwala nam wybrać atrybut najistotniejszy, to znaczy ten, który najszybciej redukuje naszą niepewność.

W naszym przykładzie cecha `czasDecyzji` przyjmuje wartości `malo`, `srednio`, `duzo`. Dokonując podziału zbioru uczącego L za pomocą konkretnych wartości tej cechy otrzymujemy

$$G(czasDecyzji) = B\left(\frac{6}{6+6}\right) - H(L^{czasDecyzji})$$

$$\begin{aligned}
H(L^{czasDecyzji}) &= \sum_{i=malo,srednio,duzo} \frac{p_i^{czasDecyzji} + n_i^{czasDecyzji}}{p+n} B\left(\frac{p_i^{czasDecyzji}}{p_i^{czasDecyzji} + n_i^{czasDecyzji}}\right) \\
&= \frac{p_{malo}^{czasDecyzji} + n_{malo}^{czasDecyzji}}{p+n} B\left(\frac{p_{malo}^{czasDecyzji}}{p_{malo}^{czasDecyzji} + n_{malo}^{czasDecyzji}}\right) \\
&+ \frac{p_{srednio}^{czasDecyzji} + n_{srednio}^{czasDecyzji}}{p+n} B\left(\frac{p_{srednio}^{czasDecyzji}}{p_{srednio}^{czasDecyzji} + n_{srednio}^{czasDecyzji}}\right) \\
&+ \frac{p_{duzo}^{czasDecyzji} + n_{duzo}^{czasDecyzji}}{p+n} B\left(\frac{p_{duzo}^{czasDecyzji}}{p_{duzo}^{czasDecyzji} + n_{duzo}^{czasDecyzji}}\right) \\
&= \frac{3+0}{6+6} B\left(\frac{3}{3+0}\right) + \frac{3+3}{6+6} B\left(\frac{3}{3+3}\right) + \frac{0+3}{6+6} B\left(\frac{0}{3+0}\right) \\
&= \frac{3}{12} B\left(\frac{3}{3}\right) + \frac{6}{12} B\left(\frac{3}{6}\right) + \frac{3}{12} B\left(\frac{0}{3}\right) \\
&= 0,25B(1) + 0,5B(0,5) + 0,25B(0) = 0,25 \cdot 0 + 0,5 \cdot 1 + 0,25 \cdot 0 = 0,5.
\end{aligned}$$

Zatem ostatecznie

$$G(czasDecyzji) = B\left(\frac{6}{6+6}\right) - H(L^{czasDecyzji}) = 1 - 0,5 = 0,5.$$

Z kolei cecha `rodzaj` przyjmuje wartości `bajaderka`, `drozdzowka`, `paczek`, `pierniczki`. Dokonując podziału zbioru uczącego L za pomocą konkretnych wartości tej cechy otrzymujemy

$$G(rodzaj) = B\left(\frac{6}{6+6}\right) - H(L^{rodzaj})$$

$$
\begin{aligned}
H(L^{rodzaj}) &= \sum_{i=bajaderka,srednio,duzo} \frac{p_i^{rodzaj}+n_i^{rodzaj}}{p+n} B\left(\frac{p_i^{rodzaj}}{p_i^{rodzaj}+n_i^{rodzaj}}\right) \\
&= \frac{p_{bajaderka}^{rodzaj}+n_{bajaderka}^{rodzaj}}{p+n} B\left(\frac{p_{bajaderka}^{rodzaj}}{p_{bajaderka}^{rodzaj}+n_{bajaderka}^{rodzaj}}\right) \\
&+ \frac{p_{drozdzowka}^{rodzaj}+n_{drozdzowka}^{rodzaj}}{p+n} B\left(\frac{p_{drozdzowka}^{rodzaj}}{p_{drozdzowka}^{rodzaj}+n_{drozdzowka}^{rodzaj}}\right) \\
&+ \frac{p_{paczek}^{rodzaj}+n_{paczek}^{rodzaj}}{p+n} B\left(\frac{p_{paczek}^{rodzaj}}{p_{paczek}^{rodzaj}+n_{paczek}^{rodzaj}}\right) \\
&+ \frac{p_{pierniczki}^{rodzaj}+n_{pierniczki}^{rodzaj}}{p+n} B\left(\frac{p_{pierniczki}^{rodzaj}}{p_{pierniczki}^{rodzaj}+n_{pierniczki}^{rodzaj}}\right) \\
&= \frac{2+2}{6+6} B\left(\frac{2}{2+2}\right) + \frac{2+2}{6+6} B\left(\frac{2}{2+2}\right) + \frac{1+1}{6+6} B\left(\frac{1}{1+1}\right) + \frac{1+1}{6+6} B\left(\frac{1}{1+1}\right) \\
&= \frac{4}{12} B\left(\frac{2}{4}\right) + \frac{4}{12} B\left(\frac{2}{4}\right) + \frac{2}{12} B\left(\frac{1}{2}\right) + \frac{2}{12} B\left(\frac{1}{2}\right) \\
&= \frac{1}{3} B(0,5) + \frac{1}{3} B(0,5) + \frac{1}{6} B(0,5) + \frac{1}{6} B(0,5) \\
&= \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 1 + \frac{1}{6} \cdot 1 + \frac{1}{3} \cdot 1 \\
&\approx 0,33 + 0,33 + 0,16 + 0,16 = 0,98.
\end{aligned}
$$

Zatem ostatecznie

$$G(rodzaj) = B(\frac{6}{6+6}) - H(L^{rodzaj}) = 1 - 0,98 = 0,02.$$

Otrzymane wyniki zgodne są z naszą intuicją: atrybut `czasDecyzji` wydawał nam się ważniejszy niż `rodzaj`.