

Zpráva o porovnání výsledků faktorové analýzy pomocí slovních mraků

Faktorová analýza je statistická technika používaná k identifikaci základních vztahů mezi proměnnými v datovém souboru. Pomáhá snižovat dimenzionalitu dat tím, že seskupuje korelované proměnné do faktorů. Tato technika se široce používá v různých oborech, včetně psychologie, marketingu a sociálních věd, k odhalení latentních konstrukcí, které vysvětlují pozorované vzory v datech. Výsledky faktorové analýzy však mnohdy poskytují různá množství interpretovatelných faktorů, které lze navíc různými způsoby rotovat. Jinými slovy, jeden dataset může přinést i vyšší desítky faktorů k porovnání. Ty je třeba hodnotit kvalitativně, což může znamenat velmi špatnou přehlednost výsledků.

Primárním účelem tohoto projektu bylo vytvořit vizuální nástroj, který usnadňuje porovnání výsledků faktorové analýzy. Tradiční metody prezentace faktorových nábojů, jako jsou tabulky a sloupcové grafy, mohou být těžkopádné a obtížně interpretovatelné při práci s více analýzami. Slovní mraky nabízejí intuitivnější a poutavější způsob vizualizace významu proměnných v každém faktoru. Použitím slovních mraků se snažím zlepšit interpretovatelnost, zjednodušit porovnání a zlepšit komunikaci výsledků širšímu publiku. Slovní mraky poskytují vizuální reprezentaci důležitosti proměnných v každém faktoru, což usnadňuje interpretaci a porovnání výsledků mezi různými analýzami.

Projekt dále poskytuje kód ze statistického programu R, který může čtenář zapracovat do svého projektu a tvořit si tak vlastní mráčky.

Metodologie

Generování slovních mraků: Slovní mraky byly generovány pro vizualizaci faktorových nábojů pro každý faktor zvlášť. Barva, centralita a velikost slov v slovním mraku odráží velikosti faktorových nábojů.

(kód využívá knihovny "ggplot2", "ggwordcloud" a "haven")

```
factor_index <- 2 # Index mění vizualizovaný faktor
factor_loadings <- fa_result$loadings[, factor_index]
words <- rownames(fa_result$loadings)
df <- data.frame(words = words, loadings = factor_loadings)

# Filtrujte top 20 nábojů
df_filtered <- df[order(-abs(df$loadings)), ][1:20, ]

# Upraví náboje umocněním na druhou pro lepší funkci v mraku, s mocninou lze
experimentovat i řádek odebrat
df_filtered$size_loadings <- abs(df_filtered$loadings)^2

# Tvorba slovního mraku
ggplot(df_filtered, aes(label = words, size = size_loadings, color = loadings)) +
  geom_text_wordcloud_area(family = "serif", shape = "circle", rm_outside = TRUE, area_corr
= TRUE) +
  scale_size_area(max_size = 20) + # Zvyšuje maximální velikost a tím rozdíly ve velikosti
```

```
scale_color_gradient(low = "orange", high = "darkred") +
theme_void() + # Použití theme_void() k odstranění všech prvků pozadí
theme(plot.margin = margin(0, 0, 0, 0)) # zmenšuje okraje grafu
```

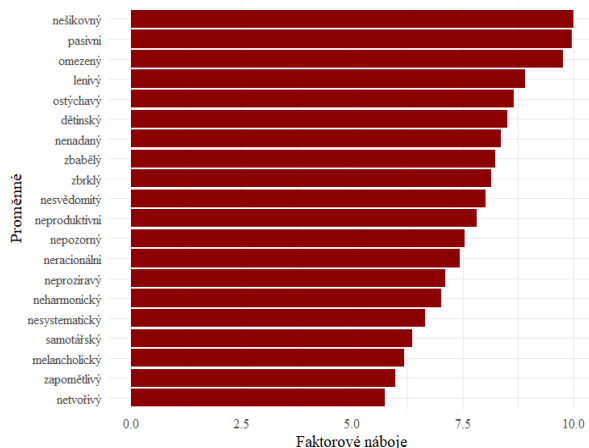
Vytvoření konvenčních sloupcových grafů: Konvenční sloupcové grafy byly vytvořeny, aby poskytly podrobnější pohled na faktorové náboje pro ověření funkčnosti slovních mraků a pro srovnání interpretovatelnosti.

```
ggplot(df_filtered, aes(x = reorder(words, loadings), y = loadings)) +
geom_bar(stat = "identity", fill = "darkred") +
coord_flip() +
labs(x = "Proměnné", y = "Faktorové náboje") +
theme_minimal() +
theme(text = element_text(family = "serif"))
```

Srovnání slovních mraků a konvenčních sloupcových grafů

Demonstrace je provedena na datech z bakalářské práce jednoho anonymního studenta. Pro bezvýznamnost použitých dat a pro jeho svolení není citován v zájmu zachování anonymity. Data se týkají osobnostní struktury.

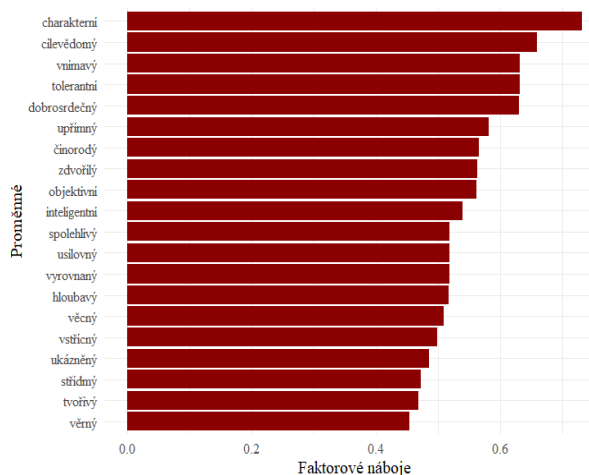
Graf č.1 Profil jednoho faktoru



Graf č.2 Profil téhož faktoru



Graf č.3 Profil jiného faktoru



Graf č.4 Profil téhož jiného faktoru



Závěr

Tento projekt úspěšně demonstroval použití slovních mraků k zjednodušení porovnání výsledků faktorové analýzy. Vizualizací důležitosti proměnných v každém faktoru slovní mraky zlepšují interpretovatelnost, zjednodušují porovnání a zlepšují komunikaci výsledků. Kombinace slovních mraků a konvenčních sloupcových grafů poskytuje přesvědčivý důkaz o větší přehlednosti nového řešení. Lze samozřejmě argumentovat, že ve specifických kontextech je zapotřebí interpretovat faktory na základě přesnější informace o jejich nábojích. Jen mě teď žádné takové konexy nenapadají.

Metodologie popsaná v této zprávě může být aplikována na různé datové soubory a analýzy, což nabízí všestranný nástroj pro výzkumníky a analytiku.